Table of Contents

Foreword

Peter Kühnlein, Hannes Rieser, and Henk Zeevat
http://www.uni-bielefeld.de/BIDIALOG

BI-DIALOG 2001 continues a glorious tradition of dialogue workshops. The founders of these series were Gerhard Jäger and Anton Benz, at the time students at the CIS, University of Munich. The first meeting (MunDial 97) took place at the CIS. MunDial 97 was followed by annual workshops in Twente (Twendial '98), Amsterdam (Amstelogue 99), and Gothenburg (Götalog 2000).

Being originally devoted to formal description of dialogue, the workshop seems to have gained considerable acceptance among scholars working on discourse and related foundational, empirical or applicational domains. This is evident from the fact that this time the local organisers were handed in fifty proposals for talks. Each of them was reviewed by at least three people. The reviewing was done by the invited speakers (Simon Garrod, Isabel Gómez Txurruka, Alois Knoll, Alex Lascarides, David Sadek, and Rob van der Sandt) and the members of the advisory board (Ellen Bard, Anton Benz, Peter Bosch, Robin Cooper, Claire Gardent, Joris Hulstijn, Yasuhiro Katagiri, Ian Lewin, Massimo Poesio, Uwe Reyle, and Henk Zeevat). In addition, all the proposals were surveyed by the local organisers as well. In the end, the present arrangements were agreed upon in overwhelming harmony, which is perhaps due to the inspiration emanating from Bosch's "Paradise", the omnipresent logo of BI-DIALOG 2001.

The proceedings were set up and prepared for print by Peter Kühnlein, Manja Nimke, and Jens Stegmann.

1 Why "Paradise"?

By and large we found positive resonance concerning the choice of Hieronymus Bosch's (∗1450–†1516) painting "Paradise" as the logo for BI-DIALOG 2001[1]. At least, most people agreed that the painting is beautiful, although from time to time the connection to the workshop on semantics and pragmatics of dialogue was overlooked. So some words seem to be in place to defend this choice a little stronger than simply saying "It's a really beautiful painting".

The chosen painting has a vertical structure by which it is divided into five or six regions. Following this structure from top to bottom, the content of the regions mirrors the historical development of the world according to the predominant medieval occidental mythology. It

[1] We want to explicitly thank the Akademie der Bildenden Künste at Vienna, Austria, and most prominently Mrs. Koch, for the kind permission to use that painting. The original painting has even more beauty than our reproduction.

is, if anything, sure that this is what Bosch intended: "Paradise" is the left inner wing of the "Last Judgement" triptych, obviously it was painted for the decoration of a church. The top region of the painting shows the christian creator in his solitude at the beginning of everything. According to mythology, he created light, literally by *fiat*. Taking the mythology serious, the first words ever spoken were "Let there be light", according to some a special kind of declarative speech act that does not have to rely on linguistic convention. Now, obviously, this is not part of a dialogue, simply because the required dialogue partners are missing.

The next one or two regions show the creation of various (kinds of) objects and the ordering of those objects into some *cosmos*. In the region that is located in the middle, there is light and harmony among the existing things.

In the forth region, matters change in this respect. The preparing step for the fifth region is laid out: The fallen angel is expelled from paradise by the arch angel, from then on being the eternal evil. It is this eternal evil that plays a central role in the fifth region, where Eve offers an apple to Adam. Both are persuaded to misbehave by the aforementioned fallen angle, being a snake according to mythology, but depicted in it's quasi-human shape by Bosch.

The last, sixth, region then shows the christian salvator in the company of a female and a male human, all three living in harmony again. That this could indeed be the bottom part of a depiction of the development of the world according to christian tradition is evident from the fact that this is what the salvator is good for: Restoring the divine regency over the world, and thereby in principle returning to the third region from the top as far as harmony is concerned. A little more acquaintance with Bosch's way of painting, however, suggests that for this unique artist harmony as a finaly state would be unsatisfactory. And indeed, the right inner wing of the triptych that complements "Paradise" and is called "Hell", shows how sinful mankind suffers pains forever. This is typical for Bosch. But we explicitly chose "Paradise" as our logo, and hence restrict ourselves to that one.

Why, and what does all this have to do with dialogue, except that the creator utters a speech act that has a very special status? To begin with: The first speech act, as mentioned above, surely is not part of a dialogue, as there is no dialogue partner. In short, it is clear that no dialogue takes place in the upper three regions (until the expellation of the fallen angel) just because one precondition is not fulfilled: There is only one agent in the setting.

The bottom region shows a different and richer scenario. Here we see *three* agents, but again no dialogue. So a different constraint seems to be violated: All agents co-exist in harmony, so maybe there is nothing to be talked about. (It would only be speculation to suppose that Bosch had this in mind, indeed.)

Only the two regions in the lower half that are left are candidates for dialogue scenarios, if this interpretation of Bosch's is right. Obviously, there is at least some monologue taking place between the evil, Eve and Adam, as the latter is persuaded to eat the apple. Probably there is no verbal exchange between the arch angle in the region above that and the evil, but this is not clear. Anyway, there could be dialogue because in these regions there is indeed something that permits clarification by verbal interaction.

Of course, this is not a really elaborate defense of the choice of Bosch's "Paradise" as a logo for the workshop. And it is even much less a comprehensive or scholarly interpretation of it. Indeed, this is not intended to be an interpretation of the painting. But this holds in any case: The painting really is beautiful and interesting, and it inspires to think about the conditions under which silence reigns or dialogue takes place.

2 Schedule & List of Participants

2.1 Schedule

Time	Thursday	Time	Friday	Saturday
9:30	Roggenhofer: Welcome			
9:45	Lascarides: Imperatives in Dialogue	9:30	Gómez Txurruka: NL Disjunction in Discourse	Garrod: Toward a mechanistic psychology of dialogue
10:30	Capone: Presuppositional Clitics, Propositional Attitudes...	10:15	Gustafson-Čapková: What are accented personal pronouns...	Newlands: An Exploration of the Complex Structure and...
11:15	Break	11:00	Break	Break
11:30	Ginzburg, Sag, & Purver: Integrating Conversational Move...	11:15	Karagjosova: Modal particles and the common ground...	Mann: Models of Intentions in Language
12:15	Lemon: Information States in a Multi-modal Dialogue System...	12:00	Tenbrink & Schilder: (Non-)Temporal Concepts Conveyed By...	Oishi: What does 'X is a Y' mean?...
13:00	Break	12:45	Break	Break
14:30	Mattheson & Kreutel: Context-Dependent Interpretation and...	14:15	Zeevat, Gustafson-Čapková, & Spenader: Constraining Pronouns	Terkourafi: The distinction between generalised and...
15:15	Schlangen, Lascarides, & Copestake: Resolving Underspecification...	15:00	Knoll & Glöckner: A Basic System for...	van der Sandt: Presuppositional denials
16:00	Spenader: Between Binding and Accommodation	15:45	Ludwig: Dialogue Understanding in Dynamic Domains	**End**
16:45	Break	16:30	Break	
17:15	Benz: On Coordinating Interpretations...	17:00	Isard: An XML architecture for the HCRC Map Task Corpus	
18:00	Sadek	17:45	Larsson, Lewin, Cooper, & Ericsson: An information state...	
18:45	Break	18:30	Break	
19:00	Poster Session & System Demo Araki: VoiceXML generator of slot-filling transaction dialogue; Berthelin & Girard: A Java Toolkit for Dialogue Evaluation; Fiedler & Horacek: Towards Understanding the Role of Hints in Tutorial Dialogs; Fischer: How much Common Ground Do we Need for Speaking?; Keizer: A Bayesian Approach to Dialogue Act Classification; Larsson, Cooper, & Ericsson: GoDiS: flexible dialogue in multiple domains; Sassen: An HPSG-based representation model for illocutionary acts in crisis talk			
20:00	Conference Dinner			

2.2 List of Participants, Advisory Board & Invited Speakers

List of Participants

Masahiro Araki	Akinremi Samuel Babatunde	Ellen Bard
Anton Benz	Jean-Baptiste Berthelin	Sofia Gustafson-Čapková
Alessandro Capone	Robin Cooper	Ann Copestake
Stina Ericsson	Kerstin Fischer	Malte Gabsdil
Simon Garrod	Jonathan Ginzburg	Alexander Gruenstein
Marie Hayet	Kirsten Hebel	Christian Hecht
Helmut Horacek	Joris Hulstijn	Christian Hying
Amy Isard	Elena Karagjosova	Simon Keizer
Alois Knoll	Jörn Kreutel	Peter Kühnlein
Alex Lascarides	Staffan Larsson	Oliver Lemon
Bernd Ludwig	William Mann	Colin Matheson
Alison Newlands	Manja Nimke	Etsuko Oishi
Massimo Poesio	Laurent Prevot	Matthew Purver
Hannes Rieser	David Sadek	Claudia Sassen
Rob van der Sandt	Frank Schilder	David Schlangen
Jennifer Spenader	Jens Stegmann	Thora Tenbrink
Marina Terkourafi	Isabel Gómez Txurruka	Henk Zeevat

Advisory Bord	**Invited Speakers**	**Local Organization**
Ellen Bard	Simon Garrod	Peter Kühnlein
Anton Benz	Isabel Gómez Txurruka	Hannes Rieser
Peter Bosch	Alois Knoll	Henk Zeevat
Robin Cooper	Alex Lascarides	
Claire Gardent	David Sadek	
Joris Hulstijn	Rob van der Sandt	
Yasuhiro Katagiri		
Ian Lewin		
Massimo Poesio		
Uwe Reyle		
Henk Zeevat		

3 Acknowledgements

We wish to thank all those who have made BI-DIALOG possible. Besides the contributors these are the following institutions and persons:

- The directorate of the ZiF (Zentrum für interdisziplinäre Forschung) for the financial and organisatorial aid they gave. In the first places we want to thank Prof. Lübbe-Wolff (Acting director) and Dr. Johannes Roggenhofer (Executive director) for giving us the opportunity to hold BI-DIALOG 2001 at ZiF at all. This, however, only constituted the background for the professional help we got from Mrs. Trixi Valentin and Mrs. Daniela Mietz. Their experience with the organization of conferences plus their endurance and kindness were presumably the strongest preconditions for the workshop to run.

- The DFG (Deutsche Forschungsgemeinschaft) who funded the workshop generously. We wouldn't have been able to invite such a number of outstanding researchers for a workshop, wouldn't it have been by the help of the DFG. The student helpers we had, Mrs. Manja Nimke and Mr. Jens Stegmann, and whom we want to thank for their help, were in part funded by the DFG, too.

- The University of Bielefeld, especially the Faculty for Linguistics and Literary Studies and the CRC SFB 360, contributed in an equally important way to the funding of the workshop. They took over the costs for the student helpers that were not covered by the DFG, and supported the invitation of our speakers.

Part I

Aspects of Dialogue

Paper 1

Imperatives in Dialogue

Alex Lascarides
Division of Informatics, University of Edinburgh
alex@cogsci.ed.ac.uk

Abstract

In this paper, we offer a semantic analysis of imperatives. We explore the effects of context on their interpretation, particularly on the content of the action to be performed, and whether or not the imperative is commanded. We demonstrate that by utilising a dynamic, discourse semantics which features *rhetorical relations* such as *Narration*, *Elaboration* and *Correction*, we can capture the discourse effects as a byproduct of discourse update (i.e., the dynamic construction of logical forms). We argue that this has a number of advantages over static approaches and over plan-recognition techniques for interpreting imperatives.

1 Introduction

An adequate theory of dialogue interpretation requires a satisfactory account of imperatives. In this paper, we will address two inter-related questions. What is their compositional semantics? And how does the discourse context affect their content?

There are several puzzles which need to be addressed. The first concerns compositional semantics. Ross (1941) observed that imperatives aren't closed under logical consequence: *post the letter* does not entail *post or burn the letter*, even though the proposition that the letter is posted entails that it is posted or burned. This makes a straightforward analysis within modal logic problematic: if ! is a 'standard' modal operator and $!A$ means A *is commanded*, then $A \models B$ will incorrectly entail $!A \models !B$, regardless of the accessibility constraints on the !-worlds.[1] Segerberg (1990) bypasses the paradox via a modal logic of action. But the semantics is static and the base language is propositional. We will find that by making the semantics dynamic, the account can be significantly simplified.

The second puzzle concerns the interaction between context and imperatives. How does context—both linguistic and non-linguistic—affect the content of imperatives, particularly the content of the action, and whether or not the imperative is commanded? Consider, for example, the discourses (1.1), adapted from Webber et al. (1995):

[1] A similar problem holds for deontic statements: *You must post the letter* doesn't entail *You must post the letter or burn the letter.*

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

(1.1) a. Go to Fred's office and get the red file folder.

b. Go to Fred's office and refile the red file folder.

c. John went to Fred's office. He got the red file folder.

d. John went to Fred's office. He refiled the red file folder.

Discourses (1.1ab) both implicate that the actions should be performed in the order described and the second action is performed in Fred's office. (1.1a) implicates that the red file folder is in Fred's office whereas (1.1b) doesn't implicate this. Similar spatio-temporal implicatures hold of the indicatives versions (1.1cd).

Segmented Discourse Representation Theory (SDRT, Asher (1993); Lascarides and Asher (1993)) accounts for the implicatures in (1.1cd) by stipulating within a dynamic semantic setting how one computes the *rhetorical relation* which connects the propositions (namely, *Narration* for (1.1cd)), and stipulating how such relations constrain the content of its arguments (e.g., the spatio-temporal content described above follows from the semantics of *Narration*). We aim to model imperatives in a similar manner. That is, we aim to account for their implicatures by identifying their rhetorical role. This involves specifying the semantics of the relations that take imperatives as arguments, and stipulating a precise default axiomatisation of how such rhetorical relations are computed on the basis of both linguistic and non-linguistic knowledge sources.

We will show that SDRT can provide an entirely uniform analysis of the imperative vs. indicative examples in (1.1), which is desirable given their similar implicatures. The uniformity rests on the fact that the SDRT axioms of interpretation that apply to these discourses are neutral with respect to sentence mood, instead relying on other compositional and lexical semantic features. In contrast, it would be hard to achieve such a uniform analysis with with plan recognition approaches (e.g., Grosz and Sidner (1986, 1990); Litman and Allen (1990); Lochbaum (1998)), where interpreting the current utterance utilises only the *goals* of the prior utterances, rather than their compositional and lexical semantics directly. This is because the goals of indicatives (typically, that the interpreter believe the proposition) are radically different from imperatives (typically, that the interpreter perform the action). The similar interpretations of (1.1a/c) and (1.1b/d) suggest that the goal of the prior clause isn't primary in these cases.

This is not to deny the importance of beliefs and goals in interpretation, however. The fact that falling downstairs is undesirable whereas going to the hardware store is not underlies the difference between (1.2a) (where the imperative is not commanded) and (1.2b) (where it is):

(1.2) a. Go straight on and you'll fall down the stairs.

b. Come home by 5pm and we can go to the hardware store before it closes.

(1.3) a. A: How does one make lasagne?

b. B: Chop onions and fry with mince and tomatoes, boil the pasta, make a cheese sauce, assemble it, and bake in the oven for 30 minutes.

(1.4) a. A: Go straight on for 5cm.

b. B: That takes me right into the crevasse.

c. A: Go left then.

Similarly, the inference that the rhetorical role of (1.3b) is to provide sufficient information that A can compute an answer to his question (1.3a) is calculable from Gricean style principles of rationality and cooperativity (e.g., Cohen and Levesque (1990); Lascarides and Asher (1999)). And whether or not such responses to questions are commanded depends on the content of the question: (1.3b) is not commanded; but an imperative *is* commanded if it serves as a response to a question whose answers all implicate that the questioner is the agent of a deontic attitude (e.g., *Where should I go now?*). Finally, in (1.4), taken from the HCRC map task corpus, the undesirability of falling into the crevasse helps one infer that the request (1.4a) is 'cancelled' and replaced by (1.4c).

Our hypothesis is that for all these examples, the interplay between content, domain knowledge and cognitive states can be captured within the semantics of the rhetorical relations, and the axioms one uses to compute them during the construction of the discourse's logical form. We will test this by incorporating a semantic analysis of imperatives into SDRT. In an attempt to do justice to the complexity of interaction between the different information sources that contribute to interpretation—both conventional and non-conventional—many theories assume a radically unmodular framework, so that a single reasoning process can access the different kinds of information at any time (e.g., Hobbs et al. (1993)). SDRT takes a different approach, assuming a high degree of modularity: reasoning with conventional clues about interpretation is kept separate from reasoning with non-conventional clues, but there are interactions between them.

2 The Compositional Semantics of Imperatives

Segerberg (1990) offers a semantics of imperatives which bypasses Ross' paradox. He augments a propositional language with two operators. First, the action operator δ takes formulae into *action terms*: e.g., if p is a propositional variable, then δp is an action term, corresponding to the action of seeing to it that p is true. Second, the command operator ! takes action terms into *practical formulae*; these are essentially the imperatives. So $!\delta p$ is a well-formed practical formula standing for "making p true is commanded"; $q \rightarrow !\delta p$ is also a practical formula standing for the conditional imperative "if q is true then making p true is commanded"; and $!p$ is ill-formed.

An action term δp denotes a set of pairs of possible worlds. Intuitively, the first world of each pair corresponds to a possible state of affairs in which the action can be performed, and the second world describes a possible outcome of performing the action in that first world. Furthermore, for each action a there is a corresponding modal operator $[a]$: $[a]p$ is true in a model M at a world w just in case p is true at all worlds w' such that $\langle w, w' \rangle \in [\![a]\!]^M$ (as we'll see shortly, $[\![a]\!]^M$ is a rigid designator). In other words, p is a necessary postcondition of a. p' is a precondition if $\neg p' \rightarrow [a]\bot$.

Plans are also terms, constituting a *sequence* of actions $a_1; a_2; \ldots a_n$. These also denote sets of pairs of worlds: $\langle w, w' \rangle \in [\![a_1; \ldots a_n]\!]$ iff $\exists w_1, \ldots, w_{n-1}$ such that $\langle w, w_1 \rangle \in [\![a_1]\!], \ldots, \langle w_{n-1}, w' \rangle \in [\![a_n]\!]$. So the possible consequences of doing a_i must be compatible with the preconditions of a_{i+1}. Finally, one express *free choice*: $[\![a_1 + a_2]\!] = [\![a_1]\!] \cup [\![a_2]\!]$.

The formula δp receives its model-theoretic denotation via a function D in the model which takes propositions (i.e., a set of worlds) to actions; i.e., $[\![\delta p]\!]^M =_{def} D[\![p]\!]^M$. D satisfies the following constraint:

$$D[\![p]\!]^M \subseteq \{\langle w, w' \rangle : w' \in [\![p]\!]^M\}$$

This makes $[\delta p]p$ true at all worlds in all models; i.e., making p true guarantees that p is true.

Since the above constraint on D uses $\subseteq$ rather than $=$, the logical relationships among actions is almost entirely impotent, in the sense that it's *not* the case that if $A \models B$, then $[\![\delta A]\!] \subseteq [\![\delta B]\!]$ (i.e., all the actions for making A true aren't necessarily also actions for making B true). This is problematic, because reasoning with these action terms, and hence planning, becomes impractical because of their weak logic.

The semantics for theoretical formulas (i.e., formulas that contain no ! operator) is essentially Kripkean, and a logic of satisfaction $\models$ for theoretical formulas is defined in the usual way. Practical formulae have their own distinct but related logic: the logic $\models_r$ of requirement or 'commanding'. This logical consequence relation exploits the notion of a command system Γ *requiring* a formula; written $\Gamma \models_r^{M,w} A$, where A now is either theoretical or practical. A command system Γ is a semantic primitive, which stipulates which actions the authority commands; or more accurately, which action he commands in which situations.

More formally, a command system Γ is a set of *command sets*, one for each possible world in the model. And a command set Γ_w is a set of actions; intuitively, the actions that the authority commands in the world w (and any one of these actions may in fact only be describable by several imperatives). One can now define the logical consequence relation $\Gamma \models_r^{M,w} A$, of the command system Γ *requiring* a formula A at the world w in the model M. We present highlights here:

1. $\Gamma \models_r^{M,w} p$ iff $w \in [\![p]\!]^M$.
2. $\Gamma \models_r^{M,w} A \rightarrow B$ iff if $\Gamma \models_r^{M,w} A$, then $\Gamma \models_r^{M,w} B$.
3. $\Gamma \models_r^{M,w} [a]B$ iff for all w' such that $\langle w, w' \rangle \in [\![a]\!]^M$, $\Gamma \models_r^{M,w'} B$.
4. $\Gamma \models_r^{M,w} !a$ iff $[\![a]\!]^M \subseteq \Gamma_w$.

Note that for any theoretical formula A, $\Gamma \models_r^{M,w} A$ iff $M \models^w A$, reflecting the intuition that wishful thinking can't make things true. Furthermore, even if $A \models^M B$ in the logic of satisfaction and $[\![\delta A]\!]^M \in \Gamma_w$, it does *not* follow that $[\![\delta B]\!]^M \in \Gamma_w$. So $\Gamma \models_r^{M,w} !\delta p$ doesn't entail $\Gamma \models_r^{M,w} !\delta(p \vee q)$, thereby bypassing Ross' paradox. Unfortunately, it's also the case that $\Gamma \models_r^{M,w} !\delta(p \wedge q)$ doesn't entail $\Gamma \models_r^{M,w} !\delta p$, indicating that the logic of commanding, as well as of actions and plans, is perhaps weaker than it should be.

This account is at best incomplete. It cannot be used to analyse imperatives with quanitifiers, since its base language is propositional. And it cannot be used to explore the interaction between content and anaphora since it's static. In fact, the static semantics would yield a highly complex translation from natural language imperatives in discourse into logical form. For note that the content of (1.5) is not adequately expressed by (1.5′), where A represents the proposition that you go to the traffic lights and B represents the proposition that there's a roundabout to your right.

(1.5) Go to the traffic lights. There's a roundabout to your right.

(1.5′) $!\delta A \wedge B$

(1.5″) $\delta A \wedge [\delta A]B$

This is because (1.5′) entails that the roundabout is to your right *now* (i.e., before the action is performed), rather than being conditional on the action being performed. In fact, its intuitive

interpretation is captured in Segerberg's semantics by the formula (1.5″). But constructing such a formula on the syntax/semantics interface is impractical.

We'll shortly see that incorporating action terms into a dynamic discourse semantics simplifies this analysis. Not only will we achieve a uniform semantic construction procedure within the grammar (cf. example (1.5) above). But we can also abandon altogether Segerberg's command system and the logic of requirement. Whether or not an imperative is commanded will not be determined by a semantic primitive (i.e., the command system), but rather by the semantic consequences of its rhetorical connection to the context, which in turn is inferred from a wide variety of knowledge sources, both linguistic and non-linguistic.

Finally, as we mentioned before, avoiding Ross' paradox by sacrificing the capacity to reason about actions and commands is problematic. Hare (1967) takes a different view, arguing that one shouldn't avoid Ross' paradox at all. He suggests that $A \vdash A \vee B$ *is* in fact valid when A and B are requests. But it doesn't appear to be valid because of Gricean-style scalar implicatures. But this is unsatisfactory too, because no details are given of how scalar implicatures would have the desired effect. An alternative solution is to include some contextually determined formula ϕ within the postconditions of the action:

$$(1.6) \qquad [\![\delta p]\!]^M = \{\langle w, w' \rangle : w' \in [\![p \wedge \phi]\!]\}$$

The problem now is to compute the value of ϕ in different contexts. We'll argue that reasoning about the rhetorical role of the imperative goes some way towards systematically stipulating the content of ϕ in (1.6); for its rhetorical function will encode why it was uttered, and for what purpose.

3 Going Dynamic

The whole notion of meaning is reconstrued in dynamic semantics as a *relation* between an input context and an output context; this is known as the *context change potential* or CCP of a formula. These contexts can be characterized extensionally as assignment functions, which map the formula's variables to individuals in the model. However, to analyse imperatives and modal action operators, we need an intensional dimension. So we make contexts a world assignment pair (w, f). Thus the truth definition of a formula K will define exactly when K relates an input context (w, f) to an output context (w', g).

In Discourse Representation Theory (DRT, Kamp and Reyle (1993)), a discourse is represented by a *discourse representation structure* or DRS, which is a pair consisting of a set of discourse referents (i.e., the individuals and events that the discourse is about) and a set of DRS-conditions (these convey properties and relations among the discourse referents). Since DRS-conditions can themselves include DRSs, DRSs are recursive. The syntax and semantics is as follows:

Syntax of DRT

Suppose $U \subseteq$ *Discourse-Referents*. Then the well-formed DRSs K and DRS conditions γ are defined recursively:

$$K \quad := \quad \langle U, 0 \rangle \mid K^{\frown}\gamma$$

Let $R \in$ Predicates be an n-ary predicate and $x_1, \cdots, x_n$ be discourse referents.

$$\gamma \quad := \quad R(x_1, \cdots, x_n) \mid \neg K \mid K_1 \Rightarrow K_2 \mid K_1 \vee K_2.$$

The Semantics of DRSs

The truth definition involves embedding DRSs into a standard Tarskian model M; so $M = \langle A_M, W_M, I_M \rangle$, where A_M is a set of individuals; W_M is a set of worlds, and I_M is a function which assigns n-ary predicates at a world w a set of n-tuples of A_M. We define simultaneously the model theoretic transition P and the satisfaction of conditions V relative to the model M. From a dynamic logic perspective, P yields a change in the assignment function, extending the input over newly introduced discourse referents, while V treats the other DRS elements as tests.

The Truth Definition:

$$\begin{array}{rll}
(w,f)P_M\langle U,\emptyset\rangle(w',g) & \textit{iff} & w = w' \wedge f \subseteq g \ \wedge \ dom(g) = dom(f) \cup U \\
(w,f) \in V_M(R(x_1,\cdots,x_n)) & \textit{iff} & (f(x_1),\cdots,f(x_n)) \in I_M(R)(w) \\
(w,f) \in V_M(\neg K) & \textit{iff} & \neg\exists g\ (w,f)P_M(K)(w,g) \\
(w,f) \in V_M(K \Rightarrow K') & \textit{iff} & \forall g\ ((w,f)P_M(K)(w,g) \rightarrow \exists\, h\ (w,g)P_M(K')(w,h)) \\
(w,f) \in V_M(K \vee K') & \textit{iff} & \exists\, g\ (w,f)P_M(K)(w,g) \vee \exists h\ (w,f)P_M(K')(w,h) \\
(w,f)P_M(K^{\cap}\gamma)(w',g) & \textit{iff} & w = w' \wedge (w,f)P_M(K)(w,g) \wedge (w,g) \in V_M(\gamma)
\end{array}$$

As yet, we've not used the possible world component. But whereas extensional formulae transform the variable assignment function, action terms will transform the possible world (as well). And the dynamic semantics of DRS conditions of the form $[a]K$ invokes quantification over worlds.

More formally, we extend the language as follows:

1. If K is a DRS formula (e.g., K could be a DRS), then δK is an action term;
2. If a_1 and a_2 are action terms, then so are $a_1; a_2$ and $a_1 + a_2$.
3. If a is an action term and K is a DRS formula, then $[a]K$ is a DRS formula.
4. If K is a DRS formula and a is an action term, then $K \rightarrow a$ is a DRS condition.

The truth conditions of these new action terms and formulae are again defined in terms of a model theoretic transition P_M and satisfiability conditions V_M. The characteristic CCP of action terms is that they change the world parameter (see clause 1. below):

1. $(w,f)P_M(\delta K)(w',g)$ iff $(w',f)P_M(K)(w',g)$
2. $[\![a_1; a_2]\!]^M = [\![a_1]\!]^M \circ [\![a_2]\!]^M$
   (i.e., $(w,f)P_M(a_1;a_2)(w'',h)$ iff there is a pair (w',g) such that $(w,f)P_M(a_1)(w',g)$ and $(w',g)P_M(a_2)(w'',h)$).
   $[\![a_1 + a_2]\!]^M = [\![a_1]\!]^M \cup [\![a_2]\!]^M$
   (i.e,. $(w,f)P_M(a_1 + a_2)(w',g)$ iff $(w,f)P_M(a_1)(w',g)$ or $(w,f)P_M(a_2)(w',g)$).
3. $(w,f)P_M(K \rightarrow a)(w',g)$ iff either:

(a) $\neg\exists h(w,f)P_M(K)(w,h) \wedge (w',g) = (w,f)$; or

(b) $\exists h(w,f)P_M(K)(w,h) \wedge (w,h)P_M(a)(w',g)$

4. $(w,f) \in V_M[a]K'$ iff for every g and for every w' such that $(w,f)P_M(a)(w',g)$, $\exists h(w'g)(K)^M(w',h)$.

Note that, thanks to condition 2 above, the denotation of the complex action (1.7) is one where the individual who talks is the same as the individual who walks:

(1.7) δ [x | walk(x)] ; δ [| talk(x)]

And guarded actions (i.e., formulae of the form $K \rightarrow a$) can be the basis of conditional commands:

(1.8) If you want to get an A, study hard.

Overall, then, we will represent imperatives in DRT as action terms. For example, we assume that the grammar generates the action term (1.9′) for (1.9) (we've simplified slightly by ignoring temporal information):

(1.9) Walk!

(1.9′) δ [u,e | walk(e,u)]

The discourse referent u is the addressee: we assume a conventional default within the grammar which generates u from the imperative sentence mood. This conventional default can be over-ridden when the subject is explicitly given (e.g., *Someone close the door!*) or in a sufficiently rich discourse context, as described in Lascarides and Copestake (1998).

Semantically, the defining characteristic of a discourse which includes a commanded imperative is that its CCP changes the input world into an output one where the action has been performed. (1.9′) changes the world this way, and thus it represents a discourse where the imperative (of walking) is commanded. The CCP of the DRS representing (1.8) also captures the command status of the imperative in the required way: the dynamic semantics of guarded actions means that the imperative is commanded if you want to get an A, and it's not commanded otherwise. This dynamic characterisation of an imperative makes introducing a semantic primitive for stipulating what's commanded and what's not redundant: we can rely instead on the signature CCP that the world is transformed into one where the action has been performed.[2] And so it can replace Segerberg's notion of a command system and the accompanying logic $\models_r^M$ of requirement, thereby considerably simplifying the semantics. We would need to bypass Ross' paradox, however, by including contextually specified information ϕ in the semantics of action terms (cf. the definition (1.6)). We return to this shortly.

A further advantage of the dynamic view is that, unlike Segerberg's static analysis, the semantics of the formula (1.5′) now captures the intuitive interpretation of (1.5) (as before, A stands for *you go to the traffic lights*, B stands for *there's a roundabout to your right*, and ; now stands for *dynamic and*):

[2]This can even apply to imperatives where one seemingly doesn't have to do anything to discharge the command; e.g., *Don't move!* and *Keep quiet!*

(1.5) Go to the traffic lights. There's a roundabout to your right.

(1.5′) $\delta A; B$

However, further investigation shows that this DRT-based analysis is flawed. As we saw earlier, not all imperatives are commanded, *even when* there are no linguistically explicit clues present to indicate this (e.g., (1.3)). In fact, maintaining the DRS language as it stands leads to one of two undesirable consequences. We either maintain a simple DRS-construction procedure for imperatives (i.e., use an action term, followed by ; if subsequent clauses are present), and therefore predict the wrong semantics of dialogues like (1.3). Similarly, this construction procedure would predict the wrong interpretation of (1.10a):

(1.10) a. Go to Fred's office. Take the file with you.

b. John went to Fred's office. He took the file with him.

The natural interpretation of (1.10a) is one where the actions are to be performed at the same time, rather than in sequence. The undesirable alternative is to generate several DRSs when updating the context with an imperative, one for each possible semantic contribution to the discourse. But proliferating ambiguity is undesirable.

These problems are in fact similar to the problems with DRT's analysis of temporal discourse. Kamp and Reyle (1993) note that their rules for DRS construction handle only those simple past-tensed discourses where event sentences move the time line forward (e.g., (1.1cd)). But not all discourses behave this way (e.g., (1.10b)). Here, we see that the DRT semantics of imperatives handles just those discourses where the imperative is commanded and the subsequent utterances should be interpreted with respect to a context where the action has been performed. But not all imperatives have this effect on content, as (1.2a), (1.3) and (1.10b) attest.

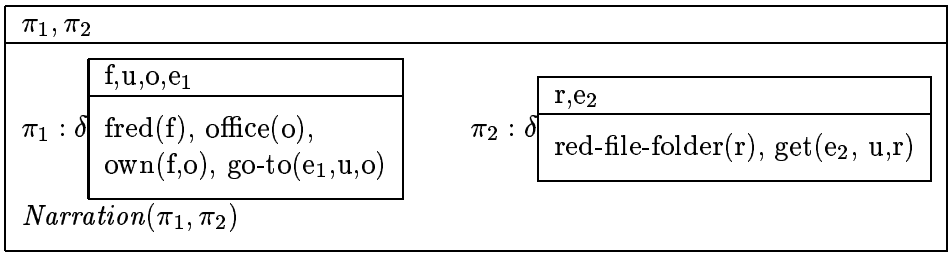
In view of these problems, we will maintain the analysis of imperatives as dynamic action terms, but incorporate it into SDRT. We'll then use SDRT's semantics of rhetorical relations to capture the various contributions imperatives can make to the overall content of the discourse.

4 Imperatives and Rhetorical Relations

Discourse is represented in SDRT as a recursive structure of labelled DRSs with rhetorical relations between the labels. For example, the logical forms of (1.1a) and (1.1c) are shown below (in slightly simplified form, since we have ignored presuppositions, tense and anaphora):

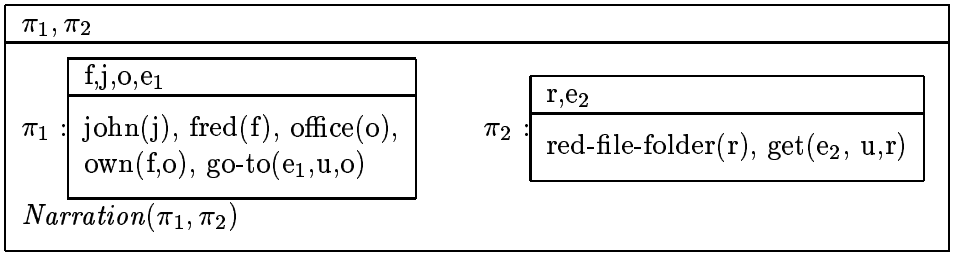
(1.1) a. Go to Fred's office and get the red file folder.

a′.



c. John went to Fred's office. He got the red file folder.

c′.



Note that the rhetorical relation *Narration* is used in both logical forms.

But how do these rhetorical relations affect the CCP of SDRSs? Well, unlike other non-logical predicates, rhetorical relations are assigned a truth definition with the capacity to *change* the world assignment pair. Moreover, we say that a relation $R(\alpha, \beta)$ is *veridical* iff $(w, f)P_M(R(\alpha, \beta))(w', g)$ entails $(w, f)P_M(K_\alpha)(w', g)$ and $(w, f)P_M(K_\beta)(w', g)$, where α labels the content K_α and β labels K_β. In other words, $R(\alpha, \beta)$ is veridical if it entails K_α and K_β. Similarly, R is left-veridical iff $(w, f)P_M(R(\alpha, \beta))(w', g)$ entails $(w, f)P_M(K_\alpha)(w', g)$, and right-veridical if $(w, f)P_M(R(\alpha, \beta))(w', g)$ entails $(w, f)P_M(K_\beta)(w', g)$.

Now, *Narration* is veridical; in fact, its CCP satisfies the content of its arguments *in dynamic sequence*, as defined by the connective ';'. So its truth definition in SDRT is as follows:

- `Semantics of Narration`
  $(w, f)P_M(Narration(\pi_1, \pi_2))(w', g)$ iff:
  1. $\langle f(\pi_1), f(\pi_2)\rangle \in I_M(Narration)(w)$ and
  2. $(w, f)P_M(K_\alpha; K_\beta)(w', g)$

This semantics ensures that the imperatives in (1.1a′) are commanded, for thanks to clause 2., the CCP of (1.1a′) transforms the input world w into an output world w' where the actions have been performed (in sequence). Similarly, clause 2. also ensures that (1.1c′) is true only if the propositions expressed by the indicative clauses are true.

There are also *meaning postulates* on *Narration* which capture its spatio-temporal effects:[3]

- `Axiom on Narration`
  $Narration(\alpha, \beta) \wedge actor(x, e_\alpha) \wedge actor(x, e_\beta) \rightarrow$
  $overlap(loc(x, prestate(e_\beta)), loc(x, poststate(e_\alpha)))$

In words, this stipulates that an actor x that is a participant in both events is in the same place, in space and time, at the end of the first event and at the beginning of the second. So,

[3]In fact, this axiom is stated here in slightly simplified form, because it ignores the role of frame adverbials.

in (1.1ac), the agent is in the same place once he's finished going to the office as he is when he starts to get the file; i.e., he must start to get the file in Fred's office. And therefore, the file must be in Fred's office too, thanks to the lexical semantics of *get*.

One can think of such meaning postulates as constraining the admissible models in SDRT. More formally, consider (1.1a′). Because of the *Narration* relation, the CCP of this SDRS relates the world assignment pair (w, f) to (w', g) only if (w', g) verifies that (a) both actions have been performed, such that (b) $e_1 \prec e_2$ (i.e., e_1 preceded e_2), and (c) the red file folder is in Fred's office at the time when you get it. Similar constraints are imposed on the CCP of (1.1c′) by the very same axioms. These constraints on rhetorical relations thus account for implicatures.

This illustrates how in SDRT, implicatures are computed as a byproduct of computing discourse update: If one infers that a particular rhetorical relation must be used to connect the content of the current clause to the discourse context (we'll outline shortly how one does this), and if neither the content of that context nor the compositional semantics of the current clause verify the consequences of the rhetorical relation's meaning postulates (e.g., the spatio-temporal information of *Narration*), then this content is in essence accommodated, for it constrains the CCP of the updated SDRS. So SDRT predicts implicatures which are brought about by the demands of discourse coherence (i.e., the demand that we connect every bit of information in the discourse to some other bit of information with a rhetorical relation).

Computing implicatures via discourse update has two desirable consequences. First, it means that we can go some way towards axiomatising inferences about the value of ϕ in the formula (1.6), which we suggested earlier as a basis for bypassing Ross' paradox. Inferences about ϕ will essentially be part and parcel of inferences about the rhetorical relations that hold and their semantic effects. For example in discourse (1.1a), the implicature that the file is in Fred's office would be part of the contextually specified postconditions ϕ of the action. Thus rhetorical relations provide a first step towards avoiding Ross' paradox without sacrificing logical relationships among actions in the way that Segerberg does.

The second desirable consequence is that using the same rhetorical relation in the logical forms of (1.1a) and (1.1c) helps to explain why they have similar (spatio-temporal) implicatures. The discourses (1.1b) and (1.1d) are also similar: `Axiom on Narration` entails that the file is refiled in Fred's office. The uniform analysis of these discourses actually goes further than this: the logical forms of (1.1a) and (1.1c) are *constructed* in the same way as well. To see this, consider the way in which SDRSs are constructed in SDRT. This is done within a glue logic, which consists of default axioms for inferring which rhetorical relation one uses to attach the new information to the logical form of the discourse context that's been constructed so far (see Asher and Lascarides (1995) for details). These default axioms encapsulate how a variety of knowledge sources provide clues about which rhetorical relations holds. So the axioms feature a default connective: $A > B$ means *If A then normally B*. The general schema for the axioms is given in (1.11): $\langle \tau, \alpha, \beta \rangle$ means that β (which labels an (S)DRS) is to be attached to a label α with a rhetorical relation, where α is part of the SDRS τ which represents the discourse context so far; and $Info(\tau, \beta)$ is a gloss for formulae that tell us properties of τ and β.

(1.11) $\quad (\langle \tau, \alpha, \beta \rangle \wedge Info(\tau, \beta)) > R(\alpha, \beta)$

Lexical semantics, domain knowledge and maxims of conversation essentially instantiate rules like (1.11). But typically, the rule itself has an antecedent which contains information that's derivable from the SDRSs that τ, α and β label. In other words, even if the justification of the

rule resides in, for example, the model of discourse participants as rational and cooperative agents, the rule itself may appeal only to *linguistic* information in the antecedent. We will see an example of such a rule in section 4.2.

The axiom for inferring *Narration* is treated as a 'basic' default in Asher and Lascarides (1995), and it captures aspects of Grice's (1975) Maxim of Manner (i.e., *be orderly*):

- `Narration`: $\langle \tau, \alpha, \beta \rangle > \mathit{Narration}(\alpha, \beta)$

This default axiom together with `Axiom on Narration` stipulates that by default, people describe things in the order in which they occur, or are to occur.

Note that this default rule is neutral with respect to sentence mood. In particular, it applies when attempting to construct the logical forms of both (1.1a) and (1.1c). And in both cases, the consequent of the rule is consistent with the monotonic information that's available. So the underlying logic for $>$ yields the inference that *Narration* holds. Hence not only do the SDRSs (1.1a′) and (1.1c′) capture the implicatures of (1.1a) and (1.1c) in a uniform way, but also, in spite of the different sentence moods, the *way* in which these logical forms are constructed is uniform.

This contrasts with the plan-recognition approach to discourse interpretation (e.g., Grosz and Sidner (1990); Litman and Allen (1990)). These theories reason about the way new information updates the meaning of the discourse by reasoning about how the communicative intention of the current utterance relates to the communicative intentions of the prior utterances. The communicative intentions that are conventionally associated (by default) with indicatives vs. imperatives are quite different. And so it's unclear how these theories could use the same axioms and proofs to explain their interpretations.

This semantic uniformity of imperatives vs. indicatives extends to other rhetorical relations as well. For example, *if...then* is a monotonic linguistic clue that the clauses are connected with the rhetorical relation *Condition*. So the axiom that encapsulates this will introduce *Condition* into the logical forms of both (1.12a) and (1.12b):

(1.12) a. If Ewan's in his office, then tell Johan the meeting is at 2pm.

b. If Ewan was in his office, then John told Johan that the meeting was at 2pm.

Unlike *Narration*, *Condition* isn't veridical. Rather, the following holds:[4]

$$\mathit{Condition}(\alpha, \beta) \rightarrow (K_\alpha \rightarrow K_\beta)$$

So *Condition* correctly predicts that (1.12a) is a conditional command, and it also conveys the correct semantics of the indicative discourse (1.12b).

The rhetorical relation *Elaboration* can also account for the semantics of the imperatives in (1.10a) and the indicatives in (1.10b): i.e., its semantics ensures that the action of taking documents with you is *part of* the action of going to the meeting.

(1.10) a. Go to Fred's office. Take the file with you.

b. John went to Fred's office. He took the file with him.

This is captured in `Axiom on Elaboration`:

[4] In fact, *Condition* is the SDRS equivalent of $\Rightarrow$ in DRT.

- `Axiom on Elaboration`:
  (a) $\mathit{Elaboration}(\alpha, \beta) \rightarrow K_\alpha \sqcap K_\beta$
  (b) $\mathit{Elaboration}(\alpha, \beta) \rightarrow e_\alpha \subseteq e_\beta$

This stipulates that the CCP of $\mathit{Elaboration}(\alpha, \beta)$ includes the *intersection* of the CCPs of the constituents that α and β label (hence *Elaboration* is veridical, and imperatives connected with *Elaboration* are commanded). Furthermore, the events are in a part-of relation (and so $\mathit{Elaboration}(\alpha, \beta)$ and $\mathit{Narration}(\alpha, \beta)$ are mutually inconsistent). So representing (1.10ab) with *Elaboration* captures the desired implicatures, and makes the temporal properties distinct from (1.1).

One doesn't infer *Narration* (via `Narration`) for connecting the constituents in (1.10), because a more specific conflicting default axiom applies in the glue logic, namely `Elaboration`:

- `Elaboration:` $(\langle \tau, \alpha, \beta \rangle \wedge \mathit{part\text{-}of}_D(\alpha, \beta)) > \mathit{Elaboration}(\alpha, \beta)$

In words, `Elaboration` states that if you're connecting β to α, and there's evidence within the discourse that they're in a part-of relationship, then normally the rhetorical connection is *Elaboration*. Discourse evidence of a part-of relation is typically modelled via *monotonic* rules which feature *linguistic* information about the constituents in the antecedent; i.e., they're axioms of the form $\mathit{Info}(\alpha, \beta) \rightarrow \mathit{part\text{-}of}_D(\alpha, \beta)$ (see Asher and Lascarides (1995) for details). In (1.10), the monotonic rule which applies instantiates $\mathit{Info}(\alpha, \beta)$ with the information that α describes movement, and β also describes a causative movement performed by the same agent. Note that this is neutral with respect to sentence mood, and so the SDRSs for (1.10ab) are constructed via the same glue logic axioms.

4.1 Defeasible Conditionals and Metatalk Relations

The rhetorical relations we've considered so far constrain the *contents* of the constituents they connect. Moore and Pollack (1992) argue convincingly that rhetorical relations can also reveal information about intentions and speechacts. They use (1.13) to observe that recognising a content-level relation is sometimes necessary for recognising the intentional one and *vice versa*:

(1.13) a. Come home by 5pm.
b. Then we can go to the hardware store before it closes.
c. That way, we can finish the shelves tonight.

At the content level, the clauses in (1.13) describe events which are in *consequence* relations: doing the action described in (1.13a) normally results in (1.13b) being true; and (1.13b) being true normally means (1.13c) is true too. Elsewhere we have used the non-veridical relation $\mathit{Def\text{-}Consequence}(\alpha, \beta)$ to mark this connection between propositions (e.g., Asher and Lascarides (1998b)):

- `Axiom on Def-Consequence`
  $\mathit{Def\text{-}Consequence}(\alpha, \beta) \rightarrow (K_\alpha > K_\beta)$

When the first constituent is a request, however, the corresponding action term cannot be an antecedent to $>$ *directly*, because it's not of the right semantic type. Rather, where α labels $\delta K'_\alpha$, the appropriate consequence relation is $[\delta K'_\alpha]\top > K_\beta$. Or, in words, any

situation where the action described by the imperative α is performed is normally one where the proposition β is true as well. We encode this content-level relationship in the rhetorical relation *Def-Consequence*$_r$ (α :! means that α labels an imperative and β : . means that β labels an indicative):

- `Axiom on Def-Consequence`$_r$
  (a) $\mathit{Def\text{-}Consequence}_r(\alpha, \beta) \rightarrow (\alpha :! \wedge \beta : .)$
  (b) $(\mathit{Def\text{-}Consequence}_r(\alpha, \beta) \wedge \alpha :! \delta K'_\alpha) \rightarrow [\delta K'_\alpha](\top > K_\beta)$

Like *Narration*, *Def-Consequence*$_r$ encodes information about what results from doing the action described by α. But unlike *Narration*, *Def-Consequence*$_r$ is not veridical: the imperative is not commanded since the CCP of the SDRS will not have the characteristic transformation of the world index.

Def-Consequence$_r$ is part of the semantic representation of the discourses (1.2), (1.13) and (1.14):

(1.2) a. Go straight on and you'll fall down the stairs.

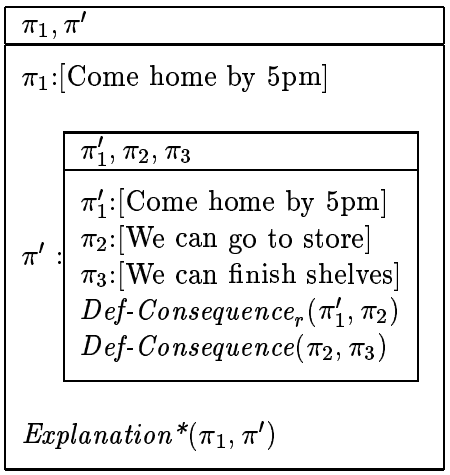
b. Come home by 5pm and we can go to the hardware store before it closes.

(1.14) Smoke 20 cigarettes a day and you will die before you're 30.

Discourse (1.14) doesn't implicate that the imperative *smoke 20 cigarettes a day* is commanded, largely because the consequent state (death) is undesirable. Similarly for (1.2a).

However, in contrast to (1.14) and (1.2a), the request is commanded in (1.13) and (1.2b). As Moore and Pollack (1992) point out informally, the consequence relations at the 'content-level' in (1.13), together with the background knowledge that (1.13c) is a desirable state, yield further inferences: these consequence relations *explain why* the speaker made the request. This is an example of what Polanyi (1985) calls a 'meta-talk' relation, for it connects the content of one utterance to the performance of uttering another. In words, the meta-talk relation *Explanation**(α, β) means that β explains why *Agent*(α) (i.e., the person who uttered α) performed the speechact α (e.g., *Explanation** is part of the representation of *Close the window. I'm cold*). Like *Narration* and *Elaboration*, it's a veridical relation. So (1.13) is represented as (1.13′) (for simplicity, we haven't stipulated the action terms and DRSs that represent the clauses):

(1.13′)



In words, (1.13′) stipulates that the following explains why the request (1.13a) is commanded: doing the action described by (1.13a) normally leads to being able to go to the hardware store before it closes, which in turn normally leads to being able to finish the bookshelves tonight. Note that (1.13′) represents the content of the imperative at two levels in the discourse structure: it's the first argument in an *Explanation** relation; and its also part of the representation of the second argument π' to this relation. This reflects the fact that rhetorically, the imperative plays a 'dual role': its content and its default consequences motivate its own command status. Since *Explanation** is veridical, the request expressed by (1.13a) is commanded. The SDRS labelled π' must be true as well; but because *Def-Consequence isn't* veridical, this doesn't mean that the propositions expressed by (1.13b) or (1.13c) are true.

The difference between (1.13) and (1.14) is (1.14) does not feature the veridical *Explanation** relation, but only the $\textit{Def-Consequence}_r$ one. So the imperative in (1.14) isn't commanded. There's a similar difference between (1.2a) and (1.2b). But how can one infer these differences in the glue logic for constructing logical forms? Well, the SDRS (1.13′) can be inferred via monotonic axioms which take account of the cue phrases *that way*, *then* and punctuation. Similarly, when the cue phrase *and* connects an imperative to an indicative, it monotonically generates $\textit{Def-Consequence}_r$; this applies to (1.2ab) and (1.14). Finally, *Explanation** would be inferred via a default axiom which states: if α is a request, $\textit{Def-Consequence}_r(\alpha, \beta)$ holds, and β is a desirable state, then normally $\textit{Explanation*}(\alpha', \pi)$ holds, where α' labels a 'repeat' of the content of the request α, and π labels α's and β's content and the $\textit{Def-Consequence}_r$ relation between them.

4.2 Imperative Answers

We suggested earlier that an imperative that's an (indirect) answer to a question isn't necessarily commanded. We follow the SDRT analysis of indirect answers from Asher and Lascarides (1998a), using the relation *IQAP* (standing for Indirect Question Answer Pair) to represent the connection between a question and its indirect answers. Semantically, $IQAP(\alpha, \beta)$ holds only if α is a question, and the speaker of α can infer a direct answer to his question (according to the compositional semantics of questions and answers) from the content that's labelled by β. This relation *IQAP* will feature in the logical form of (1.3), for it reflects the fact that an adverbial of manner, which constitutes the semantic type of direct answers to *how*-questions, can be computed from the contents of the imperatives.

(1.3) a. A: How does one make lasagne?

b. B: Chop onions and fry with mince and tomatoes, boil the pasta, make a cheese sauce, assemble it, and bake in the oven for 30 minutes.

Now, we must encode within the truth definition of $IQAP(\alpha, \beta)$ that imperative answers aren't always commanded: i.e., $IQAP(\alpha, \beta) \rightarrow K_\beta$ isn't valid when K_β is a request. However, this is in contrast to *IQAP* when it relates propositions, for these are right-veridical. I.e., $IQAP(\alpha, \beta) \rightarrow K_\beta$ *is* valid when K_β is a proposition (see Asher and Lascarides (1998a).

In fact, whether or not $IQAP(\alpha, \beta)$ makes the imperative β commanded depends on the compositional semantics of the question. The imperatives in (1.3) aren't commanded. But if the question is about what plans should be on the questioner's agenda, the imperative answer does seem to be commanded (e.g., (1.15) and (1.16)):[5]

[5] Actually, it's not clear whether or not (1.16) is a fragment answer where its message type isn't imperative at all, and thus not commanded.

(1.15) a. A: Where do I go now?

b. B: Go to platform 1.

(1.16) a. A: What should I do now?

b. B: Own up to the police.

We need to reflect this in the semantics for *IQAP*. First, we must have a monotonic axiom which makes $IQAP(\alpha, \beta)$ right-veridical when β is a proposition:

- `Veridicality of Propositional Answers`:
  $(IQAP(\alpha, \beta) \wedge \beta : .) \rightarrow K_\beta$

Second, when β is an imperative, β is commanded only if it is an indirect answer to a question whose *direct* answers implicate a deontic modality on the questioner. We can represent such implications as a $>$-statement. And so the axiom below captures the required information (where *QAP* stands for Question Answer Pair, and $QAP(\alpha, \beta)$ means that β is a true direct answer to the question α, according to the compositional semantics of questions and answers):

- `Veridicality of Imperative Answers: Deontic`
  $(IQAP(\alpha, \beta) \wedge \beta :! \wedge (QAP(\alpha, \gamma) > [deontic]_{Agent(\alpha)}(\phi))) \rightarrow K_\beta$

So, the logical form of (1.3) is (1.3′):[6]

(1.3′)

π_1, π

π_1:[How make lasagne?]

π :

$\pi_2, \pi_3, \pi_4, \pi_5, \pi_6$

π_2:[Chop onions], π_3:[fry with mince]
π_4:[boil pasta], π_5:[make cheese sauce]
π_6:[assemble it], π_7:[bake]
$Narration(\pi_2, \pi_3)$, $Narration(\pi_3, \pi_4)$, $Narration(\pi_4, \pi_5)$,
$Narration(\pi_5, \pi_6)$, $Narration(\pi_6, \pi_7)$

$IQAP(\pi_1, \pi)$

Since *IQAP* isn't right-veridical, the imperatives in (1.3b) aren't commanded. However, the *Narration* relations ensure that the complex action described in π has the desired temporal properties; e.g., the individual actions would be performed in the order they were uttered.

Note that all direct answers to (1.16a) are propositions that the the questioner should ϕ, for some value of ϕ. So the semantics of *IQAP* correctly predicts that the imperative answer (1.16b) is commanded. Direct answers to the question (1.15a) don't *semantically* entail a deontic proposition, but they do implicate one. Assuming that this implicature is captured in SDRT, the imperative (1.15b) is commanded according to the above semantics of *IQAP*. *IQAP* will also deal adequately with the command-status of an imperative to a conditional question:

[6]The axiom `Narration` is used to connect the imperatives with *Narration*, and the axiom `IQAP` which we'll specify shortly is used to connect the segment of imperatives to the question. We forego giving details here, however, of how one chooses which labels are connected to which other labels (but see Asher and Lascarides (1998a,b)).

(1.17) a. A: If the exam is tomorrow, then what should I do?

b. B: Revise your notes tonight!

That is, it correctly predicts that B's imperative is a conditional command; conditional on whether the exam is tomorrow.

The glue logic axiom for inferring $IQAP$ exploits Morgan's (1975) notion of *short-circuiting* calculable implicatures. In Lascarides and Asher (1999), we demonstrate that a logical model of discourse participants as rational and cooperative agents validates the following: if β is a response to a question α, then normally $IQAP(\alpha, \beta)$ holds; i.e., β supplies sufficient information that the questioner can infer a direct answer to α from it. This is represented in the glue logic via the following axiom:

- `IQAP`: $(\langle \tau, \alpha, \beta \rangle \wedge \alpha :?) > IQAP(\alpha, \beta)$

Note that although the *justification* for this axiom involves inferences that are derived from reasoning about the beliefs and intentions of the dialogue participants, the rule `IQAP` itself has an antecedent which includes only information about the sentence moods; something that is given by the grammar. So in essence, `IQAP` short-circuits the calculable inferences about when the speechact of providing an indirect answer is performed, because it allows the interpreter to entirely bypass reasoning with cognitive states, using just the sentence mood of α instead. This axiom plays a central role in constructing the SDRSs for (1.3) and (1.15–1.17).

4.3 Corrections

Consider (1.4), where intuitively A's second imperative 'replaces' the first one:

(1.4) a. A: Go straight for 5cm.

b. B: That will take me straight into the crevasse

c. A: Go left then.

We must model how rhetorical relations can yield such non-incremental interpretations: whereas the logical form for the discourse context entails that an imperative is commanded, the logical form of the updated discourse context cancels this entailment.

Commanding the imperative (1.4a) is *not* incompatible with its (undesirable) outcome (1.4b). And yet (1.4b) functions as a corrective move, since it brings into dispute that the imperative is commanded (or, more accurately, that it should be commanded). Now, in earlier work (e.g., Asher and Lascarides (2001)), we have used the relation *Plan-Correction* to model this: *Plan-Correction*(α, β) holds if β indicates that the goal which lay behind uttering α is incompatible with *Agent*(β)'s goals. This is analogous to Searle's speechact of rejection: it features in the analysis of (1.4) and (1.18) (taken from Searle (1969)).

(1.18) a. A: Let's go to the movies tonight.

b. B: I have to study for an exam.

Plan-Correction(α, β) is right-veridical but not left-veridical, thereby providing the non-incremental interpretation we require: if α labels an imperative, then an SDRS that contains *Plan-Correction*(α, β) does *not* have a CCP with the characteristic transformation of the world index, indicating that this SDRS is one where the imperative is not commanded. This then

leaves A free to issue a further command in response to the *Plan-Correction*. In fact, A requests (1.4c) as a *result* of B's utterance (1.4b), as indicated by the cue phrase *then*. But this result relation is at the speech-act level (i.e., it's a metatalk relation, connecting the content of B's utterance to A's speechact of uttering the request (1.4c)). And so (1.4) is represented as (1.4′), where *Result** encodes the appropriate metatalk relation:

(1.4′)

π_1, π_2, π_3
π_1:[Go straight for 5cm] π_2:[π_1 takes me into crevasse] *Plan-Correction*(π_1, π_2) π_3:[Go left then] *Result**(π_2, π_3)

Like *Explanation**, *Result** is veridical. Therefore, (1.4′) entails that the imperative (1.4c) is commanded, but it does not entail that (1.4a) is commanded.

5 Some Concluding Remarks

We have examined the content of imperatives in dialogue, paying particular attention to their compositional semantics and to the ways in which the discourse context affects their interpretation. We argued that dynamic semantics provides an elegant account of their compositional semantics based on action terms: the defining characteristic of a discourse which features a commanded imperative is that its context change potential (CCP) transforms the world index into one where the action has been performed.

We observed that context can affect whether imperative is commanded, and also the content of the action term—for example, the time at which the action is to be carried out. We argued that these contextual effects are best explained through capturing the *rhetorical role* of the imperative. Indeed, representing the content of discourse in terms of the rhetorical connections between the propositions and requests partly contributes to the simplicity of the compositional semantics of the imperatives, in that it makes a semantic primitive for stipulating what's commanded unnecessary, in contrast to Segerberg's analysis. The command status of an imperative is instead represented via the veridicality of its rhetorical connection to the rest of the dialogue. And since this rhetorical relation is inferred on the basis of both linguistic and non-linguistic information, SDRT provides a framework in which the information flow between the content of an imperative, domain knowledge and goals can be modelled.

We also observed similarities in the implicatures of imperatives and indicatives. SDRT is distinct from plan-recognition approaches to discourse interpretation, in that the rhetorical relations are inferred on the basis of axioms which have direct access to the linguistic form of the context. We argued that this allows for a uniform analysis of these implicatures which would be hard to achieve through plan recognition.

There are many outstanding issues. For instance, we need to examine more closely the semantic relationships between imperatives and adverbials of manner; compare (1.19a) and the semantically similar (1.19b):

(1.19) a. Go to the kitchen and make a cup of coffee.

b. Go to the kitchen to make a cup of coffee.

The interaction between imperatives, presuppositions and anaphora also deserves closer attention, as does the link between interpreting imperatives and planning (see, for example, Stone (in press)). We will examine these issues in future work.

Acknowledgements

This paper could not have been written if it weren't for extensive discussions with Nicholas Asher, Johan Bos and Ewan Klein. Nicholas Asher pointed out to me that a dynamic semantics makes Segerberg's command system redundant. Much of the work presented here will also appear in Asher and Lascarides (forthcoming). Alex Lascarides is supported by an ESRC (UK) research fellowship.

Bibliography

Asher, N. (1993). *Reference to Abstract Objects.* Kluwer AP.

Asher, N. and Lascarides, A. (1995). Lexical disambiguation in a discourse context. *Journal of Semantics*, 12(1):69–108.

Asher, N. and Lascarides, A. (1998a). Questions in dialogue. *Linguistics and Philosophy*, 23(3):237–309.

Asher, N. and Lascarides, A. (1998b). The semantics and pragmatics of presupposition. *Journal of Semantics*, 15:239–99.

Asher, N. and Lascarides, A. (2001). Indirect speech acts. *Synthese*, 128(1).

Asher, N. and Lascarides, A. (forthcoming). *The Logic of Conversation.* Cambridge UP.

Cohen, P. and Levesque, H. (1990). Rational Interaction as the Basis for Communication. in: Cohen et al. (1990).

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

Cole, P. and Morgan, J. L., editors (1975). *Speech Acts.* Academic Press.

Grice, H. P. (1975). Logic and conversation. in: Cole and Morgan (1975).

Grosz, B. and Sidner, C. (1986). Attention, Intentions and the Structure of Discourse. *Computational Linguistics*, 12:175–204.

Grosz, B. and Sidner, C. (1990). Plans for Discourse. in: Cohen et al. (1990).

Hare, R. M. (1967). Some alleged differences between imperatives and indicatives. *Mind.*

Hobbs, J. R., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as Abduction. *Artificial Intelligence*, 63(1–2):69–142.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic.* Kluwer AP.

Lascarides, A. and Asher, N. (1993). Temporal Interpretation, Discourse Relations and Commonsense Entailment. *Linguistics and Philosophy*, 16:437–93.

Lascarides, A. and Asher, N. (1999). Cognitive states, discourse structure and the content of dialogue. In *Proceedings from Amstelogue 1999*, pages 1–12, Amsterdam.

Lascarides, A. and Copestake, A. (1998). Pragmatics and word meaning. *Journal of Linguistics*, 34(2):387–414.

Linnell, P. (1998). *Approaching Dialogue: Talk interaction and contexts in a dialogue perspective*, volume 3. John Benjamins.

Litman, D. and Allen, J. (1990). Discourse Processing and Commonsense Plans. in: Cohen et al. (1990).

Lochbaum, K. (1998). A Collaborative Planning Model of Intentional Structure. *Computational Linguistics*, 24(4):525–72.

Mann, W. and Thompson, S. (1987). Rhetorical structure theory: A framework for the analysis of texts. *IPRA Papers in Pragmatics*, 1:79–105.

Moore, J. D. and Pollack, M. (1992). A problem for rst: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–44.

Polanyi, L. (1985). A theory of discourse structure and discourse coherence. In Eilfor, W. H., Kroeber, P. D., and Peterson, K. L., editors, *Papers from the General Session a the Twenty-First Regional Meeting of the Chicago Linguistics Society*, Chicago.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, 77:97–131.

Ross, A. (1941). Imperatives and Logic. *Theoria*, pages 53–71.

Searle, J. (1969). *Speech Acts*. Cambridge UP.

Segerberg, K. (1990). Validity and Satisfaction in Imperative Logic. *Notre Dame Journal of Formal Logic*, 31:203–21.

Stone, M. (in press). Towards a computational account of knowledge, action and inference in instructions. *Language and Computation*.

Webber, B., Badler, N., DiEugenio, B., Geib, C., Levinson, L., and Moore, M. (1995). Instruction, Intentions and Expectations. *Artificial Intelligence*, 73:253–69.

Paper 2

On Co-ordinating Interpretations - Perspectives and Optimality

Anton Benz
toni.benz@german.hu-berlin.de
http://anton-benz.de

Abstract

In this paper we investigate some questions about co-ordination and interpretation which have been addressed by *bidirectional* Optimality Theory (OT). We especially look at anaphora resolution, and there at the role of world knowledge and expectations, and at the role of perspectives, i.e. the partial information of interlocutors about the world, and about each other.

1 Introduction

Bidirectional Optimality Theory (OT)[1] has been suggested as a framework which explains especially the way how speaker and interpreter co-ordinate their choice of preferred forms and preferred interpretations. In this paper we are interested in the speaker's preferences for *economic* forms, and in the role of (nonmonotonic) inferences on the side of the interpreter. Recently[2] this theory has been applied to anaphora resolution. If we assume that it is more economic for the speaker to produce a pronouns than a name, and better to repeat the same name than to produce a definite description, then OT can explain why he should prefer in **(1)** *her* over *Andrea*, and *Andrea* over *the woman*.

(1) Yesterday, Andrea called me up. I know *her/Andrea/the woman* from the party last weekend.

In **(2)** the presence of two referents which denote probably female persons block the use of *her* and *the woman*.

(2) A: Do you know when the guests Andrea and Maria will arrive? B: I've phoned with *her/*the woman/Andrea. They arrive tomorrow.

[1] (Blutner, 1998, 2000; Blutner, Jäger, 2000; Zeevat, 2000; Beaver, 2000)

[2] (Beaver, 2000). Beaver's version of a two–sided OT is in some respects different from the version cited above.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

toni.benz@german.hu-berlin.de
http://anton-benz.de.

In **(2)** it is crucial that *Andrea* is interpreted as a name of a female person. But this is only a defeasible inference. It will normally hold if the conversation takes place in a German community but not if Andrea and Maria are known to be Italians. In the latter case it will be common knowledge that the hearer will assume that Andrea is male, hence, the speaker can refer with *HER* or *the woman* to Maria. World knowledge, and common expectations about the domain talked about enter at this point into the resolution process.

In bidirectional OT it is usual to assume that there is a set $\mathcal{F}$ of *forms*, and a set $\mathcal{M}$ of *meanings* (Blutner, 2000). The speaker has to choose for his next utterance a form which then must be interpreted by the hearer. It is further assumed that the speaker has some ranking on his set of forms, and the hearer on the set of meanings. Blutner (2000) introduced the idea that the speaker and interpreter co-ordinate on form–meaning pairs which are most preferred from both perspectives. In (Jäger, 2000) the mechanism which leads to optimal form–meaning pairs is discussed in greater detail[3]. The speaker has to choose for a given meaning M_0 a form F_0 which is optimal according to his ranking of forms. Then the interpreter has to choose for F_0 a meaning M_1 which is optimal according to his ranking of meanings. Then again the speaker looks for the most preferred form F_1 for M_1. A form–meaning pair is optimal, if ultimately speaker and hearer choose always the same forms and meanings. If $\langle F, M\rangle$ is optimal in this technical sense, then the choice of F is the optimal way to express M such that both speaker's and interpreter's preferences are matched.

The OT–mechanism which allows to calculate the optimal form–meaning pairs does not make any reference to knowledge, or perspectives of participants. In fact, it presupposes that all the inferences are part of the common ground. But in a normal dialogue situation the participants have only partial knowledge about the described situation and about each other. The following example shows that this poses some problems for bidirectional OT. It was first discussed by J. Mattausch (2000, pp. 33–36).

(3) Assume that Marion is a male person, and Jo a female one. The speaker wants to express with the second sentence that Jo was pulling Marion's hair out:

a) Marion was frustrated with Jo. She was pulling his hair out.

b) Marion was frustrated with Jo. He was pulling her hair out.

c) Marion was frustrated with Jo. Jo was pulling Marion's hair out.

Intuitively, c) is the right way to put it. We assume that pronouns have to agree with the natural gender of the person referred to, and that the hearer prefers an interpretation where *Marion* is female and *Jo* male. These constraints lead into a circle: The speaker starts with the meaning *pulling–hair–out*(*Jo*, *Marion*), hence, he has to choose the form *She was pulling his hair out.* The hearer will interpret this form according to his preferences as *pulling–hair–out*(*Marion*, *Jo*). But this content should be expressed by the speaker as *He was pulling her hair out.* For this form the hearer should prefer the interpretation *pulling–hair–out*(*Jo*, *Marion*). And here the circle closes. We never reach a situation where speaker and hearer will always choose the same form and meaning. This means that bidirectional OT can't provide for an optimal form–meaning pair, and if the speaker wants to communicate that *Jo was pulling Marion's hair out*, then it fails to predict that exactly this sentence is the optimal one.

[3]We describe the procedure which provides for a *strong z–optimal* form–meaning pair. (Blutner, 1998, 2000) introduced in addition *weak* optimality, also called *superoptimaliy*, see (Jäger, 2000, p.45).

Examples **(2)** and **(3)** show that world knowledge and expectations about the world enter into anaphor resolution. If there is more than one possible resolution, then it can be made unique by *accommodation* of expected facts. We make this idea precise in Section 2. It has soon be noted that optimal form–meaning pairs can be seen as *Nash equilibria* in the sense of game theory[4]. I.e. one can look at the situation as a problem of rational choice where the speaker has to choose the best form and the hearer the most preferred meaning. Then, optimal form-meaning pairs are the possible candidates which rational agents can agree to choose. Hence, we can look at the interpretation problem as a *co-ordination problem*. It is solved, if speaker and hearer can make sure that it is *common knowledge* that they both get the same interpretation for an asserted natural sentence. This move allows us to make use of theories about co-ordination and knowledge in multi–agent systems. In Section3 we define a simple framework for our examples. In Section 4 we show that the co-ordination problem is always solved if the interlocutors adhere to the rules of semantics and a number of pragmatic constraints.

2 Interpretations and Expectations

All our examples are assertions, and we assume that it is the goal of an assertion to inform the interpreter that ψ is the case for some formula ψ chosen by the speaker. Let $\mathcal{L}$ be a first order language which contains representations for all predicates the interlocutors can use to talk about a described situation, and NL the set of sentences of a natural language. Let $\mathcal{C}$ be a set of *contexts*. We assume that there are two structures which define the semantics of $\mathcal{L}$ and NL:

$\langle \mathcal{C}, \mathcal{L}, \models \rangle$ defines the *static* semantics for $\mathcal{L}$ in the usual way.

$\langle \mathcal{C}, NL, \mathcal{L}, ^* \rangle$ where $^* : \mathcal{C} \times NL \longrightarrow \mathcal{L}$ translates natural sentences into formulas.

The contexts should contain enough information to make the translation unique. E.g. it should always be clear which variable the interpreter must choose for a pronoun, if he has full knowledge about the situation. A context c divides into three components: Two for the interlocutors, and one for the *environment* including the situation talked about. This means that a context is of the form $\langle e, c_S, c_H \rangle$, where e denotes the state of the environment, c_S the state of the speaker, and c_H the state of the hearer. Basically, the information states of speaker and hearer are sets of possible worlds together with a partial assignment function for the free variables. We assume that sentences with anaphoric NPs translate into formulas where the argument position for this NP is filled with a variable which is already interpreted. Normally, the set of epistemically possible contexts will contain more than one dialogue situation. But this implies that the set of possible translations for a natural sentence F may contain different formulas φ for different contexts, i.e. the translation is underspecified. The first sentence of Example **(3)**, *Marion was frustrated with Jo*, restricts the possibilities to the set of all world–assignment pairs where a formula of the form $\mathit{frustrated\text{–}with}(x, y) \,\&\, \mathit{Marion}(x) \,\&\, \mathit{Jo}(y)$ is true. This means that no information with respect to the sex of Marion and Jo is added. Hence, in some possible contexts the pronouns *she* and *he* translate into the variables x and y for Marion and Jo, in others into y and x. It is common knowledge that the models where Marion is female and Jo male are highly preferred. In such a situation, we assume that the use of the

[4](Dekker & v. Rooy, 2000)

pronouns *she* and *he* by the speaker triggers an *accommodation* of $female(x)\,\&\,male(y)$. What was after the first sentence only a defeasible expectation becomes thereby part of the common ground. If this is correct, then the versions **(4)**b) and **(5)**b) should be better because in the a) versions the third sentence contradicts the information which must be accommodated in order to interpret the second one.

(4) **a)** The doctor kissed the nurse. She is beautiful. The doctor there is a woman.

b) The doctor kissed the nurse. The doctor there is a woman. She is beautiful.

The same holds for cross speaker anaphora.

(5) *A* was told that a doctor kissed a nurse. He has no evidence whether the doctor is male or not. *B* knows that.

a) *A*: C told me that the doctor kissed a nurse. *B*: Did C tell you her name? All doctors there are woman.

b) *A*: C told me that the doctor kissed a nurse. *B*: All doctors there are woman. Did C tell you her name?

More generally, this means that:

If it is common knowledge

1. that the interpreter can find possible contexts where a natural sentence F translates into a formula φ_1 and contexts where it translates into a different formula φ_2,

2. that the (defeasible) expectations based on common knowledge imply a fact χ such that after an update with χ only the contexts remain where F translates into φ_1, and

3. that the speaker knows whether this fact χ holds

then the assertion of F triggers the accommodation of the fact that only φ_1 is a possible translation.

That it must be common ground that the speaker knows whether χ can be seen in:

(6) *A*: Marion was pulling Jo's hair out. *B*: Why did he do that?

If *A* knows that Marion is a boy then *A* should relate the pronoun *he* to him only if he knows that *B* knows this fact too. Example **(2)** shows that the expectations must be based on *common knowledge*. If **(2)** takes place in a hotel which is regularly frequented by German and Italian guests, then *B* can't use *she* or *the woman* even if *B* knows that Andrea and Maria are an Italian couple.

The interpreters preferences reflect *nonmonotonic reasoning*. We represent them by a function epc which provides for each information state σ a set $\mathrm{epc}(\sigma)$ of *preferred models*[5]. Hence, if σ represents the information state of the hearer, then $\mathrm{epc}(\sigma)$ represents his expectations. If the speaker asserts a sentence F and if F can't get a unique translation for all contexts in σ but one for all contexts in $\mathrm{epc}(\sigma)$, then this triggers under the conditions stated above an update with $\mathrm{epc}(\sigma)$.

[5]For the basic connections between nonmonotonic reasoning and preferences we can refer to introductory books on nonmonotonic reasoning, e.g. (Brewka et al., 1997).

These considerations show how defeasible expectations define a preference relation $\preceq_{c,F}$ on the set of possible translations of a form F in a context c. This provides for the connection with bidirectional OT. There, these preferences then interact with the speaker's preferences on forms such that they choose optimal form–meaning pairs. But Mattausch's Example shows that the search mechanism for (z–optimal) form–meaning pair leads into difficulties in cases where the interpreter has only partial knowledge about the actual context. Hence, we have to explain how the co-ordination works. We look at this problem as a co-ordination problem for a joined project. Here, make use of an idea which was put forward by *H.H. Clark*. According to Clark (1996, pp. 140-153) a communicative act comes in a hierarchy of joined actions, a so-called *action ladder*. He distinguishes four levels, where we are especially interested in the two highest levels. At the lower of the two levels (level 3) the speaker presents a signal, and the hearer has to recognise it. For our examples this means that the speaker presents a sentence of *natural language* which is a signal for some formula φ, and the hearer has to recognise this formula. We call this level the *interpretation level*. At the higher level (level 4) the speaker proposes a mutual update, and the hearer has to uptake this project. In our case this means that the speaker proposes the joined update with φ. We call this level the *update level*. Success at the higher level implies success at the lower level. But the co-ordination problem for the two levels can in general not be solved independently from one another. We describe a joined project by a multi–agent system together with a joined goal. Hence, we have to define two systems for each level, and show how they are related to one another.

3 Dialogues as Multi–Agent Systems

We think of a dialogue situation as an example for what is known as a *multi–agent system*. A multi–agent system[6] consists of the following components

1. A set $\mathcal{C}$ of *global states*.
2. A set ACT of possible dialogue acts.
3. A function P which tells us which dialogue acts can be performed in which dialogue situations. Hence, $P : \mathcal{C} \longrightarrow \mathcal{P}(\mathrm{ACT})$.
4. A (partial) transition operation τ with domain $\{\langle \mathtt{act}, c\rangle \mid \mathtt{act} \in P(c)\}$ and values in $\mathcal{C}$. It models the effect of the performance of dialogue acts.
5. A set of initial dialogue situations $\mathcal{C}_0$.

We identify dialogues with sequences $D = \langle c_0, \mathtt{act}_0, \ldots, \mathtt{act}_{n-1}, c_n\rangle$ where c_0 is an initial dialogue situation, and:

- $\mathtt{act}_i \in P(c_i)$, i.e. $\mathtt{act}_i$ is possible in c_i.
- $c_{i+1} = \tau(\mathtt{act}_i, c_i)$.

We will confine our considerations to minimal exchanges, hence, we assume that all D are of the form $\langle c_0, \mathtt{act}, c_1\rangle$. We denote the set of all dialogues by $\mathcal{D}$.

It is usual to identify the knowledge of an agent in an multi-agent system with the set of all global states which are *indiscernible* from the actual state. It is assumed that two global states

[6]Our presentation of multi–agent systems closely follows (Fagin e.al., 1995).

are indiscernible for an agent X, iff his local states are identical. This is essentially a possible worlds approach. We don't want to call it *knowledge* but more neutrally *information* what we represent in this way. This means that we should identify the information of a participant X in a context c with the set of all contexts c' where the local state c_X is identical with c'_X. But this would mean that we must represent all necessary information about the history of the actual dialogue in the local states. For example, we would have to represent all former local states and all utterances of the speaker. This is not a principle problem but it leads to cumbersome representations. Instead we put this information into the indiscernability relation. I.e. a participant X should not be able to discern dialogues D and D' where the sequence of his local states and the publicly performed acts are the same. This induces an equivalence relation on dialogues. We assume that all dialogue acts divide into a speaker's and a hearer's act, where only the first one becomes public knowledge. We represent every act as a pair $\langle \mathtt{act}_S, \mathtt{act}_H \rangle$. Hence, we identify the *information of agent* X after dialogue $D = \langle c_{D,0}, \mathtt{act}_D, c_{D,1} \rangle$ with:

$$I(X, D) := \{D' \in \mathcal{D} \mid \mathtt{act}_{D,S} = \mathtt{act}_{D',S} \,\&\, \forall i = 1, 2 \; c_{D,i,X} = c_{D',i,X}\},$$

where $c_{D,i,X}$ denotes the local state of X in $c_{D,i}$, and where $\mathtt{act}_D = \langle \mathtt{act}_{D,S}, \mathtt{act}_{D,H} \rangle$. This leads directly to the following representation of the *common information* $\mathrm{CI}(D)$ after Dialogue D: Let $M^0 := \{D\}$, $M^{n+1} := \bigcup\{I(X, D') \mid X = S, H \,\&\, D' \in M^n\}$, and

$$\mathrm{CI}(D) := \bigcup_{n \in \mathbf{N}} M^n.$$

If M characterises some property of dialogues, i.e. if $M \subseteq \mathcal{D}$, then it will be *common information* that the actual Dialogue D has this property, iff $\mathrm{CI}(D) \subseteq M$.

We now want to be more precise about the contexts which represent a dialogue situation. We use here a framework which is known as *possibility approach*[7]. We are not interested in the explicit representations of an agents knowledge. Therefore, we will identify his beliefs with a set of epistemic possibilities. We confine our considerations to the case where there are only two participants S and H. The outer situation may contain information about the immediate environment but most importantly it provides information about a situation talked about.

Let W be a set of models for $\mathcal{L}$ representing the possible states of affairs. In dynamic semantics it is usual to identify an information state with a set of world–assignment pairs $\langle w, f \rangle$, where f is a *partial* function from the set of variables into the set of objects in w. The arguments of f are intended to represent *discourse referents* introduced in the common ground. We represent dialogue situations as structures of the form $w = \langle (s_w, f), (\psi_w, w(S)), (\Sigma, w(H)) \rangle$. We call them *possibilities*, and denote the set of all possibilities by $\mathcal{W}$[8]:

- (s_w, f) is a world–assignment pair,
- $(\psi_w, w(S))$ is the speaker's state where $w(S)$ is a subset of $\mathcal{W}$, and ψ_w represents the speaker's *goal*. We represent it by a formula in $\mathcal{L}$.
- $(\Sigma(w), w(H))$ is the hearer's state where $w(H)$ is a subset of $\mathcal{W}$, and $\Sigma(w)$ represents a place where he can store his interpretation of a sentence. Hence, $\Sigma(w)$ may be empty or contain a single $\mathcal{L}$–formula.

[7](Gerbrandy & Groeneveld, 1997)

[8]The following definition can be made precise in (AFA) set theory (Aczel, 1988; Barwise & Moss, 1996).

The information states $w(S)$ and $w(H)$ are sets of possibilities, hence, they again represent information about each other. We assume that there are some restrictions on the class of possibilities. First, that all information states are non–contradictory, that *full introspection* holds, i.e. that all participants know what they believe, and that this is common knowledge. We denote the class of all these possibilities by $\mathcal{I}$. More precisely, let $w = \langle (s_w, f^w), (\psi_w, w(S)), (\Sigma(w), w(H)) \rangle \in \mathcal{I}$, then for $X = S, H$

(1) $w(X) \neq \emptyset$, (2) $\forall v \in w(X)\, v(X) = w(X)$, and (3) $w(X) \subseteq \mathcal{I}$.

Similarly we assume for goals and interpretations that the participants know them, and that this is common knowledge. Hence, it holds in addition

(4) $\forall v \in w(S)\, \psi_v = \psi_w$, and (5) $\forall v \in w(H)\, \Sigma(v) = \Sigma(w)$.

Then, our basic dialogue situations will always be situations where both dialogue participants have only *true* beliefs, and where this is common knowledge. We denote the class of all such situations by $\mathcal{T}$. Hence, if $w \in \mathcal{T} \subseteq \mathcal{I}$, then

(1) $w \in w(S) \cap w(H)$, and (2) $w(S) \cup w(H) \subseteq \mathcal{T}$.

As the domains of the assignment functions should represent discourse referents in the *common ground*, we assume furthermore that for $w \in \mathcal{T}$

(3) $\forall v \in w(S) \cup w(H)\, \mathrm{dom}\, f^w = \mathrm{dom}\, f^v$.

As a last condition we make a *sincerity* condition for the speaker's goal ψ_w. We assume that the goal of an assertion is to inform the hearer about some fact about s_w. ψ_w represents this fact. We confine our considerations to cases where the speaker does not mislead the hearer. Hence,

(4) $\forall v \in w(S)\, (s_v, f^v) \models \psi_v$.

Let M be any subset of $\mathcal{W}$. If a participant learns that the world is an element of M, then he can eliminate all his epistemic possibilities which are not elements of M. If the participants mutually learn this fact, then they should mutually restrict their belief states to this set. In the following we want to make our life easier and assume that the new assertion does not introduce new discourse referents. Hence, we really can represent the update as an eliminative update. Let w be a possibility. Then the updated possibility $M(w)$ is defined by[9]:

1. $M(w)(X) = \{M(v) \mid v \in w(X) \cap M\}$, $X = S, H$.
2. $v = M(w) \Rightarrow (s_v, f^v) = (s_w, f^w) \wedge \psi_w = \psi_v \wedge \Sigma(w) = \Sigma(v)$.

4 Co-ordination of Interpretation

We make the idea precise that dialogue exchanges are joined projects, and that they are organised in an action ladder. We distinguish two levels, the interpretation and the update level. A multi–agent system describes the possible acts and their effects. A joined project

[9]For the mathematics behind this definition we refer to (Barwise & Moss, 1996), or (Gerbrandy, 1998).

is defined by the task to reach a common goal. We represent such a goal by a set G of possibilities, where the goal is reached if the agents choose their actions in such a way that their effects lead to a situation in G. It is not necessary that both interlocutors *know* the set G. We describe the levels as separate joined projects, i.e. as two multi–agent systems with own common goals.

The Interpretation Level

We consider only dialogues with assertions. We can identify the set of possible actions the speaker can perform with the set of all natural sentences and the corresponding acts of the interpreter with their associated interpretation as formulas in $\mathcal{L}$. ACT_I represents all possible joined actions on the interpretation level. Hence,

$$\mathrm{ACT}_I := \{\langle F, \varphi\rangle \mid F \in NL \,\&\, \exists c\, \varphi = {}^*(c, F)\}.$$

The interpretation level is intended to represent the system defined by the pure semantics of the language NL. Hence, the speaker is allowed to assert F in a given context c, iff its translation is true, i.e.:

$$P_I(c) := \{\langle F, \varphi\rangle \in \mathrm{ACT}_I \mid c \models \varphi \,\&\, \varphi = {}^*(c, F)\}.$$

Both participants should know in the resulting state that F has been uttered. We represented this information in the indiscernability relation on dialogues. If the hearer interprets F as φ, then he should store φ in the set Σ. We assume that in the initial state this set Σ is empty. Hence, $\mathcal{C}_{0_I} = \{w \in \mathcal{T} \mid \Sigma(w) = \emptyset\}$. Then $\Sigma(\tau_I(\langle F, \varphi\rangle, w)) = \{\varphi\}$. We get:

$$\mathrm{MAS}_I = \langle \mathcal{T}, \mathrm{ACT}_I, \tau_I, P_I, \mathcal{C}_{0_I}\rangle.$$

This describes the possible assertions and their interpretations. In order to define a joined project we need a common goal. It is the task at this level to interpret the asserted natural sentence. This aim is reached for F in w if the resulting state is an element of a set

$$G_{F,w} := \{v \in \mathcal{T} \mid \Sigma(v) = \{\varphi_{F,w}\}\}, \text{ where } \varphi_{F,w} \text{ is the translation of } F \text{ in } w.$$

This goal is always reached by definition of the rules in MAS_I.

The Update Level

At the update level both interlocutors should mutually update with a meaning $[\![\varphi]\!]$. More generally, we can assume that they update with some information state $M \subseteq \mathcal{T}$. Hence, we set $\mathrm{ACT}_u := \{M \mid M \subseteq \mathcal{T}\}$. We consider only updates where the real situation supports the new information. Hence, $P_u(w) := \{M \in \mathrm{ACT}_u \mid w \in M\}$. The transition operation is defined by the mutual updates for possibilities : $\tau_u(M, w) = M(w)$. We set $\mathrm{MAS}_u := \langle \mathcal{T}, \mathrm{ACT}_u, \tau_u, P_u, \mathcal{T}\rangle$.

Now, it is easy to see how the solution for the co-ordination problem on the update level depends on the solution of the co-ordination problem on the interpretation level. If it is common knowledge that an asserted sentence F has to be interpreted by a formula φ, then both dialogue participants should update with $[\![\varphi]\!] = \{w \in \mathcal{T} \mid (s_w, f^w) \models \varphi\}$. This defines the combined system of MAS_u and MAS_I.

As φ is determined by F and w we can identify the action $\langle F, \varphi_{F,w}\rangle$ for $\varphi_{F,w} = {}^*(w, F)$ with F. This means that we can describe the induced update multi–agent system by:

$$\begin{aligned} \mathrm{ACT}_u &= \{F \in NL \mid \exists w \in \mathcal{T}\, w \models \varphi_{F,w}\}, \\ P_u(w) &= \{F \mid w \models \varphi_{F,w}\}, \\ \tau_u(F, w) &= [\![\varphi_{F,w}]\!](\tau_I(F, w)). \end{aligned}$$

We denote this system by $\mathrm{MAS}_u(I)$, and call it the *induced* update system.

The goal connected to an assertion is to inform the hearer that ψ for some $\psi \in \mathcal{L}$. This aim is reached for ψ if the resulting state is an element of a set $G_\psi := \{v \in \mathcal{T} \mid v(H) \subseteq [\![\psi]\!]\}$. But this goal is identical with the speaker's goal and does not need to be known to the interpreter.

Pragmatic Constraints

The rules for use and interpretation in the multi–agent systems MAS_I and $\mathrm{MAS}_u(I)$ are directly defined by the underlying semantics. Hence, they don't account for any pragmatic constraints. We will show that the co-ordination problem is solved if we assume that the interlocutors adhere to a number of pragmatic constraints. Of course, we want to have a minimal number of such constraints. First, we assume that it is rational to perform an action relative to a public goal, if the resulting state belongs to the set of desired situations. This means, an action $\mathtt{act} \in \mathrm{ACT}$ is rational in context w relative to a goal $G(w)$ represented as a subset of $\mathcal{T}$, iff

$(\mathbf{A_1})$ $\tau(\mathtt{act}, w) \in G(w)$.

These constraints hold on the interpretation level by definition. But they impose new restrictions on the update level. Then, we assume that the interpreter must know how to perform his task, i.e. he must know which action to choose. Hence, if the speaker asserts some sentence F, then there should be only one possible interpretation for his set of epistemic possibilities. This means for the interpretation level that[10]:

$(\mathbf{A_2})$ $\langle F, \varphi\rangle \in P_I(w) \Rightarrow \exists!\varphi\, \exists v \in w(H)\ \langle F, \varphi\rangle \in P_I(v)$.

Let $P^{-1}(F, \varphi) := \{w \in \mathcal{T} \mid \langle F, \varphi\rangle \in P(w)\}$. This denotes the set of all situations where the joined action $\langle F, \varphi\rangle$ is defined. The pragmatic conditions impose additional restrictions on these sets. We want to provide for an explicit description. Therefore, we introduce the following operators. We define them for the participant S and H. They are closely related to the modal operators $\Box_X$ and $\Diamond_X$ and we denote them by the same symbols:

$$\begin{aligned} \Box_S M &:= \{w \in \mathcal{T} \mid w(X) \subseteq M\} \\ \Diamond_H M &:= \{w \in \mathcal{T} \mid w(X) \cap M \neq \emptyset\} \end{aligned}$$

These operators mean: S is convinced that the actual world w belongs to M iff $w \in \Box_S M$; H believes that it is possible that w belongs to M iff $w \in \Diamond_H M$. Let w be given with the common goal $G_\psi = \{w \in \mathcal{T} \mid w(H) \subseteq [\![\psi]\!]\}$. Then, it is provable that a joined action $\langle F, \varphi\rangle$ belongs to $P_I(w)$, and the axioms $(\mathbf{A_1})$ and $(\mathbf{A_2})$ hold for $\langle F, \varphi\rangle$, if w belongs to the intersection of the following to sets:

[10] where $\exists!$ means *there exists exactly one.*

$$M_u := [\![\varphi]\!] \cap [\![\psi]\!] \cap \Diamond_H([\![\varphi]\!] \cap [\![\psi]\!]) \setminus \Diamond_H([\![\varphi]\!] \cap [\![\neg\psi]\!]).$$

$$M_I := P_I^{-1}(F,\varphi) \cap \Diamond_H(P_I^{-1}(F,\varphi)) \setminus \Diamond_H \bigcup_{\varphi' \not\equiv \varphi} P_I^{-1}(F,\varphi').$$

M_u relates to the update level and M_I to the interpretation level. They are determined by F, φ and ψ. Hence, we can write $B(F,\varphi,\psi) := M_u \cap M_I$. Now, we add as a last constraint: The speaker must know that $\langle F,\varphi\rangle$ is a possible joined act, and that $(\mathbf{A_1})$ and $(\mathbf{A_2})$ hold, i.e. we postulate:

$(\mathbf{A_3})$ $\langle F,\varphi\rangle \in P(w) \Leftrightarrow w \in B(F,\varphi,\psi_w) \cap \Box_S B(F,\varphi,\psi_w)$.

This defines a new sub–multi–agent system MAS_P of $\mathrm{MAS}_u(I)$, where MAS_P is as $\mathrm{MAS}_u(I)$ except for $P_p(w)$

$$P_P(w) := \{\langle F,\varphi\rangle \mid w \in B(F,\varphi,\psi_w) \cap \Box_S B(F,\varphi,\psi_w)\}.$$

The central claim is, that the co-ordination Problem is always solved for this multi–agent system.

Lemma 1 *Let* MAS_P *be as above. Let* $\mathcal{D}_P$ *be the set of dialogues generated by* MAS_P. *Let* $w \in B(F,\varphi,\psi_w) \cap \Box_S B(F,\varphi,\psi_w)$. *Hence, there is an* $v \in \mathcal{T}$ *such that* $D := \langle w, \langle F,\varphi\rangle, v\rangle \in \mathcal{D}_P$. *Then:*

$$\forall D' = \left\langle w', \left\langle F',\varphi'\right\rangle, v'\right\rangle \in I(X,D) : \left\langle F',\varphi'\right\rangle = \langle F,\varphi\rangle \ \&\ w' \in B(F,\varphi,\psi_{w'}) \,\&\, v' \in G_{\psi_w} \cap G_{\psi_{w'}},$$

where $G_\psi := \{v \in \mathcal{T} \mid (s_v, f^v) \models \psi\}$. *It follows that:*

$$\mathrm{CI}(D) \subseteq \{D = \langle w, \langle F,\varphi\rangle, v\rangle \in \mathcal{D}_P \mid v \in G_{\psi_w}\}.$$

$\{D = \langle w, \langle F,\varphi\rangle, v\rangle \in \mathcal{D}_P \mid v \in G_{\psi_w}\}$ is the set of all dialogues where the speaker's goal, which is also the common goal, is reached. Hence, the last equation means that it is also common information that the interlocutors reached the goal successfully.

If the translation of a form is underspecified, then the joined project on the interpretation level fails. If it can be made unique by accommodation of expected facts, then this accommodation leads to a situation where $(\mathbf{A_2})$ holds. Hence, the hearers preferences on meanings enter at the interpretation level and allow for the use of forms in cases where pure semantics would not guarantee a unique translation.

Mattausch's Example Reconsidered

Mattausch's Example **(3)** shows that we must consider the knowledge of interlocutors if we want to determine optimal form–meaning pairs in general. The first sentence of the example, *Marion was frustrated with Jo*, restricts the possibilities to the set of all world–assignment pairs where a formula of the form *frustrated–with*(x,y) & *Marion*(x) & *Jo*(y) is true. In some contexts the pronouns *she* and *he* translate into the variables x and y for Marion and Jo, in others into y and x. But this means that the use of the pronouns *she* and *he* violates $(\mathbf{A_2})$ and the joined project on the interpretation level can't succeed. This is common knowledge, and we argued in Section 2 that this triggers an *accommodation* of *female*(x) & *male*(y) because then we are in a situation where $(\mathbf{A_2})$ holds for $\varphi_1 \equiv$ *pulling-hair-out*(x,y). If we would assume that this formula is true and that the speaker wants to say that *Marion was pulling*

Jo's hair out, then the conditions of Lemma 1 hold, and the use of pronouns should be successful. Then the speaker's preferences for economic forms considered in OT should in fact lead him to prefer them instead of repeating names.

But, the assumption is that Marion is male and Jo female. Hence, the conditions of $(\mathbf{A_3})$ are not met because the real situation is not an element of $[\![\varphi_1]\!]$, and also because the speaker knows this. This means that *She was pulling his hair out* does not meet the pragmatic conditions of MAS_P. For the same reasons *The girl was pulling his hair out* is not a possible choice. Hence, only *Marion was pulling Jo's hair out* remains as desired. Of course, we have to check the conditions. The sentence has only one translation, and the interpreter knows that there is a possibility where this sentence is true, hence, $(\mathbf{A_2})$ is met and the speaker knows this too. This guarantees that the co-ordination problem on the interpretation level is solved. It is also easy to check that the conditions hold which guarantee successful co-ordination at the update level.

5 Summary

Partiality of knowledge poses some problems if we want to explain how interlocutors manage to co-ordinate their choice of form–meaning pairs. Bidirectional OT assumes that preferences of speaker and hearer play here an essential role. Following (Beaver, 2000) and (Mattausch, 2000) we looked at examples for anaphora resolution. Here, the considered forms are sentences of natural language and the meanings are their translations into a formal language. Our basic move was to consider an assertion as a joined project. Following (Clark, 1996) we divided this project into two dependent subprojects. We could show that pure semantics plus some pragmatic conditions always guarantee that it is mutual knowledge that these projects are successful. At one level, the interlocutors have to agree on the translations of uttered sentences. On another level, they have to reach the conversational goal by a mutual update. We considered examples where world knowledge, and expected (defeasible) facts about the world define the preferences of the hearer for translations. These enter at the interpretation level. Here, expected facts were accommodated if this was needed to make an interpretation task unambiguous. For no example we needed to consider the speaker's preferences for economic forms on the two levels of the joined project. They entered only in order to explain the choice between forms where the successful interpretation was already guaranteed.

Bibliography

P. Aczel (1988): *Non–Well–Founded Sets;* CSLI–Lecture Note 14, Stanford.

J. Barwise, L. Moss (1996): *Vicious Circles;* Stanford, CSLI,.

D. Beaver (2000): *The Optimization of Discourse;* ms, Stanford.

G. Brewka, J. Dix, K. Konolige (1997): *Nonmonotonic Reasoning;* CSLI, Stanford.

R. Blutner (1998): *Lexical Pragmatics;* Journal of Semantics 15, pp. 115–162.

R. Blutner (2000): *Some Aspects of Optimality in Natural Language Interpretation;* to appear in *Journal of Semantics.*

R. Blutner, G. Jäger (2000): *Against Lexical Decomposition in Syntax;* to appear in A.Z. Wyner (ed.): *Proceedings of IATL 15*, University of Haifa.

H.H. Clark (1996): *Using Language;* Cambridge.

P. Dekker, R. v. Rooy (2000): *Optimality Theory and Game Theory: Some Parallels;* to appear in *Journal of Semantics.*

R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi (1995): *Reasoning About Knowledge;* MIT–Press, Cambridge, Massachusetts.

J. Gerbrandy, W. Groeneveld (1997): *Reasoning about Information Change;* Journal of Logic, Language and Information 6, pp. 147–169,.

J. Gerbrandy (1998): *Bisimulations on Planet Kripke;* ILLC Dissertation Series, Institute for Logic, Language and Computation, Universiteit van Amsterdam.

G. Jäger (September 2000): *Some Notes on the Formal Properties of Bidirectional Optimality Theory;* ms, ZAS Berlin.

J. Mattausch (November 2000): *On Optimization in Discourse Generation;* master thesis, Univeriteit van Amsterdam.

H. Zeevat (2000): *Semantics and Optimality Theory;* H. de Hoop, H. de Swart (eds.): *Optimality Theoretic Semantics*, OTS preprint, University of Utrecht.

Paper 3

Towards Understanding the Role of Hints in Tutorial Dialogs

Helmut Horacek, Armin Fiedler
Universität des Saarlandes, Postfach 151150
D-66041 Saarbrücken, Germany
{horacek|afiedler}@cs.uni-sb.de

Abstract

Major driving forces underlying dialogs have been examined from a variety of perspectives, including, among others, elaborations of the semantics underlying several kinds of speechacts. For one such speech act, hints, the most crucial dialog contributions in tutorial sessions, no adequate formal treatments of their underlying semantics has been offered so far. Motivated by the currently low degree of understanding about the nature of hints in tutorial dialogs, we attempt to capture major aspects of the semantics of hints, comprising their contextual embeddings and effects on the course of the dialog. Some of the concepts driving the effectiveness of hints in such an environment are increasing degrees of accuracy, varying perspectives, and the provision of partial enablements on behalf of the student in a tutorial session. We intend to exploit the insights gained for the generation of suitable hints in a learning environment for theorem proving techniques in certain areas of mathematics.

1 Introduction and Motivation

Major driving forces underlying dialogs have been examined from a variety of perspectives, including, among others, elaborations of the semantics underlying several kinds of speechacts. For one such speechact, hints, the most crucial dialog contributions in tutorial sessions, no adequate formal treatments of their underlying semantics has been offered so far. In our view, the ignorance of hints in formal treatments of dialog acts is a consequence of the concentration on task-oriented conversations, where aspects specific to tutorial dialogs are irrelevant. In particular, information required by one conversant would be provided in a straightforward manner, or some subtask would be carried out by the more informed conversant rather than attempting to assist the less informed conversant indirectly by well-chosen hints.

Motivated by the currently low degree of understanding about the nature of hints in this context, we envisage an examination of their properties and functionality. We attempt to capture major aspects of the semantics of hints, which we intend to exploit for the generation

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

of suitable hints in task-oriented tutorial sessions. Our domain of application is learning of theorem proving techniques, elaborated for certain areas of mathematics, as advocated in the ambitious architecture in Melis and Horacek (2000).

This paper is organized as follows. We provide an informal characterization of the properties of hints. Then we explain how we intend to arrive at a sufficiently detailed formal account of the semantics of hints. Finally, we discuss how we plan to exploit these insights for generating appropriate hints in a tutorial session.

2 Properties of Hints

In order to examine the properties on an informal basis, let us consider the following situation from our domain, which is an excerpt from a simulated tutorial session. In this dialog fragment, applying the definitions of set equality and subset is required (the theorem to be proved states that A being a subset of B and B being a subset of A implies that A and B are equal):

System: What does set equality mean in our example?

User: A and B should be equal.

System: OK, and when are A and B equal?

User: They are equal if A is a subset of B and B a subset of A.

System: That is the problem definition. What does it mean that A is a subset of B, resp. B a subset of A.

User: That set A is contained in set B and vice versa.

System: Apply the definition of subsets.

This short dialog fragment illustrates some of the properties of hints:

- *Surface forms*

  Hints may appear in any surface; compare the alternative phrasings "the definition of subsets is relenvant", "apply the the definition of subsets", and "what is the definition of subsets?", which can all be considered as similar hints in a tutorial session about mathematical sets and related topics. In addition, some of these forms may entail superficial paradoxons, such as the question form, which presupposes the asking person knows the answer precisely when using the question as a hint.

- *Reference to pieces of knowledge*

  Hints may address some piece of knowledge required for task accomplishment in its generic form, in degrees of partial or full instantiation, and also in alternative forms. A suitable repertoire of hints comprises different perspectives and (usually increasing) degrees of precision.

- *Effectiveness of hints*

  Effectiveness is harder to judge for hints, in comparison to direct information conveyance, where shortness of a dialog seems to be a suitable indication for effectiveness. Longer dialog sequences, though reaching some discourse goal slower, may have an additional learning effect on behalf of the hint-receiving conversant, especially for persons with less trained domain skills. Moreover, hints are typically more likely to fail—at least, temporarily—than other dialog acts.

In the following, we examine these properties in some more detail.

3 Towards Capturing the Semantic of Hints

To characterize hints in task-oriented tutorial sessions, we consider two aspects to be the crucial properties worth considering (in accordance with other work in the area of dialogs, comprising theoretical Cohen and Levesque (1990) and related practical approaches Sadek et al. (1997)):

- the propositional attitudes of the agents involved
- the contextual embedding made up by the problem solving task

At the current state of our work, we are not yet able to give a formal account of the semantics underlying hints at a reasonable degree of accuracy. However, based on the example given and on other examinations and discussions we had so far, we have the following conjectures about the propositional attitudes of the agents involved:

- If p is a piece of information referred to by a hint, then the tutor generally assumes that $KNOW(student, p)$ holds or, at least, believes this to be very plausible. In some cases (as in our dialog fragment above), conveying p is embedded in a question, which makes the hint a test whether the student indeed knows the required piece of information.
- However, there is generally no point in merely addressing information that is known to the audience without some other purpose behind. Hence, the point in providing this piece of information p must lie elsewhere, and we think it serves the intention of the tutor that the student recognizes the *relevance* of p with respect to the problem solving context C or some part of it currently focused in the tutorial session.

However, the relation between p and C can only be captured when considering the situational context. Regarding this aspect, we believe to have identified some crucial properties and concepts with respect to the contextual embedding:

- *vagueness in expressing the relation* between p and C
- *increasing precision* in which the relation is reexpressed in case the student fails to make progress

Typically, the information p addressed by a hint is embedded in some larger activity P, which can be broken down into nested sequences p_1, p_2,..., p_n of subactivities, as in the theory of *shared plans* Grosz and Kraus (1999). In several domains including ours, individual

steps p_i can be richer actions whose substructures cannot be broken down into independently treatable subactions, as argued in Horacek (2000). In the domain of mathematics, an individual inference step, such as applying a definition comprises its terminological expansion and eventually non-trivial substitutions, which should not be separated unless there is explicit focus on one of these. A hint in such a situation may then address such an inference step in some but usually not all of its aspects.

The motivation for producing dialog contributions with these properties lies in the purpose underlying good tutorial sessions: the student should be given a minimal amount of help, depending on the behavior and skill exhibited, so that he successively recognizes and constructs the relations between known information and the problem solving context in the accuracy needed to accomplish the associated task. This comprises instantiation of generic knowledge and decomposition of vaguely articulated relations into conceptually or operationally managable relations.

4 Generating Hints

Dialog capabilities are considered a major weakness of today's tutorial systems, as opposed to the work done by human tutors Moore (2000). Within tutorial dialogs, *hints* for supporting the student in discovering a solution to some given problem by himself play a central role. The state of affairs regarding the problem solving context, the evidence or assumptions about knowledge, and the experience of the student should provide a source for determining an appropriate level of specification for a hint to be produced. In that respect, the semantics of hints should contribute to answer questions about their usage:

- Which hints are sensible in some situation; in particular, when certain hints have already been employed in that session?
- Which of the applicable hints can be expected more effective than others in view of assumptions about the audience within a given context?
- Under what circumstances are hints less useful than other tutorial techniques, that is, when is a strategy change indicated?

According to our current understanding, major ingredients in actually composing hints are:

- The discrepancy between a model of the problem solving course, including variations over individual steps, and pieces of knowledge and information attributed to the audience. A hint should address one of the gaps assessed to exist between the student's view and the completed model. If a gap seems to be too large, breaking the underlying relation down into conceptually better managable pieces should be addressed by a hint.
- The current focus in the tutorial session, which should be maintained as long as some progress in this part can be expected. At the beginning, this must be established through some systematic strategy—for example, top down (in our domain, starting from the theorem), or bottom up, (in our domain, starting from the premises).

In the course of a project we are currently investigating, we intend to address these issues.

Bibliography

Cohen, P. and Levesque, H. (1990). A theory of rational action and interaction. In Berwick, R. C., editor, *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics (ACL-90)*, Pittsburgh. Association for Computational Linguistics.

Grosz, B. and Kraus, S. (1999). The evolution of shared plans. In Rao, A. and Wooldridge, M., editors, *Foundations of Rational Agency*, pages 227–262. Kluwer Acadamic.

Horacek, H. (2000). Tailoring inference-rich descriptions through making compromises between conflicting cooperation principles. *Int. J. Human-Computer Studies*, 53:1117–1146.

Melis, E. and Horacek, H. (2000). Dialog issues for a tutor system incorporating expert problem solvers. In Rosé, C. and Freedman, R., editors, *AAAI-2000 Fall Symposium — Building Dialogue Systems for Tutorial Applications*, Falmouth, MA.

Moore, J. (2000). What makes human explanations effective? In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pages 131–136, Hillsdale, NJ. Earlbaum.

Sadek, D., Bretier, P., and Panaget, F. (1997). Artimis: Natural dialogue meets rational agency. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, pages 1030–1035. Morgan Kaufmann.

Paper 4

Integrating Conversational Move Types in the Grammar of Conversation

Jonathan Ginzburg, Ivan A. Sag, and Matthew Purver
{ginzburg,purver}@dcs.kcl.ac.uk
http://www.dcs.kcl.ac.uk/{staff/ginzburg, pg/purver}
sag@csli.stanford.edu, http://www-csli.stanford.edu/~sag

Abstract

Analyses of dialogue that incorporate the insights of speechact theory presuppose that an utterance gets associated with a conversational move type (CMT). Due to difficulties that beset attempts to integrate CMTs into grammar in early generative work, as well as the perceived problems concerning multifunctionality, CMT information is typically not included in most formal grammatical analyses. We provide arguments as to why CMT does need to be integrated in grammatical analysis of conversation. We offer a proposal for such an integration couched in Head Driven Phrase Structure Grammar (HPSG). We sketch explanations as to why our proposal does not run into the foundational and empirical pitfalls that have beset previous proposals.

1 Introduction

Categorizing utterances in terms of a notion of illocutionary force or conversational move type (CMT) is common in corpus-based work (for some recently proposed CMT taxonomies, see Carletta et al. (1996), Core and Allen (1997)). Indeed any analysis of dialogue that incorporates the insights of speech act theory presupposes that an utterance ultimately gets associated with a CMT. Nonetheless, there exist few attempts to integrate such notions into contemporary formal grammatical work. In part, this is due to the fact that most grammatical formalisms to date have been designed with monologue or text in mind, where this issue is easier to put aside than in conversational settings. A more principled reason for this lacuna is perhaps the phenomenon of *multifunctionality* (see e.g. Allwood (1995)): it is often the case that a given utterance serves more than one purpose—an assertion can function also as an offer, a query as a suggestion etc. This has often led to the feeling that issues pertaining to CMT belong entirely to the realm of pragmatics. Although no worked out pragmatic theory as to how CMTs get assigned to utterances has emerged to date, the one influential series of

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

{ginzburg,purver}@dcs.kcl.ac.uk
http://www.dcs.kcl.ac.uk/{staff/ginzburg, pg/purver}
sag@csli.stanford.edu, http://www-csli.stanford.edu/~sag.

attempts to subsume CMT into the grammar, based on the Performative Hypothesis (PH) is generally viewed to have been a resounding failure (see Levinson (1983), pp. 247-263).

In this paper we argue that CMT can and should be integrated in the semantic analyses provided by the grammar. That is, CMT is a parameter of meaning conventionally associated with certain words and classes of phrases. For instance, in hearing an utterance by A of a sentence such as (1a), we claim that a competent interlocuter B knows that its meaning is the template schematically given as (1b), *not* simply the proposition (1c). That is, B knows that in order to ground A's utterance she must try to instantiate the parameters A, t, l, P within the template given in (1b) in such a way as to satisfy the constraints provided by the grammar (e.g. A must be the speaker, t must be a time the day after utterance time, P ranges over a set that includes {assert, threaten, promise,...}, but not over, for instance, {ask, exclaim, apologize,...})[1]

(1) a. A: I will leave tomorrow.

b. $P(A, B, leave(leaver : A, time : t, location : l))$

c. $leave(A, time : t, location : l)))$

The paper is structured as follows: we start by providing a couple of concrete arguments as to why CMT does need to be integrated in grammatical analysis of conversation. We then offer a proposal for such an integration couched in Head Driven Phrase Structure Grammar (HPSG). We sketch explanations as to why our proposal does not run into the problems associated with the PH, or other foundational and empirical pitfalls.

2 Motivation for integrating CMT in grammatical analysis

Although there are a variety of versions of the PH, they essentially boil down to positing that all (English) matrix sentences have the form *I illoc-verb S*, where *I* is the first person singular pronoun and *illoc-verb* is a verb from the class of performative verbs (e.g. *assert*, *ask*, *order*, *bet*, ...). For all matrix sentences which do not have this form overtly, the PH involves the assumption that the 'illocutionary prefix' *I illoc-verb* is not realized at the surface but is represented at some other syntactic level. In its formulations in the 1970s, at least, the PH ran into a variety of problems, the most serious of which revolved around the difficulty of maintaining a coherent definition of truth for declaratives. The difficulty arises from the parallelism that the PH enforces between sentences that lack an overt illocutionary prefix (e.g. (2a)) and explicit performatives (e.g. (2b)):

(2) a. Snow is black.

b. I claim that snow is black.

[1] *How* any of these values get instantiated, if indeed B manages to do so, can involve highly complex reasoning (involving e.g. domain-specific knowledge, reasoning about intentions etc) with which of course the grammar as such provides no assistance. However, the use of such reasoning to resolve the value of a constituent of content also affects constituents of content (e.g. tense and anaphora) that lie uncontroversially within the realm of semantics. Hence, this cannot be used as an argument against integrating CMT within grammatical analysis.

Such a parallelism is untenable because it either conflates the truth conditions of quite contingent sentences such as (2a) with those of (2b), which, essentially, become true once they are uttered. Alternatively, the parallelism requires a mysterious filtering away of the semantic effect of the illocutionary prefix. Despite the difficulties for the PH, we argue that in fact there are good reasons for assuming that the contents specified by the grammar do contain CMTs as a constituent. Our first argument concerns the existence of words that actually carry their CMT on their sleeve. Examples of such words are given in (3):

(3) a. [Context: A sees B as she enters a building] A: Hi.

b. [Context: A enters train carriage, sees B leave] A: Bye.

c. [Context: in a bus queue A slips and unintentionally pushes B] A: Sorry.

d. [Context: B is a bus conductor who gives A a ticket.] A: Thanks.

A competent speaker of English might paraphrase each of these utterances as in (4):

(4) a. A *greeted* B.

b. A *bid farewell* to B.

c. A *apologized* to B (*for having pushed her*).

d. A *thanked* B (*for giving her a ticket*).

This can be used as evidence that these words are associated with meanings schematized as in (5). In these representations, the main predicate constitutes the CMT associated with the utterance, whereas *m(es)s(a)g(e)-arg* indicates the semantic type of the propositional/descriptive content selected by the CMT. Note a contrast illustrated in (4): whereas both [the relations denoted by] *apologize* and *thank* select clausal complements (whose denotations) constitute the descriptive content, there is no such selection by *greet* and *bid-farewell.* This provides some of the motivation for assuming that these latter should not specified for a msg-arg, in other words that such speechacts have no descriptive content.

(5) a. Hi: **greet***(speaker,addressee,msg-arg:none)*

b. Bye: **bid-farewell***(speaker,addressee,msg-arg:none)*

c. Sorry: **apologize***(speaker,addressee,msg-arg: event)*

d. Thanks: **thank***(speaker,addressee,msg-arg: event)*

If we assumed the existence of a 'post-semantic module' which associates CMTs with the (descriptive) contents provided by the grammar, we would run into significant problems. To get the right result for *hi*, we would need to assume that a null descriptive content however represented somehow gets associated with the CMT **greet**. But this would result

in a problem with *bye*, utterances of which equally lack a descriptive content.[2] Assuming underspecification—e.g. null descriptive content associates with, say, **greet** $\vee$ **bid-farewell**—would lead to the unintuitive expectation that *hi* and *bye* potentially allow for multiple CMTs. Assuming that eventive descriptive contents are associated with the CMT of **apologize** or alternatively with **thank** or are underspecified between, say, **apology** and **thank**, would lead to similar problems *mutatis mutandis*. Thus, in their representation in the lexicon such words must have a CMT associated with them.

A second argument concerns reprise utterances. It has been argued (see e.g. Ginzburg and Sag (1999); Ginzburg and Cooper (2001)) that utterances such as B's in (6a,b) can be understood (on the 'clausal' reading, where the addressee verifies she has understood the content of the utterance correctly) as in the respective parenthesized paraphrases; whereas B's utterance in (6c) unambiguously involves the adjacent parenthesized content:

(6) a. A: Who left? B: Who left? (clausal reading: *Are you asking who left?*)

b. A: Go home Billie. B: Go home? (clausal reading: *Are you ordering me, of all things, to go home?*)

c. A: Did Belula resign? B: Did WHO resign? (unambiguously: *Who$_i$ are you asking whether i resigned?*)

If such paraphrases are the correct basis for an analysis of such utterances, this indicates that in reprise utterances at least CMT (the CMT of the preceding utterance, to be precise) can become a constituent of the *descriptive* content of an utterance.[3] In other words, CMT becomes a constituent of the content the grammar incontrovertibly needs to build up.

In fact, following Ginzburg and Sag (2000), we suggest that reprise utterances provide a probe that allows one to filter away the indirect force of an utterance and establish a single direct CMT with a given utterance.[4] Consider (7), uttered outside a West End theater currently showing a best selling musical:

(7)(1) Stina: I have a ticket for tonight's performance.

(2) Padraig: You have a ticket for tonight's performance?

(3) Stina: Yes.

[2] An anonymous reviewer for BIDIALOG expresses skepticism about this argument on the grounds that our assumption that *hi* and *bye* lack descriptive content is dubious. Before turning to consider this assumption, we should point out that our argument here is actually *independent* of this assumption, as it applies equally to pairs such as *sorry* and *thanks*, which clearly do possess a descriptive content. The reviewer questions our assumption that *hi* and *bye* lack descriptive content by pointing to the existence of expressions such as *good morning*, *good afternoon*, and *good night*. According to the reviewer '[these] all have the same CMT but a different content'. We agree with the reviewer that, at least to a first approximation, *hi*, *good morning*, and *good afternoon* all involve the same CMT, namely greeting (*good night* is actually akin to *bye*, as it is used to bid a nocturnal farewell by conversationalists who will not speak again before the morrow.). Where these words differ is in terms of their presuppositions—*good morning* presupposes that the utterance time is basically before noon, *good afternoon* that the utterance time is basically before sundown, whereas *hi* carries no temporal presupposition. Encoding these varying presuppositions does not require postulating a descriptive content for the act of greeting (see footnote 8 for exemplification.).

[3] This claim was originally made, independently, by Ginzburg (1992) and Jacobs (1991).

[4] Using reprises as such a probe was first suggested to us by Richmond Thomason in an oral discussion that followed presentation of Ginzburg and Sag (1999).

(8) a. I'm offering to sell a ticket for tonight's performance.

b. Are you claiming that you have a ticket for tonight's performance?

c. Are you saying that you wish to sell a ticket for tonight's performance

d. I'm claiming that I have a ticket for tonight's performance.

e. I'm offering to sell a ticket for tonight's performance.

Stina's utterance (7[1]) could naturally be understood to convey (8a). However, Padraig's reprise—(7[2])—merely requests clarification of the claim Stina made; it can be understood solely as (8b), not as (8c). This can be further demonstrated by noting that *yes* in (7[3]) conveys (8d) in this context, but cannot convey (8e), despite the salience of the offer.[5]

Indeed, far from casting doubt on the assumption that grammatically associated CMTs exist, we believe that the phenomenon of multifunctionality *strengthens* the need for the assumption. In order to deal with indirectly conveyed messages such as (8a), one will need to state domain axioms whose antecedents will often involve a content with a gramatically associated CMT. For instance, If agent A states to B that he has a ticket, he might wish to sell it to B, rather than simply If agent A has a ticket, he might wish to sell it to B. Programming a robot with the latter axiom is a recipe for disaster, as the robot will hassle any approaching theatre-goer, rather than solely loudly declaiming touts.

3 Integrating CMT into a constraint-based grammar

We adopt a version of HPSG developed in Sag (1997); Ginzburg and Sag (2000). The content associated with signs, phrasal or lexical, is drawn from a situation theoretic ontology. The ontology distinguishes *inter alia* questions, propositions, facts, situations/events, and outcomes. Information about phrases is encoded by cross-classifying them in a multi-dimensional type hierarchy. Phrases are classified not only in terms of their phrase structure schema or X-bar type, but also with respect to a further informational dimension of CLAUSALITY. Clauses are divided into *inter alia* declarative clauses which denote propositions, interrogative clauses denoting questions, exclamative clauses denoting facts, and imperative clauses denoting outcomes. Each maximal phrasal type can inherit from both these dimensions. This classification allows a specification of systematic correlations between clausal construction types and types of semantic content.

[5] Our discussion of these data is of necessity all too brief. As discussed in Ginzburg and Cooper (2001), reprises exemplify an additional reading dubbed the *constituent-reading*, which involves a request for reformulation of the import of the reprised (sub)-utterance. Thus, for an referential NP utterance, as in (i), this will be understood as a request for reference resolution:

(i) Andie: Did Jo leave?
Bo: Jo?
Andie: Your cousin.

Given this, reprises such as (7[2]) will also yield readings paraphrasable as (ii), where the inferred component of content is *not* necessarily filtered away:

(ii) Shi: What do you mean by saying *you have a ticket for tonight's performance*?

yes, however, is an inappropriate response to this reading.

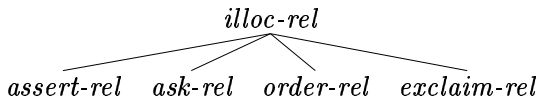
We note two considerations that an account integrating CMT information into the grammar needs to heed:

- In order to avoid the problems associated with the PH, one has to ensure that the way in which CMT information enters into the content of a sign does not affect the assignment of (non-CMT) content. One must also ensure that a sign that has CMT information (of the current utterance) cannot be embedded as a daughter of another sign.
- In order to describe reprise utterances, one must have the means to let signs with CMT information be inputs to grammatical constraints, e.g. to build questions whose queried proposition contains CMT information.

We will satisfy these requirements by making a finer grained distinction than usually made with respect to "matrix" (non-embedded) signs. Whereas all signs that cannot be complements of an embedding predicate bear the specification I(NDEPENDENT)C(LAUSE):+, we will introduce a further partition among such signs, depending as to whether or not they can play a role in recursive operations of the grammar. Those that cannot will be designated as ROOT:+. Before we can illustrate how this actually works, we need to bring CMTs into the picture.

Our approach is consistent with various ontologies of CMTs. The *minimal* such ontology one could posit involves a 1–1 relationship between what is often called the CONTENT of a sign, i.e. entities of type *message* (*proposition*, *question*, *outcome*, *fact*, ...) and CMTs: propositions are associated with the CMT of asserting, questions with asking, outcomes with ordering, and facts with exclaiming. This involves positing a type *illoc(utionary)-rel* as the immediate supertype of these four CMTs:

(9)



Each of these types introduces its own constraint on the type of its MSG-ARG value:

(10) a. *assert-rel* ⇒ $\left[\text{MSG-ARG} \quad \textit{proposition}\right]$

b. *ask-rel* ⇒ $\left[\text{MSG-ARG} \quad \textit{question}\right]$

c. *order-rel* ⇒ $\left[\text{MSG-ARG} \quad \textit{outcome}\right]$

d. *exclaim-rel* ⇒ $\left[\text{MSG-ARG} \quad \textit{fact}\right]$

Arguably, such a relationship between message types and CMTs constitutes something like a default. But each of the afore-mentioned subtypes of *message* clearly does have other uses: questions can be used 'rhetorically' (also known as a *reassertion* of a resolved question), outcomes can be suggested, propositions can feature in threats and so on. Thus, an adequate view of utterance content needs to allow for a richer ontology of CMTs and for the CMT associated with a given message-type to be underspecified. This refinement is easy to implement by (a) positing more maximal subtypes of *illoc-rel* (e.g. *threat-rel, promise-rel, reassert-rel* etc) and (b) positing types intermediate between *illoc-rel* and the leaves of the hierarchy in (9)

(e.g. a type *prop-illoc-rel* which would subsume all propositional CMTs—*assert-rel*, *threat-rel*, *promise-rel* etc.). In this abstract, as in our implementation at present, we maintain the more simplistic view, enshrined in (9).

The final ingredient we need as far as phrases go is a constraint that determines the appropriate CONTENT value for utterances, i.e. for root clauses. We propose that the content of every root clause be a proposition whose SOA value is of type *illoc-rel*. This proposition represents the belief an agent forms about the (full, direct illocutionary) content of an utterance. More specifically, this is the content a speaker will assign to her utterance, as will an addressee in case communication is successful. Given (9), this will mean that a root clause will be resolved so as to have as its content a proposition whose SOA value is of one of the subtypes of *illoc-rel*.

In order to ensure that root clauses have contents in which CMT information is represented, we posit a type *root-cl* and propose a constructional treatment of root utterances in terms of a non-branching phrasal type (*hd-only-ph*) that embeds message-denoting sentences as arguments of an *illoc-rel*. The constraints idiosyncratic to this construction, akin to a 'start' symbol in a context free grammar, are illustrated in (11):[6]

(11) *root-cl*:

$$\begin{bmatrix} \textsc{root} & + \\ \textsc{head} & \begin{bmatrix} v \\ \textsc{vform} & \textit{fin} \end{bmatrix} \\ \textsc{cont} & \begin{bmatrix} \textit{proposition} \\ \textsc{soa} & \begin{bmatrix} \textit{illoc-rel} \\ \textsc{utterer} & \boxed{1} \\ \textsc{addressee} & \boxed{3} \\ \textsc{msg-arg} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \textsc{cntxt}|\textsc{c-indices} & \begin{bmatrix} \textsc{c-spkr} & \boxed{1} \\ \textsc{c-addr} & \boxed{3} \\ \textsc{c-time} & \boxed{4} \end{bmatrix} \\ \textsc{slash} & \{\,\} \\ \textsc{store} & \{\,\} \\ \textsc{wh} & \{\,\} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} \textsc{root} & - \\ \textsc{ic} & + \\ \textsc{cont} & \boxed{2} \end{bmatrix}$$

Note that the arguments of the *illoc-rel* are identified with the appropriate individuals in the context of utterance. As mentioned above, we now distinguish root clauses from other independent clauses in terms of positive versus negative specifications for the feature ROOT.[7]

Let us illustrate the effects of the constraint in (11). (12a) has an analysis as a polar question in which it expresses the question in (12b). Therefore, given (10) and (11), the content such a clause gets as a root utterance (ignoring tense) is (12c):

[6]The constraint here relates the mother to its (sole) daughter, denoted with a large bold faced **H**.

[7]On this view, signs are [ROOT −] by default. Since this is the case, we will suppress [ROOT −] specifications on all phrases other than instances of the type *root-cl*.

(12) a. A: Did Jo leave?

b.
$$\begin{bmatrix} \textit{question} \\ \text{PARAMS} \quad \{\,\} \\ \text{PROP} \quad \begin{bmatrix} \textit{proposition} \\ \text{SIT} \quad \boxed{2} \\ \text{SOA|NUCL} \quad \begin{bmatrix} \textit{leave-rel} \\ \text{LEAVER} \quad \boxed{3} \\ \text{TIME} \quad \boxed{4} \end{bmatrix} \end{bmatrix} \\ \text{BCKGRD} \quad \left\{ \textit{named}(\boxed{3}, \textit{Jo}),\ \textit{t-precede}(\boxed{5}, \boxed{4}) \right\} \\ \text{CNTXT|C-INDICES} \quad \begin{bmatrix} \text{C-SPKR} \quad \boxed{1} \\ \text{C-ADDR} \quad \boxed{6} \\ \text{C-TIME} \quad \boxed{5} \end{bmatrix} \end{bmatrix}$$

c.
$$\begin{bmatrix} \textit{proposition} \\ \text{SIT} \quad \boxed{0} \\ \text{SOA} \quad \begin{bmatrix} \text{QUANTS} \quad \langle\,\rangle \\ \text{NUCL} \quad \begin{bmatrix} \textit{ask-rel} \\ \text{UTTERER} \quad \boxed{1} \\ \text{ADDRESSEE} \quad \boxed{6} \\ \text{MSG-ARG} \quad \begin{bmatrix} \textit{question} \\ \text{PARAMS} \quad \{\,\} \\ \text{PROP} \quad \begin{bmatrix} \textit{proposition} \\ \text{SIT} \quad \boxed{2} \\ \text{SOA|NUCL} \quad \begin{bmatrix} \textit{leave-rel} \\ \text{LEAVER} \quad \boxed{3} \\ \text{TIME} \quad \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{BCKGRD} \quad \left\{ \textit{named}(\boxed{3}, \textit{Jo}),\ \textit{t-precede}(\boxed{5}, \boxed{4}) \right\} \\ \text{CNTXT|C-INDICES} \quad \begin{bmatrix} \text{C-SPKR} \quad \boxed{1} \\ \text{C-ADDR} \quad \boxed{6} \\ \text{C-TIME} \quad \boxed{5} \end{bmatrix} \end{bmatrix}$$

So far we have focussed on phrases. However, one can within this approach equally describe words such as those discussed in (3)–(5). For instance, the word *hi* can be described as follows:[8]

[8] A lexical entry for e.g. *good morning* would involve adding the assumption in BCKGRD that the time k is located before noon.

(13)
$$\begin{bmatrix}
\textsc{phon} & \langle\ \text{hi}\ \rangle & & & & \\
\textsc{ss|loc} & \begin{bmatrix}
\textsc{cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textit{interj} \\ \textsc{ic} & + \end{bmatrix} \end{bmatrix} & & & \\
\textsc{cont} & \begin{bmatrix} \textit{proposition} \\ \textsc{sit} & \boxed{0} \\ \textsc{soa} & \begin{bmatrix} \textsc{quants} & \langle\ \rangle \\ \textsc{nucl} & \begin{bmatrix} \textit{greet-rel} \\ \textsc{utterer} & \text{i} \\ \textsc{addressee} & \text{j} \\ \textsc{time} & \text{k} \end{bmatrix} \end{bmatrix} \end{bmatrix}
\end{bmatrix} & & & \\
\textsc{arg-st} & \langle\ \rangle & & & \\
\textsc{ctxt|c-indices} & \begin{bmatrix} \textsc{c-spkr} & \text{i} \\ \textsc{c-addr} & \text{j} \\ \textsc{c-time} & \text{k} \end{bmatrix} & & &
\end{bmatrix}$$

Note that *hi* is specified as IC:+, which means that it cannot be embedded. However, it is underspecified for ROOT. This means, as we will soon see, that this entry can be the head-daughter of a reprise construction (reflecting cases such as [Context: A is a crusty brigadier, B a raw recruit] B: Hi. A (growls): Hi? (= Are you greeting me)).[9,10]

Finally, we explain briefly how the CMT of the previous utterance enters as a constituent of the content of certain reprise utterances. We assume the account developed in Ginzburg and Cooper (2001) of how clarifications arise during attempted integration of an utterance in a conversationalist's information state (IS). Simplifying somewhat, on this view a necessary condition for B to ground an utterance by A is that B manage to find values for the contextual parameters of the meaning of the utterance. What happens when B cannot or is at least uncertain as to how he should instantiate in his IS a contextual parameter i? In such a case B needs to do at least the following: (1) perform a partial update of the existing context with the successfully processed components of the utterance (2) pose a clarification question that involves reference to the sub-utterance u_i from which i emanates. Since the original speaker, A, can coherently integrate a clarification question once she hears it, it follows that, for a given utterance, there is a predictable range of < partial updates + consequent clarification questions>. These we take to be specified by a set of *coercion operations* on utterance representations.[11] Indeed we assume that a component of dialogue competence is knowledge of these *coercion operations*.

One such operation is dubbed parameter focussing by Ginzburg and Cooper (2001). This involves a (partially updated) context in which the issue under discussion is a question that

[9]As with all CE utterances, this one can be understood in a number of ways. In this case, the constituent reading alluded to in footnote 5 is possibly even more prominent. It would yield a reading paraphrasable as **what do you mean by saying *hi* to me**.

[10]Underspecifying *hi* for ROOT might suggest that it could function as the head daughter in (11), thereby yielding an unwanted reading *I assert that I greet you*. However, in the framework of Ginzburg and Sag (2000) all headed phrases are subject to the Generalized Head Feature Principle (GHFP), which involves the SYNSEM value of the mother of a headed phrase and that of its head daughter being identical by default. This means that the head daughter of a *root-cl* is specified to be CAT|HEAD:*v[fin]*; *hi* (as its relatives *bye, sorry, thanks* etc) is specified as CAT|HEAD: *interj*, and hence cannot serve as the head daughter of a *root-cl*.

[11]The term *coercion operation* is inspired by work on utterance representation within a type theoretic framework reported in Cooper (1998).

arises by instantiating all contextual parameters except for i and abstracting over i. In such a context, one can *confirm* that i gets the value B suspects it has by uttering with rising intonation any apparently co-referential phrase whose syntactic category is identical to u_1's (see (6a,b) above). One construction type appropriate for this context are reprise interrogative clauses (*repr-int-cl*). In the framework of Ginzburg and Sag (2000) they are described by means of the following schema:

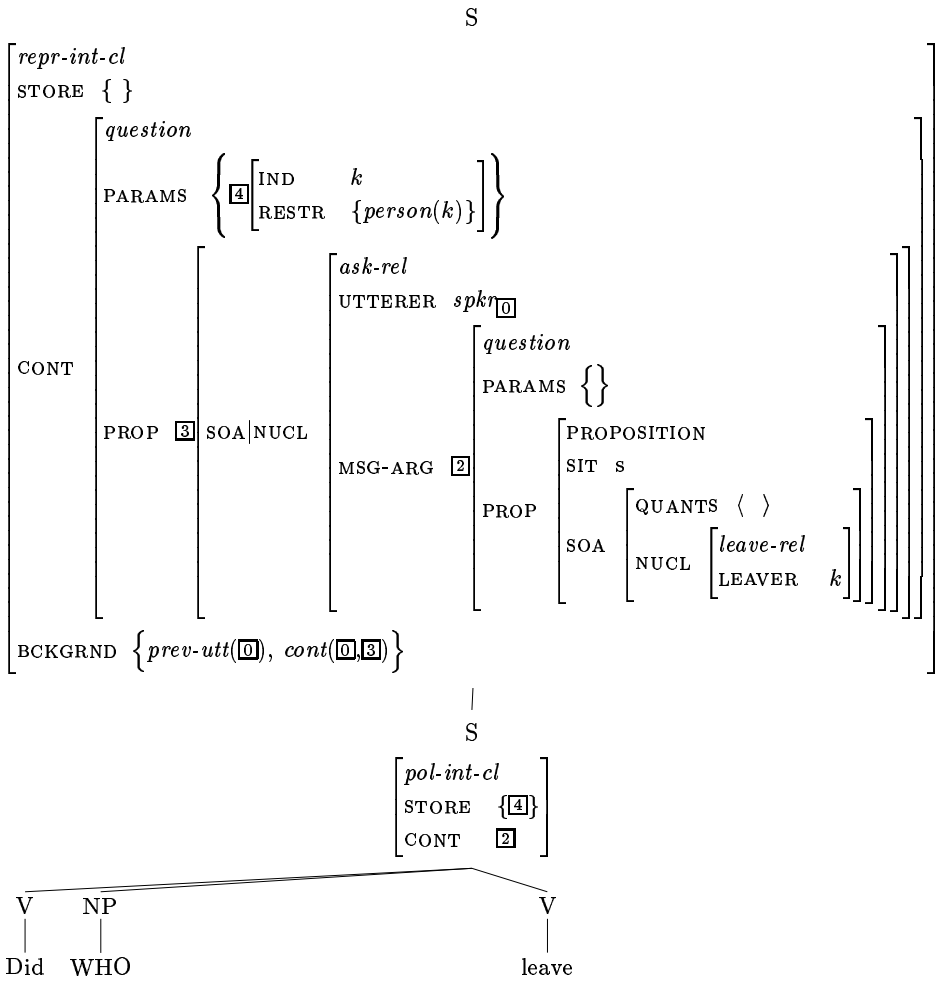
(14)
$$\left[\begin{array}{ll} \textit{repr-int-cl} & \\ \textsc{cont} & \left[\begin{array}{ll} \textit{question} & \\ \textsc{params} & \boxed{\Sigma_3} \\ \textsc{prop} & \boxed{1}\left[\textsc{soa}|\textsc{nucl}|\textsc{msg-arg} \quad \boxed{2}\right] \end{array}\right] \\ \textsc{store} & \boxed{\Sigma_1} \\ \textsc{bckgrnd} & \left\{\textit{prev-utt}(\boxed{0}),\ \textit{cont}(\boxed{0},\boxed{1})\right\} \uplus \boxed{\Sigma_2} \end{array}\right] \rightarrow \mathbf{H}\left[\begin{array}{ll} \textsc{cat} & \left[\textsc{head} \quad \left[\begin{array}{ll} \textsc{ic} & + \\ \textsc{vform} & \textit{fin} \end{array}\right]\right] \\ \textsc{cont} & \boxed{2} \\ \textsc{store} & (\boxed{\Sigma_1} \uplus \boxed{\Sigma_3}) \\ \textsc{bckgrd} & \boxed{\Sigma_2} \end{array}\right]$$

To illustrate this: a reprise of (12a) can be performed using (15a). This can be assigned the content in (15b) on the basis of the schema in (14):[12]

[12]Note that the previous utterance identified the utterer of the *ask-rel* with the speaker of that utterance (this is ensured by the constraint in (11) on the type *root*). Hence, the utterer of the *ask-rel* in the content of the reprise must also be that individual, indicated as $spkr_{\boxed{0}}$ in (15).

(15)

Did WHO leave?



wh-less reprises, as in (6a,b), are accommodated as a special case of no parameters being abstracted over. Reprise uses of *hi* can be similarly analyzed, using an instantiation of (13) with ROOT:-.

4 Conclusions and Future Work

In this paper, we have presented a number of arguments that indicate the need to integrate CMT information in grammars intended to analyze conversational interaction. One such argument concerns the proper analysis of words such as *hi, thanks, sorry* which can stand alone as complete utterances. Another arguent derives from the consideration of reprise utterances. We have sketched briefly the basics of an HPSG in which CMT information is integrated. This grammar has been implemented as part of the SHARDS system Ginzburg et al. (2001). In future work we hope to show how grammars of this type can, when integrated with domain knowledge, offer insightful solutions to the many puzzles posed by multifunctionality.

Acknowledgements

We would like to thank three anonymous BIDIALOG reviewers for very useful comments. The research described here is funded by grant number R00022269 from the Economic and Social Research Council of the United Kingdom and by grant number GR/R04942/01 from the Engineering and Physical Sciences Research Council of the United Kingdom.

Bibliography

Allwood, J. (1995). An activity based approach to pragmatics. In *Gothenburg Papers in Theoretical Linguistics*, number 76. Dept. of Linguistics, University of Göteborg.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson, A. (1996). Map Task coder's manual. *HCRC Research Paper*, RP-82.

Cooper, R. (1998). Mixing situation theory and type theory to formalize information states in dialogue exchanges. In Hulstijn, J. and Nijholt, A., editors, *Proceedings of TwenDial 98, 13th Twente workshop on Language Technology*. Twente University, Twente.

Core, M. and Allen, J. (1997). Coding dialogs with the DAMSL scheme. *Working notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines.*

Ginzburg, J. (1992). *Questions, Queries, and Facts: a semantic and pragmatics for interrogatives.* PhD thesis, Stanford University.

Ginzburg, J. and Cooper, R. (2001). Resolving ellipsis in clarification. In *Proceedings of the 39th Meeting of the Association for Computational Linguistics.*

Ginzburg, J., Gregory, H., and Lappin, S. (2001). SHARDS: Fragment resolution in dialogue. In Bunt, H., editor, *Proceedings of the 1st International Workshop on Computational Semantics.* ITK, Tilburg University, Tilburg.

Ginzburg, J. and Sag, I. (1999). Constructional ambiguity in conversation. In Dekker, P., editor, *Proceedings of the 12th Amsterdam Colloquium.* ILLC, Amsterdam.

Ginzburg, J. and Sag, I. (2000). *Interrogative Investigations: the form, meaning and use of English Interrogatives.* Number 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California.

Jacobs, J. (1991). Implikaturen und 'alte information' in w-fragen. In Reis, M. and Rosengren, I., editors, *Fragesätze und Fragen.* Niemayer, Tübingen.

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

Sag, I. (1997). English relative clause constructions. *Journal of Linguistics*, 33:431–484.

Paper 5

Information States in a Multi-modal Dialogue System for Human-Robot Conversation

Oliver Lemon, Anne Bracy, Alexander Gruenstein, and Stanley Peters
CSLI, Stanford University, Stanford, CA 94305
lemon,bracy,alexgru,peters@csli.stanford.edu
http://www-csli.stanford.edu/semlab/witas/

Abstract

We discuss the dialogue modelling techniques in the ongoing development of a dialogue system for multi-modal conversations with autonomous mobile robots. The dialogue system is implemented using Nuance, Gemini, and Festival language technologies under the Open Agent Architecture. This paper focusses on our general-purpose dialogue manager which implements a dynamic information state model of dialogue.

1 Introduction

We present modelling techniques in a dialogue system for multi-modal conversations with autonomous mobile robots – in this case a robot helicopter, or UAV ('Unmanned Aerial Vehicle') – see Doherty et al. (2000)[1]. The system operates over a dynamic environment, which supercedes the standard travel-planning dialogue system domain in its complexity. In particular, interactions with such a system are not scriptable in advance, and rely on mixed task and dialogue initiatives in conversation. This setting presents new challenges for dialogue modelling, in that conversations must be asynchronous, mixed-initiative, open-ended, and involve a dynamic environment. The system is implemented using Nuance, Gemini, and Festival language technologies under the Open Agent Architecture (see e.g. (Stent et al., 1999)). The interface's main feature is its dialogue manager. The dialogue manager interprets spoken language and map-gesture inputs as commands, queries, responses, and declarations to the robot, and generates synthesized speech and graphical output to express the robot's responses, questions, and reports about situations as they unfold in the environment. Our

[1]This research was (partially) funded under the Wallenberg laboratory for research on Information Technology and Autonomous Systems (WITAS) Project, Linköping University, by the Wallenberg Foundation, Sweden.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

currently implemented model supports ambiguity resolution, presupposition checking, processing of anaphoric and deictic expressions, command revision, report generation, and a confirmation backchannel.

This paper focusses on our general-purpose dialogue manager which implements a dynamic information state model of dialogue (e.g. Bohlin et al. (1999); Cooper and Larsson (1998); Ginzburg (1996a,b); Groenendijk and Stokhof (1991)). In contrast to other recent state-based approaches (e.g. Xu and Rudnicky (2000); Roy et al. (2000)) our dialogue manager implements a model (see Section 4) involving an *Issues Raised stack*, a *System Agenda*, a *Salience List*, and a *Modality Buffer*.

We close the paper by discussing current developments of the dialogue model, which allow conversations about tasks (both current and planned) and the changing abilities of the robot given its own state and that of the environment.

2 Dialogues with mobile robots

Various dialogue systems have been built for use in contexts where conversational interactions are largely predictable and can be scripted, and where the operating environment is static. For example, a dialogue for buying an airline flight can be specified by filling in certain parameters (cost, destination, and so on) and a database query, report, and confirmation cycle. In such cases it suffices to develop a transition network for paths through the dialogue to recognizable completion states. Now consider an operator's conversation with a mobile robot in a environment which is constantly changing. As argued by Elio and Haddadi (1999), dialogues with such a device will be very different. There will be no predictable course of events in the dialogues. The device itself may "need" to communicate urgently with its operator. There may not be a strictly defined endpoint to conversations, and relevant objects may appear and disappear from the operating environment.

Conversational interaction with robots places the following requirements on dialogue management (see also Clark (1996)):

- *Asynchronicity*: events in the dialogue scenarios can happen at overlapping time periods (for example, new objects may enter the domain of discourse while the operator is giving a command).
- *Mixed task-initiative*: in general, both operator and system will introduce issues for discussion.
- *Open-ended*: there are no clear start and end points for the dialogue and sub-dialogues, nor are there rigid pre-determined goals for interchanges.
- *Resource-bounded*: participants' actions must be generated and produced in time enough to be effective dialogue contributions.
- *Simultaneous*: participants can produce and receive actions simultaneously.

In particular we note that simple form-filling or data-base query style dialogues (e.g. the CSLU Toolkit, McTear (1998)) will not suffice here (see Roy et al. (2000); Elio and Haddadi (1999) for similar arguments). We do not know in advance what all the possible paths through

successful dialogues are in robot interaction scenarios. Dialogues with a robot will be more open and flexible – interactions which are more akin to conversations between humans.

In our current application, the robot is a UAV ('unmanned aerial vehicle') – a small autonomous helicopter with onboard planning and deliberative systems, and vision capabilities (for details see e.g. Doherty et al. (2000)). Mission goals are provided by a human operator, and the planning system then generates a list of suitable waypoints for the UAV to navigate by. An on-board active vision system interprets the scene or focus below to interpret ongoing events, which are reported (via NL generation) to the operator.

3 Dialogue Processing

As argued above, robot interaction scenarios present a number of challenges to designers of dialogue systems. Such systems require a particularly flexible architecture – one which can coordinate multiple asynchronous communicating processes. For these reasons we currently use the Open Agent Architecture (OAA2, see Martin et al. (1999)), with the following agents (see Figure 5.1);

- NL (natural language): a wrapper to SRI's Gemini parser and generator using a grammar for human-robot conversation developed at CSLI

- SR: (speech recognizer) a wrapper to a Nuance speech recognition server using a language model compiled directly from the Gemini grammar (with the consequences that every recognized utterance has a logical form, and that every logical form can be mapped to a surface string)

- TTS: (text-to-speech) a wrapper to the Festival 1.4.1 speech synthesiser, for robot speech output

- GUI: an interactive map display of the current operating environment which displays route plans, waypoints, locations of vehicles including the robot, and allows deictic reference (i.e. mouse pointing) by the user

- DM: (dialogue manager): co-ordinates multi-modal inputs from the user, interprets dialogue moves made by the user and robot, updates and maintains the dialogue context, handles robot reports and questions, and sends speech and graphical outputs to the user

- Robot Control and Report: translates commands and queries from the dialogue interface into commands and queries to the robot, and vice-versa for reports and queries received from the robot. Uses a realtime CORBA layer.

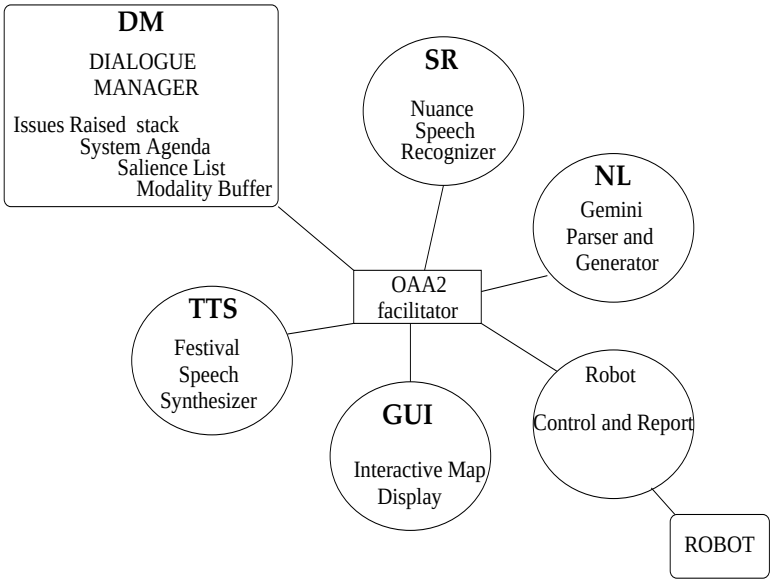


Figure 5.1: Dialogue system architecture

The dialogue segments shown in Figure 5.2, p. 61, illustrate many of the capabilities of the multi-modal interface, as of November 2000. They can be carried out in continuous sequence using spoken voice input and mouse clicks on a map screen. We now explain how the above agents act in concert so as to support conversations.

3.1 Interpretation and Generation

The operator's speech is recognized by Nuance and parsed into logical forms by Gemini. If these forms do not already indicate the speech act of the user, the dialogue manager inspects the current dialogue Information State (see Section 4) to determine how best to incorporate the utterance into the dialogue. Gemini's Semantic-Head-Driven Generation method (Shieber et al. (1990)) is used to produce English sentences for the UAV's queries and reports. Reports are sent through Gemini via the Robot Controller and on to Festival, which synthesizes the robot's speech.

For instance, given the logical-form (output by the UAV)

```
wh_query(which([n(phobj(static(landmark(house))),sg)]))
```

the NL agent (i.e. Gemini plus the grammar for human-robot interaction) returns the string "Which house do you mean", which can be spoken by the Festival text-to-speech agent.

Variants of some of these components have been used in other dialogue systems, notably SRI's CommandTalk (Stent et al. (1999)), the NASA Personal Satellite Assistant (Rayner et al. (2000)), and the system of Guzzoni et al. (1996). However, our system stands apart from these in its particular combination of complex dialogue capabilities (including Natural Language generation) with multimodality over a dynamic operating environment. The core of our system, and its most innovative feature, is the dialogue manager, described in section 4.

Multi-modal Utterances	**Dialogue Moves**
Operator (O): *Where are the buildings?*	Raise question
Robot (R): *Here you are.* [displays buildings on GUI]	Answer question
O: *Go to this one.* [click]	Deictic command
O: *Fly to Main Street.*	Command
O: *No, I meant Circle Road.*	Elliptical revision
R: *Okay, I changed that.* [Route changed on GUI]	Confirmation backchannel
O: *Hide the buildings.*	GUI Command
R: *Okay* [buildings disappear from GUI].	
O: *Next, continue to the bridge and then here* [click].	Complex command with deictic
O: *Then look at the parking lot, the hospital, and the building.*	Ambiguous command
R: *Where is the parking lot?*	Presupposition check, raise question
O: *Here* [click]	Graphical ambiguity resolution
U: *Okay.*	Confirmation, update database
U: *Which building do you mean?* [buildings highlighted on map]	Presupposition check, raise question
O: *The airport.*	Verbal ambiguity resolution
U: *Okay, the airport.*	Confirmation backchannel
O: *Go to the ground station, hover here* [click], *and go back to it.*	Anaphoric and deictic command
O: *Then land at the parking lot.*	Command, previously added referent
O: *No, not the parking lot, the waypoint.*	Revision
R: *Which waypoint do you mean?* [waypoints highlighted on GUI]	Presupposition check, raise question
O: *This one* [click].	Clarification
R: *Okay, I changed that.* [Route changed on GUI]	Confirmation backchannel
O: *Plan the route.*	Command
R: *Planning Route. Route planned.* [Route displayed on GUI]	Confirmation backchannel
O: *Roger. Proceed.*	Command
R: *Executing route.*	Confirmation backchannel
R: *Way-point two reached.*	Robot report generation
R: *Truck 8 is turning left onto Circle Road.*	Robot report generation
R: *The truck is passing the warehouse.*	Robot report generation
O: *Follow it.*	Anaphoric reference to Robot's NP

Figure 5.2: A Sample Dialogue with the Robot, using the system

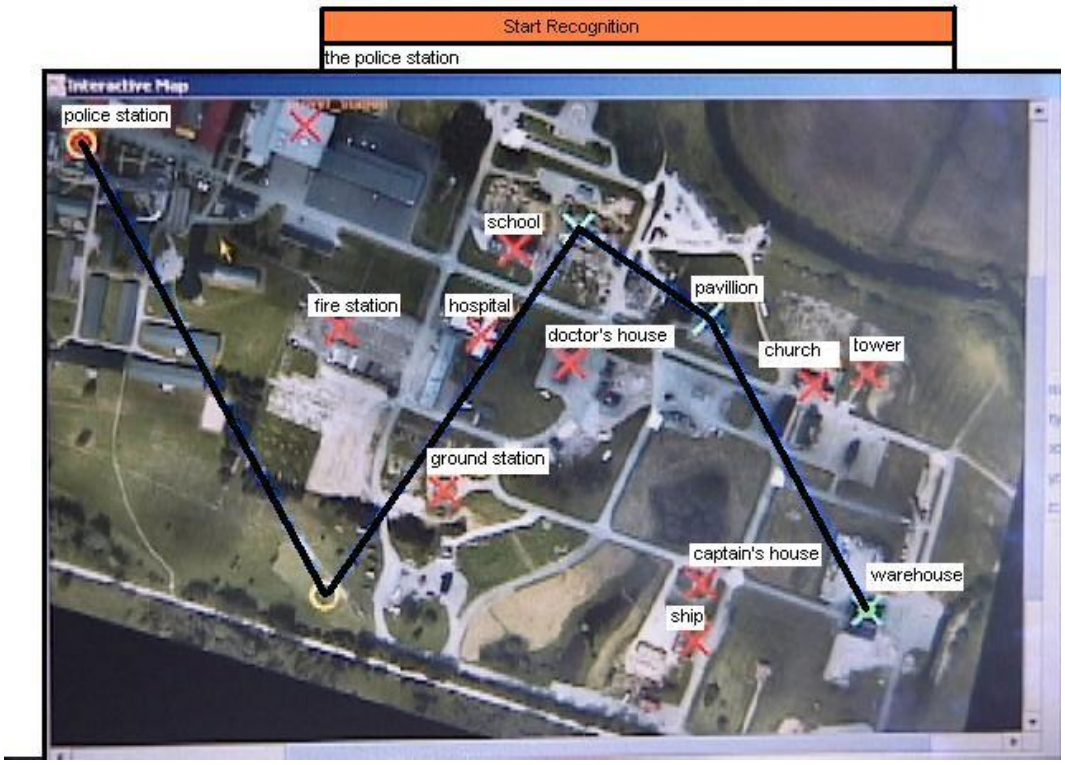


Figure 5.3: Part of the Graphical User Interface

4 Information States

Our dialogue manager embodies several recent theoretical ideas in dialogue modelling. It creates and updates an *Information State* (IS) corresponding to a notion of dialogue context. Dialogue moves have the effect of updating information states, and moves can be initiated by both the operator and the robot. A dialogue move might cause an update to the GUI, send an immediate command to the UAV, elicit a spoken report, or prompt a clarifying question from the UAV. Subdialogues can be arbitrarily nested.

Central parts of these information states are an *IR stack* – a stack of public unresolved *issues raised* in the dialogue thus far, and a *System Agenda* – a private list of issues which the UAV has yet to raise in the conversation. Under certain conditions, items from the System Agenda are made public by an utterance from the UAV (e.g. "Which building do you mean?"), moving the issue onto the IR Stack. Such an operation is a Dialogue Move (in this case by the UAV). The dialogue manager contains a collection of rules which interpret (multi-modal) input from both operator and UAV as dialogue moves with respect to the current information state, and update the state accordingly. Similarly, there are rules which process UAV responses, reports, or questions, again updating the context accordingly.

Logical-form outputs from the parsing process are often already interpreted as speechacts of various kinds (e.g. "Fly to the hospital" is parsed as a COMMAND). For example, the operator utterance "fly to the temple and the river" is assigned the logical form:

```
command([go],[param_list([pp_loc(to,arg(conj,
```

```
[np(det([def],the),[n(phobj(static(landmark(temple))),sg)])],
[np(det([def],the),[n(phobj(static(landmark(river))),sg)] )]))])])
```

The dialogue manager then interprets this structure as a dialogue move involving certain presuppositions which must be checked (e.g. uniqueness and existence of 'the temple' and 'the river') and various context-update functions (e.g. add 'temple' and 'river' to the Salience List – see below).

Certain utterances do not have a specific illocutionary force, and these are simply specified as DECLARATIONs. The dialogue manager then decides, on the basis of the current IS, what speechact such utterances constitute. This is akin to the robust parsing strategy described in Allen et al. (1996).

Another important part of the information state is a *Salience List* consisting of the objects referenced in the dialogue thus far, ordered by recency (see e.g. Fry et al. (1998)). This list also keeps track of *how* the reference was made (i.e. by which modality) since this is important for resolving and generating anaphoric and deictic expressions in the dialogues.

A related structure, the *Modality Buffer*, keeps track of mouse gestures until they are either bound to deictic expressions in the spoken input or, if none such exists, are recognized as purely gestural expressions. Other aspects of updating the dialogue context are database maintenance tasks.

To recap, our Information States consist of:

- Issues Raised (IR) stack
- System Agenda
- Salience List
- Modality Buffer
- Databases: dynamic objects, planned routes, geographical information, names.

The dialogue manager acts in the following cycle:

1. Multimodal inputs arrive from the NL agent, the robot interface, or the GUI agent.
2. This information is examined and the Information State is updated accordingly (see section 4.1). For instance, if a logical form arrives, then it is pushed onto the IR stack.
3. The dialogue manager then enters into a cycle of examining the contents of the information state, taking appropriate action, and then looping again until no action should be taken without further input. In each iteration, it examines the System Agenda to determine whether there are any issues that should be pushed on to the IR stack. It then peeks at the top of the stack in order to determine which set of rules should be applied.

Note that dialogue capabilities can be added in a modular way, due to the architecture of the dialogue manager. We now give informal examples of the interpretation, generation, and update rules corresponding to dialogue moves.

4.1 Example dialogue rules

When the system receives an utterance from the user, candidate referential phrases (X) can be retrieved via parsing. In order to generate dialogue moves correctly and interpret such phrases in an IS, the following sorts of rules are employed (here noun phrases refer to physical objects with locations):

- Resolve(X): attempt to process X using resolve-deixis(X), resolve-anaphora(X), and lookup(X), in that order. If all of these fail, move into the resolve-ambiguity dialogue state and put *resolve-ambiguity(X)* on the system agenda.

- Resolve-deixis(X): when X is "here", look at the modality buffer for the last resolved gestural expression (mouse click) and bind to that. If none exists, give up. If the referential term is "there" look at the salience list for the last resolved referential expression (gesture or spoken) and bind to it. If the expression is "that Y" or "this Y" and the user has gestured, match the points. If the user has not made a gesture then move into the resolve-ambiguity state – i.e. put *resolve-ambiguity(Y)* on top of the system agenda.

- Resolve-anaphora(X): when X is "it", look at the salience list for the last spoken resolved noun-phrase (NP) and bind X to the value of that NP. If no such NP exists put a presupposition failure report on the System Agenda. (e.g. "I don't know what 'it' refers to.") Update the information state.

- Resolve-ambiguity(X): if X is unknown, ask "Where is the X?" and wait for a GUI-gesture. If X is an object type (e.g. "the building") ask "Which X do you mean?" Display the Xs on the GUI. Switch on speech recognition. Wait for either an utterance or GUI-gesture to select one of the Xs. Pop *resolve-ambiguity(X)* off the IR stack.

- Revisions (e.g. "Not the X the Y"): Look for the specified object (X), remove it from the current command (or report a presupposition failure if X was not specified in the current command), and replace it with the new referential term (Y), which can be a gesture (e.g. "Not the tower, here [click]") or a spoken phrase. Try to resolve the new referential term – put *resolve(Y)* on top of the IR stack. If no object is specified for removal, delete the last spoken object (anaphoric revision). If no object is specified as a replacement, delete the removed object from the plan.

These sorts of rules, taken with the information state structures, constitute the dialogue system in abstract.

Note that multi-modal aspects of the system can be used in disambiguation. For example, if the operator says "Fly to that car" without a corresponding deictic gesture on the map screen, reference resolution will be attempted by looking at the salience list for an NP previously spoken about by the operator. If the user makes (or has made) a gesture, reference will be resolved deictically.

5 Summary

We explain the dialogue modelling techniques which we implemented in order to build a real-time multi-modal conversational interface to an autonomous robot.

A general point of distinction between our system and many others is that it is not restricted to plan-based dialogues. In other words, paths through dialogues need not be specified in advance, as is necessary in some other systems. Our approach, based on updates over Information States, allows us to be much more flexible in the way we process conversation.

Our current demonstration system has the following features:

- a dynamic information state model of dialogue
- support of commands, questions, revisions, and reports, over a dynamic environment
- mixed-initiative, open-ended dialogues
- Semantic-Head-Driven Generation (see Shieber et al. (1990)) of robot reports
- asynchronous, real-time, multi-modal operation
- CORBA interface to real-time UAV simulator
- Solaris or Windows NT/2000 implementations available, using Java, Prolog, and CORBA.

A demonstration of the system is available at `www-csli.stanford.edu/semlab/witas/demo1/`. We are now enhancing the dialogue model and system to handle conversations concerning negotiation of tasks, resources, and abilities.

6 Evaluation and extension

One of our first observations has been that the adoption of *stack* structures to drive dialogue move processing (see e.g. Section 4) has proven to be too restrictive in general. In particular it has made navigation back and forth between different sub-dialogues and topics difficult, since some information is lost when issues are popped off the IR stack (see also Xu and Rudnicky (2000)). For these reasons the latest version of our system employs *Dialogue Move Trees* as navigable records of the conversation.

Another problem is that the current system does not take task-initiatives (e.g. "Shall I land now?") – it takes dialogue-initatives (e.g. "Which building do you mean?") when necessary for continuation of the conversations. For this reason we now also employ an abstract task model (e.g. Elio and Haddadi (1999)) in the form of a *Task Tree*, which brings our system greater flexibility in recognizing user intentions and taking task initiatives.

We have also recently implemented a more complex world-state model (c.f. Rayner et al. (2000)) in order to check the operator's proposed actions before execution and generate system task-initiative utterances via an error-report stream. This agent currently employs JTP (the Java Theorem Prover).

6.1 Future work

The system described here is only the prototype of more general dialogue system for interaction with intelligent agents in dynamic environments. As well as technical improvements (e.g. context-sensitive language models, hands-free operation, video inputs) we are developing a model for dialogues concerning negotiations with autonomous agents, concerning their tasks, resources, goals, and abilities. Innovations in version II of the system (now in development) concern dialogue move trees (a more structured approach to the IR stack), task

trees (a dynamic hierarchical representation of the system's tasks and their status), and the use of automated reasoning modules (e.g. JTP and SNARK, Stickel et al. (2000)) to handle common-ground as well as application-specific aspects of negotiation of tasks, resources, and system abilities in conversations.

Recent work at CSLI includes the development of a tutorial dialogue system Fry et al. (2001)[2] using the same software base (i.e. OAA, Gemini, Nuance, Festival). In future work, we also plan to explore how well our dialogue manager handles tutorial dialogues.

Acknowledgements

We wish to thank Erik Sandewall and Patrick Doherty of WITAS, David Martin, Liz Bratt, and Didier Guzzoni of SRI, and John Dowding and Beth-Ann Hockey of NASA/RIACS.

Bibliography

Allen, J. F., Miller, B. W., Ringger, E. K., and Sikorski, T. (1996). A robust system for natural spoken dialogue. In *Proceedings of ACL*.

Bohlin, P., Cooper, R., Engdahl, E., and Larsson, S. (1999). Information states and dialog move engines. *Electronic Transactions in AI*. Website with commentaries: `www.etaij.org`.

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

Cooper, R. and Larsson, S. (1998). Dialog moves and information states. Technical Report 98-6, Goteborg University. Gothenburg papers in Computational Linguistics.

Doherty, P., Granlund, G., Kuchcinski, K., Sandewall, E., Nordberg, K., Skarman, E., and Wiklund, J. (2000). The WITAS unmanned aerial vehicle project. In *European Conference on Artificial Intelligence (ECAI 2000)*.

Elio, R. and Haddadi, A. (1999). On abstract task models and conversation policies. In *Workshop on Specifying and Implementing Conversation Policies, Autonomous Agents'99*, Seattle.

Fry, J., Asoh, H., and Matsui, T. (1998). Natural dialogue with the Jijo-2 office robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems IROS-98*, pages 1278–1283, Victoria, B.C., Canada. (See `www-csli.stanford.edu/semlab/juno`).

Fry, J., Gintzon, M., Peters, S., Clark, B., and Pon-Barry, H. (2001). Automated tutoring dialogues for damage control training. Unpublished Manuscript, CSLI, under review for SIGdial 2001.

Gibbon, D., Mertins, I., and Moore, R. (2000). *Handbook of Spoken and Multi-modal Dialogue Systems*. Kluwer.

Ginzburg, J. (1996a). In Lappin, S., editor, *Interrogatives: Questions, facts and dialogue*, chapter The Handbook of Contemporary Semantic Theory.

Ginzburg, J. (1996b). Dynamics and the semantics of dialogue. In Seligman and Westerstahl, editors, *Logic, Language, and Computation*.

Groenendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100.

[2]Information on the tutorial dialogue system is available at `www-csli.stanford.edu/semlab/muri/`

Guzzoni, D., Cheyer, A., Julia, L., and Konolige, K. (1996). Many robots make short work. In *AAAI Robotics Contest*, Menlo Park, CA. SRI International, AAAI Press.

Lemon, O. (1996). *States in Flux: Logics of Change, Dynamic Semantics, and Dialogue*. PhD thesis, Centre for Cognitive Science, Edinburgh University.

Lemon, O. (1998). First Order Theory Change systems and their Dynamic Semantics. In Ginzburg, Khasidashvili, Vogel, Levy, and Vallduvi, editors, *The Tbilisi Symposium on Language, Logic, and Computation: selected papers*, pages 85 – 100. SiLLI and CSLI Publications, Stanford.

Lemon, O., Bracy, A., Gruenstein, A., and Peters, S. (2001). A multi-modal dialogue system for human-robot conversation. In *Proceedings of North American Association for Computational Linguistics (NAACL 2001)*.

Litman, D., Kearns, M., Singh, S., and Walker, M. (2000). Automatic optimization of dialogue management. In *Proceedings of COLING 2000*.

Martin, D., Cheyer, A., and Moran, D. (1999). The Open Agent Architecture: a framework for building distributed software systems. *Applied Artificial Intelligence: An International Journal*, 13(1-2).

McTear, M. (1998). Modelling spoken dialogues with state transition diagrams: Experiences with the CSLU toolkit. In *Proc 5th International Conference on Spoken Language Processing*.

Moran, D., Cheyer, A., Julia, L., Martin, D., and Park, S. (1997). Multimodal user interfaces in the Open Agent Architecture. In *Proc IUI 97*, pages 61 – 68.

Pittman, J., Smith, I., Cohen, P., Oviatt, S., and Yang, T.-C. (1996). Quickset: a multimodal interface for military simulation. In *Proceedings of the Sixth Conference on Computer Generated Forces and Behavioral Representation, Orlando*, pages 217–224.

Rayner, M., Hockey, B. A., and James, F. (2000). A compact architecture for dialogue management based on scripts and meta-outputs. In *Proceedings of Applied Natural Language Processing (ANLP)*.

Roy, N., Pineau, J., and Thrun, S. (2000). Spoken dialog management for robots. In *Proceedings of ACL 2000*.

Shieber, S. M., van Noord, G., Pereira, F. C. N., and Moore, R. C. (1990). Semantic-head-driven generation. *Computational Linguistics*, 16(1):30–42.

Stent, A., Dowding, J., Gawron, J. M., Bratt, E. O., and Moore, R. (1999). The CommandTalk spoken dialogue system. In *Proceedings of the Thirty-Seventh Annual Meeting of the ACL*, pages 183–190, University of Maryland, College Park, MD. Association for Computational Linguistics.

Stickel, M. E., Waldinger, R. J., and Chaudhri, V. K. (2000). A Guide to SNARK. Technical Note Unassigned, AI Center, SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025.

Xu, W. and Rudnicky, A. (2000). Task-based dialog management using an agenda. In *Proceedings of ANLP/NAACL 2000 Workshop on Conversational Systems*, pages 42–47.

Paper 6

Context-Dependent Interpretation and Implicit Dialogue Acts

Jörn Kreutel and Colin Matheson
joern.kreutel@semanticedge.com, http://www.cogsci.ed.ac.uk/~jorn
colin.matheson@ed.ac.uk, http://www.ltg.ed.ac.uk/~colin

Abstract

A common assumption in current dialogue models is that the actions of dialogue participants (DPs) can be analysed in terms of DIALOGUE ACTS. Recent research in dialogue has suggested that the most appropriate way to describe the DPs' behaviour is in terms of multiple actions which affect various levels of information structure (see for instance (Poesio and Traum, 1998; Kreutel and Matheson, 2000; Matheson et al., 2000)). Treating dialogue acts as the basic units of analysis has been shown to provide more powerful expressive means than, for instance, the traditional concepts of DIALOGUE MOVES or SPEECHACTS. However, employing different levels of dialogue acts for interpreting the DPs' actions is generally still an unresolved issue in dialogue research in general. In this context, the current paper presents an algorithm which assigns context dependent dialogue acts using INFORMATION STATE UPDATE SCENARIOS, which are defined in terms of DISCOURSE OBLIGATIONS, and this allows us in particular to provide what we feel is an intuitive analysis of implicit acceptance acts.

1 Introduction

In recent years, the notion of dialogue acts as basic units in terms of which the contributions of dialogue participants (DPs) can be analysed has provided a powerful form of expressive means covering more concepts such as dialogue moves and speech acts. One influential taxonomy of dialogue acts is the one proposed by Poesio & Traum (Poesio and Traum (1998)), who classify dialogue acts as CORE SPEECH ACTS, describing basic actions like assertions or questions, and ARGUMENTATION ACTS, which encode the various functions a core speechact may have in a wider discourse context. Additionally, dialogue acts are classified as FORWARD-LOOKING or BACKWARD-LOOKING depending on how they contribute to topic management in discourse. However, relating this classification of dialogue acts to the task of interpreting the actions performed by DPs is still a widely unresolved issue in dialogue research.

With the desideratum of describing the relationship between acts and interpretation in mind, this paper presents an algorithm for interpreting actions in discourse which is based

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

on the assumption that the effects of utterances can be described in terms of INFORMATION STATE UPDATES, an approach which underpinned much of the work of the TRINDI consortium (see (Cooper, 1998; Bohlin et al., 1999; Traum et al., 1999) and (Cooper and Larsson, 1999)). Formalising discourse context in terms of INFORMATION STATE UPDATE SCENARIOS, we will show that scenarios not only provide means for context-dependent interpretation of dialogue acts like assertions and questions, but also allow for an intuitive analysis of implicit dialogue acts.[1]

After a general outline of our notions of information state and update scenario, we will motivate the algorithm for information state update which we are assuming and sketch how implicit dialogue acts can be analysed systematically in this framework. We will then briefly present a formalisation of the algorithm which underlies our implementation of the proposed update model.

2 Information States and Update Scenarios

An analysis of interactions in terms of information states (ISs) allows us to view dialogues in terms of the relevant information that the dialogue participants have at each stage in the discourse. The main effect of utterances is thus to change this information. Viewing the DPs' contribution in terms of the way they update ISs allows a decomposition of the classical notion of dialogue move (see for example (Carletta et al., 1996)), which can be seen as performing multiple actions affecting various levels of the information structure.

Our use of ISs adheres to the principles established in the TRINDI project, in which feature structures displayed as attribute-value matrices (AVMs) were used to represent the components which constitute the DPs' knowledge of 'the state the dialogue is in', and we continue this usage here. For current purposes, we assume a simple information state model with no grounding mechanism[2], in which information is either private (here just BEL) or part of the common ground (G). The main components of G, as shown in Figure 6.1, are a stack of obligations and a structured representation of the dialogue history (DH), which represents the move currently being processed (LM) as well as all the previous moves as sets of dialogue acts (DA). The acts themselves are discussed at length below. Some acts give rise to conditionals, and these are also represented here as part of the dialogue history using the CONDS attribute.

Further components of the IS are a set of propositions which capture the DPs' commitments as these arise in the course of the interaction (SC), and a representation of the intentional structure (INT). The latter structure contains a set of intentions associated with dialogue acts (I), information on intentions which have been satisfied or dropped (*sat* and *drop*), and a representation of the intentional hierarchy ($\gg$) as a set of pairs of intentions in which the first member of the pair immediately dominates the other.

information states also contain the CSC attribute, which describes the particular scenario which the IS as a whole represents. Scenarios specify certain 'constellations' of IS, corresponding to situations such as the turnholder's responding to a question, or evaluating an assertion with respect to its assertive or answerhood properties. As they therefore constitute

[1] Clearly the set of acts used in this paper is not intended as a complete inventory; the acts we employ are assumed only to be sufficient for the task at hand.

[2] So the DPs' private beliefs only consist of sets of propositions and do not include assumptions about the course of the interaction. With respect to the latter, we assume that all the information arising during the dialogue will belong to the shared beliefs of the DPs represented in the common ground.

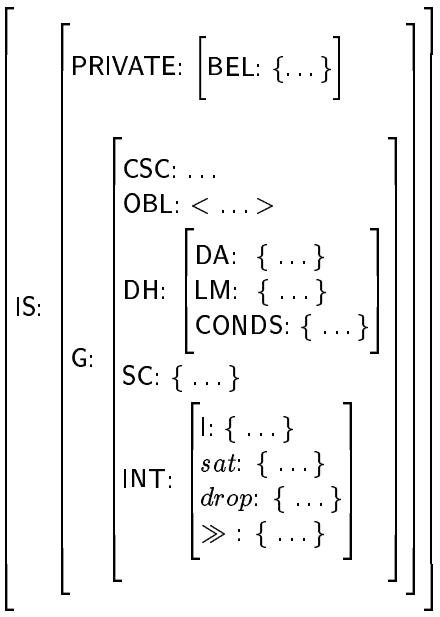


Figure 6.1: Information State Structure

a formalisation of 'discourse context', scenarios provide a basic expressive means for assigning context-dependent interpretations to the core speech acts performed by the DPs.[3]

It is important to keep in mind that each scenario is determined only in terms of the overall structure of the DPs' obligations and the history of the dialogue. Hence it is not necessary to take into account the state of the intentional structure when updating the scenario, and our model is therefore compatible with the assumptions in (Kreutel and Matheson, 2000), according to which in cooperative dialogues the behaviour of DPs can be modelled on the basis of their observable actions and does not require reasoning over intentions. Note that Figure 2 contains an overview of the dialogue acts assumed in our model.[4]

3 Incremental Updates and Context Accommodation

Our model distinguishes between context-independent and context-dependent updates, where the former operate independently of the current scenario and consist of a sequence of three stages involving context-independent interpretation, the downdating of obligations and intentions, and finally the updating of the latter structures depending on what kind of act has been performed. The stage of context-independent interpretation may affect several substructures of an information state. For example, we assume that conditionals which are associated with requests for evidence can be inferred at this point. Here we also determine the effects that acceptance acts have on the DPs' commitments. As for managing obligations, we assume that *assert* and *ask* acts introduce *address* and *answer* obligations respectively, and that these will be satisfied as soon as an act which matches their content appears in DH.

Our update rules for context-dependent interpretation, on the other hand, use the current

[3]Formally, the CSC attribute simply maintains a representation of the particular scenario which the IS represents as shorthand for the full definition.

[4]The acts marked with † in Figure 2 are backward-looking core speech acts.

Core Speech Acts

$assert(A, p)$	A asserts that p holds.
$ask(A, B, q)$	A asks B a question with content q.
$accept(A, m)$	A accepts the assertion performed in move m.†
$reject(A, m)$	A rejects the assertion performed in m.†
$accept\_answer(A, m, n)$	A accepts m as an answer to the question performed in n.†
$drop\_question(A, m)$	A drops the question performed in m.†

Argumentation Acts

$address(A, m)$	A expresses a statement regarding the assertion performed in m
$correct(A, m)$	A corrects the assertion performed in m.
$request\_evid(A, m)$	A requests evidence for m.
$answer(A, m)$	A means to provide an answer to m.
$info\_request(A, m)$	A asks a question in order to come up with an answer for m.
$reject\_answerhood(A, m, n)$	A rejects m as an assertion that resolves n.

Figure 6.2: Dialogue Acts

scenario when determining the argumentation acts performed in a move. In addition to the scenario and the core speech act *csa*, the rules can refer to the propositional or interrogative content of *csa* to assess the beliefs of the respective DP. This way, our rules allow us to distinguish the different ways the addressee of an assertion may incorporate a statement with respect to the latter's propositional content, for example by requesting evidence (*request_evid*) or by asserting the contrary (*correct*).

These two kinds of update rules are employed in the principle of incremental IS update outlined in (1) below in a way that implies a bottom-up/top-down management of processing: starting with the core speechact in a move m, we first determine the way the latter relates to the wider discourse context in which it occurs, assigning additional dialogue acts to m (bottom-up) by means of applying context-dependent updates. Once we have determined all the acts that have been performed we then apply the context-independent updates from top to bottom to each of the acts in m thus taking account of the fact that the way a core speechact influences obligations or intentional structure can only be determined accurately when we have considered the previous effects of a 'higher-order' argumentation act on the context. For example, the effect of an assertion can only be determined when we know whether it was also performing an answer act, or any other argumentation act, and when we already know the effect of this argumentation act on the context.

(1) **Incremental Update of Information States**

For any move m that occurs in a given scenario sc:

I Determine $CSA(m)$, the core speechact performed in m.

II If $CSA(m)$ is a forward-looking act: Interpret $CSA(m)$ in the context of *sc*.

III Apply the context independent update rules to any argumentation act $AA(m)$ that might result from the occurrence of m in *sc*.

IV Apply the context independent update rules to $CSA(m)$.

V Determine the new context that results from the occurrence of m.

Given this update strategy we can handle the two moves performed by B in [4] and [5] below in an intuitive way, assuming a two-fold evaluation of assertions that are meant to answer a question which involves the assessment of its 'assertive' and then its 'answerhood' properties:

(2) A[1]: Helen did not come to the Party.
B[2]: How do you know that?
A[3]: Her car wasn't there.
B[4/5]: Ok. But she could have come by bicycle.

In the context of A's assertion in [3], [4] is interpreted as an *accept* core speechact (**I**). As this is a backward-looking act, context-dependent interpretation is skipped. Applying the context-independent updates results in the obligation to *address* [3] being dropped, as well as in the satisfaction of A's intention that the absence of Helen's car be shared belief (**IV**). Having thus evaluated [3] as an assertion, the evaluation of its answerhood properties is still pending. In this new context, B's assertion (**I**) that Helen could have come by bicycle counts as a *reject_answerhood* argumentation act (**II**) which reintroduces the obligation to answer [2] (**III**), which had temporarily been dropped due to [3]. Additionally, [5] introduces an obligation on A to *address* it (**IV**), thus providing A with the possibility of initiating a discussion about its propositional content before getting back to dealing with B's request for evidence (**V**).

However, whereas the update strategy proposed in (1) works well for examples such as the one above in which for each scenario there is a move which fits in the given context, it fails to provide an appropriate account of cases like those in (3a) and (3b) below in which an assertion is acknowledged implicitly (without a move such as *ok* in example (2), which we assume expresses the addressee's acceptance of the propositional content):

(3) a. A[1]: Helen didn't come to the party.
B[2]: Did you see Jack?

b. A[3]: Helen's car wasn't there.
B[4]: She could have come by bicycle.

Given that the scenario created by an assertion is determined by the obligation on the addressee to respond to the content of the assertion (see Figure 2 above), the context dependent update rules will check whether the follow-up move expresses the addressee's acceptance, rejection or doubting of the truth of the relevant proposition. In the discourse in (4) below, for example, we can clearly assign a context-dependent interpretation to the dialogue acts in [2a] and [2b], classifying them as a correction and a request for evidence, respectively:

(4) A[1]: Helen didn't come to the party.
B[2a/b]: But I'm sure I saw her there/How do you know that?

In contrast to this, moves [2] and [4] in (3a) and (3b) above do not allow for a similar interpretation because the utterances do not evaluate the assertive content of [1] and [3],

respectively. However, reconstructing an intuitive reading of these examples, we can assume that [2] and [4] do in fact accept the preceding assertions simply because there is no reason to assume the contrary. In a similar way, in cases such as (5) below, in which an asker responds to an answer with a single *ok*, we can infer acceptance of the answerhood properties of the assertion from the acceptance of its propositional content and the absence of any hint that the asker continues to consider the question as unresolved:

(5) A[1]: Helen didn't come to the party.
B[2]: How do you know that?
A[3]: Her car wasn't there.
B[4]: Ok.

Where context dependent interpretation fails to assign an interpretation to a reply in terms of an evaluation of its assertive or question-resolving content, we can therefore assume that the act performed by the addressee expresses an implicit acceptance of the relevant aspect of the assertion. The question is then whether this assumption should be treated as a default rule to be included in the set of context-dependent update rules used in the model. With respect to this issue, consider once again our example (2), repeated with a slight change below:

(6) A[1]: Helen did not come to the Party.
B[2]: How do you know that?
A[3]: Her car wasn't there.
B[4]: She could have come by bicycle.

Here, after assuming that [4] implicitly accepts [3], a change of context takes place: having dealt with [3] as an assertion, the evaluation of its answerhood properties is still pending. It is in this context that a context-dependent interpretation (*reject_answerhood*) can be assigned to B's assertion in [4], and the effect of the explicit content of [4] has to be determined here. In the situation where acceptance is assumed for the sake of achieving an interpretation for some move m in a given context, m is thus retained for interpretation and is subject to our incremental update rules in the new context. We formulate this interpretation strategy in a generalised way as a principle of CONTEXT ACCOMMODATION:

(7) **Context Accommodation**

For any move m that occurs in a given scenario sc_i: if assignment of a context-dependent interpretation to m in sc_i fails, try to accommodate sc_i to a new context sc_{i+1} in an appropriate way by assuming implicit dialogue acts performed in m, and start interpretation of m again in sc_{i+1}.

Apart from being able to deal with implicit acceptance acts, we assume that the principle of context accommodation subsumes the process of QUESTION ACCOMMODATION which is described in (Cooper et al., 2000) as a formal means of dealing with the phenomenon of 'overanswering', as in the following example:

(8) A[1]: Where would you like to fly to?
B[2/3]: To Toronto. From Miami.

According to Cooper et al. (2000), B's move [3] can be interpreted appropriately if one reads it as an answer to a question such as *Which airport are you departing from?*. The context in which interpretation of [3] succeeds thus results from accommodating the context after [2] in a manner which fits the update procedure discussed above, namely by assuming an implicit *ask* act from A to which [3] is meant to provide an answer. The principle of context accommodation can thus be seen as a general means of interpretation which copes with the fact that DPs tend to produce smooth and concise expressions in natural discourse, missing out 'unessential' information.

4 Formalising the Update Model

We can summarise the results of the informal analysis outlined above as shown below, unifying the mechanisms of incremental IS update and context accommodation in a single update algorithm. The basic IS component we employ is the field IS.G.DH.LM ('latest move'; the pathnames refer to the IS fields shown in Figure 6.1), which holds all the dialogue acts assigned to a move. Note that the suggestion above that context accommodation involves retaining an action for which interpretation has failed, with interpretation pending, is reflected in the algorithm by keeping the action in LM after accommodation has taken place and then calling the algorithm once again from the beginning. In contrast to this, successful interpretation will result in the content of LM being merged with the set of previous dialogue acts:

(9) **The Update Algorithm**

1. Interpret m in the context of IS.G.CSC
   (a) Unless $CSA(m)$ has already been assigned:
      i. Determine $CSA(m)$
      ii. IS.G.DH.LM := $CSA(m)$
   (b) If $CSA(m)$ is a forward-looking act: Apply Context-Dependent Interpretation Rules‡
2. **if** interpretation succeeds:
   (a) $CI\_update(CSA(m))$
      i. Apply Context-Independent Interpretation Rules to $CSA(m)$
      ii. Apply Downdates Rules for obligations and intentions
      iii. Apply introduction Rules for obligations and intentions to $CSA(m)$
   (b) Resolve Conditionals‡
   (c) Update IS.G.CSC
   (d) $merge$(IS.G.DH.LM, IS.G.DH.DA)
   (e) IS.G.LM := nil

   else Apply Context Accommodation Rules‡
   (a) Resolve Conditionals‡
   (b) Update IS.G.CSC
   (c) goto step 1

‡ For any dialogue act DA added to IS.G.DH.LM: apply $CI\_update(DA)$.

The algorithm applies the context-independent update rules *CI_update* to each dialogue act separately in the reverse order of their introduction. This means that the context-independent update rules apply to acts which have been determined inferentially *before* the acts from which the latter have been inferred are processed. As in (Kreutel and Matheson, 2000), obligation and intention downdate and update rules only operate on OBL and INT respectively; however, context-independent interpretation can affect several substructures of an IS. For example, we assume that a conditional is associated with a request for evidence, and that this can be inferred at this point. Requests for evidence, acceptances and corrections are also interpreted here as acts which *address* an assertion and express the DP's attitude to its propositional content. Finally, this stage of interpretation determines the effects that acceptance acts have on the DPs' commitments. The current model uses four context-independent interpretation rules:

(10) **Context-Independent Interpretation**

$$\frac{\text{DH.LM} \vdash m : accept(A, n)}{\begin{array}{l} add(\text{DH.LM}, m : address(A, n)) \\ add(\text{SC}, shared\_belief(n')), \text{ where } n' \text{ is the content of the assertion in } n. \end{array}}$$

$$\frac{\text{DH.LM} \vdash m : correct(A, n)}{add(\text{DH.LM}, m : address(A, n))}$$

$$\frac{\text{DH.LM} \vdash m : request\_evid(A, m, n)}{\begin{array}{l} add(\text{DH.LM}, m : address(A, n)) \\ add(\text{DH.CONDS}, p : accept\_answer(A, o, m) \rightarrow p : accept(A, n)) \end{array}}$$

$$\frac{\text{DH.LM} \vdash m : accept\_answer(A, n, o)}{add(\text{SC}, resolved(o')), \text{ where } o' \text{ is the content of the question in } o.}$$

Particular dialogue acts trigger specific context dependent updates, and this is done on the basis of the update scenario which represents the context in which the move to be interpreted has been performed. The update algorithm specifies the current update scenario after processing each move or after context accommodation has taken place; as noted above, the CSC attribute stores the scenario information, and its contents are determined by the rules in (11) below:

(11) **Update Scenarios**

$$\frac{\text{OBL.}[1] = address(A, m)}{\text{IS.CSC} := respond\_assert(A, m)}$$

$$\frac{\text{OBL.}[1] = answer(A, m)}{\text{IS.CSC} := reply\_question(A, m)}$$

$$\frac{\begin{array}{l} \text{DH.LM} \vdash m : accept(A, n) \\ \text{DH.DA} \vdash n : answer(B, o) \end{array}}{\text{IS.CSC} := reply\_answer(A, n, o)}$$

Notice that in processing an *assert* act the update scenario which applies is *respond_assert*, and that this is true whether or not the assertion was discourse initial, or meant as an answer to a question, or for any other purpose. However, in order to deal with the acceptance of an

assertion it is necessary to check if the accepted act also constitutes an *answer* act, and if this is the case the current scenario becomes one where the ASKER evaluates the answerhood properties of the assertion.

We now can refer to the above scenarios in the process of determining the argumentation acts which have been performed, and the context-dependent interpretation rules in (12) are generally simpler as a result. Note that, in addition to referring to the scenario and the core speechact, the rules can also access the propositional or interrogative content of a core act in the process of assessing the beliefs of the relevant DP. This allows us to distinguish the different ways the propositional content of an assertion determines how the information should be incorporated into the IS by the addressee, for instance by requesting supporting evidence or by asserting the contrary proposition. Five context-dependent interpretation rules are currently assumed:

(12) **Context-Dependent Interpretation**

$$\frac{\text{IS.CSC} = reply\_question(A, m) \quad \text{DH.LM} \vdash n : assert(A, p)}{add(\text{DH.LM}, n : answer(A, m))}$$

$$\frac{\text{IS.CSC} = reply\_question(A, m) \quad \text{DH.LM} \vdash n : ask(A, B, q)}{add(\text{DH.LM}, n : info\_request(A, m))}$$

$$\frac{\text{IS.CSC} = reply\_assert(A, m) \quad \text{DH.LM} \vdash n : ask(A, B, q) \quad q \vdash Bel_A \diamond \neg m', \text{ where } m' \text{ is the content of the assertion in } m}{add(\text{DH.LM}, n : request\_evid(A, m))}$$

$$\frac{\text{IS.CSC} = reply\_assert(A, m) \quad \text{DH.LM} \vdash n : assert(A, p) \quad p \vdash Bel_A \neg m', \text{ where } m' \text{ is the content of the assertion in } m}{add(\text{DH.LM}, n : correct(A, m))}$$

$$\frac{\text{IS.CSC} = reply\_answer(A, m, n) \quad \text{DH.LM} \vdash o : assert(A, p) \quad p \vdash Bel_A \neg resolved(n'), \text{ where } n' \text{ is the content of the question in } n}{add(\text{DH.LM}, o : reject\_answerhood(A, m, n))}$$

The rules for context-dependent update above all require the performance of a core speech act, whereas the rules for context accommodation in (13) below simply assume that some move m has occurred. Given this basci assumption, we subsequently attempt to exclude the conditions which suggest that implicit acceptance may have taken place. If a core speechact *csa* has successfully been assigned to m this check is done based on the propositional or interrogative content of *csa*. Otherwise, if speechact assignment has failed, the check will always succeed. There are therefore two ways in which the interpretation of a move in a given context can fail (failure of speech act assignment or failure of context-dependent interpretation). We employ three accommodation rules in the current model, cf. (13).

(13) **Context Accommodation**

$\textsc{is.csc} = respond\_assert(A, m)$
$\textsc{dh.lm} \vdash o$
$o \not\vdash Bel_A \neg m'$
$o \not\vdash Bel_A \diamond \neg m'$

$add(\textsc{dh.lm}, n : accept(m))$

$\textsc{is.csc} = reply\_answer(A, m, n)$
$\textsc{dh.lm} \vdash o$
$o \not\vdash Bel_A \neg resolved(n')$, where n' is the content of the question in n
---
$add(\textsc{dh.lm}, n : accept\_answer(A, m, n))$

$\textsc{is.csc} = reply\_answer(A, m, n)$
$\textsc{dh.lm} \vdash nil$
$\textsc{dh} \vdash o : accept(A, m)$, where o is the last move that has been performed by A

$add(\textsc{dh.lm}, o : accept\_answer(A, m, n))$

The last rule applies in situations where no move has been performed (or where the turn is simply released by the turnholder even when the scenario predicts that the last speaker should retain the turn). In these circumstances this accommodation rule allows us to assume that the answerhood properties of an assertion are accepted, given the acceptance of its assertive properties as in example (14) below. In such cases a single move which accepts both aspects of an assertion seems a more natural option than the alternative in [3b/4] in which two utterances are used (although [3b/4] is of course perfectly possible):

(14) A[1]: Did Pete show up at the party?
B[2]: I don't know.
A[3a]: Ok.
A[3b/4]: Ok. Thanks.

5 Summary

In this paper we have outlined a formal framework for incremental information state updates in a dialogue model which uses discourse obligations as the basic means for dialogue control in subdialogues initiated by questions and assertions. Apart from accounting for the fact that the participants in a dialogue do act even in situations in which their behaviour cannot be explained in terms of intentions (see (Traum and Allen, 1994)), our definition of information state update scenarios has shown that representing the obligations imposed on the DPs as a stack structure can provide the necessary expressive means for determining characteristic states in the course of a dialogue in which DPs plan their own actions and interpret those of their conversational partner. In particular, we have shown that our update model allows us to reconstruct the reasoning processes that are involved in the interpretation of implicit acceptance acts in a well-defined fashion.

The main aspects of our analysis have been implemented using the TrindiKit dialogue move engine development environment (Larsson et al., 1999), where the actual algorithm employed is very close to the schematic process described above. In general, the TrindiKit allows a fairly faithful rendering of the theoretical approach outlined here, and as a result the implementation serves to verify a number of important aspects of the update model.

Bibliography

Bohlin, P., Cooper, R., Engdahl, E., and Larsson, S. (1999). Information states and dialogue move engines. In *IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson, A. (1996). HCRC dialogue structure coding manual. Research Paper 82, Human Communication Research Centre, University of Edinburgh.

Cooper, R. (1998). Information states, attitudes, and dialogue. In *ITALLC-98*.

Cooper, R., Engdahl, E., Larsson, S., and Ericsson, S. (2000). Accommodating questions and the nature of QUD. In *Götalog 2000, the 4th Workshop on the Semantics and Pragmatics of Dialogue*. University of Göteborg.

Cooper, R. and Larsson, S. (1999). Dialogue moves and information states. In *The Third IWCS, 1999*.

Kreutel, J. and Matheson, C. (2000). Obligations, intentions, and the notion of conversational games. In *Götalog 2000, the 4th Workshop on the Semantics and Pragmatics of Dialogue*. University of Göteborg.

Larsson, S., Bohlin, P., Bos, J., and Traum, D. (1999). TrindiKit 1.0 manual. TRINDI Deliverable 2.2, University of Göteborg, Sweden.

Matheson, C., Poesio, M., and Traum, D. (2000). Modelling grounding and discourse obligations using update rules. In *NAACL 2000*.

Poesio, M. and Traum, D. (1998). Towards an axiomatisation of dialogue acts. In *Twente Workshop on Language Technology*.

Traum, D. and Allen, J. (1994). Discourse obligations in dialogue processing. In *32nd Annual meeting of the Association for Computational Linguistics*.

Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999). A model for dialogue moves and information state revision. TRINDI Deliverable 2.1, University of Göteborg, Sweden.

Paper 7

Resolving Underspecification using Discourse Information

David Schlangen, Alex Lascarides, and Ann Copestake
University of Edinburgh; University of Cambridge and Stanford University
{das|alex}@cogsci.ed.ac.uk, aac@cl.cam.ac.uk
http://www.cogsci.ed.ac.uk/~das

Abstract

This paper describes RUDI ("Resolving Underspecification with Discourse Information"), a dialogue system component which computes automatically some aspects of the content of scheduling dialogues, particularly the intended denotation of the temporal expressions, the speechacts performed and the underlying goals. RUDI has a number of nice features: it is a principled approximation of a logically precise and linguistically motivated framework for representing semantics and implicatures; it has a particularly simple architecture; and it records how reasoning with a combination of goals, semantics and speech acts serves to resolve underspecification that's generated by the grammar.

1 Introduction

Our aim in this work is to investigate formally the interaction between compositional semantics, goals, and discourse structure in task-oriented dialogues. Specifically, we look at how an underspecified semantic representation may be instantiated by discourse information, and we investigate the extent to which we can preserve principled approximations of a general theory of dialogue semantics in a practical implementation for a restricted domain. To this end, we designed an experimental software dialogue system, RUDI.

As a testbed for this dialogue modelling, we chose the domain of fixing appointments, because we had access to a range of realistic dialogues that had been collected as part of the Verbmobil project Wahlster (2000) and to a parser which was capable of producing semantic representations for them (see below). We concentrate on dialogues that deal with the subtask of identifying a mutually agreed time to meet, ignoring other subtasks such as agreeing on a *place* to meet. The particular kind of underspecification we are investigating arises from the use of definite temporal descriptions in such dialogues. Example (1) shows an excerpt from such a dialogue:

(1) A: Can we meet on Friday?
B: How about 4pm?

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

We analyse such definite descriptions as requiring a bridging relation to an antecedent in the context.[1] Neither the bridging relation nor the antecedent are determined by the compositional semantics of the utterance, however. Thus, we take the semantic representation of such expressions to contain an underspecified relation between an underspecified antecedent and the referent for the expression.

A task that's co-dependent on resolving this underspecification is computing how the utterance contributes to a *coherent* dialogue. Following Segmented Discourse Representation Theory (SDRT, cf. e.g. Asher (1993); Lascarides and Asher (1993)), we assume that a dialogue is coherent just in case every proposition (and question and request) is *rhetorically connected* to another proposition (or question or request) in the dialogue, and all anaphoric expressions can be resolved. The rhetorical relations can be viewed as *speech act types* (see Asher and Lascarides (2001) for details), and they constrain both the semantic content of the utterances they relate, and what we call *speech act related goals* or SARGs.

Our thesis is that information can flow either from resolving the semantic underspecification to computing the rhetorical relation, or *vice versa* (and hence we're claiming rhetorical relations are an essential source of information for resolving semantic underspecification that's generated by the grammar). For example, the rhetorical relation which connects the utterances in (1) is inferred on the basis of the sentence moods (justification for this is given shortly), and the semantics of this rhetorical relation constrains the interpretation of *4pm* to be 4pm on Friday (as opposed to the alternative, which is the next 4pm to now).

The inference from linguistic form to the rhetorical relation (or equivalently, the speechact) is a *default* inference, however. Although the sequence of sentence moods in (3) is the same as in (1), the speechact's semantics is incompatible with all the possible resolutions of the temporal underspecification in (3).

(3) A: Let's meet next Saturday.
B: How about Sunday?

In such cases, RUDI has the capacity to explore whether an *indirect* speechact (ISA) has taken place; in this case, it will correctly predict that the illocutionary contribution of B's utterance is not simply that of the question, but it also conveys a *rejection* of A's proposal (to meet next Saturday).[2] So in this case, information flows from resolving the underspecification to

[1] We are interested here only in this one class of definite descriptions. (For a general classification cf. Hawkins (1978).) The term *bridging* was introduced by Clark (1975) for definite descriptions which lack a unique antecedent that is present on the basis of what has been explicitly said, and where thus the interpreter "... is forced to construct an antecedent, by a series of inferences, from something he already knows. [...] The listener must therefore bridge the gap from what he knows to the intended antecedent." (Clark, 1975, p.413)

In unrestricted domains, these bridging inferences can be quite involved and the reasoning is thus difficult to formalise generally (for an overview, see Vieira and Poesio (2000), but see also Asher and Lascarides (1998a)). We chose this domain partly so that we can exploit conventional constraints on the possible bridging relations among temporal expressions; for unlike other domains, the possibilities are finite. For example, a complicated nonce-relation like *the first interval in the antecedent that satisfies the description in the anaphoric expression* doesn't seem to be a possible bridging relation, even though it would provide us with a unique antecedent for the example below:

(2)

A: How about meeting in May?
B: #The Monday is good for me.

In fact, we believe that using only *temporal inclusion* and *next* as candidates for bridging relations is sufficient.

[2] We treat this as an ISA because rejections and questions are incompatible at the level of *semantic value*

inferring the type of speech act that B has performed (or equivalently, the rhetorical relation which connects his utterance to A's).

Dialogue (4) shows another example where the resolution of anaphoric expressions yields inferences about the speechacts. If *now* is Monday 12th February 2001, then *next week* is the interval from the 19th to the 25th, and from this we conclude B's speechact is to reject A's SARG. If, however, *now* is the 7th February 2001, so next week is the 12th to the 18th, then B's speechact narrows the temporal parameter in A's SARG, viz. the 12th to the 15th. Inferring these different speech acts thus requires knowledge of the times denoted (and the relationship between them).

(4) A: Can we meet next week?
B: I'm busy from the 16th to the 25th.

This work is part of a larger project, whose aim is to provide a computationally tractable and formally precise theory of how non-sentential fragments (e.g., *Not Tuesday*) are interpreted and generated. Therefore, we also need to *predict* when one can leave content implicit and when one can't. E.g., in (5), B's second utterance is odd. On the one hand, linguistic constraints on antecedents to anaphora stipulate that *4pm* should be resolved to Saturday 4pm Kamp and Reyle (1993). But on the other hand, one cannot infer any of the candidate rhetorical relations to attach this resulting interpretation of the question to the context. Details are given shortly, but roughly speaking, no rhetorical relation can be computed in this case because the semantics of the relations capture the intuition that B should not ask whether A can meet him on Saturday afternoon, when he knows (because A has told him already) that he can't meet him then.

(5) a. A: Can we meet next weekend?

b. B: How about Saturday afternoon?

c. A: I am busy then.

d. B: ??How about 4pm?

This contrasts with the question *Even at 4pm?*, which ameliorates the incoherence in (5). In contrast to *How about 4pm*, *Even at 4pm?* can be interpreted as a question which addresses the communicative goal of 'belief transfer' that underlies A's prior utterance; namely, the goal that B believe that A is busy on Saturday afternoon. This shows that reasoning about the linguistic constraints on the interpretation of anaphora, rhetorical relations and communicative goals are all necessary for an adequate account of the coherent interpretation of temporal expressions.

RUDI adopts a dynamic semantic approach to dialogue interpretation: First, a compositional semantic representation of the current clause is constructed via a large HPSG (the English Resource Grammar built in the LinGO project, as parsed by the LKB).[3] This representation is then used to update the semantic representation of the discourse context. The

(see Asher and Lascarides (2001) for details): a rejection is conveyed via a proposition, whereas a question denotes a set of propositions (ie. its direct answers; see Groenendijk and Stokhof (1984)). We give more details of this analysis in section 3.2. Note that we abstract away from intonational clues that a contrast is intended here (stress on Sunday), which presumably would be present if B's utterance were *spoken*.

[3]Henceforth, we will refer to this grammar/parser combination as ERG/LKB. The LinGO project is described on `http://www-csli.stanford.edu/hpsg/lingo.html`, the LKB on `http://www-csli.stanford.edu/~aac/lkb.html`. See also Copestake and Flickinger (2000).

co-dependent tasks of computing speech acts and goals and resolving semantic underspecification are a byproduct of computing this update. For this, we approximate SDRT.

In the next section, we will briefly introduce the relevant bits of this theory, and then explain in section 2.2 how we can derive a body of simpler domain-specific rules from this theory in a principled way. Section 3 describes the implementation of these rules. We close with a brief discussion of related work and some conclusions.

2 Theoretical Background

2.1 SDRT

SDRT represents discourse content as an SDRS, which is a recursive structure of labelled DRSs, with rhetorical relations between the labels. In contrast to traditional dynamic semantics (e.g., DRT, Kamp and Reyle (1993)), SDRT attempts to represent the *pragmatically preferred* interpretation of a discourse. Discourse update is formulated within a precise nonmonotonic logic, in which one computes the rhetorical relation (or equivalently, the speechact type) which connects the new information to some antecedent utterance. As mentioned in the introduction, this speechact places constraints on content and the speech act related goals or SARGs; these in turn serve to resolve semantic underspecification. Note that SARGs are goals that are either conventionally associated with a particular type of utterance or are recoverable by the interpreter from the discourse context; this distinguishes the goals that interact with linguistic knowledge from goals in general.

The rhetorical relations which are relevant to us here are the following:

Q-Elab(α, β) (Question Elaboration): β is a question where any possible answer to it elaborates a plan for achieving one of the SARGs of α. Eg. *A: Let's meet on Monday. How about 2pm?*

IQAP(α, β) (Indirect Question Answer Pair): α is a question and β conveys information from which the questioner can infer a direct answer to α. Eg. *A: Can we meet next week? B: I'm free on Monday.*.

Plan-Correction(α, β)**:** the speaker of β rejects the SARG of α. Eg. (4) in the first setting above.

Plan-Elaboration(α, β)**:** β elaborates a plan to achieve a SARG of α. Eg. (4) in the second setting.

Note that these speechact types are *relations* (cf. Searle (1967)), to reflect that the successful performance of the current speechact is logically dependent on the content of an antecedent utterance (e.g., successfully performing the speechact *IQAP*, as with any type of answering, depends on the content of the question α).

The default rules for computing speech acts have the form (6) ($A > B$ means *If A then normally B*):

(6) $(\langle \tau, \alpha, \beta \rangle \wedge \mathit{Info}(\tau, \beta)) > R(\alpha, \beta)$

$\langle \tau, \alpha, \beta \rangle$ means β is to be attached to α with a rhetorical relation (α and β label bits of content) where α is part of the discourse context τ; $\mathit{Info}(\tau, \beta)$ is a gloss for information about the content that τ and β label; and R is a rhetorical relation. This rule schema contrasts with the plan-recognition approach to computing speechacts (e.g. Lochbaum (1998)), which uses

only the goals of the antecedent utterance, rather than its compositional and lexical semantics directly, to constrain the recognition of the current speech act.

There are a number of advantages to allowing direct access to the content of τ in these inferences. For example, the successful performance of the current speechact is often dependent on the *logical structure* of the antecedent utterances, and goals don't reflect this logical structure; rather compositional semantics does (following DRT, Kamp and Reyle (1993)). In fact, dialogue (5) demonstrates this. Given the context, a SARG for (5d) is to find a time to meet that's next weekend but not on Saturday afternoon. So computing the speechact solely on the basis of the prior goals and the current linguistic form would predict that *4pm* successfully refers to 4pm on Sunday and the speech act $Q\text{-}Elab(5c, 5d)$ is performed. The fact that (5d) is odd indicates that recognising its speechact is constrained by something else. On our approach, the logical and rhetorical structure of (5a-c) plays a central role, for according to *linguistic* constraints defined within dynamic semantics (e.g., Kamp and Reyle (1993)), (5a-c) make Sunday inaccessible, thereby forcing *4pm* to denote 4pm on Saturday.

Some of the axioms of the form (6) are in fact derived via a formally precise model of cognitive reasoning, which encapsulates general principles of rationality and cooperativity (see Lascarides and Asher (1999) for details). For example, such cognitive modelling validates `Q-Elab` and `IQAP` (where $\alpha :?$ means that α is an interrogative):

- `Q-Elab`: $(\langle \tau, \alpha, \beta \rangle \wedge \beta :?) > Q\text{-}Elab(\alpha, \beta)$
  `IQAP`: $(\langle \tau, \alpha, \beta \rangle \wedge \alpha :?) > IQAP(\alpha, \beta)$

`Q-Elab` stipulates that the default role of a question is to help achieve a SARG of a prior utterance. `IQAP` stipulates that the default contribution of a response to a question is to supply information from which the questioner can infer an answer. Thus inferences about speechacts, and hence about implicit content and goals, can be triggered (by default) purely on the basis of sentence moods.[4] This justifies our analysis of (1) we gave above. Per default we take B's utterance to attach via *Q-Elab* to A's because it is a question. The semantics of this relation, viz. that the utterance helps elaborating a plan, is only met in this domain if it is true that the time β specifies is temporally included in the time α proposes. We add this information in discourse update so as to ensure that the updated logical form is consistent; and this thereby resolves the underspecification.

In an attempt to do justice to the complexity of interaction between the different information sources that contribute to dialogue interpretation—both conventional and non-conventional—many researchers have assumed a radically unmodular framework, so that a single reasoning process can access the different kinds of information at any time (eg. Hobbs et al. (1993)). In contrast, SDRT assumes a highly modular framework: reasoning about beliefs and goals is separate from, but interacts with, reasoning about content and speechacts. We will exploit this modularity so as to gain a particularly simple architecture to the implemented system.

Of course, world knowledge (WK) also affects interpretation. In this domain, relevant WK includes knowledge of which plans/actions when performed at time t are (in)compatible with meeting at t, and temporal reasoning with intervals and calendar terms. We'll discuss the former knowledge in the next section, and the latter in section 3.

[4]Since `IQAP` and `Q-Elab` are derived from axioms which model dialogue participants as rational and cooperative agents, one can view these rules as *short-circuiting* calculable implicatures about the content that the speakers intended to convey Morgan (1975).

2.2 Approximation

As we said in the introduction, our aim is to investigate the extent to which we can preserve principled approximations of the underlying theory, while maintaining a relatively good degree of robustness and precision. To this end, we make the assumption that the dialogue participiants (DPs) don't digress from trying to reach their main goal, which is to meet at a time t.[5] This means that we assume that all utterances address this goal, so that we can say that the main SARG of all utterances is to provide information about available times for a meeting.[6] The domain-level plan to reach this goal now can be specified as follows: the DPs have to "zero in" on a time, by narrowing down the range of times that are available for a meeting.

Having made this assumption, we can make approximations to the general theory on two levels. First, we approximate knowledge of which events permit meeting at time t and which don't via postprocessing the underspecified semantic form (the MRS[7]) generated by the ERG/LKB.

The result is an expression in a discourse input language (DIL), that preserves information about the temporal description of the time variable t, the sentence mood, and whether t was a good time or a bad time. Hence we abstract over information which is irrelevant to the task at hand, such as, for example, whether the utterance was about going to the dentist or going on vacation; they both generate *bad-time*(t).

This kind of postprocessing rule simply encapsulates knowledge of actions in the domain. Others are derived logically "off line" (ie. manually) in SDRT: for example, in this domain, SDRT validates the inference that asking a question about a time t implicates that it's a *good-time*(t) for the speaker to meet. The reasoning goes as follows. By default, a question attaches as *Q-Elab*. The semantics of this relation, namely that the question helps achieve a SARG of a prior utterance, is only met, given our additional assumption, if the utterance serves as a suggestion of a good time. This reasoning is 'hard-wired' into the post-processing rules, and thus we 'short-circuit' some SDRT inferences in the translation from MRS to DIL.

Approximation also occurs at the discourse level. First, we assume that the dialogue participants always believe the content of the other participants' utterances (i.e., the SARG of belief transfer that's conventionally associated with assertions is always successful). This means that questions which attach with *Q-Elab* to prior utterances are never interpreted as questions which elaborate a plan for achieving the SARG of belief transfer. In essence, this means that we assume that B won't utter *Even at 4pm?* in response to A's utterance (5d). Of course, this approximation is unjustified in general, but is acceptable in the restricted Verbmobil domain, since it is indeed the case that a dialogue agent assumes that the other agent is competent with respect to his assertions about when he can and can't meet.

Secondly, we utilize the assumptions about the overall purpose of these dialogues and the above approximations manually within SDRT, to yield the valid inferences that follow. In particular, the default rules of the form (6) yield monotonic rules of a similar form, since proviso the 'non-digression' assumption, exceptions to the defaults can be exhaustively enumerated.

By turning default rules into monotonic rules, we avoid computationally expensive con-

[5] This non-digression assumption is of course unfounded in the general case, but can be justified in our simple restricted domain.

[6] In the following, we will simply talk of these SARGs *being* a time t, which is to mean that the goal is to meet at a time within t.

[7] MRSs Copestake et al. (1999) are similar to (Reyle, 1993) UDRSs.

sistency checks. Also, fixing the main goal allows us to specify the semantics of the relations for this domain as follows (cf. the general rules above in Section 2.1 and the actual update rules the system uses in Fig. 7.3):

Q-Elab(α, β)**:** β is a question (which means it proposes a *good_time*, see above) and t_β at least overlaps with SARG_α, which makes sure that any possible answer addresses α's SARG.

IQAP(α, β)**:** α is a question and β talks about a time that overlaps with SARG_α.

Plan-Correction(α, β)**:** the speaker of β rejects the SARG of α, by marking a time as *bad_time* that includes SARG_α.

Plan-Elaboration(α, β)**:** β elaborates a plan to achieve a SARG of α, either by marking a time which overlaps with SARG_α as *good_time*, or by marking only parts of SARG_α as *bad_time*.

Another valid SDRT inference is a default rule for attaching to the previous utterance, because otherwise SARGs are left unaddressed, contrary to the cooperativity assumption (see Lascarides and Asher (1999)); we'll exploit this in RUDI when choosing the site to which the new information connects. Overall, then, we hope that this method of system development will ensure that all rules encoded in the software are logically and linguistically principled.

3 The System

3.1 Overview

RUDI's information state is shown in Fig. 7.1. Its main components are CONTEXT, which holds all information about the discourse context, and CUR-UTT, which represents the current utterance with which the context is to be updated.

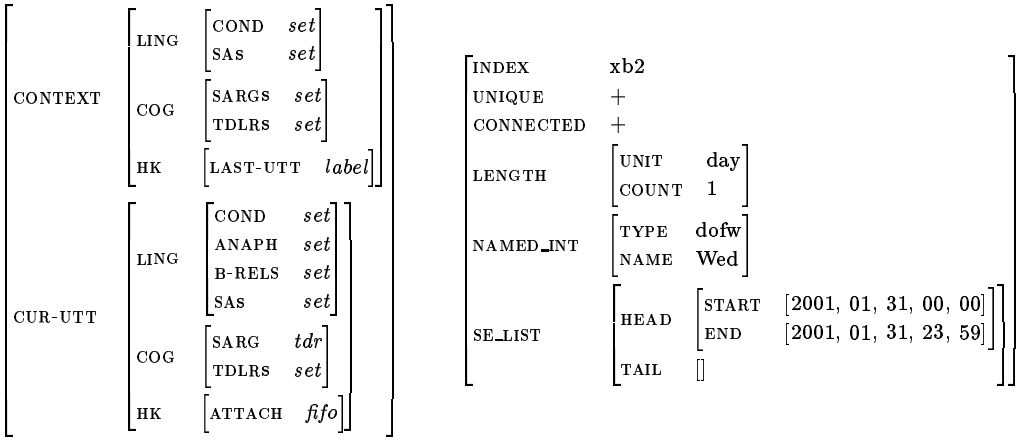


Figure 7.1: RUDI's information state (left) and a TDL-representation (right)

Both representations consist of a linguistic part (LING) and a cognitive part (COG). The linguistic part contains a set of conditions, ie. labelled predicates, and a list of the speechacts performed.[8] CUR-UTT additionally has fields to keep track of the anaphora and possible resolutions. The cognitive part represents information about cognitive states, viz. the SARGs

[8]This can be seen as being a notational variant of SDRT-style labelled boxes.

and the intended denotations in the domain, in the form of TDLRs. These are representations in a domain specific language, the temporal domain language (TDL). We encapsulate all knowledge about calendars, durations and intervals in this language; all domain specific reasoning takes places on these structures. Fig. 7.1 shows as an example a TDL representation for Wednesday 31st January 2001. The start- and end-points of the interval are specified in a list-structure, so as to allow representation of non-connected intervals. The feature HK in both parts finally holds information that is strictly speaking not part of a semantic representation; it is rather "housekeeping" information needed during the update process.

The modular architecture of the update process in RUDI reflects the high degree of modularity within SDRT. In particular, the update process is divided into different stages at which different classes of update rules are applied, as is shown schematically in Figure 7.2 below.

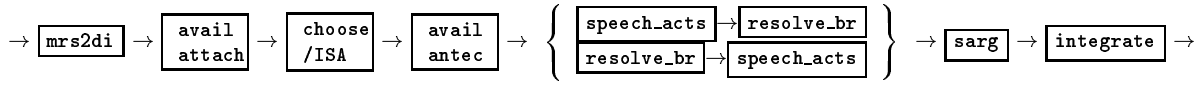


Figure 7.2: The algorithm

The initial stage translates the MRS of the chosen parse into the DIL semantic representation, which abstracts away from certain semantic details, as described above in Section 2.2.[9] At the next stage, an utterance in the context is chosen to which the current utterance can be attached via a rhetorical relation, and this in turn determines which antecedents are available. The preference is to attach to the prior utterance, as explained above. Under certain circumstances, the system tries to add content of an indirect speech act; we'll show how this works in the analysis of example (3) in the next section. The following two modules, `speech_acts` and `resolve_bridging`, interact in a special way: the system first tries to infer the speech-act and then uses this information to constrain the temporal bridging relation; if this doesn't succeed, RUDI tries to resolve the bridging relation first, using this additional information to then infer the speechact. Fig. 7.3 shows some of the update rules RUDI uses.[10] *qelab* and *iqap* are rules which don't need access to the intended model (as computed in the TDL), while *plan-correction* and *plan-elaboration* do. In the two latter rules, the function *resolve* relates the anaphor to an antecedent, and *tdr* tries to extend the model built so far (as represented by the TDLRs) so that it satisfies the new set of conditions. The speechact is inferred only if the model can be extended this way.

Including information about anaphora resolution in the antecedent to the rule for inferring *Plan-Elaboration* contrasts with the default rule in SDRT for inferring this speechact, which lacks this information. Adding this information to the antecedent of the monotonic rule is necessary for ensuring that all monotonically derived inferences about speechacts are mutually consistent; the temporal information we've included in the antecedent of the rule ensures that exceptions to inferring *Plan-Elaboration* are stipulated.

[9] At the moment, we have only implemented a few postprocessing rules that deal with our small test corpus. We expect to need lots more of these domain specific rules to extend coverage. Note, however, that the core *logical* rules in the other modules described here are complete as they are.

[10] Note that the rules are *monotonic*, as justified in section 2.2 above. The rule for infering *Plan-Elaboration* deals with the case where β expresses a strict interval of SARG_α as a bad time; a further rule for inferring *Plan-Elaboration* deals with the case where β expresses a good time.

Name:	**Question-Elaboration**
Preconditions:	CUR-UTT.HK.ATTACH $= \langle \alpha, \ldots \rangle$
	CUR-UTT.LING.COND $\supseteq int(\beta)$
Effects:	CUR-UTT.LING.SAS $\supseteq qelab(\alpha, \beta)$
	CUR-UTT.LING.COND $\supseteq temp\_overlap(\text{SARG}_\alpha, t_\beta)$
Name:	**Indirect Question-Answer-Pair**
Preconditions:	CUR-UTT.HK.ATTACH $= \langle \alpha, \ldots \rangle$
	CONTEXT.LING.COND $\supseteq int(\alpha)$
Effects:	CUR-UTT.LING.SAS $\supseteq iqap(\alpha, \beta)$
	CUR-UTT.LING.COND $\supseteq temp\_overlap(\text{SARG}_\alpha, t_\beta)$
Name:	**Plan-Correction**
Preconditions:	CUR-UTT.HK.ATTACH $= \langle \alpha, \ldots \rangle$
	CONTEXT.LING.COND $\supseteq good\_time(\alpha)$
	CUR-UTT.LING.COND $\supseteq bad\_time(\beta)$
	CUR-UTT.LING.COND $\supseteq prpstn(\beta)$
	CUR-UTT.LING.COND $\supseteq temp\_inc(t_\beta, \text{SARG}_\alpha)$
	$resolve$(CUR-UTT.LING)
	$\Gamma =$ CONTEXT.LING.COND $\cup$ CUR-UTT.LING.COND
	tdr(CONTEXT.COG.TDLR$s, \Gamma,$ CUR-UTT.COG.TDLRs)
	CUR-UTT.COG.TDLR$s \neq \perp$
Effects:	CUR-UTT.LING.SAS $\supseteq plan - correct(\alpha, \beta)$
Name:	**Plan-Elaboration**
Preconditions:	CUR-UTT.HK.ATTACH $= \langle \alpha, \ldots \rangle$
	CONTEXT.LING.COND $\supseteq good\_time(\alpha)$
	CUR-UTT.LING.COND $\supseteq bad\_time(\beta)$
	CUR-UTT.LING.COND $\supseteq prpstn(\beta)$
	CUR-UTT.LING.COND $\supseteq temp\_overlap(\text{SARG}_\alpha, t_\beta)$
	$resolve$(CUR-UTT.LING)
	$\Gamma =$ CONTEXT.LING.COND $\cup$ CUR-UTT.LING.COND
	tdr(CONTEXT.COG.TDLR$s, \Gamma,$ CUR-UTT.COG.TDLRs)
	CUR-UTT.COG.TDLR$s \neq \perp$
Effects:	CUR-UTT.LING.SAS $\supseteq plan - elab(\alpha, \beta)$

Figure 7.3: The `speech_act`-update rules

Finally, the goal behind the utterance is constructed from the resolved content, including the speechacts (cf. the rules in Table. 7.1). For example: $R(\alpha, \beta)$ (where R isn't plan-correction) and *plan-correction*(β, γ) entail that the SARG behind γ is to meet at a time which is: the time in the SARG of α minus the time specified in γ. The discourse update is coherent only if temporal information generated by linguistic content (e.g., `avail_antec`, `speech_acts` and `resolve_br`) is *consistent* with these 'cognitive' inferences in `sarg`. This consistency check fails for (5), since `avail_antec` constrains *4pm* to be on Saturday, and `speech_acts` constrains the speechact to be *Q-Elab*, but this speech act triggers an inference in `sarg` that *4pm* is 4pm on Sunday. Such inconsistency triggers backtracking, which may ultimately mean choosing an alternative parse for the current utterance ((5) is incoherent because no alternative parse is available): One of the goals of this system is to provide a principled symbolic way of choosing a sentence parse that, statistically, may be dispreferred.

3.2 Highlights of a few worked examples

This section shows RUDI at work for a few examples. The first is (7) below. The labels of the utterances are given in brackets, and the name of the main temporal referent and what it eventually resolves to according to the rules given in Figure 7.3 is also given:

(7)

($h1$)	A:	Can we meet next week?	$x1$	*next week*
($h2$)	B:	How about Tuesday?	$x2$	*Tuesday of next week*
($h3$)	A:	Two pm is good for me.	$x3$	*2pm on Tuesday of next week*
($h4$)	B:	I'm busy then.	$x4$	*2pm on Tuesday of next week*

We will show here how the context is updated by $h4$. At the point of processing $h4$, RUDI has computed the interpretations $x1$–$x3$ indicated above (which in the system are represented by TDLRs). It attached $h2$ to $h1$ via *iqap* and *qelab*, which means it computed that $h2$ gives an (implicit) positive answer to A's question and at the same time elaborates A's proposal (cf. the analysis of (1) we gave in section 1). Similarly, $h3$ is attached via *iqap* to $h2$. The SARG of $h3$ is to meet at 2pm on Tuesday of next week.

Figure 7.4: The MRS for "I'm busy then" (simplified)

Fig. 7.4 now shows the MRS representation of the compositional semantics of sentence $h4$ that is fed into the system, while Fig. 7.5 shows RUDI's information state after applying the update rules in `mrs2di`, `avail_attach`, `choose/ISA` and `avail_antec`.

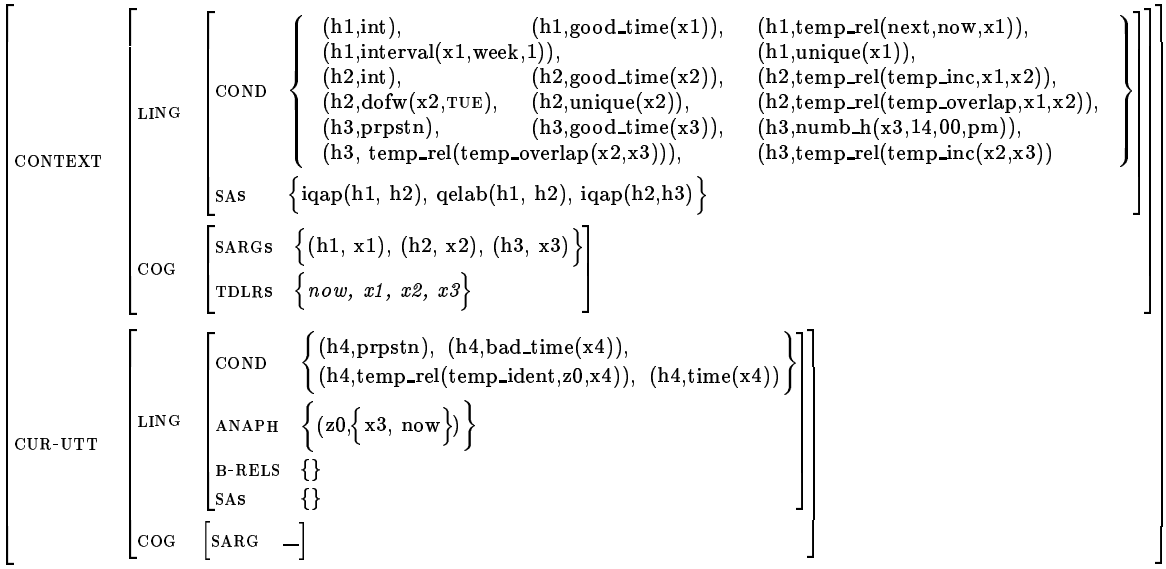


Figure 7.5: Adding "I'm busy then" to the information state

The 'previous utterance' attachment rule means RUDI attempts to attach $h4$ to $h3$, making $x3$ the available antecedent to $x4$. The lexical semantics of *then* constrains the bridging relation to be identity; thus the rule `Plan-Correction` applies, for $h4$ specifies a bad time that includes the good time from $h3$ (cf. Fig. 7.3 above). This updated content triggers inferences in `sarg` (cf. Table 7.1): $h4$'s SARG is $h2$'s SARG (since $h3$ was attached to this) minus $x4$; i.e., the SARG of $h4$ is to meet on Tuesday of next week, but not at 2pm (cf. Table 7.1). The interpretation of (4) is similar in the situation *now* is Monday 12th February 2001 (so *next week* is 19th–25th):

(8) 4 A: Can we meet next week?
B: I'm busy from the 16th to the 25th.

Name:	**Q-Elab**
Preconditions:	CUR-UTT.LING.SAS $\supseteq qelab(\alpha,\beta)$
Effects:	$t_\nu = \text{SARG}_\alpha \cap t_\beta$
	CUR-UTT.COG.SARGS $\supseteq \langle\beta, t_\nu\rangle$
Name:	**IQAP-good**
Preconditions:	CUR-UTT.LING.SAS $\supseteq iqap(\alpha,\beta)$
	CONTEXT.LING.COND $\supseteq good\_time(\beta)$
Effects:	$t_\nu = \text{SARG}_\alpha \cap t_\beta$
	CUR-UTT.COG.SARGS $\supseteq \langle\beta, t_\nu\rangle$
Name:	**IQAP-bad**
Preconditions:	CUR-UTT.LING.SAS $\supseteq iqap(\alpha,\beta)$
	CONTEXT.LING.COND $\supseteq bad\_time(\beta)$
Effects:	CUR-UTT.COG.SARGS $\supseteq \langle\beta, t_\nu\rangle$
	$t_\nu = \text{SARG}_\alpha \cap \overline{t_\beta}$
Name:	**Plan-Correction**
Preconditions:	CUR-UTT.LING.SAS $\supseteq plan\text{-}correct(\alpha,\beta)$
	CONTEXT.LING.SAS $\supseteq R(\gamma,\alpha)$
	$R \neq plan\text{-}correct$
Effects:	$t_\nu = \text{SARG}_\gamma \cap \overline{t_\beta}$
Name:	**Plan-Elab good**
Preconditions:	CUR-UTT.LING.SAS $\supseteq plan\text{-}elab(\alpha,\beta)$
	CONTEXT.LING.COND $\supseteq good\_time(\beta)$
Effects:	$t_\nu = \text{SARG}_\alpha \cap t_\beta$
	CUR-UTT.COG.SARGS $\supseteq \langle\beta, t_\nu\rangle$
Name:	**Plan-Elab bad**
Preconditions:	CUR-UTT.LING.SAS $\supseteq plan\text{-}elab(\alpha,\beta)$
	CONTEXT.LING.COND $\supseteq good\_time(\beta)$
Effects:	$t_\nu = \text{SARG}_\alpha \cap \overline{t_\beta}$
	CUR-UTT.COG.SARGS $\supseteq \langle\beta, t_\nu\rangle$

Table 7.1: The `sarg`-update rules

Let's now look at an example where the content of an indirect speech act has to be computed explicitly. We proposed earlier that B's response in (3) should be analysed as an implicit plan-correction; ie. B's question tacitly rejects A's SARG.

(9) 3 A: Let's meet next Saturday.
B: How about Sunday? (↖ ISA: *That is bad for me.*)

To account for this analysis, we have to compute the content of this implicit speechact. This is done as follows. Suppose we have labelled A's utterance $h1$ and B's $h2$, and their temporal referents $x1$ and $x2$. Based on the linguistic clue "sentence mood", `speech_acts` infers *q-elab* (as it does for example (1)). However, this speechact has as monotonic consequence that the temporal relation $temp\_overlap(x1, x2)$ must hold (see above section 2.2), and this is rejected by the `tdr`. None of the other rules fire, and therefore we have to backtrack. On entering `choose/ISA` via backtracking, we call the indirect speech act module. In this module we have update rules specifying that two *good_time*s that do not temporally overlap trigger the introduction of a new label, which we will here call $h2'$, with the content that an explicit *That is bad for me* would get.[11] The dialogue is then processed with this additional content, which means that we infer $plan\text{-}correct(h1, h2')$ and $q\text{-}elab(h2', h2)$.

The reason that we compute the (labelled) content of the indirect speech act explicitly in this case is because plan-corrections are constrained to take propositions as their second arguments (since they are a kind of assertion); see Asher and Lascarides (2001) for details. Generating this content explicitly allows us to capture rhetorical relations between the indirect speechact and the 'direct' one that could not be captured otherwise. This contrasts with the case of indirect answers, which can be of any sentence type (e.g., a question or a request can entail content from which the interpreter can compute a direct answer, as demanded by the semantics of *IQAP*).

B's question in (3) contrasts with the question (5d), which cannot be interpreted as an indirect speechact of plan-correction. This is because A has already stipulated that Saturday afternoon is a bad time (for him). And, informally, the module `choose/ISA` fails to generate a coherent interpretation in this case, to reflect the fact that when B wants A to revise his assessment of t as a bad time, he needs to do this explicitly (we forego stipulating the formal rule here). So, for example, inserting the plan-corrective move *B: But I would much prefer to meet you on Saturday afternoon* between (5c) and (5d) ameliorates the incoherence (note that (5d) would attach to this explicit *plan-correction* with *q-elab*). `choose/ISA` failing to provide a discourse update triggers further backtracking; an alternative parse of the sentence must be chosen, but there isn't one, thereby yielding discourse incoherence.

4 Related Work

Stede et al. (1998) compute the temporal content of scheduling dialogues in German. Their approach to representing the temporal domain is similar, but they don't offer principled constraints for resolving anaphora. Wiebe et al. (1998) adopt a data-intensive approach to interpreting temporal expressions. We are, however, also interested in predicting when a definite description is coherent and when it's not, which this approach doesn't do.

Interpreting questions and their answers is crucial in this domain. Traum et al. (1999) analyse questions and answers by implementing the QUD-model within the TRINDI dialogue

[11] These ISA-rules have to be constrained carefully, since there seem to be strong conventional constraints on how such an indirect plan-correction can be conveyed. For example, there must be contrasting elements present, which explains why substituting B's utterance in (3) with "How about the 15th?" would make the dialogue sound a lot worse. Investigating the exact nature of these constraints remains as future work.

management system. The QUD model constructs an ordered stack of *questions under discussion*, which determines what utterances would be (currently) felicitous. Cooper et al. (2000) develop a method of "question accommodation" to deal with cases where felicitous, indicative utterances provide information that doesn't answer any question on the stack. It seems, however, that even in our domain, additional mechanisms to this are needed to account for some implicatures. Even for a simple exchange like (1), the QUD model as it stands predicts that two questions are on the stack; however, it fails to model that B's intention in (1) was not simply to ask a question, but also to implicate an *answer* to A's question (in the positive). It fails to detect this because the QUD model doesn't reason about the second question's rhetorical function in the context of the first question. This gap in the theory also means that the rule for accommodating questions overgenerates. Because the accommodated question need not be rhetorically linked to the existing QUDs, B's utterance in (10) can trigger the accommodation of a question like "On which day can we meet?", thereby predicting (10) is acceptable, contrary to intuitions:

(10) A: [said on the 1st] Can we meet next week?
B: ??The 20th is fine.

It is quite likely that the QUD-model could be extended to overcome these problems. However, we hope that by allowing access to a richer discourse structure than a stack of questions, we will constrain the necessary inferences in a more effective manner.

5 Conclusion

We have developed a system which explores the information flow between recognising speechacts, inferring the underlying goals of utterances and resolving semantic underspecification that's generated by the grammar within the domain of scheduling dialogues. The main feature of the system was to approximate a logically precise theory of the semantic and pragmatic interpretation of discourse, by making assumptions that DPs don't digress from the main goal, that they always believe each other, and by 'short-circuiting' reasoning about domain-level plans to meet (e.g., that you can't meet and go to the dentist at the same time) within a post-processing module. This allowed us to encode within the system the simpler and more computationally tractable axioms that are derived (manually) from these assumptions within the underlying logical theory. We aim eventually to test the extent to which the nonmonotonic reasoning that generally underpins computing implicatures can be made monotonic in relatively restricted domains, and to apply the result to the processing of fragments. We actually believe that the monotonic approximation of the theory will be pushed to its boundaries even in the very simple domain we've chosen here, thereby demonstrating default reasoning is an essential component to any realistic, rule-based dialogue system.

Acknowledgements

We would like to thank the Dialogue Systems Group Edinburgh for helpful discussion. This research was partially supported by the National Science Foundation, grant number IRI-9612682 to Stanford University. Alex Lascarides is supported by an ESRC (UK) research fellowship.

Bibliography

Allen, J. F. (1984). Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–152.

Allen, J. F. (1991). Time and time again: The many ways to represent time. *International Journal of Intelligent Systems*, 6(4):341–355.

Allen, J. F. and Ferguson, G. (1994). Actions and events in interval temporal logic. Technical Report 521, University of Rochester, Computer Science Department.

Asher, N. (1993). *Reference to Abstract Objects in Discourse.* Studies in Linguistics and Philosophy. Kluwer Academic Publisher, Dordrecht.

Asher, N. and Lascarides, A. (1998a). Bridging. *Journal of Semantics*, 15(1):83–113.

Asher, N. and Lascarides, A. (1998b). Questions in dialogue. *Linguistics and Philosophy*, 23(3):237–309.

Asher, N. and Lascarides, A. (2001). Indirect speech acts. *Synthese.* to appear.

Busemann, S., Oepen, S., Hinkelman, E. A., Neumann, G., and Uszkoreit, H. (1994). Cosma - multi-participiant nl interaction for appointment scheduling. Research Report RR-94-34, Deutsches Forschungzentrum für Künstliche Intelligenz GmbH.

Clark, H. (1975). Bridging. In Schank, R. and Nash-Webber, B., editors, *Theoretical Issues in Natural Language Processing.* MIT Press, Cambridge, Mass.

Cooper, R., Engdahl, E., Larsson, S., and Ericsson, S. (2000). Accommodating questions and the nature of QUD. In *Proceedings of Götalog 2000*, Gotheburg.

Copestake, A. and Flickinger, D. (2000). An open-source grammar development environment and broad-coverage english grammar using HPSG. In *Proceedings of the 2nd Linguistic Resources and Evaluation Conference*, pages 591–600, Athens, Greece.

Copestake, A., Flickinger, D., Sag, I., and Pollard, C. (1999). Minimal recursion semantics: An introduction. Stanford University, Stanford, CA.

Ginzburg, J. (1995). Resolving questions i. *Linguistics and Philosophy*, 18:459–527.

Groenendijk, J. and Stokhof, M. (1984). *Studies on the Semantics and Pragmatics of Questions.* PhD thesis, Centrale Interfaculteit, Amsterdam.

Haas, S. (1999). Getypte Merkmalsstrukturen und Unifikation zur Verarbeitung temporaler Ausdrücke in Verbmobil. Verbmobil Report 234, Technische Universität Berlin.

Hawkins, J. (1978). *Definiteness and Indefiniteness.* Croom Helm.

Hobbs, J. R., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as abudction. *Artificial Intelligence*, 63:69–142.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: Introduction to Model-theoretic Semantics, Logic and Discourse Representation Theory.* Kluwer Academic Publishers.

Lascarides, A. and Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16(5):437–493.

Lascarides, A. and Asher, N. (1999). Cognitive states, discourse structure and the content of dialogue. In *Proceedings to Amstelogue 1999*.

Lochbaum, K. (1998). A collaborative planning model of intentional structure. *Computational Linguistics*, 24(4):525–572.

Morgan, J. L. (1975). Some interactions of syntax and pragmatics. In Cole, P., editor, *Syntax and Semantics Volume 9: Pragmatics*, pages 261–280. Academic Press.

Poesio, M. and Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216.

Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics*, 10:123–179.

Searle, J. (1967). *Speech Acts*. CUP.

Stede, M., Haas, S., and Küssner, U. (1998). Tracking and understanding temporal descriptions in dialogue. Verbmobil Report 232, Technische Universität Berlin.

Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999). A model of dialogue moves and information state revision. Trindi Deliverable D2.1, University of Gothenburg.

TRINDI consortium (1998). Trindi: Task oriented instructional dialogue. http://www.ling.gu.se/research/projects/trindi/.

Uszkoreit, H., Backofen, R., Busemann, S., Diagne, A. K., Hinkelman, E. A., Kasper, W., Kiefer, B., Krieger, H.-U., Netter, K., Oepen, S., and Spackman, S. P. (1994). Disco - an hpsg-based nlp system and its application for appointment scheduling. Research Report RR-94-38, Deutsches Forschungzentrum für Künstliche Intelligenz GmbH.

Vieira, R. and Poesio, M. (2000). An empirically-based system for processing definite descriptions. *Computational Linguistics*, 26(4).

Wahlster, W., editor (2000). *Verbmobil: Foundations of Speech-to-Speech Translation*. Artificial Intelligence. Springer, Berlin, Heidelberg.

Wiebe, J. M., O'Hara, T. P., Öhrström-Sandgren, T., and McKeever, K. J. (1998). An empirical approach to temporal reference resolution. *Journal of Artificial Intelligence Research*, 9:247–293.

Paper 8

Constraining Pronouns with Optimality Theory in Several Languages

Henk Zeevat, Sofia Gustafson-Čapková, and Jennifer Spenader
Henk.Zeevat@hum.uva.nl
www.hum.uva.nl/computerlinguistiek/henk Sofia@ling.su.seWWW.LING.SU.SE/STAFF/SOFIA
Jennifer@ling.su.se WWW.LING.SU.SE/STAFF/JENNIFER

Abstract

This paper describes a pilot study that looks at how 5 different languages generate and interpret pronouns and proper names. The languages studied (Chinese, Czech, English, Finnish and Japanese) are typologically diverse and we attempt to assess the relevance of an recent OT-treatment of pronouns in English. The study is part of a new project on comparative discourse in Optimality Theory[1]

1 Introduction

We propose a systematic investigation of discourse phenomena from an empirical and comparative angle. Recently there have been a number of studies on either discourse phenomena themselves (Zeevat(1999), Blutner (2000), Beaver(MS), Mattausch (2001)) or studies that touch on discourse phenomena (Bresnan (MS), Aissen, De Hoop) all using optimality theoretic principles in the belief that this gives better insights into the processes studied. These studies have used traditional methodology of theoretical linguistics, often restricted to English and with made up examples (the first group) or taking a comparative point of view (the second group). One can say that the first group misses the necessary comparative view for establishing OT- constraints and their ordering possibilities whereas the second group though touching on discourse phenomena as such (mostly pronouns and topicality) fails to give an account that in itself is a proper discourse theory, in the sense of a theory of interpretation for the relevant phenomena to the standards of e.g. discourse representation theory.

But there are two further omissions. A large part of the data in this area is essentially soft: there are preferences for expressing things in one way rather than another and there are preferences for interpreting things one way rather than another. It is not a question of one thing being good and the other bad. This is not made clear in either of the two approaches sketched. And there is a second problem: discourse phenomena are intimately connected.

[1]Contact the authors for more information about the project.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

Henk.Zeevat@hum.uva.nl
www.hum.uva.nl/computerlinguistiek/henk Sofia@ling.su.sewww.ling.su.se/staff/sofia
Jennifer@ling.su.se www.ling.su.se/staff/jennifer

It is not possible to isolate tense from pronouns, or from discourse relations, topic-focus articulation from pronouns or pronominal reference from presupposition inducers including particles. The latter group are, for example, inseparable from the expression of speechacts and their meaning varies considerably with the speech act expressed.

Our project aims to collect a range of data gathering that could support both types of work. We want to arrive at a suitably aligned translingual corpus about the discourse phenomena in question that supports the gathering of comparative statistical data (e.g. in this situation in 67% percent of the cases language X uses a pronoun and in 20% drops it and in language B this is 23% versus 26%). These data represent a new class of facts that linguistic theories should have an explanation for.

For obtaining the most accurate data here, it is necessary to use elicitation techniques[2] But we will investigate the possibility of using translated dialogue by a systematic comparison between data obtained that way and data coming out of suitable elicitation methods. To the extent that data serve as a test bed for theoretical explanations, the accuracy of the statistics may be less important, as is indeed the standard assumption in much theoretical work in which statistics is not even taken into account.

The pilot study we present in the next pages takes the key examples of Mattausch 2001 -a further development of Beaver (MS) which is an OT reconstruction of the centering approach to pronoun resolution and generation- and tries to see how the approach fares in a number of other languages. The languages looked at were typologically diverse: Chinese, Czech, English, Finnish and Japanese. Part of the data was originally obtained from informants through translation and discussion, and has now been validated by a web-based elicitation task. The results are nevertheless tentative: the number of respondents is quite small and the settings in which they were interviewed were not completely uniform.

2 Mattausch on English Pronouns

Mattausch (2001) gives a treatment of English pronouns. This work is an further development of Beaver (MS). We take it as a starting point here only. A brief summary of this work follows below. For full argumentation the reader is referred to Mattausch (2001).

The system is version of bidirectional OT (Smolensky (1996) , Blutner (2000) i.e. a system that consists of generation contraints (as in OT syntax, selecting the optimal realisation for a semantic input), interpretation constraints (selecting the best interpretation for a syntactic form, as proposed by De Hoop and Hoekstra (2001)) and a connection theory holding these two together. (e.g the back and forth method proposed by Smolensky 1996, the various superoptimalities proposed by Blutner, the constraint approach proposed by Beaver (MS), Mattausch (2001), the more complicated models of Smolensky & Wilson and Zeevat (MS) for dealing with asymmetries.).

Mattausch has the following system of soft generation constraints: **MARK** $\leftrightarrow$ **SHIFT** $<>$ **MARK** $\leftrightarrow$ **PAR** $>$ **ECON** $<>$ **PRON** $\leftrightarrow$ **TOP** $<>$ **SYMMARK**

MARK $\leftrightarrow$ **SHIFT** is the principle that when the antecedent of the referring expression occupies a different syntactic function, the expression must be marked. Marked expressions are the stressed version of the pronoun or the full definite with a lexical noun.

[2]For the test used here, see http://www.ling.su.se/staff/jennifer/undersoekenglish.html. The project will be coordinated through a project web site and aims to, among other things, organize workshops and share resources.

MARK $\leftrightarrow$ **PAR** enforces markedness on all NPs if the left sister is parallel, i.e. based on the same predicate. The idea is that the combination of parallelism and shift causes markedness. One constraint on its own will lead to a violation of the other. **ECON** prohibits marked NPs and **PRON** $\leftrightarrow$ **TOP** says that pronouns must be referring to the subject of the last sentence. (This is essentially Beaver's constraint **PROTOP**.) **SYMMARK** finally requires that the marking of two different NPs in the same clause is the same: either they are both stressed pronouns or both full NPs.

In the interpretation direction there is one newcomer:**FAMDEF**. It tells us to interpret definite NPs as referring to familiar entities. The soft constraints give the following system:

FAMDEF $>$ **MARK** $\leftrightarrow$ **SHIFT** $>$ **ECON** $<>$ **PRON** $\leftrightarrow$ **TOP**

MARK $\leftrightarrow$ **SHIFT** wants shifted antecedents for marked definites, unshifted antecedents for unmarked ones. **ECON** asks us to interpret full NPs as new and **PRON** $\leftrightarrow$ **TOP** tells us to interpret only pronouns as referring to the subject NP of the left sister. (This just spells out the interpretational import of the coresponding generation constraints. The constraints that disappear have no clear interpretational meaning.)

The following definitions tells us how the generation constraints interact with the interpretation constraints.

F is an *optimal form* for M is defined in the usual way, using the system of generation constraints.

M is a *properinterpretation* of F if either M is an optimal interpretation for F or otherwise, but F is an optimal form for M and no form G that is better interpreted as M is an optimal form for M.

A form F is now a *proper realisation* of M iff F is an optimal form for M and M is a proper interpretation of F.

With these generation and interpretation systems and the definitions in place, we can give explanations.

(1) John had an argument with Bill. He lost.

The *he* will be interpreted (by **PRO** $\leftrightarrow$ **TOP**) as the subject of the first sentence. If the antecedent were Bill, generation would give us instead of the pronoun *he* the full name *Bill*, which, though it violates the weaker constraint **ECON**, does not violate the stronger **MARK** $\leftrightarrow$ **SHIFT**. So by bidirectionality, Bill is not a proper antecedent of *he*. That means that $lost(b)$ is not a proper interpretation of of *He lost.* and that the possible interpretation is $lost(j)$.

(2) John caught a fish for Mary.
She cooked it for him.
Mary cooked it for him.

It is wrong to put stress on any of the pronouns, to have *John* instead of *him* or *the fish* instead of *it*. The wrong examples all violate **MARK** $\leftrightarrow$ **PAR**, since these are not parallel sentences (they mark although they are not parallel) and **ECON**, though for John, it also means one less transgression of **MARK** $\leftrightarrow$ **SHIFT**. Semantically here they are all fine thanks to agreement facts. They are wrong generations, but their optimal interpretation is the one intended. This changes in (3) where we have parallelism.

(3) John gave a book to Mary.
Mary had given the book to John.
Mary had given it to John.
Mary had given the book to HIM.
SHE had given it to HIM.
SHE had given the book to HIM.

The difference with the previous example is the parallelism, which now also allows a full form for *the book* and emphatic pronouns.

(4) A man saw Bill.
*He ran away.
*HE ran away.
Bill ran away.

HE ran away. loses out in the generation due to the absence of parallellism. Due to topicality, *A man* is the antecedent for *He ran away.*. So *Bill ran away.* is the proper realisation when Bill is the one who ran away.

3 Method

First, translation equivalents of the English data in Mattausch(2001) were obtained from at least two native informants for each of the languages studied. Thereafter an elicitation task based on the original data, but placed in a context to make it more natural, was designed. This elicitation task is meant to test some of the hypotheses about pronoun usage made after the work with informants and is still an early stage.

The elicitation task consists of 9 short stories. Each story has the following structure, a discourse segment of 2 or more sentences that sets the initial context. This is followed by a second discourse segment, that was marked in someway, for example by a rhetorical marker such as "But", or by changing the focus. This second discourse segment had two sentences. The first sentence is the set-up sentence, and the second sentence is the target sentence, but the target sentence was given in the form of a photo, and this was what subjects were asked to produce.

One example: Someone had erased Peter's hard disk on purpose. Peter was sure he knew who it was. He went to Ken's office.Peter saw Ken. ¡Target Sentence Picture¿ Competition between researchers can sometimes become violent.

4 Pronouns in Chinese, Czech, Finnish and Japanese

Chinese pronouns Two native speaker informants were consulted to produce translation data. Chinese, like many other languages, can drop the subject pronoun. This happens in topic-chaining sequences and in certain subordinate constructions.

(5) John jin lai le, tuo diao mao zi, zuo le xia lai.
John come in la-PAST, take off hat, sit la-PAST down.
John came in. He took off his hat and sat down.

It is not completely clear what governs the omission of the pronoun in Chinese. It is does not seem to be like Spanish or Italian where all topic subjects are omitted. One possible generalization may be that in (some versions of) spoken Chinese omitted subjects are the protagonist of a story. A consequence is that subject pronouns must appear if the topic chain is broken, e.g. by an elaboration, even if the topic is maintained.

(6) John zhua la tiao yu ge Mary. Mary ba yu ge ta zuo la.
John catch la CL fish give Mary. Mary let fish give ta make la.
John caught a fish for Mary. Mary cooked it for him.

Putting the pronoun for the subject in the second sentence, would make it refer to John rather than Mary.

(7) Yo ge ren kai jian la Bill. Bill ma shangpao kai la.
Yo CL man see Bill. Bill immediately run away la.
There is a man who saw Bill. Bill ran away.

If we would use *ta* instead of *Bill* in (7) or the empty subject, it would refer to the man and not to Bill.

SYMMARK is not visible, as it would seem to lead to a confusion of reference. The first three versions of the first example are fine, the fourth one is out, clearly for the reason that the first *ta* would be John.

(8) John ai Mary. (Dan zhe) Mary bu ai John.
John ai Mary. (Dan zhe) Mary bu ai ta.
John ai Mary. (Dan zhe) ta bu ai John.
John ai Mary. (Dan zhe) ta bu ai ta.
John love Mary. (But) Mary not love John.

All that we clearly confirm is the presence of the constraint **PRO** $\leftrightarrow$ **TOP** and the hard constraint **FAMDEF**, although there is a problem: unmarked NPs like *yu* (fish) have a preference for interpretation as an old discourse marker. It is also correct to assume **ECON**. But **MARK** $\leftrightarrow$ **SHIFT**, **MARK** $\leftrightarrow$ **PAR**, and **SYMMARK** are not visible. Omitting these constraints however deals well with the data. In addition, we need an explanation of the empty pronouns. An option is ***SUBJ & PROTAGONIST** as a generation constraint, disallowing overt subjects that refer to the protagonist (an input feature). This however does not do justice to the fact that subject pronouns for the protagonist are an option (even full NPs). Better is an interpretation constraint: $\emptyset$**&SUBJ** $\rightarrow$ **PROTAGONIST** with **ECON** responsible for the empty realisation. This blocks empty realisations if the referent is not a protagonist and allows additional explanations for cases where full NPs are used for subject protagonists.

Czech Pronouns

Czech is an inflecting fusional language with an elaborated pronoun system. In Czech there are full pronouns and reduced pronouns. The reduced pronouns are used in unmarked contexts and occupy the position after the first stressed item in the sentence, while the full pronouns are used where the pronoun is stressed or occurs after a preposition. Czech frequently drops first, second and third person pronouns. Person is then expressed by verb congruence. Another feature of Czech which will prove to be important in the examples is

that Czech has only one past tense, i.e. it lacks the difference between perfect and imperfect past tense which is found in e.g. English.

Pronouns can be dropped for persons and objects that are present but alos for old subjects and objects.

(9) I saw Lena yesterday. She did not look very happy.
Videla jsem Lenu vcera. Nevypadala moc štastná.
see-past-sg-fem 1sg-be-pres Lena-accusative yesterday. not-look-past-sg-feminine much happy-sg-fem.

The most salient comment from the informants was on the unnaturalness of the examples. Another important point was that in many cases the informants had difficulties in maintaining the sentence structure of the examples. They felt a need to combine single sentences into complex sentences. This is shown in example (10).

(10) John kissed Mary /John smiled/

Czech: John polibil Marii. /John se usmál/
Good: John políbil Marii. Usmál se.
John kiss-past-3sg-mas Mary-accusative. Smile-past-3sg-mas reflexive.
Preferred: John políbil Marii, pak se usmal.
John kiss-past-3sg-mas Mary, then reflexive smile-past-3sg-mas

The most favoured way to express John in the second sentence or clause was with a zero pronoun.

(11) John kissed Mary. /Mary slapped John/.
Czech: John pol´ıbil Marii. /Marie Johnovi dala facku./
John kiss-past-sg-masculine Mary-dative. Mary John-dative give-past-sg-feminine slap.
Not good: John pol´ıbil Marii. Marie mu dala facku.
John kiss-past-sg-masculine Mary-dative. Mary on-dative-nonstressed give-past-sg-feminine slapaccusative-feminine.
Preferred: John pol´ıbil Marii a za to dostal facku.
John kiss-past-sg-masculine Mary-accusative and for that get-past-sg-masculine slap-accusative-feminine.

In the bad case above **PRO** $\leftrightarrow$ **TOP** is respected in both directions, but the preferred sentence changes the structure and more strongly expresses the causal relation. There is no explanation for this in the set of constraints we are considering here. Perhaps there is a principle like ***SHIFT** at work against changes of topic. Also in the examples below the informants preferred to restructure the message.

(12) English: Fred dates Mary. /Fred loves Mary/.
Czech: Fred chodí s Marii. /Fred Marii miluje./
Fred go-pres-3sg with Mary-instrumental. Fred Mary-accusative love-pres-3sg.
Bad/Pointless: Fred chodí s Marii. Fred ji miluje.
Fred go-pres-3sg with Mary-instrumental. Fred she-accusative-nonstressed love-pres-3sg. Preferred: Fred chodí s Marii. Moc ji miluje.
Fred go-pres-3sg with Mary-instrumental. Much she-accusative-nonstressed love-pres-3sg.

Here "chodí" and *dates* are not completely equivalent.

(13) English: Fred fought Bill. /Fred won/.
Czech: Fred se pral s Billem. Fred vyhrál
Ambiguous: Fred se pral s Billem. On vyhrál.
Fred reflexive fight-past-masculine with Bill. He win-past-sg-masculine
Preferred: Fred se pral s Billem a vyhrál
Fred reflexive fight-past-sg-masculine with Bill and win-past-sg-masculine.

(13:Ambiguous) is genuinely ambiguous according to the informants. This is a clear violation of **PRO** $\leftrightarrow$ **TOP** in the interpretative direction, which would predict that Fred is the referent. If we use two sentences, the full name must be used to disambiguate. If we integrate and coordinate the message it becomes clear that Fred is the referent.

The explanation is perhaps that the use of the pronoun is less good than dropping the subject altogether as in (14).

(14) Fred se pral s Billem. Vyhral.

This is bad but it is unambiguously Fred who loses. The explanation is perhaps that on the one hand **PRO** $\leftrightarrow$ **TOP** tells us to interpret the pronoun as Fred, while on the other hand the more general version of **PRO** $\leftrightarrow$ **TOP** which tells us to reduce the old topic as much as possible leads to zero-realisation.

It seems fairly clear that there are other constraints at work in Czech apart from the ones we discussed for English and that the Czech informants are much stricter about the marking of discourse relations. The trigger for zero-pronouns is definitely not the same as in Chinese. In Czech it seems to indicate what has now been made topic. In coordinated structures however it is not possible to change topic. On the face of it, the old subject is not the preferred interpretation of a pronoun, contra interpretive **PRO** $\leftrightarrow$ **TOP**. Also —contra generative **PRO** $\leftrightarrow$ **TOP**— the old subject is frequently realised as nothing at all. But it should be clear that **PRO** $\leftrightarrow$ **TOP** needs to give way to a principle like **REDUCED** $\leftrightarrow$ **TOP** anyway.

Finnish pronouns

Finnish is an agglutenating language with inflecting elements. It has very clear and extensive case marking on proper names as well as pronouns. On the other hand, pronouns are not marked for gender. Finnish also cannot use intonation on pronouns to mark. Person

and number is marked on verb forms and Finnish also allows pro-drop, but generally this is confined to intrasentential ellipse or to first and second person.

(15) 1. Roger avosteli Kaarloa. /Roger became angry/.
Roger criticized Kaarl-PART
2. Roger oli vihainen.
Roger became angry.
3. **Hän oli vihainen.
**S/he became angry.
4. Roger arvosteli Kaarloa ja oli vihainen.
Roger criticized Kaarl and 0 became angry.

Note that these examples seem to illustrate that Finnish doesn't follow **PRO** $\leftrightarrow$ **TOP** in the generative direction. However, when there is only one actor in the discourse, Finnish pronominalizes the topic. Finnish doesn't use intonation to mark emphasis or contrast and as such doesn't follow **MARK** $\leftrightarrow$ **SHIFT** by using stress. The full proper name is the only way to communicate clearly when a shift has been made, and full proper names may also be used without shift in grammatical function as we saw above. So shift and parallellism are not required. But consider (16)

(16) 1. John rakastaa Maryaa. /Mary doesn't love John/.
John loves Mary-PART.
2. Mutta hän ei rakasta Johnia.
But s/he not love John-PART. 3. Hän ei rakasta Johnia.
s/he not love John-PART

Gender is not marked on pronouns, so in a shifted situation with two pronouns the interpretation would be ambiguous, similar to Chinese. Interesting is (16.2) though. Adding the word *but* (mutta) makes the use of a pronoun possible. This contradicts **PRO** $\leftrightarrow$ **TOP** in the interpretational direction since that predicts that hän is John and not Mary. But the explicit discourse marker seems to serve the same function that intonation can serve in English, and forces a crossover interpretation, which however must be confirmed by having at least one actor appear as a full form.

So Finnish seems to be driven more by the desire to avoid the risk of being misunderstood than by wanting to indicate who is the topic by either reduction or marking in its referential system. The contrast between (16.2) and (16.3) seems to indicate that more is going on than just purely semantic considerations since (16.3) is unambiguous.

Japanese pronouns

Japanese is a left-branching, SOV-language that marks case through the use of postpositional particles. It also has been said to be a language that has both topic and subject (Li & Thompson, 1976) marking topics, which are typically given information, with *wa,* and subjects with *ga* which is also used to introduce contextually new information. Names are generally topic-marked and the topic marker also marks constrast. It marks number, gender and in some cases also the relationship between speaker and hearer on pronouns. Japanese cannot use intonational prominence on pronouns. According to the informants, in sentences with only one actor, using the name or using a pronoun seems to be equally natural, which seems to indicate that Japanese does not follow **PRO** $\leftrightarrow$ **TOP** in the generation direction.

(17) 1. John wa Mary-ni kisu shita. /Mary slapped John/
John-TOP Mary-AT kiss did.
2. Mary wa John wo hippataita.
Mary-TOP John-OBJ slapped.
3. Kanojo wa kare wo hippataita.
She-TOP him-OBJ slapped.
4. Mary wa kare wo hippataita.
Mary-TOP him-OBJ slapped.
5. Kanojo wa John wo hippataita.
She-TOP John-OBJ slapped.
6. Karera-wa ureshikatta.
They were happy.

It seems that all of the different versions except a version where two pronouns are used are possible, and can be considered natural. At the same time there seems to be some sort of slight preference for **SYMMARK** because there is a slight preference for (17.2) over (17.4) and (17.5) . It seems that pronouns are not in anyway preferred over proper names, at least when we are dealing with third-person singular. In (17.6) , when Mary and John are referred to as a group, then there is a preference for using a pronoun over a conjoined nominal phrase, and this is probably for reasons of economy. All the other examples seem to work the same way, both pronominal and full name forms are considered equally as natural in the examples studied. And as there is no intonational marking, there does not seem to be a clear way of seeing how shifts are marked.

Concluding Remarks

The main finding seems to be that the English mechanism of intonationally marked pronouns was not attestable for any of the four languages we looked at and it seems the mechanism of parallellism and shift marking is not needed either for the description of these languages. What is confirmed is the constraint **PRO** $\leftrightarrow$ **TOP** in its more general form: **REDUCED** $\leftrightarrow$ **TOP**.

Most of what we report is based on the rather unnatural examples of Mattausch. One uniform finding was that our informants generally were unhappy with the lack of marking of discourse relations. They preferred *and*s, *but*s and *therefore*s over the full stops characteristic of those examples. Therefore in the elicitation task care has been taken to add context and to clearly distinguish discourse segments and mark discourse segments so that they sound more natural.

It seems that our pilot study indicates that we are a long way removed from understanding even the best studied discourse phenomenon (pronouns) and that we have therefore made our case: further comparative and comprehensive discourse data need to be made available, annotated for the full range of discourse phenomena, and new hypotheses are needed.

References

Aissen, J. (1999) Markedness and Subject Choice in Optimality Theory. In: *Natural Language and Linguistic Theory* **17**: 673-711.

Beaver, D. (MS) Centering and the Optimization of Discourse.

Blutner, R.(2000) Some Aspects of Optimality Theory in Interpretation. (to appear in JOS Special Issue on OT semantics).

Bresnan, J. (MS). The Emergence of the Unmarked pronoun: Chichewa Pronominals in Optimality Theory.

De Hoop,H. & H. Hoekstra. (2001) Optimality Theoretic Semantics. In: *Linguistics and Philosophy.*

Mattausch, J. (2001) *On Optimization in Discourse Generation.* MPhil Thesis University of Amsterdam. (obtainable from ILLC)

Smolensky, P.(1996) On the Comprehension/Production Dilemma in Child Language.In *Linguistic Inquiry* **27** .

Zeevat, H. (1999) Explaining Presupposition Triggers. In P. Dekker (ed.) Amsterdam Colloquium 1999.

Zeevat, H. (MS) The Asymmetry of Optimality Theoretic Syntax and Semantics. To appear in Special Issue of Journal of Semantics on Optimality Theoretic semantics.

Part II

Philosophical Background

Paper 9

Presuppositional denials

Rob van der Sandt
MPI Nijmegen

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

Paper 10

NL Disjunction in Discourse

Isabel Gómez Txurruka
ILCLI; Apdo. 220
20080 Donostia-San Sebastián; Spain
txurruka@sc.ehu.es

Following Gazdar (1979); Kamp and Reyle (1993) or Simons (2000), the semantics of NL disjunction is that of its correlate in propositional logic. In this view, *or* indicates that at least one disjunct is true. This interpretation is known as *inclusive* given that it includes an alternative in which both disjuncts simultaneously hold. The following is an example:

(1) If Fred has failed the entire test, then he has failed the practical part or he has failed the theoretical part. (cf. (Kamp and Reyle, 1993, 190))

Let us call the assumption that the inclusive reading (INC) is the semantic meaning of *or* the Traditional Assumption 1 or TA1. A great deal of occurrences of *or*, however, are not inclusive but exclusive (EXC). That is, the natural hearer understands that at most one disjunct is true—i.e., that both alternatives do not hold at the same time. Consider (2):

(2) I offer you two alternatives. We hire a second assistant so that they can work together, or we fire your stupid associate and I hire you a really good professional.

Kamp and Reyle (1993) suggest that the exclusive interpretation of *or* should be considered a default. This default would explain why English sentences of the form "A or B or both" are appropriate, that is, why inserting "or both" is not redundant. Note that their argument works only if one presupposes TA1. On the other hand, Kamp and Reyle (1993) do not develop a theory of which meanings (and what kind of logic) should be assumed in order to cancel this default. Let us call the assumption that the exclusive reading is associated with *or* by default the Traditional Assumption 2 or TA2. Gazdar (1979) also develops a view in which the exclusive interpretation is a sort of default, namely, a scalar implicature. The advantages of his approach, set in the framework of a theory of implicature (Grice, 1989), and based on the notions of scale and scalar implicature (Horn, 1972) are well known. As for the shortcomings, authors such as Cohen (1971) already pointed out that the definitions of scale and scalar implicature are too vacuous to be formalizable or testable. W.r.t. *or*, in particular, the scalar implicature of exclusivity does not hold in a variety of contexts, as for example under conditional such as in (1) above, but Gazdar's approach does not predict the cancelling. That is, a theory of cancellation of the default is also missing in this framework. A more recent approach to natural disjunction is (Simons, 2000), who introduces

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

ILCLI; Apdo. 220
20080 Donostia-San Sebastián; Spain
txurruka@sc.ehu.es.

an extension of the Stalnakerian framework to account for presupposition and anaphora in disjuncts. While she agrees that TA1 holds, she defends, with Grice and Stalnaker, that the rest of properties related to the presence of *or* must be accounted for "in terms of very general principles governing assertoric contributions to discourse, and their interaction with the truth conditional properties of *or*" (Op. cit.: 6). Thus, she departs from Kamp and Reyle's or Gazdar's accounts in that she holds that EXC is triggered by different pragmatic sources. The traditional Approach (TA1/TA2) is strongly supported by propositional logic, given the following wellknown theorem (10.1):

(10.1) $$((p \vee q) \wedge \neg(p \wedge q)) \rightarrow (p \vee q)$$

According to Theorem (10.1), all *or*-sentences that have an EXC reading are also INC. Thus, the semantic meaning of *or* might still be INC in these cases. Although natural disjunction has been mainly linked to INC and EXC truth conditions, there are other truth conditions that might also hold in the presence *or*. Consider (3) and (4):

(3) John interrupted her from time to time. He asked her about something he did not understand or requested more details.

(4) That is called ambivalence. Or, as my grandmother told my mother once: "Your father would be a great man if he were different."

The hearer constructs a plural entity in (3), namely *interruptions*, and each disjunct is understood to be true in some of them. If this analysis makes sense, then both disjuncts hold. It is a conjunctive reading (CON). A conjunctive reading is also understood in (4), which is a Reformulation. But CON can also be interpreted as a pragmatic restriction of INC, as in the case of EXC, given Theorem (10.2) below:

(10.2) $$(p \wedge q) \rightarrow (p \vee q)$$

Thus, were one to try to find counterexamples to TA1, one should look for examples where it is not true that (the speaker conveys that) at least one of the disjuncts must hold. I believe that there are such examples, that is, cases where the speaker communicates that all the possible truth combinations of the disjuncts might hold.

Consider the following example (take CAPITALS to indicate fall-rise intonation):

(5) A: What did Luisa do last night?
B: I don't know. She went to the MOVIES, or had some drinks with her FRIENDS...

B communicates that going to the movies and having some drinks are two of the possibilities that might be true in (5). But she is not conveying that at least one alternative is true. In fact, both of them could be false and she would not be accused of lying or of being mistaken. The contribution of speaker A in (6) below also suggests that some *or* sentences are used to communicate that some possibilities should be considered although the speaker is not conveying that at least one holds:

(6) [A journalist and a police detective are discussing how to capture Wood, a white-collar criminal. The former says:]
A: Let us not waste our time analyzing too much. Wood is going to be sunbathing in the Bahamas soon, or lining his pockets in NY.

Examples such as in (5) and (6) strongly indicate that truth conditions are built in the context. Although uttering an assertion normally commits the speaker to its truth in a nonmarked context (Gricean maxim of Quality), this is not necessarily so in every context. Information about discourse structure has a bearing on the building of the appropriate truth conditions. If this view makes sense, the disjunction inherits a problem that is already there in its absence. However, I believe that the meaning of *or* still needs to be compatible with all other semantic and pragmatic meanings in the way described in Theorems 1 and 2 above. I take (5) and (6) to be counterexamples to TA1. Although they do not create logical inconsistency they do not bear the right entailment relations between hard and cancellable meanings. Cancellable meanings should restrict hard meanings and not the other way around. Once it has been shown that the traditional assumption that the inclusive reading is the semantic meaning of *or* has counterexamples, the next step is to figure out how a theory of the meaning of natural disjunction might solve these problems. Given the strong interaction of semantic and pragmatic meanings in (5) and (6), a theory of discourse, namely, Segmented DRT (Asher, 1993), will be used to give a discourse-based approach to natural disjunction.

Bibliography

Asher, N. (1993). *Reference to Abstract Objects.* Kluwer AP.

Bar-Hillel, Y., editor (1971). *Pragmatics of Natural Languae.* Reidel.

Cohen, L. (1971). The logical particles of natural language. in: (Bar-Hillel, 1971).

Gazdar, G., editor (1979). *Pragmatics: Implicature, Presuppositional and Logical Form.* AP.

Grice, P., editor (1989). *Studies in the Way of Words.* First Harvard UP.

Horn, L. (1972). *On the Semantic Properties of Logical Operators in English.* PhD thesis, UCLA.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic.* Kluwer AP.

Simons, M. (2000). *Issues in the Semantics and Pragmatics of Disjunctions.* PhD thesis, Garland, NY.

Paper 11

Presuppositional Clitics, Propositional Attitudes, and Binding Theories of Presupposition

Alessandro Capone
sandro.capone@tin.it
Department of Philology and Linguistics, University of Messina

1 Introduction

I have decided to dub the clitics involved in the doubling of explicit arguments of verbs in sentences such as *L'ho visto Giovanni* presuppositional clitics. In this paper, due to restrictions of space, I will mainly sum up the results of my long term investigation of the phenomenon and I will mainly do so with reference to verbs of propositional attitude. As verbs of propositional attitude have sentential arguments, which can be dealt with syntactically as NPs, I will argue that these can be doubled as well and, if doubled, will give rise to presuppositional phenomena - albeit these phenomena involve a notion of modal subordination along the lines of Roberts (1989), but which is more interesting because it applies across utterances by different participants. In this paper, I will forestall an objection to the presuppositional view by considering verbs of propositional attitude other than factives (but I will consider factives too). I will also apply the notion of modal subordination to conditional presuppositions along the lines of Beaver (1997) as interpreted by Asher and Lascarides (1998) and, in passing, I will show that the conditional presupposition in question is a semantic and not a pragmatic phenomenon in so far as it is not cancellable. In the end, I will show that the standard arguments against the satisfaction theory of presupposition (van der Sandt, 1992; Geurts, 1999) are not correct, as explanatory adequacy rather than empirical coverage is the real merit of the binding theory and I will argue that clitics, through modal subordination, constitute strong evidence in favour of the binding theory. In the end, I will show that a notion of presupposition along the lines of Asher and Lascarides (1998) is what is needed in order to account for presuppositional clitics.

2 Theoretical background

There are at least three recent approaches to presuppositional clitics around which would deserve close attention. Uriagereka (1995) provides evidence to the effect that the NP with which the clitic is associated must be specific. Actually, in the same paper he argues that

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

sandro.capone@tin.it
Department of Philology and Linguistics, University of Messina.

in clitic doubling constructions (as in *Lo vimos o neno*) the NP with which the clitic is associated must be referential. As he considers specificity/referentiality the two sides of the same coin, he works out a Restrictive Mapping Slogan (RMS) to the effect that "Only and all material assigned VP-external scope is interpreted as specific at Logical Form". Bearing in mind Gutiérrez-Rexach's criticism of the RMS (Gutiérrez-Rexach, 2000), it might be useful to view it as applying to referentiality, rather than specificity. Under this interpretation, the RMS might still be valid. Actually, I believe that Gutiérrez-Rexach, who is right is saying that definiteness and not merely specificity is involved in the phenomenon, undervalues the importance of LF movement for the analysis of the opacity escaping phenomena associated with clitics in clitic doubling constructions. How can the clitic escape the modal effects of the verb in sentences such as *Mario lo vuole vendere il violoncello* (*Mario it wants to sell the cello*) (where *the cello* is referential)? Presumably this is possible through movement at LF, as the analysis by Uriagereka implies. According to Uriagereka (1995) the clitic moves up from the base position and ends up in a position within F'. It is clear that in this position it c-commands the verb and it is not c-commanded by it: hence it can escape the modal effects of the verb. I personally find this analysis valuable. Uriageraka's analysis of pro-clisis according to which the verb moves up together with the clitic and ends up in a position in F' which is c-commanded by the clitic is also valuable. (Gutiérrez-Rexach, 2000) is an interesting critique of (Uriagereka, 1995). He argues that definiteness, and not merely specificity, is involved in clitic doubling constructions. This critique seems to me to be justified. However, Gutiérrez-Rexach does not appreciate the extent to which the modal phenomena associated with clitics depend on the syntactic analysis by Uriagereka. Gutiérrez-Rexach develops three constraints: a) the Principal Filter Constraint; b) the Presuppositionality Constraint, c) the Context Dependence Constraint. Constraint a) says that the generalized quanitifier associated with an accusative clitic must be a Principal Filter (in other words the QP associate of a doubling clitic has to denote a group). Constraint b) says that the generator of the generalized quanitifier associated with an accusative clitic is a presupposed set. Constraint c) says that only a contextually restricted quanitifier can be doubled. Capone (1997, 1999, 2000b,a,c) has shown that in clitic doubling constructions clitics have a presuppositional nature. In particular they are associated with speaker/hearer presuppositions, in other words they constrain the context in such a way that both the speaker's and the hearer's presuppositions are involved (in other words there can be no asymmetry between the speaker's and the hearer's presuppositions). This notion of speaker/hearer presupposition clearly contrasts with the notion of speaker's presupposition which is the one used in the current literature (which is normally allied with the notion of accommodation). With presuppositional clitics, no notion of accommodation is required to neutralize possible asymmetries between the speaker's and the hearer's presuppositions, as there can be no such asymmetries. (Capone, 1997, 2000a) are important cross-linguistic studies, which show that presuppositional clitics exist in a number of European pro-drop languages, such as Italian, Spanish, Portuguese, Greek, Serbo-Croat, Czech, Polish. These languages have in common pro-drop phenomena, clitic-left-dislocation and presuppositional clitics. The other interesting development of Capone (1997, 1999, 2000b,a,c) is to have tied presuppositional clitics to verbs of propositional attitude, a move that is missing in (Uriagereka, 1995) and (Gutiérrez-Rexach, 2000). Thus it is not only a bare NP that is doubled but also a sentential NP, in the case of verbs of propositional attitude. I argued in the already cited work that with verbs such as sapere (know), sospettare (suspect), immaginare (imagine), dire (say), sperare (hope), ricordare (remember), the use of the clitic excludes psychologised or "mere report" interpretations of the verb and establishes the em-

bedded proposition as a fact mutually known by the speaker and the hearer (let me stress "known" as opposed to "assumed to be true"). In (Capone, 2000b,a) I argued that the problem of presupposition projection interacts interestingly with presuppositional clitics, which allow presuppositions to persist where current theories of presupposition projections predict them not to project. I will take up this issue in a following section.

3 Satisfaction and binding theories of presupposition and presuppositional clitics

The two most important approaches to presupposition projection are the local context satisfaction theory, propounded by Karttunen (1974); Stalnaker (1999); Soames (1982); Heim (1992) and the binding theories of presupposition defended by e.g. van der Sandt (1992); Krahmer (1998); Krahmer and van Deemter (1999); Geurts (1999). The satisfaction theories grew out of the observation that the presuppositions of a clause which were entailed by a prior clause (in conjunctions, conditionals) did not project as presuppositions of the complex sentences. Since that observation, the framework has become more and more sophisticated. The latest development of the framework is that if a presupposition is satisfied by its local context (e.g. the prior clause), then it does not ascend to become a presupposition of the complex sentence. I have given enough details about this framework in (Capone, 2000a) and they can be easily accessed though (Chierchia and McConnell-Ginet, 2000), a book which is quite advanced but surprisingly does not mention the recent developments in the theory of presupposition (let alone (Chierchia, 1995)). Presumably the authors take for granted that the objections of the proponents of the binding theories can be easily dealt with or can be largely ignored. Now, although I am persuaded that it is not on empirical grounds that the binding theories can claim to be superior, I nevertheless believe that their objections, which are in my view quite important have to be taken into account. In the end, it might be possible that a mixed approach will be necessary and that no single approach will really resolve the projection problem. However, I would like to argue that neglect of the objections of the binding approaches results in an unjustified dogmatism. Dogmas will one day or the other be destroyed, so it might be worth while taking into account the plausible objections by van der Sandt (1992). Van der Sandt (as Geurts (1999) stresses) focuses on the analogy between anaphora resolution and presupposition. I would be more circumspect than Geurts is in taking for granted the analogy. Presupposition resolution follows from pragmatic and semantic principles (the logic of connectives) which might be different from those required by anaphoric resolution. In fact, nobody believes that binding in the sense of van der Sandt should be equated with the binding of Chomsky's binding theory. Now the interesting examples which seemed to jeopardise the satisfaction theory are the following:

(1) a. *If John has grandchildren, his children will be happy*

b. *If John has an oriental girlfriend, his girlfriend will be happy.*

These are cases of partial matches, to use an interesting notion by Krahmer and van Deemter (1999). I believe that the belief that these examples cannot be dealt with by the satisfaction theory is wrong. Van der Sandt says that the binding theory has problems with example 1a, because he implies that the antecedent of the conditional entails *John has children.* In this case, the presupposition of the consequent is satisfied and cannot project. Van

der Sandt predicts two readings, a presuppositional and a non-presuppositional one. However, since the antecedent of 1a does not entail *John has children*, the belief that the satisfaction theory cannot predict the presuppositional reading is wrong. Furthermore, as the antecedent I-implicates (following assumptions by Levinson (2000)) *John has children*, the dispreferred non-presuppositional reading (as satisfaction by conversational implicatures is plausibly dispreferred with respect to satisfaction by entailments) is obtained by considering that the presupposition of the consequent is satisfied by the I-implicature of the antecedent. I cannot unfortunately dwell on this interesting issue, but I do not need to prove my considerations as they are substantiated - albeit not explained in terms of Gricean pragmatics - by many of Geurts's (Geurts, 1999) bridging examples. If we want to retain those examples, I-implicatures must play a role either in a satisfaction or binding framework for presupposition resolution. In connection with example b, the satisfaction framework can incorporate binding, as two readings emerge if one binds or fails to bind *his girlfriend* to *an Oriental girlfriend*. The presupposition *his girlfriend* need not be satisfied *ipso facto* by *an Oriental girlfriend*, given that *Oriental girlfriend* entails *girlfriend*. Here the local context satisfaction might say that the antecedent has a primary entailment (*Oriental girlfriend*) and a secondary entailment (a logical consequence of what has been said). Satisfaction by primary entailments is preferred to satisfaction by secondary entailments (as is obvious). To understand why this should be so, one must reflect on the fact that if I were to suppose that I have got an oriental girlfriend that would not amount to supposing that I have a girlfriend. Just consider the difference between *If I had an oriental girlfriend, I would be jealous/If I had a girlfriend, I need not be jealous*. Especially conditionals bring the distinction between primary and secondary entailments to the fore, as this distinction may not be of great importance in sentences other than conditionals (see (Capone, 2001a,b)). There are other objections to the local satisfaction approaches, such as the one raised by Geurts (1999). According to him, the satisfaction theories predict that the sentence If I go to movies, my son always comes along has the conditional presupposition If I go to movies, I have got a son. This would be too week. Not only that, it would be ridiculous. Although Geurts goes back to some old articles which nobody would now take into considerations (apart from him), I think it's fair to stress that nobody in the local satisfaction framework would say that nowadays. Geurts, however, says that it can be easily shown that satisfaction approaches must be committed to conditional presuppositions. It is to say the least strange that the argument according to Geurts should be so obvious that it does not deserve being spelled out in detail (or even as an upshot). In fact, there is no argument offered, one must simply guess what the argument might be. Why is Geurts so cryptic and reticent about the argument? I must be excused if I try to construct the argument Geurts has got in mind - the only one I can think of, in fact. Consider again If go to movies, my son always comes along. The presupposition of the consequent must be satisfied either by the context or by the prior clause. So either c $\rightarrow$ p (S') (p = presupposition) or S $\rightarrow$ p (S'). But if S $\rightarrow$ p(S'), then c $\rightarrow$ S $\rightarrow$ p (S'), as the satisfaction theorists are committed to the assumption that the update of c with S must be defined, so c must admit S. So far there is nothing wrong about the argument above. It is correct. However, a disjunction is a disjunction, thus the result is: c $\rightarrow$ p (S') or [S $\rightarrow$ p (S') (in which case c $\rightarrow$ S $\rightarrow$ p (S')]. Now, if the presupposition is not satisfied by the prior clause (S), the second part of the disjunction is false, hence the disjunction predicts that c must satisfy the presupposition. If the presupposition is satisfied by the prior clause (S), then c $\rightarrow$ S $\rightarrow$ p (S'), which is harmless. Now, as I argued in (Capone, 2001a,b), we should look instead at modal subordination as evidence in favour of the binding theories of presupposition. If I were to say If John has a niece, his niece will like him, it is

clear that a modal subordination is established between the antecedent and the consequent, either through semantics alone (if one accepts a possible world treatment of conditionals, then the antecedent is surely an intensional context) or through semantics as augmented by pragmatics (say if Gazdar's (Gazdar, 1979) view is accepted). Disjunctions can be reduced to conditionals due to the well-known equivalence rules of classical logic. I do not have the space to go into Roberts' (Roberts, 1989) treatment of modal subordination, which has been criticised by Geurts (1999). I simply want to redirect Geurts's most interesting argument. The case in point is a sentence such as Either the house has no bathroom or it is on the third floor. Roberts accommodates locally There is a bathroom in the second disjunct, a move criticised by Geurts who (plausibly) argues that he does not know where the accommodated materials come from. However, once we see the equivalence between Either the house has no bathroom or it is on the third floor and If the house has a bathroom, it is on the third floor, one understands where the locally accommodated materials come from. Actually, if equivalence is born in mind, no accommodation is needed, as the modal subordination follows from the conditional. I now want to deal with conditional presuppositions as dealt with by Asher and Lascarides (1998) (based on examples by Beaver (1997). Consider

(2) a. *If David wrote the article, then the knowledge that no good logician was involved will confound the editors*

b. *If David wrote the article, then the knowledge that David is a computer program running on a PC will confound the editors.*

According to Asher & Lascarides, the former example of the pair triggers the conditional presupposition

(2) c. *If David wrote the article, then no good logician was involved*,

while the latter does not but triggers a normal presupposition such as d)

(2) d. *David is a computer program running on a PC.*

The authors raise the issue of why two structurally similar sentences have different presuppositions, but since they do not see how to impute the difference to sentence structure, they take the view that the conditional presupposition is a conversational implicature. The latter example seems to me straightforward to explain: know is a factive verb and, although embedded in a conditional, it presupposes the embedded clause (which presents factual information). However the former example of the pair is more tricky. It should be glossed as: *If David wrote the article, then the knowledge (the editors's knowledge) that no good logician was involved in writing the article will confound the editors.* I believe that what is different from the latter example of the pair is the fact that the definite refers back to the article hypothetically written by David. Thus *know* interacts with a modal element and this suspends the factive status of the verb in a way which not even the embedding conditional can. Notice, furthermore, that embedding *know* in the antecedent of a conditional rather than in the consequent generally has the consequence of suspending the presupposition as in *If I know that Mary is a piano player, I will love her.* Thus, the conditional presupposition is caused by the dependence between *the article* and the antecedent. Asher & Lascarides say that the standard theory of accommodation cannot account for the intuitions about these examples.

But in a sense it must, if we want to capture the structural similarity of all conditionals. We simply need to add the consideration that if a modalised NP occurs within the scope of *know*, then the presupposition of *know* is modalised (a conditional presupposition). This addition probably derives from the interaction of modality with the principles of presupposition projection. For example I could say *I know that, in a hypothetical world with just water in it, I would have to swim.* Intuitively the presupposition is not that I would have to swim relative to any context but only relative to the context including the possible worlds where there is only water in them. I think that other examples support the view I have put forward, such as *If Mary cooks dinner, then Mary will know that dinner has been cooked well*; *If John finishes building his house, then Mary will know that his house has been properly built.* Thus my considerations seem to generalise to similar examples where a modal subordination (in the sense of Roberts (1989)) is created between an NP embedded in *...know that ...* in the consequent of a conditional and an NP of the antecedent such that this NP is the result of the action hypothesised in the antecedent of the conditional. It is possible to represent schematically the DRSs associated with sentences displaying conditional presuppositions. DRS1 is the main discourse representation structure. Presuppositions are represented by means of further embedded boxes. A second level of embedding corresponds to presuppositions. A third level of embedding corresponds to the presuppositions of presuppositions. It is clear that the presupposition of the consequent of the conditional has got one further presupposition. This further embedding is represented by two boxes embedded in DRS3. Following the method by van der Sandt, in DRS2 we merge the DRS of the presupposition of the consequent with the DRS of the antecedent. If the moved DRS is not absorbed by the DRS of the antecedent, we have to accommodate the presupposition at top level within the DRS2. But this is not possible, as the DRS of the presupposition of the consequent of DRS1 has one further embedded DRS and this can be absorbed by the DRS of the antecedent of DRS2. However, as a result of this absorption, the moved DRS will be realized as *If David wrote the article, no good logician was involved in writing the article.* This DRS can be accommodated at top level within the main DRS1. Asher & Lascarides take the view that the conditional presupposition is a pragmatic inference, one that is defeasible. However, the examples which ought to prove the defeasibility are not really persuasive. Consider: *If David is going diving, he will bring his wetsuit*; *If David is going diving, he will bring his dog.* According to the authors the former example of the pair presupposes *If David goes diving, he brings his wetsuit* while the latter does not presuppose *If David goes diving, he brings his dog.* These examples ought to prove defeasibility but I personally do not see why it should be necessary to posit conditional presuppositions in these cases. Furthermore, if the considerations on the pragmatic nature of the inference are correct, we should be able to cancel the conditional presupposition as in *If David wrote the article, then the knowledge that no good logician was involved will confound the editors. But I deny that if David wrote the article, then no good logician was involved.* This example seems to be quite bad. I believe that modal subordination is also involved in clitic doubling constructions. I do not have the space to argue at length why modal subordination is involved in the presuppositions of clitic expressions in examples such

(3) a. *Giovanni vuole vendere il violoncello che non ha* (John wants to sell the cello he has not got)

b. *Anche io lo vorrei vendere il suo violoncello* (I would like to sell it his cello too)

(4) a. *Giovanni ha un violoncello meraviglioso* (John has got a marvellous cello)

b. *E' un peccato che lo voglia vendere il violoncello* (It's a pity he wants to sell it his cello).

It's clear that, through modal suboridnation, the speaker/hearer presuppositions must coincide and speaker's A point of view/knowledge/asserted position is the basis for the interpretation of Speaker 2's stance towards an object. The same thing applies to sentential objects. Thus consider

(5) a. *Mario è andato a Parigi* (Mario went to Paris)

b. *Lo so che Mario è andato a Parigi* (I know that Mario went to Paris)

(6) a. *Mario è andato a Parigi* (Mario went to Paris)

b. *L'ho tanto sperato* (I hoped that so much)

(7) a. *Mario ha rubato una macchina* (Mario stole a car)

b. *L'ho sospettato che fosse un ladro* (I suspected he was a thief)

(8) a. *Mario è a Parigi* (Mario is in Paris)

b. *Si, me lo aveva detto e lo avevo creduto che fosse a Parigi* (Yes, he had told me, and I had believed that)

These examples show that not only factives, but genuine verbs of propositional attitude are involved in modal subordination. Now, the interesting thing is that modal subordination, in these cases, predicts the opposite of what van der Sandt predicted in the case of the conditionals I have already discussed. So modal subordination can either cancel a presupposition or it can project it to the level of the complex sentence. I have argued in (Capone, 2000b,a,c) that clitics make presuppositions persist where they ought to evaporate according to the projection rules of the satisfaction theories. Thus, one could have a sentence such as *O Giovanni lo sa che Mario è al cinema oppure non lo sa* (Either John knows it that Mario is at the cinema or he does not know it). As the negation of the second conjunct cancels the presupposition of the first disjunct, the sentence ought to have no global presupposition; but with the clitic it has one. One can say *Se Giovanni lo sa che Maria è al cinema sarà felice* (If John knows that Mary is at the cinema, he will be happy); here a scalar implicature would defeat the presupposition, but the semantics of the clitic can override the implicature. One can have sentences such as *E' possibile che Giovanni lo sappia che Maria è al cinema* (It's possible that John knows it that Mary is at the cinema). Without the clitic, in Italian (due to the subjunctive) no presupposition arises. However, the presupposition of *know* survives embedding in the modal context and defeats the implication of the subjunctive when a clitic is present, presumably through modal subordination. One could have sentences such as *Se Giovanni lo scopre che p, sarà infelice* (If John finds it out that p, he will be unhappy). Without the clitic, the speaker does not commit himself to the truth of p, but with the clitic he does through modal subordination. It's clear that presuppositional clitics are strong evidence in favour of the binding approaches, especially in favour of (Asher and Lascarides, 1998). The presuppositions of clitics have some attachment points and can be bound with a rhetorical relation (I propose this is the relation assertion/comment on the assertion).

Bibliography

Asher, N. and Lascarides, A. (1998). The semantics and pragmatics of presupposition. *Journal of Semantics*, 15(3):215–238.

Beaver, I. (1997). Presupposition. in: (van Benthem and ter Meulen, 1997).

Capone, A. A discussion of Higginbotham, Pianesi, Varzi (eds.) 'Speaking of events'. *Linguistics.*

Capone, A. A discussion of Levinson's 'Presumptive meanings'. *The Journal of Pragmatics.*

Capone, A. A long Review of Chierchia & Mc Connell-Ginet's 'Meaning and Grammar'. *Journal of Pragmatics.*

Capone, A. Modal adverbs and discourse. Two essays by Alessandro Capone. Pisa: Edizioni Tecnico Scientifiche.

Capone, A. Review of Levinson's 'Presumptive Meanings'. *Language.*

Capone, A. Review of P. Bosch's and R. van der Sandt's 'Focus'. Ms to be submitted to Language.

Capone, A. (1994). Scalar modality and linguistic typology. Paper presented at the LAGB meeting, Manchester.

Capone, A. (1997). Modality and discourse. PhD thesis, Oxford University.

Capone, A. (1999). Dilemmas and excogitations: considerations on modality, clitics and discourse. University of Oxford Working Papers, 4. 18-32.

Capone, A. (2000a). *Dilemmas and excogitations: an essay on modality, clitics and discourse.* Messina: Armando Siciliano editore.

Capone, A. (2000b). Dilemmas and excogitations: considerations on modality, clitics and discourse. *Lingua e Stile*, XXXV(3):447–470.

Capone, A. (2000c). Dilemmas and excogitations: further considerations on modality, clitics and discourse. Paper submitted to the proceedings of the Cambridge Conference on semantics and pragmatics.

Capone, A. (2001a). Review of Krahmer's 'Presupposition and Anaphora'. *Journal of Linguistics*, 37(1).

Capone, A. (2001b). Review of Turner's 'the semantics/pragmatics interface from different points of view'. *Journal of Linguistics.*

Chierchia, G. (1995). *Dynamics of meaning.* Chicago: Chicago University Press.

Chierchia, G. and McConnell-Ginet (2000). *Meaning and grammar.* Cambridge, Ma: MIT.

Davies, S., editor (1991). *Pragmatics.* Oxford: O.U.P.

Gabbay, D. and Guenthner, F. (1989). *Handbook of Philosophical Logic*, volume IV. Dordrecht: Reidel.

Gazdar, G. (1979). *Pragmatics.* New York: Academic Press.

Geurts, B. (1999). *Presuppositions and pronouns.* Oxford: Elsevier.

Gutiérrez-Rexach, J. (2000). The formal semantics of clitic doubling. *Journal of Semantics*, 16:315–380.

Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9:183–221.

Karttunen, L. (1974). Presupposition and linguistic context. in (Davies, 1991).

Krahmer, E. (1998). *Presupposition and anaphora.* Stanford CSLI.

Krahmer, E. and van Deemter, K. (1999). On the interpretation of anaphoric noun phrases: towards a full understanding of partial matches. *Journal of semantics*, 15:355–392.

Levinson, S. (2000). *Presumptive meanings.* Cambridge, Ma: The MIT Press.

Roberts, C. (1989). Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy.*

Soames, S. (1982). How presuppositions are inherited: a solution to the projection problem. (Davies, 1991).

Stalnaker, R. (1999). *Context and Content.* Oxford: OUP.

Uriagereka, J. (1995). Aspects of the syntax of clitic placement in western romance. *Linguistic Inquiry*, 26(1):79–123.

van Benthem, J. and ter Meulen, A., editors (1997). *Handbook of Logic and Language.* Oxford: Elsevier.

van der Sandt, R. (1988). *Context and presupposition.* London: Croom Helm.

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377.

Paper 12

Models of Intentions in Language

William C. Mann
SIL International

Intention is one of the significant topics in linguistics, partly because of the important roles that it is given outside of linguistics. All of the classic domains of text interpretation—law, religion, literary studies, philosophy and the communication sciences, give prominent roles to intention and the one who intends. The role of concepts from common culture, the so called "folk linguistics," turns out to be significant as well. The importance of the concepts represented by the word "intention" is not reduced by the fact that it is controversial, nor that some schools of thought would deny that it is relevant or tractable for study.

The key questions for this paper are: What models of intention are there? Where do they come from? How do they differ? It is a survey with evaluative comments, with a restricted focus. In particular, it focuses only on intention and language and only on adult human language. For convenience, the terms "intention" and "goal" are used in this paper as synonyms. The term "linguistics" represents a very inclusive notion, broader than any academic department.

What is intention? i.e. what do the various documented notions of intention have in common? The verb "to intend," is generally used to describe situations in which, minimally, some individual has in mind some situation, not yet actualized, along with a disposition to prefer that the situation be actualized. Related terms include *desire*, *want*, *wish*, *hope* and *prefer*. Comparable phrases include *to carry out an intention, to implement an intention, to act on an intention, to commit to an intention*, and many more. The term *intention* sometimes, but not always, includes the notion that the person who has that intention in mind also has in mind a commitment to try to actualize that intention. Context usually makes it clear whether commitment is present; the paper uses both senses.

A good starting point for considering notions of intention is *Intention in the Experience of Meaning*, (Gibbs, 1999). Gibbs has an excellent and extremely broad description of the uses of intention concepts in various literatures, including not only General Linguistics, Computational Linguistics, Cognitive Linguistics, Sociolinguistics, Psycholinguistics and various similar work but also art and literary studies. His general purpose is to defend the use of the relevant terms and concepts, especially showing their legitimacy and promise in Cognitive Linguistics and other cognitive studies. The early chapters lay out the controversies well, including the "death of the author" movement in literary studies. It turns out that, even recognizing that significant uncertainties arise, the legitimacy of developing and using such concepts is extremely defensible.

Intention has been a controversial idea in various schools of psychology for decades, especially since the period when behaviorism was dominant. The effect is still with us, but Gibbs

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

shows how these concerns can be used as precautions rather than prohibitions.

The first purpose of the paper is to bring together a diversity of mutually relevant concepts, partially surveying models of intention in linguistics and philosophy. A second purpose is to comment on particular ideas.

Gibbs' book is also a good place to start exploring the various models of intention. In the process of concentrating on the legitimacy of using the concept of intention and on how various models of intention can be beneficial, it identifies four different models, discussed below.

The first model is the well known "code model" of communication, which has no role for intention at all.[1] It is given to provide a starting point, and is quickly rejected. The code model of communication is in deep disrepute. It is widely rejected outside of linguistics (see (Craig, 1999)), and often within linguistics as well. Craig sees most of the various schools of communication theory as defining themselves in contrast to, and rejection of, the code model (p. 125). Gibbs rejects it as well.

The second model is called the Simple Intentionalist model. In this model, each language user has personal intentions which are used as the basis of expression. Each also has a personal world view which is used as a conceptual basis. Views and intentions of interacting persons are unrelated. One of many possible objections to the Simple Intentionalist model is that it has no possibility for accounting for the collaborative nature of dialogue. A student interacting with a teacher would simply be pursuing personal goals, not responding.

The third model is called Perspective Taking. Each language user is aware of the intentions and world view of the addressee, and always adopts the intentions and world view of the addressee in directing language to that addressee. This has the advantage of giving a kind of recognition to the interactive nature of dialogue, and a kind of democratic or peer respect to the addressee's views and biases.

However, the Perspective Taking model has some problems. If I always adopt (my view of) others' intentions and world view while addressing them, then I do not expose my own intentions and world view, and others may have no access to them. So if I always use Perspective Taking, my addressees cannot. Similarly I may have no access to their intentions based on their use of language, making true perspective taking impossible. Also, there is no obvious way to extend the Perspective Taking model to the task of addressing an audience with mixed world views and intentions.

The fourth model is called the Dialogic model. Here the intentions that shape the language of interaction are not imposed by one or both of the parties to a dialogue, but rather are arrived at jointly, in effect negotiated. The results of the negotiation are jointly determined constructs, that possibly are not quite the same as intentions held by either party before the dialogue began. They may be accompanied by individual subintentions for the individual roles of the participants.

This model of intention, like Perspective Taking, has the advantage of giving a kind of recognition to the interactive nature of dialogue, and a kind of democratic or peer respect to the addressee's views and biases. It goes farther by negotiating agreement, creating a communicative regime in which the other party's intentions and view can have effects as well

[1]The code model, sometimes called the transmission model, basically consists of the following: ¡¡Communication using language is in essence exchange of ideas, which are propositional. Language is a code, analogous to Morse code. Speakers encode their ideas by encoding the corresponding propositions in language, which is then transmitted to recipients. Recipients decode the propositions, thus recovering the ideas.¿¿ No intentions are referenced in encoding or decoding. For intention, it is the null model.

as respect. In this model, language understanding is based on a shared set of intentions, not pre-specified as in the code model but created by some sort of tacit collective bargaining.

Unfortunately the Dialogic model has problems as well. One problem concerns the negotiation of goals to be jointly held. In what language is this negotiation conducted, and, in that language, how is understanding achieved, and how can the success of the negotiation (even between adversaries) be guaranteed? The language of the negotiation cannot be the language that is used after the negotiation, since there is not yet a set of jointly held intentions. But if that pre-agreement language turns out to be adequate for the negotiation, it might be adequate for general communication as well. Possibly, not all such negotiations succeed. Empirically, participants do not appear to switch languages when they come to rapport. So the process of coming to negotiated goals may not be necessary, or may not be possible, or may not be discoverable.

We should note that kinds of intention that are implausible as the only form of intention, and thus required in all cases, can be very plausible as partial accounts. Surely Perspective Taking applies to some participants' behavior in some interactions, and the negotiation of goals as in the Dialogic model occurs in some interactions. Perhaps we should expect a combination of intention types to be found, rather than a single, presently unimaginable type that applies to all human interaction in all situations.

Gibbs should not be faulted for sketching these models in the way that he has. His purpose is to present an attractive diversity of useful looking ideas, without working out details. The ideas support the notion that interesting research can be done based on models of intention. (He does not choose between these models, and rejects only the code model.) Gibbs presents a very strong defense for using intentions as a key element in modeling language. These models are only a small part of his presentation.

Moving on, we find that many models of intention visible in the literature are not on Gibbs' list.

1 A Set of Attributes of Intentions

It is not possible to cover enough ideas by continuing one model at a time. We must leave the list-of-models approach and deal with distinctions and details at a slightly finer grain. The descriptions below are drawn from linguistics and its close neighbors, including philosophy.

For a diversity of models that are in use, we will focus on gross distinguishing characteristics, calling them attributes. These attributes are perhaps not as neat as one might expect. Some of the attributes represent distinctions that do not always exhaust the alternatives, and some represent concepts usually thought of as degree concepts. Not all are sharply defined, and even if sharp are not always discernible in accounting for natural cases. There are also a few dependencies between them, so that an intention with one of the attributes may also need to have a certain value of one of the others.

These attributes are discussed:

1. Activeness
2. Partialness
3. Priorness
4. Tacitness

5. Immediacy
6. Interaction-configuring
7. Intended to be recognized
8. Jointness
9. Sharedness
10. Structuredness
11. Complementarity
12. Conventionality

2 Attributes of Individual Intentions

Activeness

What kind of thing is intended? Perhaps the most important distinction for modeling the intentions that accompany language use is a contrast between intended actions and intended effects.[2] Intended effects typically are states of affairs that the intender desires or prefers, while intended actions typically involve some identifiable process within the capacities of the actor(s).

There has been much philosophical work on intended actions, considered without simultaneously considering intended effects (desired states, directing action toward particular results,...). Such work has strongly influenced models of language use. For example, Tuomela's book *Cooperation: a philosophical study*, is a detailed development of notions of joint actions, actions characteristically performed by a group of two or more persons Tuomela (2000). The actions studied are predominantly intentional, and intentions of joint actions may be accompanied by subintentions of participants intending actions which serve to carry out their individual roles.

Tuomela's work is not particularly focused on language, nor on groups of only two participants. For language, Tuomela's model of joint action is applied in Clark's *Using Language* Clark (1996). Clark describes a wide range of situations in which the concept of joint actions to be both necessary and powerful for accounting for interaction using language.

Past work in both linguistics and philosophy has made extensive use of notions of intention, but often the intended was some sort of effect or outcome rather than an action. So, for example, Austin, in the course of defining *perlocutionary* and eventually defining *perlocutionary act*, says "Saying something will often, or even normally, produce certain consequential effects upon the feelings, thoughts, or actions of the audience...: and it may be done with the design, intention, or purpose of producing them..." (Austin, 1975, p. 101). He is working with both intended effects and intended acts, and gives the general impression that the acts are selected in part in order to produce particular effects. In another place he says "In considering responsibility, few things are considered more important than to establish whether a man

[2] Although this might seem to be a clear and exhaustive dichotomy, it is not. Many cases and categories are not easily classified. Still, it is helpful to consider cases where the contrast is clear.

intended to do A..." and he goes on at length to distinguish the English forms: *intentionally, deliberately* and *on purpose*.

Linguistic work on intention, of both effects and acts, is actually abundant. Most work in computational linguistics, if it involves intention at all, involves both intended effects and intended actions, often in hierarchies or plans. The subfield of Text Generation (computer authorship) has a prominent topic called Text Planning, again involving both intended effects and intended actions. To pick another example, Rhetorical Structure Theory is defined in a way that requires that every part of every text analysis involve some statement about intended effects (Mann and Thompson, 1988). Planning and plans, including partial plans (for language use and also more broadly), are widely recognized as practical necessities. The notion that an action has succeeded or failed is also seen as a practical necessity. As a result, intended effects and intended actions are often considered together without a requirement that every intention be associated with an action.

In the light of this widespread use of notions of intended effect, it is curious to find recent work that focuses on actions to the exclusion of effects. It is more understandable in philosophy, where action theory is a well established focus. However, even in philosophy, some put focal attention on both intended action and intended effect, and their interactions. One of the influential philosophers with this orientation is Michael Bratman. In (Bratman, 1987, p. 5–9), he rejects "**the methodological priority of intention-in-action**," (which represents a strong philosophical tradition) making a fundamental assumption that humans are in an essential way *planning* agents.

So, for models of intention, there is a design issue of whether to recognize intended effects, whether to have an action/effect contrast, whether to have intentions representing an overlap of intention and action and whether to have a residue that is neither. Bratman notes that in such issues philosophy draws significantly on artificial intelligence (Bratman, 1990). For examples see (Cohen et al., 1990), the same volume.

Partialness

It may be possible, in studies of intentional actions by others, to restrict one's focus to fully specified intentions. This would yield a certain sort of simplicity. However, if we are to think effectively about our own actions, or write computer programs which function as agents, or otherwise model the intentions we encounter, we must represent partially specified intentions such as we find in plans. Current research must balance short term simplicity with the risk that work that does not represent partially specified intentions will not generalize.

Priorness

As Bratman notes, there is a tradition in philosophy of studying the intentions that immediately accompany actions, often called intention-in-action. (See (Anscombe, 1963; Goldman, 1970; Davidson, 1980; Searle, 1983).) There may be qualitative differences between an intention-in-action which is inseparably attached to an action that is in progress and an intention that is not so attached. There is thus a design decision in modeling intentions about how each is represented, and how the two are related. For prior intentions, but perhaps not for intention-in-action, there are also issues of how commitment to an intention is made or abandoned.

Tacitness

The intentions involved in language use are generally not consciously experienced. Some sort of contrast may be needed in models, and it may be consequential. For example, it may affect what sorts of reasoning or emotional influences might apply to the intention.

Immediacy

Varieties of intentions that inherently must be acted upon immediately are, by that attribute, subject to distinctive methods of recognition, understanding and pursuit. For example, a question may expression of an intention that the parties immediately jointly pursue the intention of the speaker knowing the addressee's response to the question. Immediacy is part of the understanding of questions, and although postponing a response may be negotiated, it constitutes a rejection of the question.

Interaction-configuring

Some actions in interaction are intended to alter the configuration of the interaction. For example, intentions may be added or removed from the set of mutually accepted joint intentions that are regulating a dialogue. These actions may form a very small class for which the recognition, acceptance and rejection methods may be specific to the class. If these methods are distinct then it is worthwhile to treat these intentions in a distinct manner.

Intended to be recognized

Ever since Grice wrote about intentions that are intended to be recognized (Grice 1975), there have been active attempts to incorporate his insight into larger frameworks. Use of such intentions appears to have far broader scope than Grice discussed. Just as for the interaction-configuring intentions, these have distinctive recognition, acceptance and rejection methods and so can benefit from special treatment.

Jointness

There are actions that are inherently those of multiple actors. They may be physical actions, such as moving a large table, or language use actions, such as negotiating a price. Clark has argued extensively for recognizing joint actions in language use. (Clark, 1996) His underlying assumptions generally follow Tuomela.

Joint actions are typically undertaken on the basis of joint intentions. If we see intentions as states of mind, and minds as personal and individual, and if we recognize interaction on a sensory basis but deny any sort of telepathy, then joint intentions cannot be directly analogous to joint actions. Joint actions typically have individual actions that are in some sense subordinated to the joint action, and such individual actions are also intentional. In certain cases, commitment to the joint intention and commitment to the individual intention may be inseparable.

3 Attributes of Collections of Intentions

The attributes below are typically involved with multiple intentions of a single intender, although the first one is not necessarily multiple, but has multiple intenders.

Sharedness

Multiple parties can share particular intentions. For example, two people may each intend to wash dishes until the pile of dirty dishes is exhausted. Sharing of an intention does not make it a joint intention, and having two parties committed to a shared intention does not produce cooperation. (For example, the dishes may be washed by the one who arrives first.) Even so, identifying intentions as shared can be consequential in directing or modeling interactions.

Structuredness

Many discussions treat intentions only in isolation. However in practice, intentions are often formed, negotiated and carried out in compatible sets, being restricted so that commitment to the various intentions is a single act. Plans may include a structure of subplans, and choices in planning may involve commitment or abandonment of large intentional combinations of this sort, combinations that may include more than one actor. Providing for this (or not) is a design decision in modeling. There is already an extensive literature that implicitly relies on structured intentions in language use. It includes the text planning literature, of course, but also various kinds of "dialogue acts" and various constructs called "dialogue games." (For example, see (Stolcke et al., 2000) and (Mann, 1979, 1988).)

Complementarity

Given a joint intention or comparable construct, individual actions to fulfill the intention may simply be additive, with each actor making some contribution to the total, and the intentional structure may reflect this.

Alternatively, there may be distinctive roles for the parties, with the intentions interlocking (complementary) so that the fulfillment of the joint intention depends on the parties fulfilling their distinct individual intentions. Such intentions are complementary.

Tutoring exemplifies this. In tutoring there is a shared joint intention that the student come to master some body of ideas or skills, and accompanying this intention are intentions about what the teacher will do and what the student will do. Teachers will periodically test students' knowledge, and on those occasions students will attempt to exhibit their knowledge. The testing and exhibition of knowledge are intentional, and their interaction makes the intentions complementary.

Conventionality

For recurrent kinds of situations, and the corresponding recurrent kinds of intentions, the intentions function as problems to be solved. As a consequence, the actions which are taken function in part as solutions to problems. Patterns arise and become acculturated or conventionalized, incorporated into conventions of interaction.

Tutoring also exemplifies this. Tutoring is not reinvented every day on demand. Rather it is developed as a collection of plans, skills, strategies, habits and choices that are created

incrementally over periods of years. These patterns of intentions affect the language used in ways that are every bit as conventional as the lexicon.

Other intentions, such as the intention to express appreciation for a favor, or the intention to acknowledge an expression of appreciation, also fall into patterns and become conventionalized. For recurrent situations that are culturally specific, such as giving and responding to praise, conventions of different cultures also differ. The Western practice of acceptance and expression of gratitude, or the Asian practice of denial and humility, are more easily treated as conventional patterns of intentions rather than as responses generated immediately on the basis of need.

The list of attributes above could surely be made longer and refined in detail, based on the extensive literature and many ongoing investigations. These attributes represent an informal core of notions that are necessary for representing any major share of human linguistic experience. There are some dependencies, so that they do not represent all conceivable combinations. Yet the projected complexity is still formidable.

4 Intentions in Philosophy and Linguistics

On the narrow subject of intentions, we have seen how linguistics and philosophy have been intertwined. There has been an impressively large amount of interaction on a broad range of subjects, represented in only a token way by this particular topic. All of the work cited above is a tiny portion of the total.

In this broad interaction, linguistics is functioning mainly in roles of observer and consumer. Philosophy is in more of a producer role, perhaps sometimes driven by a perception of being needed in linguistics, perhaps not. (Clearly the desire for accounts of personal responsibility, personal intention and intentional action, along with the need to extend these notions to social groups, have shaped the work in philosophy. But these concerns are not distinctively linguistic.)

Philosophy functions for linguistics as a source of ideas, a locus of invention. It also functions as a kind of quality control, a way of checking that conceptual schemes do not embody particular kinds of errors. Philosophy provides no guarantees, but it is a corrective force, genuinely beneficial.

It would be simplistic to try to say that one discipline leads and the other follows. Neither has enough unity for that. Sometimes linguistics has used complex concepts of intention that I have not found discussed in philosophy. An example would be joint intentions of effects, directly analogous to the joint intentions of actions that accompany joint actions discussed by Clark. Another example would be the intentions in dialogue game theory (Mann, 1979, 1988), which posited configurations of joint intentions and joint actions, structurally combined with complementary individual goals, involving intention of both effects and actions, both prior intentions and intentions-in-action, all represented in culturally shared conventions of interaction. These conventions were employed by means of a novel collection of speechact like devices. Although the paper was published in a philosophy journal, it stood disconnected from philosophy. Surely it would have benefited from a closer interaction.

We have seen that some conceptual frameworks used in linguistics are derived from chains of development in philosophy. What of the rest? There are certainly original inventions, and also cases where accounting for data leads to novel insight. The remainder seems to come from commonly held views, the so called "folk linguistics," of the home culture of the

researchers. (This suggests that more effort should be made to seek out cultural diversity among interacting researchers.)

In philosophy as elsewhere, local preferences vary, fashions of study vary, and not all work is based on the most fully developed line of work. Often, as in the case of Austin's groundbreaking work cited above, as well as many efforts related to computation, ideas are first used in an informal way, and only appear as focal philosophical topics much later.

Each of these sources of ideas, philosophy, data driven research and common culture, carry conceptual distinctions, local consistency and a kind of validation. Their interaction in linguistics seems to be a fruitful cross fertilization.

5 Conclusion

Some researchers will see particular choices and issues above as inadequately defined in the literature. Others will see them as defined but unresolved, and yet others will see them as old issues, now resolved. If the field proceeds by successive approximation, all three views may have some merit.

It seems clear that concepts of intentions can no longer be studied in isolation. They are now in the literature inseparably linked to intended effects (ends), intended actions (means), hierarchies of intentions (plans) and graded notions of likelihood, confidence or the like.

Although several of the cited titles mention "communication," they mostly take communication as a known base upon which theoretical understanding of intentions can be built. However, it is difficult to find an articulation of how that base might be defined. Defensive work like that of Gibbs, along with the constructive work represented here by many references, are both needed. Elaboration of the work toward a richer account of communication seems timely.

Bibliography

Anscombe, G. E. M. (1963). *Intention.* Cornell UP.

Asher, N. (1993). *Reference to Abstract Objects.* Kluwer AP.

Austin, J. L. (1975). *How To Do Things With Words.* Harvard UP.

Bratman, M. (1987). *Intention, Plans and Practical Reason.* Harvard UP.

Bratman, M. (1990). What is intention? in: Cohen et al. (1990).

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

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

Craig, R. T. (1999). Communication Theory as a Field. *Communication Theory*, 9(2):119–61.

Davidson, D., editor (1980). *Essays on Actions and Events*, chapter Intending. Oxford UP.

Gibbs, R. W. J. (1999). *Intentions in the Experience of Meaning.* Cambridge UP.

Goldman, A. (1970). *A Theory of Human Action.* Prentice-Hall.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic.* Kluwer AP.

Litman, D. and Allen, J. (1990). Discourse Processing and Commonsense Plans. in: Cohen et al. (1990).

Mann, W. C. (1979). Dialogue Games. Technical report, USC Information Sciences Institute, Marina del Rey, CA. ISI/RR-79-77.

Mann, W. C. (1988). Dialogue games: Conventions of human interaction. *Argumentation*, 2:511–32.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–81.

Searle, J. R. (1983). *Intentionality.* Cambridge UP.

Stolcke, A., Reis, K., Coccaro, N., Shriberg, E., Bates, R., Van Ess-Dykema, C., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–74.

Tuomela, R. (2000). *Cooperation: A Philosophical Study.* Kluwer AP.

Paper 13

What does 'X is a Y' mean?: Sentence meaning and four types of speech act

Etsuko Oishi
The Department of English Language and Culture, Fuji Women's College
Kita 16 Nishi 2-21, Kita-ku Sapporo 001-0016, Japan
etsuko@fujijoshi.ac.jp

Abstract

1 The general assumption of semantic meaning

It has been generally assumed that we have to understand two types of meaning to understand what the speaker says by uttering a sentence. One is the truth-conditional meaning of the sentence: a state of affairs which is unambiguously correlated with the sentence. The other is the non-truth-conditional meaning, which is fully dependent on the specific context in which the sentence is uttered. Truth-conditional meaning is described in semantics, and non-truth-conditional context-dependent meaning is analysed in pragmatics. In this paper, we will criticise the generally accepted assumption that a sentence is unambiguously correlated with one state of affairs as its truth-condition and other meanings are context-dependent meanings. We will claim that, owing to the assumption, we fail to see the diversity of sentence meaning. As we will show below, by uttering a sentence, the speaker can express two or more meanings, and those meanings are equally context-independent. To explain this, we have to describe not the diversity of meaning which comes from the specific context in which the sentence is uttered, (i.e., pragmatic meaning), but the diversity of meaning which comes from the sentence itself, or from the general context in which a sentence is uttered to perform a certain speechact, which we might call *extended semantic meaning*. In Sections 2–4, we will discuss generics, Donnellan's referential use and attributive use, and Burton-Roberts's A-type and B-type utterances as examples in which diversity of sentence meaning is not fully explained. Then we will propose to explain this diversity of meaning in terms of the difference in speech act by employing Austin's (1953) idea that the speaker can potentially perform four types of speechact by uttering the sentence, 'X is a Y'.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

The Department of English Language and Culture, Fuji Women's College
Kita 16 Nishi 2-21, Kita-ku Sapporo 001-0016, Japan
etsuko@fujijoshi.ac.jp .

2 Generics

Generic sentences are generally assumed to be the sentences where an entire class or kind rather than a particular individual is referred to and generalization about the class or kind is expressed[1]. For example, 'a dog', 'the dog', and 'dogs' in the sentences in (1) do not refer to a particular dog or a particular group of dogs, but to the class or kind DOG:

(1) a. A dog has four legs.

b. The dog has four legs.

c. Dogs have four legs.

However, since (Carlson, 1977), another interpretation of genericity has been generally granted. Let us see his famous example :

(2) Beavers make dams.

Krifka et al. (1995) explain Carlson's interpretation of genericity as generalization over events as follows:

> The second phenomenon commonly associated with genericity are [*sic*]propositions which do not express specific episodes or isolated facts, but instead report a kind of general property, that is, report a regularity which summarizes groups of particular episodes or facts. (Krifka et al., 1995, 2)

Krifka et al. say the sentence in (3) does not report a particular episode but a habit, some kind of generalization over events:

(3) John smokes a cigar after dinner.

It is notoriously difficult to describe generic sentences within the truth-conditional theory. Using the examples in (4)–(6), Lyons says:

(4) The lion is a friendly beast.

(5) A lion is a friendly beast.

(6) Lions are friendly beasts.

> ... The kind of adverbial modifier that suggests itself for insertion (either in initial position or immediately after the verb) in (4)–(6), is one that approximates in meaning to 'generally', 'typically', 'characteristically' or 'normally', rather than to 'essentially', or 'necessarily'; and it is notoriously difficult to specify the truth-conditions for propositions containing adverbs of this kind (cf. (Lewis, 1975)). They certainly cannot be formalized, in any straightforward fashion, in terms of either universal or existential quantification; and, so far at least, there does not seem to be available any satisfactory formalization of the truth-conditions of the vast majority of the generic propositions that we assert in our every day use of language. (Lyons, 1977, 196)

[1]See (Lyons, 1999, 179)

We cannot equate the meanings expressed in (4), (5) and (6) with a state of affairs where every lion is a friendly beast ($\forall$x[lion'(x) $\rightarrow$ friendly'(x) & beast'(x)]). This is simply too strong: the existence of a lion which is not a friendly beast does not immediately falsify the sentence. We cannot equate them with a state of affairs where there is a lion which is a friendly beast or there are some lions which are friendly beasts, either ($\exists$x[lion'(x) $\rightarrow$ friendly'(x) & beast'(x)]). The sentences in (4), (5) and (6) mean more than this. Similarly, the sentence in (3) does not mean that John always smokes a cigar after dinner: the instance in which John does not smoke after dinner does not immediately falsify the sentence. The sentence in (3) does not mean that John smokes/smoked a cigar after dinner at a particular time, either.

3 Referential use and attributive use

Donnellan (1966) distinguishes attributive use from referential use of definite descriptions. By using a definite description, the speaker either refers to a particular entity or whoever/whatever is of the description. Consider Donnellan's famous example:

(7) Smith's murderer is insane.

Donnellan (1966) explains the referential use of the description 'Smith's murderer' as follows:

> Suppose that Jones has been charged[2] with Smith's murder .... Imagine that there is a discussion of Jones's odd behavior at his trial. We might sum up our impression of his behavior by saying, 'Smith's murderer is insane'. (Donnellan, 1966, 103)

In the referential use of the definite description of (3) 'Smith's murderer', the speaker refers to a particular person, say, Jones, by 'Smith's murderer' and describes him as insane. There is another use of the definite description of 'Smith's murderer', which Donnellan (1966) calls the attributive use:

> Suppose that we come upon poor Smith foully murdered. From the brutal manner of the killing and the fact that Smith was the most lovable person in the world, we might exclaim, 'Smith's murderer is insane'. (Donnellan, 1966, 102)

Donnellan's referential and attributive uses of the definite description show that we can refer not only to a particular entity, such as a particular person, Jones, but also to a type of entity, such as a type of person who kills a lovable person in a brutal manner. The attributive use of 'Smith's murderer' is more clearly brought out by paraphrasing the sentence in (7) as:

(8) Whoever killed Smith is insane.[3]

Donnellan (1966) then shows an interesting consequence when these two types of referring fail. When the first type of referring, i.e. referential use, fails, the statement might still hold, but when the second type of referring, i.e. attributive use, fails, the statement does not

[2] It is inappropriate to call Smith's murderer someone who has been charged with Smith's murder. Presumably, Donnellan meant 'Jones has been *convicted* of Smith's murder.'

[3] See (Lyons, 1977, 185–186)

hold. Suppose that the person who is referred to by 'Smith's murderer', i.e., Jones, is in fact not Smith's murderer. This is incorrect referring in the sense in which Lyons (1977) uses the term: the entity is not Smith's murderer. However, it can be successful referring: if the hearer identifies the entity the speaker intended to refer to by 'Smith's murderer' and the entity is in fact of the type INSANE, the statement, 'Smith's murderer is insane', is true. However, the situation is different if the second type of referring, i.e. attributive use, fails. Suppose the police found out that Smith had killed himself. The speaker then fails to refer to a type by 'Smith's murderer'. The type of person who kills a lovable person in a brutal manner cannot be referred to by 'Smith's murderer' because Smith killed himself. As a consequence, the statement is about nothing. In the truth-conditional theory of meaning in which a sentence is correlated with a state of affairs as its truth-condition, these two types of referring cannot be distinguished. In such a theory, words and noun phrases are correlated with entities, and the distinction is not drawn between referring to an entity as an individual who has idiosyncratic features and referring to the attribute which is embodied in the individual.

4 A-type and B-type utterances

Burton-Roberts (1986) distinguishes two different interpretations of the following sentence (9) and describes it in terms of difference in discourse. He says that when 'Max' in the following utterance is not a theme, the utterance is either about a person who is dandy or about the attribute of 'dandiness':

(9) Max is a dandy.

The difference is clarified when the utterance is put into different discourses, (10) and (11):

(10) A: Who is a dandy?
B: Max is a dandy.

(11) A: What is a dandy?
B: Max is a dandy.

The example in (10) shows that 'Max is a dandy' is uttered as the answer to the question about the individual who has the feature of being dandy. The example in (11) shows that the same sentence can be uttered as the answer to the question about the attribute of dandiness. This point becomes clear when we compare them with a sentence, 'A dandy is what Max is'. Let us see the following examples:

(12) A: Who is a dandy?
B: !A dandy is what Max is.

(13) A: What is a dandy?
B: A dandy is what Max is.

'A dandy is what Max is', which is overtly about dandiness, can be exchangeable with 'Max is a dandy' in (11), but not with 'Max is a dandy' in (10). Burton-Roberts (1986) calls 'Max is a dandy' in (10) a Type A utterance and the one in (11) a Type B, and analyses the

difference between these utterances as a difference in discourse, i.e., Type A discourse and Type B discourse. Burton-Roberts (1986) gives a further interesting observation; a question that initiates a B-type discourse, say, 'What is a dandy?' in (11), constitutes a canonical means of requesting a definition:

> Canonical answer to that question should supply, or purport to supply, definitions, or at least partial definitions. (Burton-Roberts, 1986, 55)

In other words, to utter a sentence 'X is a Y' as a B-type utterance is to give a definition-like description about Y. Burton-Roberts (1986) does not go so far as to say that the sentence 'Max is a dandy' has two meanings: one is about an individual, Max, who happens to have a feature of being dandy; the other is about the attribute of dandiness which is embodied in Max. This would be against the assumption of the standard truth-conditional theory that a sentence can be correlated with a state of affairs of the world. The above examples of generics, referential use and attributive use of definite expressions, and A-Type utterance and B-Type utterance clearly show that the sentence is not only about a particular entity and its feature but also about a type of entity and its attribute. The truth-conditional theory of meaning does not cope with this diversity of sentence meaning, at least in a straightforward sense.

5 How do we solve these problems?

We have been criticizing the truth-conditional theory by saying that it does not explain the diversity of sentence-meaning which contributes to a proposition. We need a theory of meaning whose scope is wide enough to describe different aspects of sentence meaning. Austin's speechact theory is a promising theory to explain such diversity of sentence meaning because the theory allows for the speaker to perform different types of speechact by uttering the same sentence. As one type of speech act, the speaker asserts something about a particular entity. The speaker refers to a particular entity and asserts that its feature is of a certain type. This is a truth-conditional meaning. The speaker can perform another type of speechact. The speaker refers to a type of item and asserts that it has a certain attribute. In the following we will claim that the difference between these speechacts amounts to the differences in meaning between generic and non-generic sentences, and between attributive use and referential use of definite descriptions. A-type utterances and B-type utterances will be also explained as different speech acts. We will first explain Austin's (1953) theory in the following section.

6 Four different speechacts in Austin (1953)

In Austin (1953), which is less known than *How to Do Things with Words* (Austin, 1962), Austin claims that even in uttering a sentence as simple as 'X is a Y', the speaker can potentially perform four types of speechact. To explain basic speech acts Austin hypothesises what he calls *'Speech-situation S_o'*, a simplified model of a situation in which we use language for talking about the world. In S_o, the world consists of numerous individual items and each is of one definite type. Imagine the world consisting of numerous colour patches of the same pure red, the same pure blue or the same pure yellow, each of which has a number applied to it. Or imagine the world consisting of numerous pieces of paper in the shape of the same

triangle, the same oval or the same rhombus, each of which has a number applied to it. The language in S_o permits only sentences of one form *S*:

(14) I is a T,

where 'I' stands for an item and 'T' stands for a type. The language contains an indefinite number of words inserted in the place of the 'I' or the 'T' in form S. Each of these words is either an I-word or a T-word in the language. For example, the following sentence is a grammatical sentence in the language:

(15) 1227 is a rhombus.

There are also two sets of semantic conventions. One is an *I-convention*, or a convention of reference, which fixes the item to which an I-word is to refer. The other is a *T-convention*, or a convention of sense, which correlates a T-word with the item-type

We may inaugurate T-conventions by one or the other of two procedures of linguistic legislation: *name-giving* and *sense-giving*. Name-giving consists in allotting a certain word to a certain item-type as its 'name'. Sense-giving consists in allotting a certain item-type to a certain word as its 'sense'. For example, we might give the word 'dog' to an item-type which is an animal of canine type as its 'name'. This is name-giving because the name 'dog' is given to the item-type. We might give an item-type, 'an animal of canine type', to the word 'dog' as its sense. This is sense-giving because the sense, 'an animal of the canine type', is given to the word. When either procedure has been gone through, a specific type is attached by convention to a certain word, i.e., a T-word and its 'name', as the 'sense' of that word. Then a satisfactory utterance (assertive) on any particular occasion will be one where the item referred to by the I-word (in accordance with the I-conventions) is of the type which matches the sense which is attached to the T-word (by the T-conventions). Austin (1953) then distinguishes four different speechacts which are performed by the whole utterance of an assertion in the *Speech-situation S_o: placing* (c-identifying or cap-fitting), *stating, casting* (b-identifying or bill-filling), and *instancing*. How does this complexity arise? There is first a difference in direction of fit between fitting a name to an item and fitting an item to a name. The differences of fit here are as different as fitting a nut with a bolt and fitting a bolt with a nut. We may be given an item, and purport to produce a name whose sense matches the type. Conversely, we may be given a name and purport to produce an item whose type matches the sense of that name. There is also another distinction to be drawn. We fit the name to the item or the item to the name on the grounds that the type of the item and the sense of the name match. But in matching X and Y, there is a distinction between matching X to Y and matching Y to X. Austin calls this the difference in the onus of match. These two distinctions generate four different performances in uttering the sentence, '1227 is a rhombus'. To explain the choice of terms, we use the useful word 'identify' in two opposite ways. We may speak of 'identifying it (as a daphnia)' when you hand something to me and ask me if I can identify it. We also speak of 'identifying a daphnia (or the 'identifying the daphnia')' when you hand me a slide and ask me if I can identify a daphnia (or the daphnia) in it. In the first case we are finding a cap to fit a given object: hence the name 'cap-fitting' or 'c-identifying'. In other words, we are trying to 'place' it. But in the second case we are trying to find an object to fill a given bill: hence the name 'b-identifying' or 'bill-filling'. In other words, we 'cast' this thing as the daphnia. The terms 'stating' and 'instancing' are used in their usual senses. *Placing* and *stating* have the same direction of fit, i.e., fitting names to given items.

Also *instancing* and *casting* have the same direction of fit, i.e., fitting items to given names. *Placing* and *instancing* have the same onus of match: the type of the item is taken for granted and the question might be whether the sense of the T-word is such as really to match it. In both *stating* and *casting* the sense of the T-word is taken for granted and the question might be whether the type of the item is really such as to match it. Since Austin does not give detailed explanations of these different speechacts, we will explain *stating, placing, casting,* and *instancing* in the following examples. Let us start with the act of *stating.* Imagine there are numerous pieces of paper in the shape of the same triangle, the same oval or the same rhombus, each of which has a number applied to it. Consider the following conversation:

(16) A: 1225 is a triangle. 1226 is a triangle, too. What is 1227?
B: 1227 is a rhombus.

To speaker A's question, 'What is 1227?', speaker B thinks of the type of the item to which the number 1227 is applied, and says, '1227 is a rhombus'. This type of speechact can be explained as follows: the speaker refers to the item by '1227' and asserts that the type of this item is such that it matches the sense of the word 'rhombus'. If someone else disputes speaker B, the point of the dispute is whether the type of this item matches the sense of the word 'rhombus' (whether the type of this item is of a type RHOMBUS) as she claims it to be. This is shown in the following conversation:

(17) A: 1225 is a triangle. 1226 is a triangle, too. What is 1227?
B: 1227 is a rhombus.
C: No! Look. The sides of the item 1227 are not equal.

Next, the speech act of *placing.* Let us examine the following conversation:

(18) A: (picking out a piece of paper) Can you identify 1227?
B: 1227 is a rhombus.

When speaker A asks, 'Can you identify 1227?', speaker B thinks the sense of the word which matches this item-type embodied in the item (to which the number 1227 is applied). By uttering the sentence, '1227 is a rhombus', the speaker refers to the item-type, not the item, by '1227' and asserts that the sense of the word 'rhombus' is such that it matches this item-type. If someone disputes the speaker, the point of the dispute is whether the sense of the word 'rhombus' or the sense of other words matches the item-type. Let us examine the following conversation:

(19) A: (picking out a piece of paper) Can you identify 1227? B: 1227 is a rhombus. C: No! A rhombus has crooked corners.

The speechacts of *stating and placing* have the same direction of fit, producing a T-word to match the given item or item-type. When the speaker performs the speechacts of *casting* or *placing*, on the other hand, the direction of fit is producing an item to match the given T-word. Let us start with *casting.* Consider the following conversation:

(20) A: (looking at many pieces of paper) Can you identify a rhombus?
B: 1227 is a rhombus.

Asked if she can identify a rhombus, speaker B thinks of the item-type which matches the sense of the word, 'rhombus', while taking the sense of the word for granted. Then speaker B refers to the item-type by '1227' and asserts that the item-type 1227 (embodied in the item) is such that the sense of this word matches it. If someone disputes the speaker, the point of dispute is whether it is the item-type 1227 (embodied in the item) or another item-type that matches the sense of this word 'rhombus'. Consider the following conversation:

(21) A: (looking at many pieces of paper) Can you identify a rhombus?
B: 1227 is a rhombus.
C: No. The sides of the item 1227 are not equal

.

The last one is a speech act of *instancing*. Let us examine the following conversation:

(22) A: Which is a rhombus?
B: 1227 is a rhombus.

When speaker A asks, 'Which is a rhombus?', speaker B thinks of the sense of the word 'rhombus'. Then , speaker B refers to the item by '1227' and asserts the sense of the word 'rhombus' is such that the type of the item 1227 matches the sense of this word. If someone disputes with speaker B, the point of dispute is whether the sense of the word, 'rhombus', is really such as to match the type of the item 1227. Consider the following conversation:

(23) A: Which is a rhombus?
B: 1227 is a rhombus.
C. No! A rhombus has crooked corners.

Austin (1953) restates these four speechacts as follows:

To state we have to find a pattern to match this sample to.
To place we have to find a pattern to match to this sample.
To cast we have to find a sample to match to this pattern.
To instance we have to find a sample to match this pattern to.

Austin (1953) proposes a framework of semantic theory in which we can describe as different speechacts different meanings expressed by uttering the same sentence. This allows us to describe meaning in general in a much wider scope than that of the truth-conditional theory.

7 Difference meaning as different speech acts

In Sections 2–4, we claimed that there are meanings which are directly related to the propositional contents of a sentence, that, nevertheless, have not been described well within the truth-conditional theory. Those meanings concern generics, attributive use of definite expressions, and Type-B utterances. They are not about a particular entity or a particular situation at a certain time and place. Those meanings are about a type of entity or an attribute of a type. We have shown that we can refer to either a particular item in the act of *stating* or a particular item-type in the act of *placing*. This distinction seems to be the distinction which exists between generic sentences and non-generic sentences. For example, by uttering the following sentence:

(24) The dog is a friendly animal,

the speaker refers to a particular entity by 'the dog' and asserts that the type of this item is such that it matches the composite sense of the words, 'friendly animal'. While asserting this, the speaker describes the item he refers to. The same sentence can be uttered to perform a different type of speechact, i.e., the act of placing. The speaker refers to an item-type by 'the dog', and asserts that the composite sense of the words, 'friendly animal', is such that it matches this item-type. While asserting this, the speaker describes an attribute of this item-type. Austin (1953) does not say that there are also four different types of speechact about a situation, but it is possible to expand his theory so that we can distinguish different speech acts about a situation. Then we can describe in a similar way generic sentences about a situation such as the one in the following:

(25) Beavers make dams.

The distinction between Donnellan's referential use and attributive use of definite expressions also corresponds to the distinction between the act of *stating* and the act of *placing*. By uttering the sentence in (26):

(26) Smith's murderer is insane,

the speaker can refer to a particular person, such as Jones, and asserts that the type of this person is such that it matches the sense of the word, 'insane'. While asserting this, the speaker describes the item he refers to. This is the act of *stating*. Another act is also possible. The speaker can refer to the item-type, such as a brutal murder of a most lovable person like Smith, and asserts that the sense of the word 'insane' is such that it matches this item-type. While asserting this, the speaker describes the attribute of this item-type. Burton-Roberts's distinction between Type-A utterance and Type-B utterance seems to correspond to the distinction between *instancing* and *casting*. We can explain Type-A utterance as follows:

(27) A: Who is a dandy?
B: Max is a dandy.

As the answer to speaker A's question, 'Who is a dandy?', speaker B refers to a person, Max, and asserts that the sense of the word 'dandy' is such that the type of the item matches the sense of this word. While asserting this, the speaker exemplifies the sense of the word. This is the act of *instancing*. Type-B utterance can be explained in terms of the act of *casting*:

(28) A: What is a dandy?
B: Max is a dandy.

As the answer to speaker A's question, 'What is a dandy?', speaker B refers to the item-type which is embodied in Max, and asserts that the item-type is such that the sense of this word 'dandy' matches the item-type. While asserting this, the speaker describes the sense of the word 'dandy'. As we explained in Section 4, Burton-Roberts says Type-B utterance purports to supply definitions, or at least partial definitions. This is compatible with our idea that, while performing the act of *casting*, the speaker describes the sense of the word, that is, he supplies the definition. The speaker can perform the same speech act by uttering the sentence in (29)-B, which is the case of ostensive definition:

(29) A: What is a dandy?
B: That is a dandy.

In conclusion, the analysis provided by the traditional truth-conditional theory is not fine-grained enough to describe the diversity of meaning concerning a proposition. Austin's theory of meaning is a promising alternative because different meanings the speaker expresses by uttering one and the same meaning can be described as different speechacts the speaker performed by uttering the sentence. Within this theory, we do not have to correlate one sentence with one state of affairs. This is, of course, a well-known hypothesis in Austin's speechact theory and different versions of the speech act theory developed by others. In the present paper, however, we have shown that this hypothesis also applies to statements. That is, the meaning the speaker expresses by uttering a sentence as a statement can be described within the speechact theory: uttering a sentence to make a statement about the world can be analysed as performing one type of speech act. In this theoretical framework meaning can be explained triadically among a sentence and a speechact (an illocutionary force) and a context (as an abstract entity), rather than dyadically between a sentence and a state of affairs.

Bibliography

Austin, J. (1953). How to talk—some simple ways. In *Proceedings of the Aristotelian Society.* Reprinted in: J. O. Urmson & G. J. Warnock (eds.), Philosophical papers, Oxford: Oxford University Press, 134-153.

Austin, J. (1962). *How to do things with words.* London: Oxford University Press.

Burton-Roberts, N. (1986). Thematic predicates and the pragmatics of non-descriptive definition. *Journal of Linguistics*, 22:311–329.

Carlson, G. (1977). *Reference to Kinds in English.* PhD thesis, University of Massachusetts, Amherst. Published 1980 by Garland Press, New York.

Carlson, G. and Pelletier, F., editors (1995). *The generic book.* Chicago, University of Chicago.

Donnellan, K. (1966). Reference and definite description. *Philosophical Review*, 75. Reprinted in: D. Steinberg and L. Jacobovits (eds.), 1971. Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology. Cambridge: Cambridge University Press, 100-114.

Keenan, E., editor (1975). *Formal semantics of natural language.* Cambridge, Cambridge University Press. 3-15.

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Link, G., and Chierchia, G. (1995). Genericity: an introduction. in (Carlson and Pelletier, 1995).

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

Lewis, D. (1975). Adverbs of quantification. in (Keenan, 1975).

Lyons, C. (1999). *Definiteness.* Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics. Vol. I and II.* Cambridge: Cambridge University Press.

Lyons, J. (1981). *Language, Meaning and Context.* Suffolk: Fontana.

Oishi, E. (1999). *Indicating and Referring: A Speech Act approach to communicative meaning.* PhD thesis, University of Edinburgh.

Tsohatzidis, S., editor (1994). *Foundations of speech act theory.* London: Routledge. 156-166.

Paper 14

Between Binding and Accommodation

Jennifer Spenader
jennifer@ling.su.se
http://www.ling.su.se/staff/jennifer

Abstract

How should naturally occurring definite descriptions be resolved and represented if analyzed in the presuppositions as anaphora theory (van der Sandt, 1992)? Particulary, what should be done with those examples that lie somewhere in between binding and accommodation and depend on non-discourse information for resolution. Three types of examples are particularly difficult: 1) bridging anaphora, 2) known information, and 3) deducible or inferrable information. Conclusions are made based on a corpus analysis of 406 definite descriptions taken from three interviews in the London-Lund Corpus of Spoken English. In particular, how these types of examples are related to the discourse record, and how they could be represented is discussed.

1 Introduction

The presuppositions as anaphora theory of van der Sandt (1992) is considered to have the best empirical coverage for examples of presupposition in the literature. However, for many naturally produced examples it is not clear what the correct analysis should be. This gap is especially clear for the presuppositions triggered by definite descriptions.

Here naturally produced examples of presuppositions triggered by the definite article, demonstratives and possessives are discussed in relation to how they could be analyzed within the anaphoric theory. Problem examples are identified as groups that do not clearly fall into the categories of accommodated or bound.

The rest of the paper is structured as follows: Section 2 describes the most relevant corpus work, section 3 presents the presuppositions as anaphora theory and discusses how it deals with problematic definite descriptions. The methods used in the corpus study are given in section 4, with results in section 5. Section 6 discusses the results, and discusses them in relation to the question of how presuppositions can best be processed and represented. Finally, this work is related to the status of definite descriptions as carrier of new information and how the function of accommodation is characterized.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).
Paper 14,
jennifer@ling.su.se
http://www.ling.su.se/staff/jennifer.

2 Definite Descriptions: Empirical Work

There have been several taxonomies proposed for categorizing the relationships that can hold between a definite description and a discourse anchor[1], with the most well known among them are (Prince, 1981; Hawkins, 1978) and (Löbner, 1985), and the reader is referred to (Vieira, 1998) for a summary of the similarities and differences between the different taxonomies. Corpus work on definite descriptions has mainly looked at definite NPs in text. I am not aware of any corpus work on possessives.

Fraurud (1990) studied definite NPs in a written Swedish corpus by categorizing them either as *first mention* or *subsequent mention.* Her intention was to show that the use of definite descriptions to refer back to contextually given information (i.e. anaphoric) is not as central as generally believed, but that they are also used to introduce new information and that this function is equally important. This was confirmed in her finding that over 60 % of her definite NPs were first mentions definites, as she defined them.

Poesio and Vieira (1998) did two different large-scale annotation experiments on definite description usage in the Wall Street Journal in an attempt to find a classification that is reliable, defined operationally as a scheme that will lead to a high degree of inter-annotator agreement as defined by statistical measures such as the Kappa-value.[2] The first experiment (3 annotators, 1,040 NPs) used a classification with five categories that referred explicitly to surface characteristics of the definite descriptions. They obtained a Kappa score of 0.68 for the first experiment and identified 204 descriptions as "associative" (or bridging[3]) relationships, meaning they made up about 20 % of all the definites studied.

The second experiment (3 annotators, 464 definite descriptions) used a different annotation scheme that was more focused on the meaning of the definite descriptions. Here annotators were linguistically naive and this could, along with the different annotation categories, account for the somewhat poorer agreement score, which was K= 0.58. Of this group it is interesting to note that 164 descriptions were classified as co-referential by all three coders (with 95% agreement) and only 7 of the 464 expressions were classified as bridging by all three, though the number of bridging descriptions identified by each annotator were 40, 29 and 49, respectively, so there was quite a lot of disagreement about when a definite description was related to the discourse by bridging.

In other work, Poesio et al. (1997) tested the feasibility of using WordNet to identify when two concepts are in a bridging anaphoric relationship, using the examples and categories from the first experiment presented above The results were not very promising, indicating just how difficult these examples are.

In summary, there does not seem to be any real consensus on how many different uses of definite descriptions should be recognized nor on how their relation to the discourse context

[1] I will reserve the term antecedent for co-referential binding and anchor for a related but not identical object or individual

[2] The Kappa statistic takes the agreement as well as the number of categories in a classification task in to the equations. A Kappa value between 0.6 - 0.8 is supposed to signify some degree of agreement and over 0.8. should allow conclusions to be drawn. See also (Carletta et al., 1997) and (Poesio and Vieira, 1998) for an explanation of how the Kappa value can be calculated for an annotation task.

[3] Note that in Clark's original paper on Bridging uses the term to cover all anaphoric relationships, including presupposition. For example, even pronoun usage (in Clark, ex. 4, *the man - he*) and coreference with the same head noun (Clark: ex. 1, *a man - the man*). This is not how the term has been used in later work, where bridging anaphora or bridging inferences do not include these types of examples, and are usually limited to relationships between two objects, and usually excludes coreference relationships (but cf. (Poesio et al., 1997)

should be described, nor does it seem to be easy to categorize examples in natural data with these taxonomies.

3 Definite Descriptions as Presuppositions

The definite article, NPs with demonstrative articles, and possessives are all considered presupposition triggers that presuppose the existence of the object they modify. The anaphoric theory of presupposition, developed by van der Sandt (1992), argues that presuppositions can be treated just as anaphora are treated in DRT (Kamp & Reyle 1993). Presupposition resolution involves examining the previous discourse context for an antecedent. If an antecedent is found then the presupposition is bound to it. If an antecedent cannot be found then the presupposition is accommodated, and the ability to accommodate is what distinguishes presuppositions from other anaphoric expressions. Binding and accommodation[4] are the two potential analyses discussed. Additionally, binding is preferred to accommodation, higher levels of accommodation are preferred over lower, and binding as close to the presuppositional expression as possible is preferred as well.

The frequency, and hence, importance, of being able to model these examples in presupposition theory has been underestimated for several reasons which are worth reflecting on. One reason is that almost all work in this area has traditionally focused on constructed examples where we as interpreters are presented with totally new information in the form of fictional discourse actors. New work in presupposition theory, such as (van der Sandt, 1992) has also focused on exemplifying the theory's ability to deal with well known problems, such as presupposition cancellation, projection problems or how to deal with presuppositions presenting new information. These puzzles have all been successfully solved in the anaphoric theory. But examples where presuppositions make reference to known information are seldom discussed because this type of resolution does not illustrate any traditional problems in presupposition theory, nor does it appear in small, constructed examples which usually introduce new actors to the reader. Another reason why these examples have been avoided is that they are very difficult.

3.1 Bridging in Presupposition Theories

There are two relevant discussions on dealing with bridging examples in the literature on the anaphoric theory of presupposition.

The first is in (Geurts, 1999), and he seems to be quite ambivalent about the need to distinguish them as either binding or accommodation. He points out that for many examples, two processing strategies seem to be possible. Either a discourse record given referent can act as an anchor, an license the creating of a new discourse referent where the presupposed information can be bound, or the definite description can be accommodated, and after accommodation this information can be related to the rest of the discourse record, something that is often discussed in presupposition theory as part of the 'wish list' of what should be in an adequate representation of accommodation, but has yet to be developed. He also points out that often the interpretations that result from these two different strategies are often quite similar, and it is difficult to say that one is preferred to the other. Geurts (1999) also

[4]There are potentially 3 different levels of accommodation, global, intermediate and local, but this won't be relevant for most of the discussion so it will be ignored

suggests that perhaps the binding theory could be modified to put these examples into a third resolution strategy category, bridging, where binding would be preferred to bridging, and bridging preferred to accommodation.

Piwek and Krahmer (2000) present a method to model presuppositional bridging where the presuppositions as anaphora theory is redone in Construction Type-Theory (CTT), a proof system. They show how world knowledge could be used to infer discourse referents based on other, already given discourse referent which in turn will allow binding between the inferred referent and the presupposition. When world knowledge doesn't contain information that allows this bridge to be made, the discourse referent is accommodated. Also pointed out in (Piwek and Krahmer, 2000) is that not all bridging relationships can be analyzed in terms of lexical relationships.

3.2 Three groups between binding and accommodation

Summarizing, partial resolution seems to fall into three groups when dealing with definite descriptions and there doesn't seem to be clear operational information about how they should be analyzed, either as binding or accommodation. These are:

1. **Bridging anaphora**: the information presupposed is not explicitly present in the discourse but an inference can be made in relation to an object in the discourse record, often because the object is in an predictable relationship with an already coded discourse referent (such as Clark's (Clark, 1975) "Necessary parts" or "Probably parts").
2. **Known information**: generally recognizable objects or individuals, known or globally known, or known as part of the discourse situation.
3. **Deducible information**: this is information that can be figured out from information that has already been given in the discourse, though it is not informationally redundant.

The frequency of these types of resolutions is difficult to determine by examining previous empirical studies. The (Fraurud, 1990) study places all of these categories as first mention, so the frequency of the above groups in comparison to the use of the definite description to introduce "purely" new information[5] is obscured in her study. The (Poesio and Vieira, 1998) study instead has used a classification based more on processing considerations. The count some co-reference relations as bridging because lexical information would be needed to recognize the relationship. This means that many of the examples that would clearly be interpreted as binding in the anaphoric theory are counted together with the bridging examples. Also, both of these studies looked a written language. In order to gain more information about the frequency and form of definite description usage in spoken discourse in general, and these three groups in particular, a corpus investigation was carried out.

[5]If there is such a thing beyond true cases of repair.See the discussion section.

4 Method

Three interviews taken from the London-Lund Corpus of Spoken English[6] were used for the empirical study. The combined interviews had a total length of approximately 8972 words. Tone units have often been considered a kind of 'sentence' for spoken language and the three dialogues contained 1029 tone units. Each interview was between three participants, with two interviewers who knew each other, and a potential student who met the interviewers at the time of the recording. In the interview, the interviewers steer the conversation by asking questions. Participants rarely make comments unrelated to the task at hand, making identifying references to private common ground less of an issue. Triggers studied were definite NPs, introduced either by the definite article, demonstrative articles or possessive pronouns and nouns modified by a noun with genitive 's.

Two annotators performed the annotation task, the author and another native English speaker who could be considered 'linguistically naive'. After a small pilot study an annotation scheme using eight categories was developed. 8 categories were used because it was believed that certain categories could be collapsed in analysis later if they proved to be less useful. The categories and their definitions are presented in table 14.1 below. The examples which were categorized as caculable(C) are often considered deictic. Because there seems to be no real consensus on where the distinction between deictic items and anaphoric items should be drawn as most expressions are context dependent (see the discussion in Levinson, 1997) so the potential distinction is ignored. For coreference(=) and related(R) the annotators identified the antecedent/anchor they considered the source of the relationship, and if there was more than one, to take the linearly nearest antecedent/anchor.

5 Results

A Kappa score for inter-annotator agreement was calculated for all categories (see table 14.2). Kappa was also calculated only for possessives and this resulted in a slightly better score, K = 0.56 The latter calculation was done to see if they were in anyway different from other definites, in that possessives have not been the subject of a corpus study that I am aware of. The totals under each category refer to the number of judgments made for each type by both annotators combined.

Type	Total	=	N	R	D	C	E	K	0	Kappa
All Definite Descriptions	406	235	261	115	77	71	47	2	4	0.45

Table 14.2: Frequency of each category

[6]Information on obtaining this corpus can be be found on the ICAME website at http://www.hd.uib.no/icame.html

CATEGORY(TAG)	DESCRIPTION	ANAPHORIC THEORY?
coreference(=)	refers to something mentioned earlier in the text	*binding*
new(N)	new to you , new to text, not related, not described	accommodated
related(R)	is related to something earlier in the text, but isn't the same thing - no description nearby	partial resolution
described(D)	self-explanatory because of how it is described, also new	*partial resolution, or accommodated?*
known(K)	everyone knows what this is, its general knowledge, but this entity is new in the text	*accommodated?*
calculable(C)	using information about a reference time or a reference place you can figure out what this means	*partial resolution*
expression(E)	the NP is part of an expression or idiom and doesn't really point out a real entity	should not get analyzed
other(O)	other	*accommodated?*

Table 14.1: Categories tagged for in the Empirical work

Example 1[7] (interview 1)

Interviewer a This means that you've got somebody * lined up to live in as A:[a sort of housekeeper]*[8]
Speaker A * I've got a, yes, a a living in* - girl, a living in girl
Interviewer a Who can really take B:[your place] there ?
Speaker A Yes - she takes C:[my place] yes, she's very good indeed
Interviewer a have you . tried this . at all . so far . I mean have you * got round to anything*
Speaker A * no, I haven't* -, I haven't . I mean . I've done nothing except . you know . bring up D:[this family] since I . left school
Interviewer a Yes - it's not as though you have already tried for two or three months to see, how this works out, working
Speaker A No, no, What I did do a certain amount, I've done I did a certain amount of reading during E: [the last few months]

Examples 1 illustrates a long sequence where both annotators were in complete agreement about category. B:[your place] was considered by both annotators to be related(R) to the noun phrase marked A: by both annotators. C:[my place] was identified (i.e. coreferential(=)) with B:[your place] and D:[this family] was considered new to the discourse (though most likely known to the two interviewers, it seems. See the discussion on perspective in section 6.1). E:[the last few months] is a time that must be calculated from the time of the discourse and as such was marked as C by both annotators.

[7] In some cases the examples have been simplified from the original to make them more clear and to conserve space.

[8] Underlined words mark overlapping speech and the diacritic marks on the ends identify what overlapped with what

Example 2 (interview 1, begin tone unit 506)

> **Interviewer A** Well, what A:[your your best bet] is to go to B:[the University Library] or write for C:[the English honours syllabus] - read it and study it - do you see? Find out what D:[the course] is and then start reading in E:[the various subjects], um, reading from F:[the recommended texts] that are there printed in G:[the syllabus] and and so prepare yourself for H:[the degree].

Example 2 illustrates, on the other hand, what the Kappa value indicates - that often the categorization differed between the two annotators. The author classified B: as related(R) to an earlier mention of the university. The second annotator classified this as new(N). There was a gap of over 100 tone units between the mention of the university (tone unit 389) and the University Library above (tone unit 506), though arguably the University should be quite activated given the discourse situation. This type of difference in tagging was common: the first annotator often tagged items as related(R) where the second annotator classified them as new(N). In the above example the same thing occurred for C:[the English honours syllabus], where the author related the NP to an earlier introduction of English as a university honours subject, and for E:[the various subjects], which the author related to the course mentioned in D:[the course]. E: was also classified as new(N) by the second annotator but related(R) to D: by the author. Also H: was again tagged as related(R) to the course in D: by the author but marked as new by the second annotator.

D: was classified as coreferentional(=) with an earlier reference to the course at the very beginning of the interview (i.e. tone unit 27). Again, even though the second annotator had access to the entire transcript during the tagging, it seems that she couldn't remember whether the course had been specifically mentioned earlier. F: was classified as related by both annotators, but different anchors where given. The author chose D:[the course] whereas the second annotator chose C:, and both are arguably equally good. A correct understanding of the utterance would seem to require making both these connections. G:[the syllabus] was considered coreferential with C:[the English honours syllabus] by both, though this is quite easy as it is the same head noun.

Example 3 (interview 2)

> **Speaker B:** Do you like A:[Latin]? - six lines - Speaker A I like B:[the precision] of it.

In Example 3, the author coded this as *described(D)*, in that a representation where Latin is marked as having the quality of precision should be possible to relate to the discourse record by examining the local context, and that this relationship is also made explicitly. The groups tagged as described(D) are what Hawkins (1978) called "referent establishing relative clauses" or "associative clauses", and what Prince (1981) called "containing inferrables". The second annotator didn't recognize this category consistently, and a number of examples tagged as described(D) by the author were tagged as new(N) by the second annotator.

Example 4 (interview 1)

> **Interviewer a:** Yes, um, what about A:[your earlier reading] what's B: [the earliest author] that you've read at all?

This example seems to fall in between what Hawkin's called "Unexplanatory Modifier Use" and "referent establishing relative clauses". It was classified as related(R) by both annotators, but related to A: by the author and to an earlier mention of "your reading" (not shown) by

the second annotator. Clearly, this is an ambiguity in interpretation because "your earlier reading" in A: can be interpreted either as "reading that you did at an earlier time", in which case it is not as suitable as an anchor, or as "reading from earlier works", which would make it a potential anchor. Using that particular NP, [your earlier reading] to mean this second interpretation is sloppy, but quite typical of spoken language.

6 Discussion

6.1 Comments on low inter-annotator agreement

First, it must be pointed out that the level of annotator agreement was very low, lower than for the material and annotation schemes reported on in (Poesio and Vieira, 1998). Spoken language may be more vague and this may have lead to less clear cut examples of each type (cf. (Eckert and Strube, 2001), who found that 13.% of pronouns studied could be classified as vague. However their agreement scores were much higher for their study of pronoun resolution in spoken dialogue). There seem to be two related reasons for the low inter-annotator agreement.

First, many of the entities introduced by definite descriptions have strong relationships to the other elements in the discourse context (see example 2). Most examples seem to allow several potential interpretations and this is also attested to in (Poesio and Vieira, 1998), who comment that one of their main results was that the disagreement among annotators didn't reflect actual mistakes, but the fact that there was often more than one potential relationship to the context. Also note that Fraurud's (Fraurud, 1990) motivation for only using two categories was the fact that many of the proposed classifications for definite descriptions were ambiguous or overlapping.

The second reason has to do with individual differences in interpretation. Identifying anchors for definite descriptions as new depends on how clearly the interpreter sees a relationship with the context, and seems allow room for a great deal of interpretational freedom. Much more so than resolving coreference relationships, with definite descriptions or pronouns. If we repeated the annotation and tested for stability (as defined in (Carletta et al., 1997) as an individuals consistency of tagging over time), we would probably find that this task not only has a low degree of inter-annotator agreement, but also a low degree of *intra*-annotator agreement.

Annotator perspective seems to play a much greater role here than for resolving pronouns. By perspective I mean whether or not the annotation is attempted from the point of view of the annotator or from the view of a discourse participant. The most obvious influence is that different knowledge sources are used to resolve the same anchor-anaphoric relationships by a non-participant annotator than by a discourse participants, even when the anchor is the same. This means that according to the above categories there would be different classifications depending on the individual's relationship to the information. What is interesting however, is that the end representation in the discourse model will often be the same regardless of the interpretation procedure because of the semantic information coded in the definite description. For example, in one of the interviews, the interviewer says "your little girl". Clearly, as the speaker is introducing the individual he already knows about her, and at least one other discourse participant knows her (the mother!) but the individual is new to the annotator. But for either case, a new discourse referent must be added and integrated into the discourse.

So because the annotators perspective is taken (and this is hard to avoid in corpus work)

the number of definite descriptions that will be interpreted as new/accommodated is probably higher than for the actual discourse participants. So frequency data doesn't really reflect the participants burden of interpretation.

What does low inter-annotator agreement tell us about the nature of the things we are studying? There is an assumption that a high degree of inter-annotator agreement on a classification task tells us that the categories we have identified an accurate description of the most relevant characteristics of what we are classifying. (This is also why, for example, Poesio and Vieira (1998) do two experiments, testing two different classifications). But is definite description usage something that can be classified discretely? The great number of different theoretical classifications proposed (see (Vieira, 1998)) as well as the low degree of annotator agreement in this study and other (e.g. (Poesio and Vieira, 1998) seem to suggest that this work may be on the wrong track. Except for examples of coreference (binding) definite description are more or less related to the discourse, and often related in several different ways and it would seem that they are perhaps better described in a different way.

I can see two potential solutions to the categorization problem. The first possibility is to code more than one relationship, that is, allow all plausible relationships to be part of a set of anchors. Annotator choice of anchor can then be evaluated as to whether or not the anchor was part of the set. Or, having a great number of coders would also allow evaluations of anchor identification based how common that particular answer was.

The second possible solution would be to try to define more clearly what the representation will be used for and then evaluate the categories on how well it functions in the application. This could also help with the problem of granularity, i.e. determining how much information or how detailed your representation should be.

6.2 Natural language examples and the anaphoric theory

The aim of the empirical study was to get more information about how frequent definite description usage that could be considered to lie between binding and accommodation occurred, and to see how these examples looked in spoken language data. Very few examples were known(K) and examples that required some sort of intensive reasoning from the discourse record were difficult to pinpoint. The rest of the discussion will focus on the group called bridging anaphora.

Did examples in the data that were tagged as related(R), which corresponds to bridging examples, occur in the spoken data? The answer is not really. Only 17 examples were classified as related(R) by both annotators(cf. (Poesio and Vieira, 1998), experiment 2, of 464 NPs the three annotators agreed only 7 times that examples were bridging). Of these examples 9 identified the same linguistic strings in the discourse as the source of the anchor, but the relationship was often very difficult. Examine example 5 below which is one of the simpler examples:

Example 5 (interview 3)

> **Interviewer B** How many times does the ghost appear in Hamlet?
> **Speaker A** I played A: [the ghost], um, (laughs) I should know that. (several lines)
> **Interviewer B** Why do you think he why why does he appear in the closet scene?
> **Speaker A** Now now this is something I couldn't understand but I had to play this in haem-, B:[this part].

Both annotators categorized B: as related to A:. First, from the example (which isn't even shown in its entirety) we can see that there could of have been several other relationships

that might have been considered more central, but let us treat the example as if the anchor identified by the two annotator *is* the best possible anchor. How would this example then be resolved using the anaphoric theory of presupposition?

First, if we ignore Interviewer B's second statement in our representation, we get something like (overly simplified so that only central elements are given):

[x,y : the ghost(x), played(A,x), this-part(y), played(A,y) part-of(y,x)]

This is not your standard bridging example, based on some sort of lexical relationship. The resolution seems simple but is quite difficult to describe (see also example 1, NP A:[your place]). We must have information that playing a part means playing the part of a character, and then make the connection that Speaker A (who throughout the interview mentions how he has played other characters) means the character of the ghost. In the standard anaphoric theory of presupposition, "the part" can first be accommodated and then this accommodated information must be related to the discourse record in someway to make this connection explicit (because clearly the part is meant to refer to the part of the ghost).

The other way to resolve (not following the theory in (van der Sandt, 1992) and (Geurts, 1999)) is to, on hearing "the part", and failing to find a discourse referent to bind to, to try to infer from something else in the discourse that there must be a related "part". The obvious candidate is the ghost, (which seems to have the similar characteristic of being "played") and infer a "part" that is of the ghost. This part gets added by some sort of inference process and then the discourse referent for "part" is bound to this inferred part. In this case accommodation is avoided at the cost of backtracking and inferring new discourse referents by some non-presupposition inference process. This is basically what Piwek and Krahmer (2000) suggest.

In the first solution all the in-between examples will get classified as accommodation and some sort of integration of accommodated information to the discourse record must be done. And in the second solution all the these examples will be bound after some sort of inference process that seems to lie outside presupposition.

I prefer the first solution as it keeps to the original theory of van der Sandt (1992). But using the first solution means we need to develop strategies to augment the discourse representation with information about how the accommodated NP is related to the discourse, and as the example above shows this problems is quite big.

The main goal should be to try to obtain a coherent representation. Information about discourse referents, objects and individuals in the discourse relate to each other. Many definite descriptions were of the types already identified by Prince (1981) and Hawkins (1978) as quite explicitly related to the discourse record by virtue of their introduction. These examples follow a very predictable integration process, and these are like those identified earlier in the literature. Beyond proposals for identifying and integrating clearly lexically based bridging inferences, there doesn't seem to be much concrete work on how detailed a representation should and could be. This is an area where more work is needed.

Note also that processing these in-between examples as accommodation means that accommodation is the principle interpretation method for definite descriptions. In this case, it is difficult to defend the characterization of accommodation as a repair strategy.

6.3 How new is new(N)?

A large number of definite descriptions with the above solution will be analyzed by accommodation. Accommodated information is usually considered to be new, though actually these in-between examples are not truly just new. Usually, accommodation is associated with discourse-new *presupposed* information so it is not quite right to think of these examples as new. How new are those examples identified as new(N)? Look at the example below.
Example 6

> **Speaker A:** I sort of read A:[the play] before I go to see it usually. I like to study it a week before, and then take B:[my impressions] to the play...

Here, "my impressions" was tagged by both annotators as *new(N)*, though on reflection it is clear that it can be further modified as "the impressions I get from reading a play before I go to see it". This is similar for most of the examples tagged as new(N). And indeed, I attempted to go through all examples tagged as new(N) by both annotators and I have come to the conclusion that there are *no* examples that cannot be interpreted either as 1) related to the earlier discourse, but in a way that is difficult to describe and predict, as in the example above, or 2) seems to be something that is well known to the discourse participants, and here many of the examples of NPs modified by possessives can be found, such as references to "your application", "your little girl". Also, "your English" is clearly related to the context of the discourse, and things like "my father" are quite predictable. Possessives seem to be more integrated in the discourse record by their very nature than definite noun phrases because they not only introduce a new object/individual, but relate it to someone else in the discourse context .

It seems that, at least in the dialogues studied here, it is hard to find cases of definite description use which is not related to the discourse record, and therefore truly new. This shouldn't be a surprising result. Discourse is expected to be coherent, therefore related, and we usually understand presupposition as a way to present information that should be backgrounded, and already partly known. Assertion is the normal means of presenting given information. This result also may at first seem to contradict the findings of Fraurud (1990) and Poesio and Vieira (1998), but actually it is just a difference in how you define things (just as the number of accommodated or bound presuppositions will vary depending on the terms in which you choose to discuss their resolution). It may also be that spoken dialogue in general has a lower tendency to use definite descriptions to introduce "really new" items than text.

7 Conclusions

Naturally produced examples of definite descriptions are extremely complicated, and how they can be resolved and represented is not clear in part because there is no consensus on how they should be analyzed in the first place.

Also, examples 'in-between' binding and accommodation can be analyzed either way, depending on how you prefer to describe the resolution process. Nothing was found in the natural examples that would argue for one method over another. Also, whether definite descriptions should be considered to introduce new information, or given information is also a matter of how new and given are operationally defined. Almost all examples in the corpus

that were categorized as new(N) by both annotators could be argued to be related to the discourse record by some type of "bridging", though here again it depends on your definition.

Future work should try to make better suggestions on how to integrate accommodated information better into the discourse record.

Acknowledgments

Thank you to Elizabeth Spenader for being the second annotator! A big thank you to Natalia Nygren Modjeska and Sofia Gustafson-Čapková for commenting on an earlier version of this paper. Also, thanks to Henk Zeevat and Emiel Krahmer for their help.

Bibliography

Bonzon, P., Cavalcanti, M., and Nossum, R., editors (2000). *Formal Aspects of Context*, volume 20 of *Applied Logic Series*. Kluwer AP.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1).

Clark, H. (1975). Bridging. in: (Schank and Nash-Webber, 1975).

Cole, P., editor (1981). *Radical Pragmatics*. Academic Press.

Eckert, M. and Strube, M. (2001). Dialogue acts, synchronising units and anaphora resolution. *Journal of Semantics*. to appear.

Fraurud, K. (1990). Definiteness and the processing of noun phrases in natural discourse. *Journal of Semantics*, 7:395–433.

Geurts, B. (1999). *Presuppositions and Pronouns*, volume 3 of *Current Research in the Semantics/Pragmatics Interface*. Elsevier.

Hawkins, J. (1978). *Definiteness and indefiniteness*. Croom Helm Ltd.

Löbner, S. (1985). Definites. *Journal of Semantics*, 4:279–326.

Poesio, M. and Vieira, R. (1998). Corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216.

Piwek, P. and Krahmer, E. (2000). Presuppositions in context: Constructing bridges. in: (Bonzon et al., 2000).

Poesio, M., Vieira, R., and Teufel, S. (1997). Resolving bridging descriptions in unrestricted text. In *Proc. ACL-97 Workshop on Operational Factors in Practical, Robust, Anaphora Resolution For Unrestricted Texts*, pages 1–6, Madrid. ACL.

Prince (1981). Towards a taxonomy of given-new information. in: (Cole, 1981).

Schank, R. and Nash-Webber, B., editors (1975). *Theoretical Issues in Natural Language Processing*. Cambridge, MIT.

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–77.

Vieira, R. (1998). A review of the linguistic literature on definite descriptions. *Acta Semiotica et Lingvistica*, 7:219–58.

Paper 15

The distinction between generalised and particularised implicatures and linguistic politeness

Marina Terkourafi, University of Cambridge
mt217@cam.ac.uk
http://www.cus.cam.ac.uk/~mt217/

Abstract

The view that politeness is communicated by means of implicatures is widely held in the relevant pragmatic literature (cf. Lakoff 1973; Searle 1996 [1975]; Leech 1983: 169ff.; Brown & Levinson 1987: 37ff.). However, a different view, namely that politeness is anticipated rather than communicated, has also been defended (cf. Fraser 1990, 1999; Escandell-Vidal 1998; Jary 1998). In this paper, I shall investigate when and how implicatures of politeness might plausibly arise. Politeness will now be defined as a perlocutionary effect consisting in the addressee holding the belief that the speaker is polite. It will emerge that, when relying on the recognition of the speaker's intention, politeness is communicated by means of particularised implicatures. However, politeness may also be anticipated. It is then achieved independently of the recognition of the speaker's intention, in virtue of the speaker's utterance containing an expression which both speaker and addressee take to be conventionalised for some use relative to the context in which it is uttered. To the extent that politeness passes unnoticed (cf. Kasper 1990: 193), this is the primary way in which it is achieved.

1 Introduction

Building on an idea by Fraser (1999), politeness is defined as a perlocutionary effect consisting in the addressee holding the belief that the speaker is polite. Perlocutionary effects are, however, open-ended and indefinite: they depend upon the set of assumptions held by the addressee alone, that is, no constraint of mutual availability of assumptions applies to them, as it does to conversational implicatures. This raises the question: if achieving perlocutionary effects is very much what we may call a 'hit-and-miss' operation, how is it that, usually, when a speaker thinks s/he is being polite, the addressee thinks so too? I shall approach this question by first drawing a distinction between 'ambivalent' and 'indirect' utterances, which interacts with the degree to which an expression is conventionalised for some use in determining how politeness as a perlocutionary effect is achieved. Conventionalisation is to be understood here in a broad sense: it refers to an experientially established statistical likelihood that a

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

mt217@cam.ac.uk
http://www.cus.cam.ac.uk/~mt217/.

particular expression will be used in a particular context; it is thus a matter of degree, and subject to variation cross-linguistically as well as intra-linguistically. I shall then explore how an implicature to the effect that the speaker is being polite might be derived, and show that deriving such an implicature is only plausible in case the expression used by the speaker is not conventionalised for some use. Such implicatures being particularised, they will be drawn in different ways depending on whether the utterance is indirect or ambivalent in context. I shall then examine the role of a particular form of words (Grice's 'what is said') in achieving politeness, and end by proposing, on the basis of these findings, a slightly revised version of the three-tiered picture of utterance interpretation advocated within Neo-Gricean approaches.

2 Ambivalent vs. indirect utterances and conventionalisation of form

Building on an idea by Thomas (2000), we may distinguish (1)-(4) below in two general classes:

(1) Open some windows.
(2) I was wondering if it would be OK to open some windows.
(3) I was asking myself if it would be OK to open some windows.
(4) It's hot in here.

(1), (2), and (3) are 'ambivalent': they make clear how the speaker's utterance may be complied with — namely, by opening some windows, or explaining why this is not possible — although their illocutionary force may remain unclear. (4), on the other hand, to the extent that it is not conventionalised for some use, is 'indirect': it may be a request, a criticism, or a mere statement of fact; how it may be complied with will be different in each of these cases. The distinction between ambivalent and indirect utterances parallels Brown and Levinson's (1987) distinction between on-record and off-record strategies. The reason for adopting a different terminology here is that their reference to strategies can only too easily make one oblivious to the fact that "many of the classic off-record strategies ... are very often actually on record when used, because the clues to their interpretation ... add up to only one really viable interpretation in the context" (1987: 212). Consequently, the distinction between ambivalent and indirect applies exclusively to utterances in context. To illustrate this point, consider Brown's (1995) discussion of irony in Tzeltal. According to Brown and Levinson's classification, irony is an off-record strategy (cf. 1987: 221-2). Discussing its extensive use between Tzeltal women, however, Brown shows how, in this case, irony has been conventionalised as a positive politeness strategy. Brown claims that conventionalisation cannot account for all the instances of Tzeltal irony that she discusses, and builds on this claim to argue that recognition of the speaker's intention is always necessary for the attribution of politeness (cf. 1995: 154-5). However, it seems to me that this conclusion does not necessarily follow from her examples. In one of these, irony is used by both plaintiff and defendant in court (cf. 1995: 162-4): Brown calls this 'angry irony'. However, there is an important difference between this and the cases of conventionalised irony which she discusses as instances of positive politeness: the setting of the exchange in court, which sets this example apart from the other (informal) exchanges. Conventionalisation is essential to achieving politeness, but it cannot be understood in the narrow sense, whereby some expressions in a particular language are conventionalised in comparison with some others, that is, it is not a property of linguistic expressions, "inhering in particular linguistic forms" (cf. Brown 1995: 154). Rather, an

expression is conventionalised for some use only in relation to some context. Uttered in a different context, the same expression will no longer be conventionalised, and the recognition of the speaker's intention will then be necessary for inferring its meaning. Indeed, it will be the discrepancy between the context in relation to which the expression is conventionalised, and the one in which it is actually uttered, that will serve as a trigger for the inferential process. In Brown's court-case example, the setting of the exchange in court may well trigger the face-threatening interpretation of irony. Arguably, then, irony in Tzeltal is only conventionalised as a positive politeness strategy when used between women in informal settings. In this case, utterances which, based on their propositional content, are indirect, turn out to be ambivalent when used in a context in relation to which they are conventionalised.

Based on the preceding discussion, conventionalisation will be defined as a relationship holding between utterances and contexts, which is a correlate of the (statistical) frequency with which an expression is used in one's experience in a particular context. That is, it is a matter of degree, which may well vary for different speakers, as well as for the same speaker over time. This does not preclude the possibility that a particular expression may well be conventionalised in a particular context for virtually all speakers of a particular language, thereby appearing to be a 'convention' of the language.[1] However, it does bring into prominence the rational bases of the relevant conventions (cf. Lewis 1969). Experimental and observational data (cf. Blum-Kulka 1987; Holtgraves & Yang 1990; Weizman 1992: 125; Turner 1996: 5-6; Manes & Wolfson 1981; Rhodes 1989; Brown 1995), as well as data from language acquisition (cf. Snow et al. 1990) and L2 learning (cf. Phillips 1993) converge on the central role which conventionalised expressions play in achieving politeness. These indications are in harmony with my own findings pertaining to politeness realisations in Cypriot Greek (cf. Terkourafi, forthcoming). While the importance of conventionalised expressions has thus been repeatedly acknowledged, the possibility that this translates into essentially different inferential paths involved in achieving politeness when an expression is conventionalised for some use as opposed to when it is not, though intuitively evident, has not been previously explored.

3 Calculating implicatures of politeness

According to Grice (1989a: 39-40), for a proposition to count as a conversational implicature of an utterance, it must meet the conditions of cancellability, non-detachability, and calculability. Récanati emphasises the 'empirical' nature of this last condition: "it is the speaker and the hearer who must be able to work out the implicatures." (1998: 523). Calculability may then be understood as a condition of psychological plausibility, predicting a potential end-point for the inferential process. It is my purpose in this section to show that an implicature to the effect that the speaker is being polite is drawn at different points in the inferential process depending on whether the utterance is indirect or ambivalent in context, and is always particularised. However, such an implicature will not be drawn when an utterance is conventionalised for some use, unless there occurs to the hearer a reason to the contrary.

In line with the premises of speakers' mutually assuming each other's rationality and face wants, the indirect utterance in (4)

(4) It's hot in here.

[1]This will depend on the extent to which their experience is similar. Two factors will be essential for this: their objective conditions of existence, and communication; both are captured under Bourdieu's (1990) notion of the 'habitus'.

addressed by a guest to the host during a dinner-party at the latter's house, can give rise to the implicature in (5)

(5) The speaker wants me to somehow make it 'not-hot' for him/her.

as follows:[2] in uttering (4), the speaker is expressing the attitude of belief toward the proposition 'it's hot in here' (step I: recovery of propositional content). Interlocutors' propositional attitudes being private, the speaker would not go to the trouble of revealing his/her attitude toward any particular proposition of his/her own will, if s/he did not have some other intention which s/he intends me to recognise (step II: assumption about speaker's rationality). People do not generally like being hot (step III: background knowledge). The speaker knows that I, as the host, can somehow make it 'not-hot' for him/her (step IV: background knowledge). So, in addressing (4) to me, the host, as opposed to another guest, the speaker could be expressing his/her desire that I somehow make it 'not-hot' for him/her (step V: inference from steps I, III and IV). Yet, the speaker has not done this, but s/he has merely asserted that 'it's hot in here' (step VI: propositional content). If the speaker has not explicitly expressed his/her desire that I somehow make it 'not-hot' for him/her, it is probably because s/he is trying to give me options/ avoiding to impose on my negative face, i.e. s/he is being polite (step VII: inference from assumption about interlocutors' mutual face wants). Therefore, the speaker's intention in uttering (5) is probably to request that I somehow make it 'not-hot' for him/her (step VIII: inference from steps V, VI and VII). In this first scenario, the inference about the speaker's polite intention (step VII) precedes the recognition of the speaker's intention (step VIII) and actually plays a direct part in it. And since the recognition of the speaker's intention arguably marks the end-point of the inferential process, the implicature to the effect that the speaker is being polite will have been drawn as part of that process, i.e. at no extra cost.

The four scenarios outlined above illustrate how, in the case of indirect utterances, the implicature that the speaker is being polite is cancellable: its truth conditions are independent from those of (5) the request reading of (4) since either may be true while the other is simultaneously false (scenarios 2 and 4), or both may be true (scenario 1) or false (scenario 3). Moreover, it is non-detachable: (4) could read 'it's [rather] hot in [this room]', or 'il fait chaud dedans', and still give rise to the implicature that the speaker is being polite, along the lines of scenarios 1 and 4. Finally, it is calculable, since its derivation — subject to cultural and situational constraints — is part of, rather than an extension of, the inferential process leading to the recognition of the speaker's intention. The hearer may then plausibly draw an implicature to the effect that the speaker is being polite if the speaker's utterance in context is indirect. However, whether s/he will actually draw such an implicature — in other words, which of the four scenarios s/he will follow — is wholly dependent on the particulars of the situation. Intonational and kinesic clues, the addressee's prior impression of, and/or familiarity with, the speaker, as well as the addressee's mood of the moment will all play a part in deriving a particularised implicature as to the speaker's polite intention, which thus has to be recognised before politeness as a perlocutionary effect can be achieved.

[2]This and following accounts parallel that of Searle (1996 [1975]). However, whereas in Searle's account, the inferential process contains explicit reference to elements of Speech Act theory and the CP, whose psychological reality remains doubtful (on SAT see Sperber & Wilson 1995: 244-5; Geis 1995: 31-2; on revisions of the CP see Terkourafi, forthcoming), the accounts in this section draw on the premises of speakers' mutually assuming each other's rationality (from which their mutual face wants can be seen to emanate), background knowledge, and inferring goals from plans, whose psychological plausibility has been tested in a series of experiments (cf. Lindsay et al. 1993; Lindsay & Gorayska 1994).

A different inferential path will be followed if the speaker's utterance in context is ambivalent. Addressed by a guest to the host during a dinner party at the latter's house, (3) can give rise to the implicature in (6)

(3) I was asking myself if it would be OK to open some windows.

(6) The speaker wants some windows opened.

as follows: in uttering (3) the speaker is informing me of his/her attitude toward the proposition 'it is OK to open some windows' (step I: recovery of propositional content). Interlocutors' propositional attitudes being private, the speaker would not go to the trouble of informing me of his/her attitude toward any particular proposition out of his/her own will if s/he did not have some other intention which s/he intends me to recognise (step II: assumption about speaker's rationality). Despite having reason to believe that 'it is OK to open some windows' is true in this context (step III: background knowledge), the speaker has not asserted 'it is OK to open some windows' (step IV: propositional content). This being my house, the speaker has reason to believe not only that I am in a privileged state of knowledge over other participants as to whether 'it is OK to open some windows' is true, but also that I am in a privileged position to make this proposition true (step V: background knowledge). So, by choosing to inform me (rather than any of the others present) that s/he is entertaining 'it is OK to open some windows' as a possibility, the speaker is actually asking me to confirm whether this proposition is true (step VI: inference from steps I, IV and V). Normally if people express an interest in finding out whether some proposition is true or false, this is because this information somehow helps them further their goals (step VII: as in step II). Whether 'it is OK to open some windows' is true or false can be used in this situation to further the goal of opening some windows (step VIII: associating plan elements with desired goals). So the speaker's intention in uttering (3) must be to request that some windows be opened (step IX: inference about speaker's intention). Recognising the speaker's intention in this way amounts to establishing how his/her utterance may be complied with — in this case, by opening some windows, or explaining why this is not possible.

However, in order to derive from (3) the implicature in (7)

(7) [In uttering (3)] the speaker is being polite.

the hearer must take some further steps, as follows: if the speaker's intention is to request to open some windows, s/he could have expressed this desire (step X: background knowledge/theory of speech acts).[3] Instead, the speaker has chosen to inform me that s/he is entertaining 'it is OK to open some windows' as a possibility (step XI: propositional content). If the speaker has expended more effort than minimally required for me to recognise this intention, s/he must have some other intention that s/he intends me to recognise at the same time (step XII: assumption about speaker's rationality). Requesting for something to be done by means of expressing one's corresponding desire is normally[4] avoided, given that it encroaches on the other's negative face (step XIII: inference from step X and interlocutors'

[3]This step implies the conscious availability of a ranking of possible realisations of requests according to their degree of (in)directness, such that taking this step is made contingent on the prior understanding of (3) as a request. This will only be possible to the extent that the illocutionary force of the speaker's utterance may be jointly determined by the propositional context of his/her utterance and features of the situational context. Given, however, that this may not always possible (see fn.2 above), the possibility of itself of calculating (7) on the basis of (3) must be left open.

[4]This qualification is important: this assumption can be safely drawn only when negative face wants are given priority over positive face wants. Where positive face wants are prioritised, on the other hand, expressing one's desire for some thing/action may well constitute one's own positive face, by making one appear confident about one's desires.

mutually assuming each other's face wants). So, the speaker has avoided to express his/her desire to open some windows in order to avoid imposing on my negative face, in other words, in order to be polite (step XIV: inference from steps XI, XII and XIII).

(7) is thus cancellable, non-detachable, and potentially calculable from an utterance of (4). (7) is cancellable because (3) may be understood as conveying the request in (6), but will only give rise to the implicature in (7) if negative face wants are given priority over positive face wants in context — something which depends on cultural constraints, as well as situational ones (cf. Brown & Levinson 1987: 63-4; Terkourafi, forthcoming). (7) is also non-detachable: (2) above could still give rise to it. However, the existence of (2) may well hamper the calculability of (7) from an utterance of (3). Deriving (7) depends on the prior derivation of (6): only if (3) counts as a request can its propositional content be compared to alternative ways of making requests, and the speaker's polite intention recognised. However, it seems doubtful that the hearer would actually expend the extra effort needed to derive (7) from an utterance of (3) (steps X to XIV above), since s/he has established how (3) may be complied with once (6) has been derived (i.e. after step IX). For this, the hearer would have to be provided with a reason. The phrasing of (3) may then provide such a reason. (3) sounds rather odd in English: (2) above would be a much more natural way of expressing (roughly) the same propositional content. Given that (2) is conventionalised in English for conveying (6), the speaker's not using (2) on this occasion could serve as a trigger for the addressee to expend the extra effort needed to derive (7) based on an utterance of (3). However, the fact that (2) is thus conventionalised could also hamper the derivation of (7) from (3) via an M-type implicature (cf. Levinson 1995: 97; 2000: 135ff.): the addressee could be deterred from understanding (3) as a request in the first place, and instead take it to be an assertion of what is on the speaker's mind. The calculability of (7) depends on whether the addressee thinks that (2) is equally held by the speaker to be a conventionalised expression for conveying (6):[5] If the addressee takes it that this is so, s/he is likely to take (3) as an assertion; if, however, s/he has reason to think that the speaker does not hold (2) to be thus conventionalised — say, because s/he takes the speaker to be an L2 learner based on the latter's accent — s/he may still understand (3) as a request. (7) is, then, a potential implicature of (3), which will require a reference to the particulars of the situation as outlined above in order to be derived, i.e. it will be particularised.

So far, we have established that politeness may be communicated by means of particularised implicatures if the speaker's utterance in context is indirect, or ambivalent but not conventionalised for some use. The question which arises next is whether this is also the case when the speaker's utterance in context is conventionalised for some use — which, as pointed out in 2 above, renders it ambivalent *ipso facto*. That is, how does the inferential process proceed when a guest addresses (2)

(2) I was wondering if it is OK to open some windows.

to the host during a party at the latter's house? The fact that (2) is conventionalised in English for conveying the request in (6) above will, in my view, serve to short-circuit the inferential process from (2) to (6) (steps I to IX above). However, this time, the further proposition that the speaker is being polite would not qualify as an implicature of (2). Such a proposition would not be cancellable: the addressee's thinking that (2) is a conventionalised expression under the circumstances for conveying (6) amounts to him/her holding the belief

[5] This is a concomitant of calculability as a property of implicatures not *relative to propositions*, but *attributed to them by implicating agents* (cf. Récanati 1998: 523).

that, roughly, 'uttering (something like) (2) when one is a guest at another's house and one wishes to perform (something like) the request in (6) in English is polite'. That is, the evaluative judgement is part of this belief, rather than an *a posteriori* evaluation of the belief. To put this generally, when one learns (through experience or through explicit instruction) that this is the way to do some thing — for example, the way to eat is by holding a knife in one's right hand and a fork in one's left — what one is effectively learning is an *evaluation* by a set of agents of a particular way of doing a particular thing. This means that, if (2) counts as a request in the context above, it does so by counting as a polite request in this context. In addition to not being cancellable, the proposition that the speaker is being polite would be detachable: were (3) above to be used instead of (2), this proposition would not necessarily be derived (see above on the difficulties of deriving this on the basis of (3)). Finally, coming as it does after the derivation (6) — based on which (2) may be complied with — this proposition would, arguably, no longer be calculable in this context. It is arguable whether the speaker can plausibly take the addressee to be *able* to calculate a proposition which s/he already takes the addressee to *hold* as true. The addressee, on the other hand, will have no reason this time to expend the extra effort (infer steps X to XIV above), since, based on his/her prior experience in similar communicative situations, s/he will already hold a belief that, roughly, 'uttering (something like) (2) in a situation when one is a guest at another's house and one wishes to perform (something like) the request in (6) in English is polite'. Based on this belief, upon the speaker's uttering (2) in this situation, the addressee will form the further belief that the speaker is polite. Since politeness as a perlocutionary effect consists in the addressee holding this further belief, politeness will have now been achieved without the recognition of whatever intention the speaker may have had. That is, it will have now been not implicated, but *anticipated*: requiring no inferencing to be achieved, it passes unnoticed.

Conventionalisation was defined above as a relationship between expressions and contexts (rather than languages). (2) is thus conventionalised for conveying (6) in many contexts in English. This may make the request reading of (2) appear to be a convention of the English language. Indeed, the fact that (7) would now be non-cancellable, detachable, and non-calculable — based on which I argued above that this does not qualify as a conversational implicature of (2) — may make one think this is rather a conventional implicature of (2), i.e. part of its linguistic (conventional) meaning. However, this is not so, because this proposition can only be arrived at on the basis of the request reading of (2) as in (6), which can be contextually cancelled, e.g. if there are no windows which may plausibly be opened. The proposition that the speaker is being polite is not a conventional implicature of (2) for the further reason that on occasion (2) can be understood as conveying (6), without necessarily giving rise to this further proposition. For example, uttered by a husband addressing his wife in their house, (2) would sound odd. Two things may plausibly happen next. First, in virtue of (2) being conventionalised for conveying (6) *in a number of other contexts*, there would be a strong tendency for the inferential process from (2) to (6) to be short-circuited, that is, for the wife to recover (6). Second, (2) not being conventionalised *in this context*, the wife would now be provided with a reason to continue the inferential process beyond (6). For this, she would appeal to further features of the context: is there a reason why her husband might think windows should not be opened, even though he would so desire, for example, is one of those present ill, or is a storm expected? Are further parties present, in whose presence her husband may wish to sound more formal than usual? Has there occurred a rift between spouses prior to this utterance, such that her husband may wish to communicate this by means of his utterance? Based on what she takes it to be the case, the wife may still

attribute the intention to be polite to her husband; or she may not. However, politeness in this case, if achieved, will have been communicated rather than anticipated. The proposition that the speaker is being polite will now be cancellable: it is possible that the wife will not take it that her husband is being polite, despite understanding (2) as a request. It will also be non-detachable: had her husband said (3) above, she could still take it that he is being polite. The proposition that the speaker is being polite is not already attached to either of these two expressions in this context, with the result that either can give rise to it. Finally, this proposition will now be calculable: the husband's taking into account his wife's ability to calculate this is now plausible, since he does not take it that she already holds this proposition to be true. And the wife will now have a reason to continue the inferential process beyond (6), since the context is not one in which (2) is conventionalised, and she can therefore take it that her husband had a reason for using (2) in this context. Being cancellable, non-detachable, and calculable in this last context, (7) qualifies as an implicature of (2) in this context. In the absence of an *a priori* belief that using (2) in this context is polite, politeness as a perlocutionary effect cannot be achieved, until this proposition is arrived at by taking into account the particulars of the situation, that is, unless it is inferred as a particularised implicature of (2) in this context.

4 The argument from 'what is said'

In this section, I present evidence in support of the existence of a level of 'what is said' akin to Grice's understanding of this level as "closely related to the conventional meaning of the words (the sentence) [the speaker] has uttered" (1989: 25). This is the level at which conventionalised expressions remain distinct from each other, and as such it plays a direct role in the acquisition of, and memory for, polite forms. However, this level is not, but can be made, consciously available. The fact that the determination of Gricean 'what is said' does not always yield a truth-evaluable proposition, even after reference and disambiguation have taken place, is now widely accepted in the literature. The level at which truth conditions are assigned has correspondingly shifted from the level of Gricean 'what is said' to the level of the proposition expressed, which is also taken as the starting point for the determination of any (particularised) conversational implicatures. As a result, the motivation for distinguishing a separate (Gricean) level of 'what is said' has been brought into question. While Bach (1994: 273) defends a strict notion of 'what is said' as the (not necessarily consciously available) level of the literal meaning of the utterance, which in turn provides the grounds on which the impliciture is worked out, within the framework of RT this level is abandoned in favour of pragmatically enriched explicatures, which are derived as developments of the logical form of the sentence (cf. Carston 1999: 109).

Three types of evidence drawn from instances of using conventionalised expressions support the need for an independent level of 'what is said'. First, consider cases where the addressee's reply addresses (solely) the conventional meaning of the words uttered by the speaker, rather than the pragmatically enriched inference (when there is no indication preventing such an inference from being drawn), as in (8):

(8a) Can you open some windows?

(8b) Yes, I can. (without proceeding to do so)

We have all been confronted with similar replies on some occasion; and, while we may be upset, or amused by them, we cannot accuse the addressee of providing an irrelevant reply, merely an

unhelpful one. Nevertheless, this was clearly not the reply we had expected: our being amused or upset acts as an indication of this. Such examples suggest that a strict notion of 'what is said' may still be theoretically useful, in accounting, not for how we *normally* understand utterances, but for how we are *able* of understanding them: such a level is recovered (by back-tracking our steps, as it were) when the need arises (i.e. when our expectations were frustrated).

The second type of evidence comes from the acquisition of polite forms (cf. Snow et al. 1990). When directly teaching the child what forms to use in various situations, the emphasis is on the exact forms to be used, rather than on the pragmatically enriched inferences they normally give rise to. Consider (9)-(12) below:

(9) I was wondering if it would be OK to open some windows.

(10) Would it be OK to open some windows?

(11) It's OK to open some windows, isn't it?

(12) I don't suppose it would be OK to open some windows.

In the absence of any reason why the addressee should take into account the particulars of the situation, these may be reported to a third party by saying:

(13) S/he asked (me) if it's OK to open some windows.

(14) S/he asked me to open some windows.

Furthermore, any of (9)-(12) can be replied to in the positive by saying something like "Sure" or "Certainly", or in the negative by saying something like "No, I'm afraid not/that's not possible (because ...)". In either case, the addressee would not be responding to the conventional meaning of the words uttered, but to their reading as requests for some to be windows opened. In addition to their reading as requests, conventionalised expressions as in (9)-(12) may give rise to particularised implicatures in context (see the discussion of (2) above exchanged between spouses at home). The fact that only (13) or (14) — the request reading of (9)-(12) rather than the conventional meaning of the words uttered — can plausibly be reported or replied to strongly suggests that (13) and/or (14) are a necessary step toward guaranteeing the optimal relevance of (9)-(12) above in context. At the same time, though, (9)-(12) remain distinct from one another at a level prior to pragmatic enrichment. It is on the basis of a distinction at this level that use of an expression which is conventionalised relative to a number of contexts but not relative to the one in which it is uttered may give rise to any number of particularised implicatures, and that "direct teaching of the child about what forms to use in various situations" (Snow et al. 1990: 303) is at all possible.

The final piece of evidence comes from naturalistic experiments which showed "significant memory for the wording of polite remarks" (Holtgraves 1997: 106). Holtgraves (1997) found that 'what is said' — the conventional meanings of the words uttered — was often accurately recalled by recipients of conventionalised expressions, more so than when a hint was used, when they were asked to report a prior speaker's utterance without previous warning. The evidence, then, strongly suggests that a level of 'what is said' which is prior to pragmatic enrichment has a place in a theory of linguistic communication: the information represented at this level can be made consciously available, although it normally (so long as our expectations are met) is not. This level cannot be reduced to the level of logical form, identified with the output of grammar: utterances of sentences to which the same logical form may be applicable can remain distinct at the level of 'what is said', though not necessarily at the level of what is communicated:

(15) Would it be OK to open some windows?

(16) Would it be bad to open some windows?

A strict notion of 'what is said' (cf. Bach 1994: 277ff., 1999: 77-79) does, then, play a role in the interpretation of utterances which is not already played by the independently motivated level of logical form, and as such it must be retained.

5 A third level of meaning

We are now presented with a three-tiered picture, composed of the following levels: (i) a level of 'what is said', which is normally not, but can be made, consciously available — (9)-(12) above would remain distinct at this level; (ii) a level of pragmatically enriched/ standardised inferences, at which truth-conditions are assigned — this would capture the reading of (9)-(12) as requests; and (iii) a level of particularised implicatures, which are drawn in virtue of special features of the context — implicatures which are derived drawing on the particulars of the situation from utterances which in context are indirect, or ambivalent but not conventionalised for some use, or conventionalised relative to a context other than the one in which they are used, would be assigned to this level. Drawing the picture in this way makes it obvious that context has an input at both levels (ii) and (iii). At the same time, if, following Grice (1989: 31), 'what is said' is to be taken as the starting point for the derivation of conversational implicatures, 'what is said' should be truth-evaluable, i.e. it should be extended to include level (ii).

Two options are open to us at this point: the first is to give up a strict notion of 'what is said' in favour of a pragmatically enriched one. This is the stance taken within Relevance Theory. This option is however hard to reconcile with the evidence presented in 4 above. The second option is to argue that the two types of context that have an input at levels (ii) and (iii) are essentially distinct: recovering levels (ii) and (iii) involves different pragmatic processes, which helps explain the different behaviour of the two types of inferences (see 4 above). However, the information conveyed by these two levels, as opposed to level (i), is not explicit, and this why the three-way distinction outlined above should be retained. This last option may be defended on the basis of Bach's (1999: 72) distinction between narrow and broad context:

> There are two sorts of contextual information, one much more restricted in scope and limited in role than the other. Information that plays the limited role of combining with linguistic information to determine content (in the sense of fixing it) is restricted to a short list of variables, such as the identity of the speaker and the hearer and the time and place of an utterance. Contextual information in the broad sense is anything that the hearer is to take into account to determine (in the sense of ascertain) the speaker's communicative intention.

It follows from the definition of narrow context that contextual information falling under it will often be available prior to the making of any particular utterance. In this way, narrow, but not broad, context can give rise to expectations about a speaker's goals.

Evidence from the use of conventionalised expressions supports the claim that only inferences drawn with reference to narrow context can have an input to determining the proposition expressed (cf. Levinson 1988: 20). If we compare the derivation of (6) from (2) to that of (5) from (4) as outlined in 4 above, it turns out that the contextual information appealed to in each case is different. In the case of (2), this included a reference to the identities of the speaker (a guest), and the addressee (the host), as well as the place (the host's house)

and the time (a party at the host's house) of the utterance. Whatever background knowledge played a part in the derivation of (6) on the basis of (2) was made salient by the availability of this information alone. But in the case of (4), the inference as to the speaker's intention required an additional appeal to cultural assessments of particular behaviours, and could not actually be drawn until further contextual factors, such as intonational and kinesic clues, and idiosyncratic factors had also been taken into account. The comparison of the inferential processes involved in these two examples suggests that, when an utterance contains a form of words which is conventionalised for some use, an appeal to narrow context is sufficient to infer the speaker's intention. However, in the absence of such a form of words, an appeal to context in the broad sense is required before an inference with respect to the speaker's intention can be made.

Inferences resulting from the use of conventionalised expressions were earlier differentiated from particularised implicatures on empirical grounds. The relevant intuitions can now be theoretically accounted for by distinguishing between the two types of context (narrow vs. broad, respectively) with reference to which these two types of inferences are derived. Inferences derived with respect to narrow context fall naturally under Levinson's (1995: 93; original emphasis) third level of meaning, which he sees as

> intermediate between coded meaning and nonce speaker meaning [...] a level of systematic pragmatic inference based not on direct computations about speaker-intentions, but rather on *general expectations about how language is normally used.*

The derivation of inferences from narrow context can be accounted for with reference to two of Levinson's proposed heuristics (cf. 1995: 97; 2000: 112ff.).[6] Levinson sees these heuristics as having default application: they are applied unless the context or the content of the utterance is perceived to contain explicit indications that they should not be.[7] The resulting inferences are generalised, in the sense that they do not constitute additional propositions, the recovery of which follows from, and relies on, the recovery of the proposition expressed by the utterance, as is the case with particularised implicatures. We may illustrate this claim with the earlier example of a guest at someone's house asking for some windows to be opened by saying 'I was wondering if it would be OK to open some windows'. The speaker's utterance in this case meets the addressee's general expectations about how the English language is used given the place (someone else's house) and the time of the utterance (a dinner-party). There is therefore no need for it to be first computed as an assertion (along the lines of Searle 1996 [1975]), and only subsequently as a request, a move which seems intuitively implausible anyway. Rather

[6]Levinson proposes three heuristics (in 1995 he used the terms Q1, Q2, and M; in 2000 these are replaced by the terms Q, I, and M respectively. The latter terms will be adopted in what follows). The first of these (Q) seems to me not to be relevant in accounting for inferences associated with conventionalised expressions, since it is constrained to expression alternates, defined as such on the basis of their semantic content, rather than form (cf. Levinson 1995: 97). The two remaining heuristics, however, motivate specific inferences based on the form of the words used. These are (cf. ibid.): "Q: 'What is simply described is stereotypically and specifically exemplified' (a) unmarked expressions warrant rich interpretations to the stereotype; (b) minimal forms warrant maximal interpretations. Constraint: only of unmarked, minimal expressions. Characteristics: not fundamentally metalinguistic; invokes world-knowledge of stereotypical relations; positive inference to specific subcase. M: 'Marked descriptions warn "marked situation"' Constraint: only of marked, unusual or periphrastic expressions. Characteristics: metalinguistic (marked compared to unmarked) the inference is to the complement of the inference that would have been induced by the unmarked expression."

[7]Franken (1998: 7) also points out that default reasoning involves a narrower conception of context.

than the proposition expressed, what is recovered (by way of a Q-type inference) is simply the fact that the speaker has made a request. Politeness as a perlocutionary effect — the addressee coming to hold the belief that the speaker is polite — is then achieved just in case the addressee holds a belief to the effect that the particular form of words used ('I was wondering if it would be OK to open some windows') is a polite way of requesting for some windows to be opened in English in the particular situation (during a dinner-party at someone else's house). On the other hand, if the speaker's utterance did not meet the addressee's general expectations about how the English language is used in this context, the addressee would draw specific (M-type) inferences pertaining to the speaker's intentions (as, e.g., when (2) is uttered by a husband to his wife at home).

Remarking on the nature of the proposed heuristics, Levinson points out that the metalinguistic character of M inferences consists in their being dependent on beliefs shared between the speaker and the addressee regarding their common code (they are metalinguistic in the sense that they are about language), whereas Q inferences are independent of such beliefs, relying as they do on general world-knowledge (cf. Levinson 1995: 103). Furthermore, Q- and M-type inferences exhibit a systematic complementarity, best captured with reference to the notions of markedness and iconicity: unmarked expressions invite unmarked interpretations, while non-stereotypical expressions invite interpretations to non-stereotypical extensions (cf. 1995: 103-4). Thus, a Q-type inference will not go through if the description used is in any way marked: M inferences defeat inconsistent Q inferences (cf. 2000: 157).

An important caveat is in order here. The notion of markedness appealed to by Levinson (1995: 104; cf. 2000: 137) is, as he puts it, "very broad, covering formal prolixity, infrequent expressions or those of unusual formation." However, when a certain form of words is conventionalised for some use, the relevant notion of markedness needs to be extended: markedness would now appear to be a function not of the formal properties of a certain form of words alone, but of such properties in conjunction with the (narrow) context in which the words are used. Expressions are conventionalised (i.e. they give rise to Q-type inferences) only in relation to some (narrow) context. Blur the boundary between particularised and generalised conversational implicatures as it may do, this revision does not abolish it. The resulting inferences are still generalised because they are independent from context in the broad sense. However, they are not independent from context in the narrow sense, and are therefore universal (cf. Levinson 1995: 110) only inasmuch as the mechanism for their derivation is also universal. This move is in accordance with Hirschberg's findings (1985: 43) regarding scalar implicatures, as well as evidence presented by Barsalou corroborating the context-dependent nature of stereotypicality (cf. 1987: 104ff.). And it is perhaps not wholly unwarranted given Levinson's own observation that Q-type inferences "restrict the interpretation to what *by consensus* constitutes the stereotypical, central extensions" (1995: 103; emphasis added), i.e. they are still somehow tied to (narrow) context.

Bibliography

Bach, K. (1994). Semantic slack: What is said and more. in: Tsohatzidis (1994).

Bach, K. (1999). The semantics/pragmatics distinction. in: Turner (1999b).

Barsalou, L. (1987). The instability of graded structure: implications for the nature of concepts. in: Neisser (1987).

Blum-Kulka, S. (1987). Indirectness and politeness in requests: same or different? *Journal of Pragmatics*, 11:131–46.

Bourdieu, P., editor (1990). *The logic of practice*. Polity Press. (transl.:) R. Nice.

Brown, P. (1995). Politeness strategies and the attribution of intentions: the case of tzeltal irony. in: Goody (1995).

Brown, P. and Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge UP.

Carston, R. (1999). The semantics/pragmatics distinction: a view from relevance theory. in: Turner (1999b).

Coulmas, F., editor (1981). *Conversational routine: explorations in standardized communication situations and prepatterned speech*. Mouton.

Escandell-Vidal, V. (1998). Politeness: a relevant issue for relevance theory. *Rivista Alicantina de Estudios Ingleses*, 11:45–57.

Franken, N. (1998). The status of the principle of relevance in relevance theory. Paper presented at the 6th International Pragmatics Conference, Reims, 19–24 July 1998.

Fraser, B. (1990). Perspectives on politeness. *Journal of Pragmatics*, 14:219–36.

Fraser, B. (1999). Whither politeness? Plenary lecture delivered at the International Symposium on Linguistic Politeness, Bangkok, Thailand, 8 December 1999.

Geis, M. (1995). *Speech acts and conversational interaction*. Cambridge UP.

Goody, E., editor (1995). *Social intelligence and interaction: expressions and implications of the social bias in human intelligence*. Cambridge UP.

Grice, H. (1989). *Studies in the way of words*, chapter Logic and conversation. Harvard UP.

Hirschberg, J. (1985). *A theory of scalar implicature*. PhD thesis, University of Pennsylvania.

Holtgraves, T. (1997). Politeness and memory for the wording of remarks. *Memory & Cognition*, 25:106–16.

Holtgraves, T. and Yang, J.-N. (1990). Politeness as universal: cross-cultural perceptions of request strategies and inferences based on their use. *Journal of Personality and Social Psychology*, 59:719–29.

Jary, M. (1998). Relevance theory and the communication of politeness. *Journal of Pragmatics*, 30:1–19.

Kasher, A., editor (1998). *Pragmatics: some critical concepts*, volume IV. Routledge.

Kasper, G. (1990). Linguistic politeness: current research issues. *Journal of Pragmatics*, 14:193–218.

Kasper, G. and Blum-Kulka, S., editors (1992). *Interlanguage Pragmatics*. Oxford UP.

Lakoff, R. (1973). The logic of politeness; or minding your p's and q's. In *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*, pages 292–305. Chicago Linguistic Society.

Leech, G. (1983). *Principles of Pragmatics*. Longman.

Levinson, S. (1995). Three levels of meaning. in: Palmer (1995).

Levinson, S. (2000). *Presumptive meanings: the theory of generalised conversational implicature*. MIT Press.

Lewis, D. (1969). *Convention*. Cambridge UP.

Lindsay, R. and Gorayska, B. (1994). Towards a general theory of cognition. Unpublished paper.

Lindsay, R., Gorayska, B., and Cox, K. (1993). The psychology of relevance. Unpublished paper.

Manes, J. and Wolfson, N. (1981). The compliment formula. in: Coulmas (1981).

Martinich, A., editor (1996). *The philosophy of language*. Oxford UP.

Neisser, U., editor (1987). *Concepts and conceptual development: ecological and intellectual factors in categorisation*. Cambridge UP.

Palmer, F., editor (1995). *Grammar and meaning: Essays in honour of Sir John Lyons*. Cambridge UP.

Phillips, E. (1993). Polite requests: 2nd-language textbooks and learners of french. *Foreign Language Annals*, 26:372–81.

Récanati, F. (1998). Truth-conditional pragmatics. in: Kasher (1998).

Rhodes, R. (1989). 'we are going to go there': positive politeness in ojibwa. *Multilingua*, 8:249–58.

Searle, J. (1996). Indirect speech acts. in: Martinich (1996).

Snow, C., Perlmann, R., Gleason, J., and Hoosshyar, N. (1990). Developmental perspectives on politeness: sources of children's knowledge. *Journal of Pragmatics*, 14:289–305.

Sperber, D. and Wilson, D. (1995). *Relevance: communication and cognition*. Blackwell.

Terkourafi, M. (2001). *Politeness in Cypriot Greek: a frame-based approach.* PhD thesis, University of Cambridge. (forthcoming).

Thomas, J. (2000). The pragmatic analysis of corpora of naturally-occurring talk. Paper presented at 'Pragmatics: Corpora and Conversation'. Seminar Series, University of London, 8 March 2000.

Tsohatzidis, S., editor (1994). *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives.* Routledge.

Turner, K. (1996). The principal principles of pragmatic inference: politeness. *Language Teaching*, 29:1–13.

Turner, K., editor (1999a). *The semantics/pragmatics interface from different points of view.* Elsevier.

Turner, K., editor (1999b). *The semantics/pragmatics interface from different points of view.* Elsevier.

Weizman, E. (1992). Interlanguage requestive hints. in: Kasper and Blum-Kulka (1992).

Part III

Empirical Findings

Paper 16

What are accented personal pronouns in dialogue signalling?

Sofia Gustafson-Čapková
sofia@ling.su.se
http://www.ling.su.se/staff/sofia

Abstract

Accented pronouns are rare in dialogue, but they do however exist. What is the function of such pronouns? To investigate this question a corpus study was carried out. It was found that the accented pronouns generally signal a contrast but that the strength of the contrast differed. Cases where the contrast relation itself added important information to the discourse were interpreted as a more salient contrast, and cases where the contrast relation did not add central information were interpreted as a weaker contrast. A stronger contrast was also connected to a more explicit contrast element, whereas weaker contrast relations had vaguer contrast elements.

1 Introduction

Pronouns can be described as a type of reduced linguistic sign. They are semantically dependent, in that they need a deictic or anaphoric correlate. They are reduced in that they on average are very short words, and furthermore, they are reduced in that they seldom have phonetic prominence in the form of sentence accent. But sometimes they do carry sentence accent. How common are such accented personal pronouns in dialogue, and what is the accenting signalling?

Sentence accent can be seen as a way to highlight or mark out a certain item, so that listeners will pay attention to it (Grice, 1978). Often this attention-pointing strategy is used when introducing new discourse items, i.e. sentence accent often correlates with New information (Cruttenden, 1986). Personal pronouns, however, usually indicate items that are already in the focus of attention, i.e. Given or activated information (Gundel, 1994). Because of the anaphoric nature of most personal pronouns, accenting them generally implies accenting of already highly activated information.

Accenting such already highly activated information may give rise to a contrastive interpretation (Cruttenden, 1986). This is examplified with sentences as in Example 1 below:

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

Example 1
a) Pelle$_1$ log åt Kalle$_2$, men han$_1$såg inte riktigt glad ut ändå.
Pelle smiled at Kalle, but he didn't really look happy.
b) Pelle$_1$ log åt Kalle$_2$, men HAN$_2$såg inte riktigt glad ut ändå.
Pelle smiled at Kalle, but HE didn't really look happy.

In 1a) the unaccented pronoun han is resolved to Pelle, while in sentence 1b) the accented pronoun HAN is resolved to Kalle. In sentence 1b) there is also a contrast between Pelle and HAN (Kalle). Cases like sentence 1b) are by many researchers also classified as having a crossover interpretation, i.e. in sentence 1b the accented subject pronoun resolves to the referent of the object NP in the preceding clause (e.g. Horne and Filipson (1995); Gardent (2000)).

Gardent (2000) has applied higher order unification to crossover cases, and showed that unification holds between the deaccented parts of the utterances. This shows that the contrast relation in cases like this is dependent both on a contrastive accent, but also some kind of parallelism with the preceding context. However, in Example 1 such an approach would not be applicable. The parallelity of the deaccented parts of the sentence requires a quite creative interpretation.

Gardent (2000) argue that there are similarities between deaccenting and anaphoricity. In cases such as the accented personal pronouns, the deaccented parts of the utterance have some kind of parallel element in the preceding discourse, to which it can link back. This may have either positive or negative polarity. In both cases the result is however a contrast relation between the items given prominence. Prince (1981) however, describes cases where deaccented parts of an utterance not can be considered given or equivalent to something in a former utterance. In the examples discussed by Prince, the accented item is however not given information either, as in the case with pronouns.

Krahmer and Swerts (2001) have treated contrast accent as a presupposition trigger, using the anaphoric presupposition theory of van der Sandt (1992). This stresses the relation to some kind of antecedenthood involved in a contrast relation, and also reduces the importance of the parallellity.

As shown above, the pronouns in the examples are confined to third person pronouns, but in dialogues first and second person pronouns also occur frequently. How these pronouns should be handled with regard to e.g. givenness or activation status of the referents does not seem to be a uncomplicated issue. In analysing discourse, these word tokens are some times left out from the analysis (Eckert and Strube, 2001; Byron and Stent, 1998).

Most examples in the literature are quite simple, but what do they look like in reality? If accenting interacts with deaccenting in the way described about crossover interpretation, what is the nature of those parallel or similar elements of the deaccented parts connected to the contrasted element in the preceding context? How obviously parallel are they? Does the form say anything about the centrality of the contrast in the evolving of the discourse? The study reported in this paper is an attempt to shed light on some of these questions.

As material dialogues from the Kiel corpus (IPDS, a,b,c), were used, all with two dialogue participants. Word tokens like "I" and "you" can then be assumed to have the same degree of ground activation level throughout all dialogues. A stress on such an element would imply a relative marking in one way or another. Is this marking resulting in a perceived contrast, or could the result be something else? Another question addresses the deaccented counterpart of the contrast element. Is this counterpart, or source always obvious, or is it more difficult

to trace? If the sources of deaccenting differs between cases, are they possible to categorise in replicable categories?

2 Method

The method used was a corpus investigation, where personal pronouns with markup for four different degrees of sentence accent (degree 0 - 3) were excerpted. The corpus used was the Kiel corpus, for a description of the corpus, labelling procedure and markup, see (Simpson, 1998) and (Kohler et al., 1995).

Instances of pronouns of all four degrees of sentence accents were analysed, the accented ones with respect to contrast element and to possible source of deaccenting. In categorising the examples the centrality of the contrast relation to the speakers was also taken into consideration.

2.1 Material

As material the part of the Kiel Corpus of Spontaneous Speech (IPDS, a,b,c) was used. The Kiel Corpus consists of both read and elicited German labelled segmental, but for this investigation just the dialogue part was used. The dialogue part used had 59036 annotated entrances on the word-level, including non-verbal utterances.

Prosodic labelling for sentence accent is in the Kiel Corpus attached on the level of the word, and has four degrees (Kohler et al., 1995):

0 - unaccented

1 - partially accented

2 - accented

3 - reinforced

The dialogues used were all elicited around the task for the dialogue participants to agree on meeting times, so the dialogue was clearly task-oriented. All dialogues had two dialogue participants.

2.2 Procedure

For the investigation a set of personal pronouns was excerpted from the corpus. The set consisted of the forms of personal pronouns presented in Table 1.

	1st sg	2nd sg	3rd sg m	3rd sg f	3rd sg n	1st pl	2nd pl	3rd pl
Nominative	ich	du	er	sie	es	wir	ihr	Sie
Dative	mir	dir	ihm	ihr	ihm	uns	euch	Ihnen
Accusative	mich	dich	ihn	sie	es	uns	euch	Sie

Table 16.1: All forms of pronouns excerpted

All pronouns from the set in Table 1 were excerpted and sorted according to the prosodic labelling of the sentence accent. Pronouns carrying sentence accent of degree 1 to 3 were analysed in their context, and categorised according to the nature of the parallel element of the clause containing the contrasted pronoun and according to the interpretation of the relation to the context (contrastive/non-contrastive). "Accented pronouns" will hereafter be used to refer to the set of personal pronouns which carry sentence accent of degree 1–3. A subset of unaccented pronouns was also excerpted and used as a comparison. The context of the unaccented pronouns was examined regarding similarities with contexts of accented pronouns and behaviour with a possible accent. Instances from the different categories were compared afterwards with instances from the set of personal pronouns labelled with accent degree 0. These pronouns will hereafter be referred to as "unaccented pronouns".

3 Results

The excerption of the set of pronouns gave a result of 4201 instances of personal pronouns, of which 2796 were marked for some degree of sentence accent. In the case of "du", "er" and "ihr" were never realised as accented pronouns, and for this reason "du", "er" and "ihr" were excluded from further analysis. Similarily with "es"; just unaccented cases were found (189), because of this and because most instances were in the form of formal subjects, "es" was likewise discarded from the analysis. The frequencies of the remaining 2609 tokens totaled by degree of accentuation and type are given in Table 2:

	1st sg ich	3rd sg sie	1st pl wir	3rd pl Sie	Total
Total (0-3)	1380	10	777	442	2609
Unaccented (0)	1255	6	759	360	2380
Partially Accented (1)	41	0	12	36	89
Accented (2)	76	1	6	43	126
Reinforced (3)	8	0	0	3	11
Total accented (1-3)	125	1	18	82	226

Table 16.2: Frequencies of the distribution of accent degree for all personal pronouns excerpted. The bottom row indicates the distribution and amount of the pronouns closer analysed.

The set of accented pronouns consisted, as is shown in the bottom row of table 2, of 226 instances. A subset of 226 unaccented pronouns from the original set of 2380 was excerpted and served as comparison set.

3.1 Accented pronouns

After an initial examination the 226 cases of accented personal pronouns, were intuitively sorted according to function and according to presence or absence of a source parallel element. Five main categories crystallised:
1) Emphatic/Introductory
2) Confirmation
3) Weak contrast
4) Parallel contrast

5) Very Strange
Generally a contrastive interpretation was possible to make, but it could not be regarded as preferred in all cases, in some cases emphasis was a more plausible interpretation. Also, the nature of the contrast differed between the different categories, mainly according to where contrast element information was to be found, i.e. if it was a clear contrast to some element in the discourse, or if the contrast was interpreted to merely be between the two speakers in the discourse situation. It also differed in the dimension to which the contrast was central to the message.

1. Category: Emphatic/Introductory

In this category are cases, which are more introducing than contrasting in nature, e.g. dialogue or subdialogue initial cases, or topic change. Here no clear source parallel could be determined from the context, the only contrast possible to interpret is between the two speakers in the discourse situation. However, the contrast relation does not seem to be the point of the message. The accent is rather interpreted as emphatic rather than contrastive. In Example 2 below this is the case for "Ihnen" (the numbers after the italicised pronouns indicates the degree of accenting).

Example 2 Emphatic introductory.
WEM000 Frau Meier? Ich würde gern mit *Ihnen-2* (ah) einen Termin für (schmatzen) eine zweitägige Arbeitssitzung (A) vereinbaren. *mir-2* würde es am besten passen wenn *wir-1* (A) Mittwoch und Donnerstag nehmen könnten.

In comparing utterances from this category with utterances containing unaccented pronouns there are many unaccented similar cases.

2. Category: Confirmation

This category contain cases where an earlier message is repeated for the sake of clarity, i.e. a parallell element exists, but no negative polarity, and no clear contrastive relation. The absence of a meaningful contrast is the reason that these examples are not in the category Parallell, even though a clear source parallel could be determined from the utterance or the discourse situation.

Example 3 Confirmation
KAP003 ... ich könnte dann von Montag bis , ja , Mittwoch kann ich auch auf jeden Fall (A).
OLV004 (A) Montag , den zwölften , bis Mittwoch , den vierzehnten könnten *Sie-1* (Klicken).

In Example 3 above the speaker OLV is merely repeating the information. Later on in the dialogue it turns out that the time period suggested by KAP is possible for OLV too. The utterance could certainly be contrastively interpreted in isolation, but given the longer context the contrastive interpretation is less probable, i.e. the contrast relation does not seem to be a central part of the utterance.

Similar unaccented cases, i.e. the same clause or sentence but with an unaccented personal pronoun were not found in the material.

3. Category: Weak Contrast

This category consists of cases with a contrastive effect, but without a clearly parallel contrast element. This is the largest category. Its characteristics are that a contrastive interpretation is favoured by the analyser, but there is no clear presence of a contrast element in terms of clear overt parallelism. It is however always possible to find a vague contrast element, in terms of implicit message or message signalled in multiple places or signalled from multiple sources.

Example 4 Weak contrast
SVA007 (A) fünfundzwanzigste selber , sowie sechsundzwanzigster ist schlecht(Z) , aber dann kann ich noch fünf Tage unterbringen (Klicken)
AME008 (Klicken) ja , das kommt bei *mir-2* auch ganz gut hin , das(Z) wäre dann vom siebenundzwanzigsten bis zum einunddreißigsten März (Klicken)

Under this heading are also question-answer pairs like Example 5 categorised

Example 5 Weak contrast in a question-answer pair
AME008 (Klicken) erster Februar selbst paßt mir nicht so gut. wie sieht das denn ab dem dritten bei Ihnen aus (Klicken)
SVA009 (Klicken) dritter wäre in Ordnung für *mich-1* (Klicken)

Here, in the question "wie sieht das denn ab dem dritten bei Ihnen aus" speaker AME implies that for him the time is okay. The answer could, in that case, be interpreted as intentionally parallel, i.e. the speaker intends to communicate "the third is okay for me too". This category was the largest category and similar cases with unaccented pronouns could also be found. the centrality of the contrast relation is a relatively important part of the utterance, but still not to the same extent as for next category, Parallell contrast.

4. Category: Parallell contrast

This category contains cases where it is possible to find a clear deaccented source in the preceding context. The contrast could be of positive or negative polarity. This category is most like the classic examples of crossover interpretation.

Example 6 Parallel contrast
FRS008 tut mir leid, da bin ich noch in Dresden. Montag den einunddreißigsten.
ANS009 (Schmatzen) es tut mir leid, da bin *ich-2* leider nicht da.

In cases like Example 6 above, the two utterances are highly parallel, the big difference is who is speaking and how "not here" is expressed (Dresden/nicht da). In this example the distance between the parallel elements is quite short, in other cases it could be much longer. In this category the contrastive relation seemed to be very central to the discourse. It was hard to find equivalent examples with unaccented pronouns, mainly because the parallel elements here could have many utterances in between, i.e. the distance between the parallel elements could be quite long (in the example it is however short). However, to make a contrast over

such a long distance, seems in many cases to require accenting.

5. Category: Very Strange

This category consists of corrections, lexicalised expressions and very difficult cases, like Example 7.

Example 7
ANL005 (Klicken) (ah) das ist leider auch nicht möglich.
CHD006 ja gut, dann nehmen *wir-1* (Z) den Sonntag vorher (Klicken)

In case of "wir" (we) the natural thing to make a contrast against is "you" or "they", but in the dialogues studied there are just two speakers, so, there is no "they", and there is also no "you" to contrast against, because the listener is included in "we". So, when considering the example in isolation, a contrast seems to be a possible interpretation

The distribution of the differents degrees of sentence accent (except degree 0, no sentence accent) is shown in Table 3:

	Part. Accented (1)	Accented (2)	Reinforced (3)	Total
Emphatic/Introductory	46	38	0	84
Confirmation	5	6	2	13
Weak contrast	28	53	4	84
Parallel contrast	6	18	5	29
Very Strange	8	7	0	15
Total	93	122	11	226

Table 16.3: The distribution of different degrees of sentence accent in the different categories

A further observation was that the accented personal pronouns were unevenly spread over the dialogues in the corpus. The percent of accented pronouns used by different speakers differed from 0% of the speakers total amount of produced tokens up to 1.7%, with an average of 0.4%. This indicates that the use of accented pronouns to a certain extent could be idiolectic.

3.2 Unaccented pronouns

The subset of 226 unaccented pronouns was also examined, specially with regard to contextual similarity to the accented instances, the potential to accent, and the possibility or impossibility in terms of plausibility of a contrast relation given the discourse context. The results form the categorisation is shown in table 4 below.

	Unaccented (0)
Possible with accent	81
Impossible with accent	145

Table 16.4:

Possible to accent means that the unaccented pronoun should be able to carry sentence accent, and that that would not change the message as result, while impossible to accent means that sentence accent on these pronouns would give rise to a change in the original message.

When examining the unaccented pronouns it was found that many of the unaccented cases had a context very similar to accented pronouns in the category weak contrast. Also identical instances as in the category emphatic/introductory was found. This indicates a possibility of different degrees of a contrast, or perhaps a zone of ambiguity betwen contrast and emphasis.

4 Discussion

Generally personal pronouns with sentence accent are rare in spontaneous dialogue. However, there are examples of this usage, and these examples appear in a wide range of contexts. There is a weak tendency that a higher degree of accentuation correlates with a clearer parallel element in the text. Apart from this weak tendency the different accented pronouns did not differ much from each other.

The contrast element was in most cases found in the close preceding context, but it could also hold over long distances in the discourse. In categorising the examples it also became clear, that a single discourse relation like contrast could hold on multiple levels at the same time. I.e. a contrasted item could have a contrast element in the near context as well as a second contrast element with a long-distance contrast element at the same time. This indicates that discourse relations may hold on multiple levels of the discourse.

Example 8

TIS003
... also (ahm) am Mittwoch is Ihnen recht, wenn ich zu *Ihnen-2* komme?

HAH004
das ist mir sehr recht...
...
TIS003
...kommen Sie diesmal zu *mir-2* dann?

HAH012
ja, natürlich...
...
HAH018
ja, natürlich. wunderbar. kommen *Sie-2* zu *mir-2* ...

The categories are made up on the basis of the kind of information in the contrast element. Here the differences were clear. In the case of parallel contrast a very clear parallel element could be found in the text, i.e. in the discourse itself. In the case of weak contrast it is also possible to trace the contrast element to the discourse itself. Confirmation does have a parallel element, but it lacks contrast element in the discourse. There is no clear contrastive relation inside the discourse. In some cases it is rather a repetition, which just slows down the discourse at a certain point, but still lets the speaker keep the turn. In other cases it has a more clarifying function, and the utterance is a repetition of a message communicated

much earlier in the discourse. The reason for categorising these instances as confirmation does not, however, lie in the two parallel utterances themselves, which in some cases may clearly signal the relation contrast, but in the wider discourse context, which does not at all support a contrastive interpretation.

Introductory/emphasis are generally lacking a clear contrast element in the discourse. Because of the speech situation it is however always possible to use the speaker or hearer themselves as contrast element. This does however not seem to be the preferred function of the contrasted pronoun here. Rather these instances seem to serve as starting-points for a topic or subtopic, e.g. after both dialogue participants agreed on step one, "to meet", they could start with step two, e.g. "to decide meeting time". Such sentences would often be of the form like "mir würde es am besten...". It is thus clearly possible to interpret this as a contrast relation between the speaker and the hearer, but it may in the same time use the contrast to make a distinction between the two different steps or topics of the discourse. These cases were the easiest to find exact parallels with unaccented pronouns for, a fact that indicates that the contrast relation itself might be of lesser importance in this cases.

In the case of weak contrast, it should be stressed that it is uncertain whether subjects in an experiment situation should judge the instances as contrastive, or if they should interpret them as just emphatic. The analysist was explicitly searching for some possible contrast element, and with that perspective it was easy to find something that could be interpreted as a contrast element. With regard to the similarity in context between the accented and the unaccented pronouns in the category weak contrast, this is an issue that needs further investigation. To check for the interpretation contrast/emphasis an experiment is needed, where subjects have to judge whether the accented pronouns were contrastive.

Accent is an efficient way to signal contrast, but in spontaneous dialogue it does not seem sufficient to alone trigger a contrastive interpretation. The relation contrast is, apart from accenting, also dependent on support from the discourse itself or from the situation. The centrality and strength of the contrast relation seems to coincide with the amount of new information the contrastive relation itself adds to the discourse. The strongest contrast relations seems to be evoked when both accent and clearly parallel elements are present in the discourse, and both the preceding and following context supports a contrastive interpretation, i.e. when the contrast relation is gives new information which is of central interest both speakers. In the cases where the utterances themselves permit a contrastive interpretation, when the contrast relation does not itself add new information and when the contrastive relation is not of primary importance for the discourse or of the speakers immediate interest the contrast is perceived weaker. This holds for cases of both introductory/emphasis and confirmation, where the contrast element in the discourse is vague, and the clearest contrast element is found in the situation, i.e. where the speaker uttering "you" itself is the most prominent contrasting item. In such cases the contrast is more interpreted as contrasting against everything else, rather than contrasting against clear contrast element.

With regard to the data, it would be a hard task to apply some higher order unification between the less accented elements. In the first case, of course, because it is difficult to judge degree of similarity, but also because the contextual relevance for the contrast relation seems to play a role in the interpretation. This means, that if the interpreter searches for a contrast, it is easy to establish a weaker form of contrast to some given element. There are however cases in the material where the result should be confusing.

As a conclusion it could be said that a central contrast relation seems to add relevant new information to the discourse. Those cases seem to have both a clear contrast element and a

clear parallel to the deaccented part of the utterance in the discourse. Contrasts, which do not have this newness, are not as salient, and they also have less discourse-based contrast element, or a vaguer parallel to the deaccented parts of the utterance.

Bibliography

Bosch, P. and van der Sandt, R. (1994). Focus in natural language processing. Working Papers of the Institute for Logic and Linguistics 8, IBM, Deutschland. vol. 3.

Byron, D. and Stent, A. (1998). A Preliminary Model of Centering in Dialog. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics.*

Cole, P., editor (1978). *Syntax and Semantics.* Academic Press.

Cole, P., editor (1981). *Radical Pragmatics.* Academic Press.

Cruttenden, A. (1986). *Intonation.* Cambridge UP.

Eckert, M. and Strube, M. (2001). Dialogue acts, synchronising units and anaphora resolution. *Journal of Semantics.* to appear.

Gardent, C. (2000). Deaccenting and higher-order unification. *Logic, Language and Information*, 9(3).

Grice, H. P. (1978). Further notes on logic and conversation. in: (Cole, 1978).

Gundel, J. (1994). On different kinds of focus. in: (Bosch and van der Sandt, 1994).

Horne, M. and Filipson, M. (1995). Developing the prosodic component for swedish speech synthesis. In *Proceedings of Eurospeech.*

IPDS. The Kiel corpus of spontaneous speech, volume 1, CD-ROM #2. Kiel: Institute für Phonetik und digitale Sprachverarbeitung.

IPDS. The Kiel corpus of spontaneous speech, volume 2, CD-ROM #3. Kiel: Institute für Phonetik und digitale Sprachverarbeitung.

IPDS. The Kiel corpus of spontaneous speech, volume 3, CD-ROM #4. Kiel: Institute für Phonetik und digitale Sprachverarbeitung.

Kohler, K., Pätzold, M., and Simpson, A. (1995). From scenario to segment. the controlled elicitation, transcription, segmentation and labelling of spontaneous speech. AIPUK 29, University of Kiel.

Krahmer, E. and Swerts, M. (2001). Meaning and intonation: The cases of contrastive intonation and meta-linguistic negation. In Bunt, H., Van der Sluis, I., and Thijsse, E., editors, *Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4).* Tilburg University.

Prince, E. (1981). Toward a taxonomy of given-new information. in: (Cole, 1981).

Simpson, A. (1998). Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. AIPUK 33, University of Kiel.

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9.

Paper 17

Modal particles and the common ground: meaning and functions of German *ja, doch, eben/halt* and *auch*

Elena Karagjosova
elka@coli.uni-sb.de
http://www.coli.uni-sb.de/~elka

Abstract

In this paper, it is argued that it is not desirable to assume separate lexical items for each context in which a modal particle may occur. Instead, definitions for the basic meanings of the German modal particles *ja, doch, eben/halt* and *auch* are provided by filtering out contextual parameters which are made responsible for the meaning variants of the particles in question. A framework is proposed in which the contribution of the MPs to the utterance meaning is divided into three interrelated aspects: basic meaning, utterance illocution and function in discourse.

1 Introduction

The German modal particles (MPs) *doch, ja, eben/halt*[1] and *auch* are used in dialogue to refer to the common ground between the participants and more closely to the beliefs shared between them[2].

In the literature on MPs, it has become a habit to assume numerous contextually bound meanings for a single MP. In (Helbig, 1988), seven variants for *doch* are listed according to the sentence type in which the particle may occur or according to the presence or absence of a preceding context. Furthermore, within one meaning variant, antonymous functions of the particle are sometimes postulated. E.g., according to (Helbig, 1988), *doch* in imperatives can make an utterance sound urgent, irritated or reproachfull on the one hand and mollifying, polite or casual on the other. We find that such lexicographic approach to MPs should not be adopted in the linguistic analysis of these lexical items. In our view, it should be possible to derive the various contextual functions and senses of a MP from a basic, contextually independent meaning. We also believe that the inadequacy of the above approach lies in ascribing properties of contexts to the meaning of MPs. Therefore, it is important to filter out the relevant contextual factors in order to arrive at a sound common basis for all contextual variants of a MP.

On the account presented here, it is argued that each MP has a contextually invariant basic meaning. The basic meanings are defined in terms of a propositional attitude (speaker's belief) that the respective MP is assumed to express. The numerous contextual meaning variants often assumed in the literature are interpreted as secondary effects that can be ascribed to different instantiations of the basic meaning in a particular contextual setting. These effects are furthermore interpreted as

[1] *eben* and *halt* are synonymous and mutually replaceable.

[2] Common ground in dialogue is regarded here as subsuming the common beliefs of the participants. The terms *common belief/knowledge* and *shared belief/knowledge* are used synonymously.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

elka@coli.uni-sb.de
http://www.coli.uni-sb.de/~elka.

particular functions that the respective MP can fulfil in discourse. The contextual properties involved in producing the different functions a MP may fulfil, include sentence mood, belief states and preceding context, and the interplay of context with the basic meaning delivers the different contextual readings.

The focus will be on the German MPs *doch, ja, eben/halt, auch* in declarative sentences. First, their basic meanings will be sketched. Next, we will discuss the relevant contextual elements, their interplay with the basic meaning of the particular MP, and show how different contextual readings arise. In order to describe the different aspects of meaning and function of the MPs in question, a framework will be proposed in which the contribution of the MPs to the utterance meaning is divided into three interrelated aspects (basic meaning, utterance illocution, function in discourse). The relation between these different aspects will be explored.

2 The basic meaning of *ja, doch, eben/halt* and *auch*

MPs do not influence the truth-conditions of a sentence, but rather express a particular attitude of the speaker towards the proposition expressed by the sentence. The belief of the speaker expressed by the group of MPs addressed concerns the status of the proposition in the common ground already established by the conversants. Although the exact formulation of these meanings is a highly complex enterprise, a wide consensus has been reached in the literature with respect to the intuitions about them. E.g., it is widely accepted that *doch* in declaratives expresses that the speaker regards the propsition in its scope as common knowledge between him and his hearer. However, for *doch* in a slightly different context, e.g., in dialogue initial position, a different meaning variant is often assumed. For example, Helbig (1988) distinguishes between two variants of *doch* in declaratives: a $doch_1$, where the MP is used without reference to a preceding utterance, and a $doch_2$ used when referring to a preceding utterance. He then motivates this postulation of homonymy by the fact that for $doch_1$, a negative component of reproach is missing, whereas it is an inherent property of $doch_2$.

Although we agree with these intuitions, it does not seem sensible in our view to assume a different lexical item for each context in which a word occurs. It seems to be more desirable to view the meaning of a MP as highly contextually sensitive and to abstract a basic meaning from its numerous possible contextual occurences.

A first step towards this goal would be to consider for each MP the contextual setting in which it can be used. E.g., the contexts (in terms of sentence mood) in which *doch* can be used are declaratives, imperatives, questions, exclamatives. Some relevant contextual features for the declarative use of *doch* are (i) the position of the *doch*-utterance in discourse[3] – initial or responsive – and (ii) in the responsive case, who is the previuos speaker: the utterer of the *doch*-sentence or his dialogue partner. Then, it should be decided which is the most typical setting. Our hypothesis (which we will try to justify below) is that for *doch* it is the declarative responsive use. *ja, eben/halt* and *auch* also occur most typically in declaratives: *ja* mainly initially, *eben/halt* and *auch* only responsive. The meaning of the respective MPs in one context (to be motivated below) will be taken to be basic, and its other uses will be derived from this. E.g., the lack of the negative component of reproach in *doch* without preceding context as claimed by Helbig can simply be ascribed to the lack of a preceding context: there is nothing to be reproachful of. Nevertheless, as Helbig also points out, the basic meaning component is the same as in the case where *doch* is preceded by an utterance: a slight contradiction to the prior utterance. In Helbig's terms, *doch* creates this contradiction in the former case (with preceding context), and rules it out in the latter (without preceding context). In our terms, these facts need not be accounted for by assuming separate lexical items, but can be explained by assuming different discourse influences.

It is furthermore hypothesized that the basic meaning of a MP in non-declaratives should not be different from the one in declaratives. The meaning variation in these cases should be accounted for in terms of the different overall utterance meaning, not by way of several basic lexical items. Here, however, we will restrict our attention to the declarative uses.

[3]which mostly corresponds to the dialogue act type the utterance belongs to: forward-looking (initial position) or backward-looking (follow-up position).

2.1 *ja*

The basic meaning of the MP *ja* can be defined as expressing the belief of the speaker that the proposition in the scope of the particle is already common knowledge between speaker and addressee. This is what is expressed in (1):

(1) A: Maria ist ja verreist.
A: Maria has MP left.

It is often argued that in a follow-up utterance, *ja* has the function of marking the proposition in its scope as an explanation of the subject matter expressed by the previos utterance. Since *ja* is typically used without referring to a preceding context, we believe that the explanatory function is not part of its basic meaning but should rather be seen as an additional function it can fulfill depending on the structure of discourse (see section 4 below). As will be argued below, this is not the case for *eben/halt* and *auch* which have as part of their meaning the property to refer back to the preceding utterance.

2.2 *doch*

doch also expresses the belief of the speaker that the proposition in the scope of the particle is already common knowledge between speaker and addressee. It means, however, also that the speaker sees this consesus endangered by a proposition (which, depending on context, can be a belief of the previous speaker or a previous belief of the same speaker or only assumed) that contradicts this assumption of shared knowledge, and that the present speaker refuses to accept this contradicting belief on grounds of the information he already has. Thus, the meaning of *doch* in (2) can be paraphrased as *since I know that we both know that p, although you are suggesting that you do not know p, I cannot accept what you suggested.*

(2) a. A: Peter kommt auch mit.
A: Peter is also coming along.

b. B: Er liegt doch im Krankenhaus.
B: He is MP in the hospital (, don't you remember?).

It could be argued that the third meaning component of *doch* – that the speaker does not accept what the other agent suggests – is induced by (2) also in the absence of the MP. But this is not so in the cases where the counterevidence is only assumed, e.g. the use of *doch* in (3b) can not be explained in these terms, since in (3b), no explicit evidence against a common knowledge assumption of the proposition expressed by *A*'s utterance is provided:

(3) a. A: Peter sieht sehr schlecht aus.
A: Peter looks very bad.

b. B: Er war doch lange krank gewesen.
B: He has been MP ill for a long time (, don't you remember?).

We can now return to the question of why the responsive use of *doch* is regarded as basic. This seems motivated by the observation that also in cases where the particle does not object to a concrete preceding utterance in dialogue initial position, it nevertheless induces an opposition to a proposition which the speaker has in mind, but which is not contextually present. E.g., in (3b), the use of *doch* suggests that the speaker wants to contrast the content of his utterance with a possible counterargument which he wants to rule out as not valid to him (the purpose of this will be discussed later on in section 4).

2.3 *eben/halt*

eben/halt always occurs backward-looking as a reaction to a preceding utterance or a(n element of the) situation. Its meaning in declaratives which will be assumed to be basic is that the speaker believes that (i) the proposition in the scope of the MP is shared knowledge among a group of language users to which speaker and hearer belong, and that (ii) it is also shared knowledge in the group that there is a relation of explanation between the proposition in the scope of *eben/halt* and the preceding dialogue contribution. The explanation relation also holds without *eben/halt* between the respective propositions, but without it, there would be no implication whatsoever about an assumption that the preceding proposition is shared knowledge among the interlocutors. The meaning of *eben/halt* in (4b) can be paraphrased as *that he looks bad is what everybody would expect, since it is well known that he was ill.* This implies also that B considers A's utterance as having no informative value for him, since B already knew that Peter was ill, and that being ill leads to looking bad, hence he was already expecting Peter to look bad:

(4) a. A: Peter sieht sehr schlecht aus.
A: Peter looks very bad.

b. B: Er war eben/halt lange krank gewesen.
B: He has been MP ill for a long time (, as we all know).

Since *eben/halt* is used only backward-looking, we believe that this relation of explanation that holds between the *eben/halt* and the preceding utterance should be seen as part of the meaning of the particle (the same holds for *auch*, see further below). This relation can be formalized by a defeasible rule of the style used in (Lagerwerf, 1998): $p > q$ meaning *Normally, if p, then q*, where q is the utterance containing *eben/halt*, and p the utterance preceding it. The intuition that also what is stated in the utterance preceding the *eben/halt*-utterance is regarded as shared knowledge can be then accounted for by applying defeasible modus ponens Asher and Morreau (1991).

2.4 *auch*

auch also occurs only in backward-looking acts.[4] In the case of *auch*, common knowledge is involved only indirectly. The particle expresses that the speaker had the preceding proposition already in his belief state before it was uttered by the other conversant, and that the proposition he asserts stands in a causal relation to the preceding one. The fact that the speaker already believes the proposition uttered by the other conversant makes it part of the common knowledge of the dialogue participants. E.g., the meaning of the MP in (5b) can be paraphrased as *I know that he looks vary bad. It is because he was ill.*

(5) a. A: Peter sieht sehr schlecht aus.
A: Peter looks very bad.

b. B': Er war auch lange krank gewesen.
B': He has been MP ill long time (, after all).

Without *auch*, there is no implication that the speaker B knew what A's utterance expresses prior to A uttering it. This is also why we take the knowledge of the speaker about the causal relation between the two propositions to be part of the particle meaning.

3 Interaction with context

When a MP appears in utterances in different contexts, additional effects arise for the interpretation of the utterance. It is argued here that these effects are not due to different contextually bound meanings

[4] In questions, however, it can be used dialogue initial, but it is nevertheless understood as backward-looking, as a reaction to the situation.

of the MPs, but can be led back to and explained by way of the basic meaning of the respective MP. The relevant contexts treated are: implicatures from utterances, speechacts performed by utterances, preceding context and the belief states of the conversants.

3.1 Implicatures and speechacts

One basic observation about particles that express common knowledge of a proposition is that different inferences can be drawn from one and the same utterance with and without a MP. The reason is that the common knowledge assumption expressed by these particles conflicts with the conversational implicature that normally arises from observing the quantity implicature: that by asserting a proposition, the speaker expects it to be new to the addressee Searle (1969). Thus, declarative utterances with the MPs treated here realize a speechact different from an *assert*-act: *ja, doch, eben/halt* realize an act that can be called a *remind*-act since they express that the speaker has expected the proposition to be already common knowledge. A *remind*-act can be defined similarly to an *assert* act in (Searle, 1969) except that one preparatory condition (next to the sincerity condition) would be that it is obvious to the speaker that the hearer knows the proposition expressed by the speaker, but that the speaker nevertheless needs to state it explicitly for some reason.

auch realizes a *confirm* and *give-reason*-act, since it expresses that the speaker already knows (i) the preceding propositions and thus confirms it and (ii) that the proposition he asserts stands in a causal relation to the preceding one and is viewed as the cause of the circumstances described in the preceding proposition. It is a confirmation and not simply a case of acceptance since it expresses that the information proposed by the previous speaker is not new to the *auch*-speaker: it conveyes something like *I think so too.* In the dialogue act taxonomy in (Alexandersson and et al., 1998, pp.23,44), a *confirm* act is defined as an utterance by which *the speaker wraps up the result of the negotiation.* This wrapping up can be done by *"either giving a summary of the result or by using certain phrases that indicate a closure of this topic of negotiation, such as 'machen wir's so'"* (let's do it that way); this implies that the subject of negotiation has been present in the dialogue prior to confirming it which implies that it is not new to the hearer. A *give-reason*-act is defined (ibid.) as an utterance that *"contains the reason/justification/motivation for a statement made in the immediately preceding and/or following context"*. It can be added that a preparatory condition for this act should be that the speaker believes the information he conveyes to be new to the hearer. Accordingly, *eben/halt* do not realize in our view a *give-reason*-act since this MP presupposes that the reason should be already known by the hearer.

3.2 Preceding context

In determining the basic meaning of most MPs addressed, a reference to a preceding utterance was involved. For *doch*, the preceding utterance usually contains evidence against the common knowledge assumption, for *eben/halt* and *auch* it is marked as already known by the speaker. But counterevidence need not be manifested in the context in order for *doch* to express a contradiction as claimed in section 2. We argued that the case where counterevidence is not contextually present should not be accounted for by postulating a new meaning variant. Its contextual presence is crucial for the interpretation of the utterance containing the MP, but has no consequences for the meaning of the MP itself. This can be accounted for by assuming different functions a MP can play given a particular contextual setting. Consider (2). Since the counterevidence for the common knowledge assumption expressed by the *doch*-utterance is contextually present, B's utterance should be interpreted as a correction of A's view about what is common knowledge (remember the definition of *doch*'s basic meaning in section 2). Thus in our view, *doch* fulfills in this case the (meta-communicative) function of correcting the other speaker's view about the common ground between the two interlocutors.

Counterevidence need not always be found in another agent's preceding utterance: it can be part of the agent's own information state. E.g., in (6), the first clause expresses a new belief the speaker has acquired that speaks against his old belief expressed in the second clause:

(6) A: Ich habe wieder Schnupfen. Dabei lebe ich doch ganz vernünftig.
A: I have again a cold although I live MP quite healthy.

If there is no contextually present counterevidence to the assumed common knowledge as in (3b)[5], the speaker does not have reasons to correct the addressee's utterance. In (3b), the meaning of the utterance can be paraphrazed as *since I know that we both know that p, even if you were suggesting not p, I would not accept it.* In this case, we believe, *doch* can be interpreted as a means for the speaker to increase the argumentative power of his utterance Karagjosova (2001). It has the function to make the hearer believe the proposition in its scope by the speaker "disguising" it as something that is already common knowledge between him and hearer and thus already believed by the hearer. This is what we call the *persuasive* use of a MP.

As already argued, no consideration of the preceding context is needed for the definition of the meaning of *ja*. That is why it can be used even in cases where the preceding context contains counterevidence for the common knowledge assumption[6]:

(7) a. A: Wo ist denn das Bier?
A: But where is the beer?

b. B: Ich bringe es ja schon.
B: I am bringing it MP already.

In (7b), *ja* does not have the function of correcting since its meaning is just a common knowledge assumption. Its function here can be said to be to reassure the hearer that the action in question will be performed. This reassurance is achieved by suggesting that the speaker regards the proposition as common knowledge which the hearer needs to be reminded of.

auch, *eben/halt*, on the other hand, require preceding context. Their function is also to correct the view or the behaviour of the co-conversant concerning the common ground. They can also have other functions as will be discussed in the next section.

3.3 The speaker's belief state

It was claimed that the MPs treated here express a belief of the speaker that refers to the common ground between speaker and addressee. In our view, this still holds even if in reality the speaker does not hold any assumptions whatsoever with regard to the common ground. The reason is that the meaning of MPs does not contribute to the truth-conditions of the proposition expressed by the MP-utterance. By using a MP, the speaker *commits* himself to the respective belief. This also explains the persuasive use of the MPs discussed in 3.2.

4 How do the basic meanings of *ja, doch, eben/halt, auch* account for their functions

It was argued that MPs have a basic meaning that influences (i) the utterance meaning (inferences that can be made from utterances), (ii) the speechact that can be performed by the utterance. It was also claimed that the same basic meaning can be the basis of different interpretations when the particle is used in various contexts. The question of how the different functions of a MP are related to its basic meaning that was touched upon in section 3.2 will be pursued here in more detail.

First of all, a taxonomy of functions should be adopted. We assume two basic types of functions in connection with the MPs which we term *metacommunicative* and *rhetoric*. The metacommunicative function refers to the ability of a MP to (i) correct the previous speaker's idea of what is in the common ground (as in the case of *doch, eben* and *auch*) or (ii) just emphasise the status of the proposition as an element of the common ground or to highlight it for whatever purpose (*ja*). It is

[5] In this case the counterevidence is only presupposed.
[6] The following example is taken from (Dahl, 1985).

called metacommunicative since it does not concern the content of the utterance, but its status in the common ground.

The rhetoric functions concern the rhetoric structure of the discourse. E.g., one can emphasise the status of a proposition as part of the common ground in order to provide an argument for some other proposition (this would be the *argumentation* function) or in order to create a salient basis for a follow-up utterance (*elaboration* function).

The functions we assume MPs to perform are connected with their illocutive force. E.g., a *doch*-utterance is used to remind the hearer on a proposition; this can be done for different purposes, e.g. to (i) correct the hearer's view about the common ground (in follow-up utterances), (ii) provide an argument for a claim (follow-up utterance without counterevidence) or (iii) elaborate on a claim (initial utterances). This relation between MP-function and illocutive force of the MP-utterance corresponds to the notion of dicsourse acts hierarchy proposed by Traum (1994) in which lower level acts like *core speech acts* are comprised in higher level discourse acts called *argumentation acts*, e.g. a core speechact like *inform* may be used in order to *summarize*, *clarify*, or *elaborate* prior conversation.

The functions MPs perform are titely connected to their basic meaning. E.g., by virtue of its basic meaning, *ja* focusses the attention of the hearer of a proposition which is (claimed to be) part of the common ground. This can be done for different purposes: to underpin an explanation for a fact associated with it (8b) or to provide further information about it, elaborate on it (9b), or just to sound more convincing, argue in favour of it.

(8) a. A: Peter sieht sehr schlecht aus.
A: Peter looks very bad.

b. B: Er war ja lange krank gewesen.
B: He has been MP ill long time (, as you know).

(9) a. A: Peter ist ja im Krankenhaus.
A: Peter is MP in the hospital (, as you know).

b. A: Er wird morgen operiert.
A: He will be operated tomorrow.

eben/halt and *auch* are always understood as a reaction to the preceding (linguistic or extralinguistic) context. The function involved in these cases is a correction of the hearer who asserts something that is already part of the common ground (*eben/halt*) or of the belief state of the hearer (*auch*). E.g., by expressing that he already believed the proposition in the contribution of the previous speaker, the utterer of the *auch*-sentence corrects the communicative behaviour of the previous speaker who has uttered a sentence with no informative value whatsoever for the hearer[7]. But *auch*-utterances can also fulfill an argumentative function when the speaker does not have reasons to assume common knowledge (as argued in section 3.3).

[7]Except that the hearer learns that the speaker also knows the proposition in question.

Utterance	Particle meaning	Utterance illocution	Discourse function
doch(p)	(i) S believes that p shared with H (ii) there is counterevidence q for p (iii) S does not accept q	S reminds H of p	(i) S corrects H that p shared (ii) S argues for p (iii) S elaborates on p
ja(p)	(i) S believes that p shared with H	S reminds H of p	(i) S emphasizes on p (ii) S argues for p (iii) S elaborates on p
eben(p) *halt*(p)	(i) S believes p shared in group G (ii) S believes $p > q$ shared in group G $\{S, H\} \subseteq G$,q uttered by H prior to p	(i) S reminds H of p (ii) S reminds H of $p > q$	(i) S corrects H that p and q shared (ii) S argues for p
auch(p)	(i) S believes q (ii) S believes $q > p$ q uttered by H prior to p	(i) S confirms p (ii) S gives reason p for q	(i) S corrects H that q not new (ii) S argues for p (iii) S elaborates on p

Table 17.1: The aspects of MP-contribution to utterance meaning

Table 17.1 summarizes the different interrelated aspects of meaning and function of the MPs treated. S and H denote speaker and hearer, respectively, p and q denote propositions. *Utterance illocution* refers to the speech act realized by the utterance containig a particular MP. *Discourse function* refers to the function(s) the respective MP may fulfill depending on the context.

The different combinations of context factors with the basic meaning can be seen to yield the different discouse functions of the particles. It is left for future work to algorithmize this. Some considerations may however be noted. For instance, for the different functions of *doch*, following conditions hold: In the case of an utterance p uttered by another speaker A prior to an utterance of the form **doch** q, if there is an adversative relation between the propositions ($q > \neg p$), the function of **doch** in q is a correction, if there is no adversative relation, the function is to argue.

Other effects of using MPs like politeness, reproachfulness etc. can also be tracked down to the levels of meaning discussed. E.g. *ja* is said to make the utterance sound more polite. This can be traced back to its basic meaning: by implying that the proposition is already known to the hearer, even if this is not the case, the speaker makes the hearer feel more comfortable.

5 Summary and conclusion

In this paper, it was argued that it is not desirable nor plausible to assume separate lexical items for each context in which a MP may occur. It was suggested that a basic meaning for a MP can be abstracted from its numerous possible contextual occurences and that the meaning variation should be ascribed to different instantiations of the basic meaning in a particular contextual setting. These effects were interpreted as functions that the respective MP can fulfill in discourse. A framework was proposed in which the contribution of the MPs to the utterance meaning is divided into three interrelated aspects: basic meaning, utterance illocution and function in discourse.

Bibliography

Alexandersson, J. and et al. (1998). Dialogue acts in verbmobil-2. Technical report, Verbmobil-Report 226.

Asher, N. and Morreau, M. (1991). Commonsense entailment: a modal theory of nonmonotonic reasoning. In *IJCAI'91, Proceedings of the Ninth International Joint Conference on Artificial Intelligence.*

Dahl, J. (1985). *Ausdrucksmittel für Sprechereinstellungen im Deutschen und Serbokroatischen.* PhD thesis, München.

Helbig, G. (1988). *Lexikon deutscher Partikeln.* Verlag Enzyklopädie, Leipzig.

Karagjosova, E. (2001). Interpreting utterances with modal particles. In *Proceedings of the IWCS-4 (Fourth International Workshop on Computational Semantics*, Tilburg.

Lagerwerf, L. (1998). *Causal connectives have presuppositions.* PhD thesis, Tilburg.

Searle, J. R. (1969). *Speech acts.* Cambridge university press.

Traum, D. (1994). *A computational theory of grounding in natural language conversation.* PhD thesis, University of Rochester.

Paper 18

A Bayesian Approach to Dialogue Act Classification

Simon Keizer
Parlevink Language Engineering Group
University of Twente
skeizer@cs.utwente.nl
http://parlevink.cs.utwente.nl/~skeizer/

Abstract

The paper proposes a probabilistic approach to the interpretation of natural language utterances in terms of dialogue acts. Bayesian networks can be used to combine partial linguistic information with knowledge about the state of the dialogue and about the speaker, in order to find the most probable dialogue act performed, using standard probabilistic inference in the network. The proposed approach is illustrated with a simple example.

1 Introduction

This paper deals with a conversational agent (the 'SERVER'), which is a participant in a dialogue with another agent (the 'CLIENT'). Our conversational agent perceives utterances of the client and tries to react to these utterances in an appropriate way. In Figure 18.1, a possible architecture for the conversational agent is given.

The dialogue manager coordinates the various steps involved in the interpretation of an incoming user utterance and the planning of what action to perform next. First of all, an incoming utterance is submitted to the components of speech recognition (in case of spoken dialogue), and syntactic/semantic analysis. Next, the pragmatic aspect of identifying what *communicative action* was performed by the user in uttering the sentence is dealt with. This part is what we will be concentrating on in this paper, and in our approach will take the form of *dialogue act classification*. Finally, the resulting interpretation will have to lead to a decision on what actions to take, including communicative actions addressed to the user.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

Parlevink Language Engineering Group
University of Twente
skeizer@cs.utwente.nl
http://parlevink.cs.utwente.nl/~skeizer/.

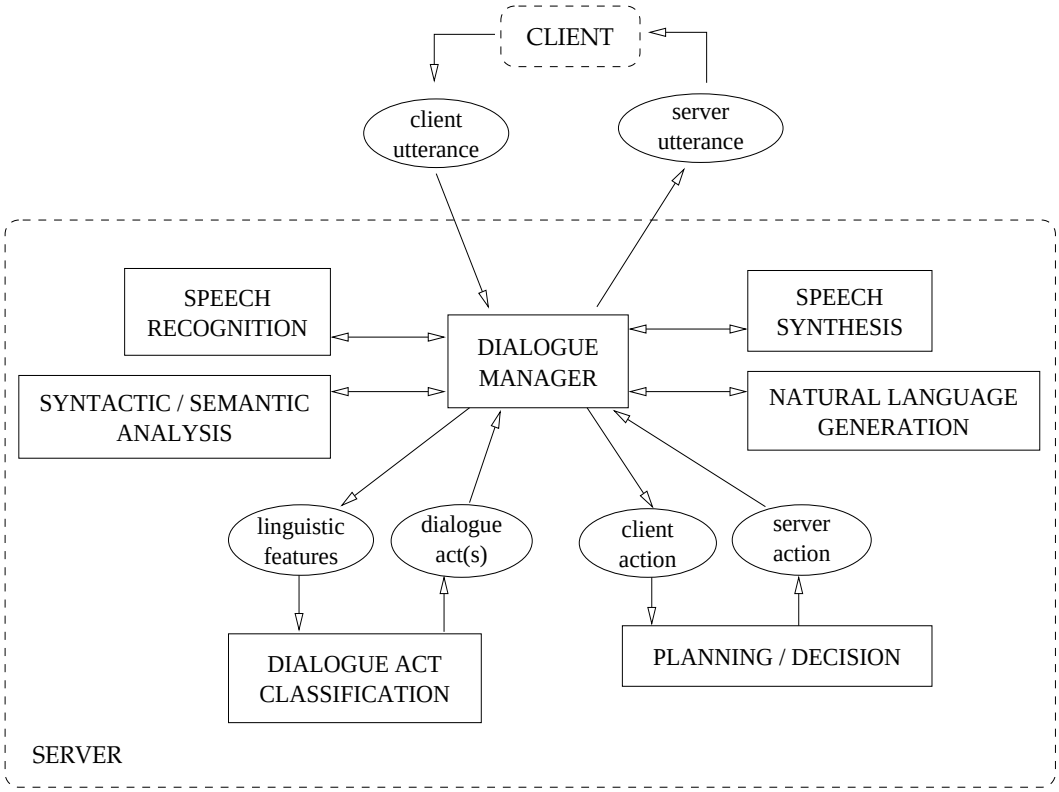


Figure 18.1: An architecture for a conversational agent

In identifying the communicative action performed in an utterance, the server has to deal with uncertainty. This uncertainty arises because of the incompleteness of the information that is provided by the components of speech recognition and syntactic/semantic analysis. In general, such components cannot provide all linguistic information that is contained in an utterance, especially because we are dealing with human speakers that may produce utterances that are partly unrecognised through bad pronunciation, sentences that are ungrammatical, or that contain unknown words. This is especially the case in mixed-initiative dialogues, where a client may tend to use more complex utterances.

In order to deal with the uncertainty, the server will have to make *educated guesses* in the interpretation of the client's utterances. Therefore, we will take a probabilistic approach to dialogue modelling, in the form of Bayesian networks.

In Section 2, we will discuss our approach to modelling dialogue acts. In Section 3, we will introduce the notion of Bayesian networks and how they can be used in the process of *dialogue act classification*. In Section 4, we discuss some related work on dialogue modelling where some notion of communicative act is central, and also other work particularly on dialogue act classification. Finally, in Section 5, some conclusions are drawn and an indication of further research is given.

2 Dialogue Acts

The notion of dialogue acts is originated in the work of Austin (1962) and Searle (1969). They observed that utterances are not merely sentences which can be either true or false, but should be seen as (communicative) actions. Searle introduced the theory of *speech acts*, which he defined as the whole act of uttering a sentence. He also gave a categorisation of types of speech acts, including e.g. REQUEST, ASSERTION, and ADVICE.

When we speak of *dialogue acts* however, we emphasise the importance of looking beyond the boundaries of an utterance itself when analysing that utterance (Traum, 1999). Besides linguistic information of the utterance in isolation (prosodic features, syntactic/semantic features, surface patterns, keywords, etc.), the meaning of dialogue utterances may also be determined by information concerning two other aspects:

1. the state of the dialogue: concerning e.g. the current topic or the communicative act(s) performed in the previous utterance(s);
2. the mental state of the speaker: concerning e.g. what the speaker believes or what his goals are.

Based on these three aspects, a categorisation of different dialogue act types can be made. We have developed a dialogue act hierarchy, which is based on the hierarchy of the DAMSL annotation scheme (Allen and Core, 1997). This scheme has been developed as a standard for annotating task-oriented dialogues and has been extended to a scheme that supports the specific features of the SCHISMA dialogue corpus. The SCHISMA project (SCHouwburg Informatie Systeem) concerns the development of a natural language dialogue system, in which users can get information about theatre performances and if desired, order tickets. The SCHISMA corpus consists of mixed-initiative, typed (i.e. non-spoken) dialogues in Dutch, which have been obtained from Wizard of Oz experiments.

The dialogue act hierarchy is subdivided into a number of layers, each of which (more or less independently) addresses different aspects of the communication. For this paper it will suffice to understand two of these layers. The layer of forward-looking Functions concerns dialogue acts which effect the future dialogue, and the layer of Backward-looking Functions concerns dialogue acts which refer to previous parts of the dialogue. In the annotation, the basic units, to which dialogue acts may be assigned, are called *segments*. Each Backward-looking Function that is assigned to a segment carries an explicit reference to another segment from the previous dialogue. Below, some of the dialogue acts from both layers in our hierarchy are given.

- Forward-looking Functions:
  - ASSERT: the speaker makes a claim about the world ("Othello starts at 8pm").
  - REQUEST: the speaker requests the hearer to perform some non-communicative action ("two tickets please").
  - REF-QUERY: the speaker requests the hearer for certain information in the form of references satisfying some specification ("what operas are on next week?").
  - IF-QUERY: the speaker asks the hearer whether something is the case ("do you want to make reservations?").
  - CONVENTIONAL: the speaker performs a conventional act like greeting or thanking the hearer ("thank you").
- Backward-looking Functions:
  - ACCEPT: the speaker positively responds to a REQUEST of the hearer ("I will take care of it").
  - REJECT: the speaker negatively responds to a REQUEST of the hearer ("I'm afraid I can't do that").
  - POS-ANSWER: the speaker gives the references asked for by the hearer, or responds affirmatively to an IF-QUERY of the hearer ("The following plays are performed this week: ...").
  - NEG-ANSWER: the speaker indicates, often in response to a REF-QUERY of the hearer, that no references satisfy the given specification ("there are no operas scheduled for next week").
  - FEEDBACK: the speaker gives feedback to the hearer, e.g. by indicating that he does not have the information requested for ("I have no information on payment").

3 Bayesian Dialogue Act Classification

Because of the need to be able to reason under uncertainty as explained in Section 1, we propose the use of probability theory in modelling the various relationships involved in interpreting dialogue utterances. Our model will consist of three interrelated components:

1. the Belief State of the server S: this state is determined by beliefs concerning:
   (a) the course of the dialogue, and
   (b) the beliefs, desires and intentions of the client C.
2. the Dialogue Act(s) performed in C's utterance;
3. the relevant Linguistic Features that C's utterance may contain.

Using this model, the server S can calculate what most probably must have been the dialogue act performed in an utterance of the client C, given new information w.r.t. linguistic features of that utterance, obtained from the speech recognition and syntactic/semantic components. We will describe how this can be done by using probabilistic inference in a Bayesian network.

A probabilistic model is described by a set of discrete *Random Variables* (RVs), i.e. variables that describe events with different possible outcomes. Such an event could be the outcome of throwing a pair of dice, the measurement of the current temperature, but also the event of a linguistic parser concluding that a typed utterance had a question mark at the end. Table 18.1 shows a set of RVs, used for a simple example that will illustrate the approach proposed in this paper. These two-valued, Boolean RVs are given together with their meaning and also which of the three components of our model they are associated with. It should be noted that it is not a requirement that the all RVs are Boolean; we could also have a RV that represents the previous dialogue act of the server, its values ranging over all possible dialogue act types.

RV	Meaning	Component
PSN	The previous dialogue act of S was a NEG-ANSWER.	Belief State
PCQ	The previous dialogue act of C was a REF-QUERY.	
CQ	C performed a REF-QUERY in the current utterance.	Dialogue Acts
CR	C performed a REQUEST in the current utterance.	
CONT	The current utterance shows a continuation pattern, e.g. in Dutch, if it starts with the word "en".	Linguistic Features
QM	the current utterance contains a question mark.	

Table 18.1: The RVs of the example network of Figure 18.2.

The given RVs are clearly interrelated, i.e., getting to know the value of one variable gives us more information on the value of other variables. This will be illustrated by considering the following dialogue passage, taken from the SCHISMA corpus (see Section 2). We are particularly interested in the dialogue act performed in utterance (3).

(1) C: Wat gebeurt er komend weekend 19 maart in de schouwburg? (What is happening in the theatre next weekend March 19?)
(2) S: Op deze datum is er geen uitvoering. (On this date no performances have been scheduled.)
(3) C: En op 18 maart? (What about March 18?)

The belief that a REF-QUERY was performed in (3), may be determined by the observation that C previously performed a REF-QUERY in (1), and that S's previous dialogue act in (2) was a NEG-ANSWER. In this context, C has decided to continue his previous dialogue act, but with a different

specification accompanying that REF-QUERY. Note that all this cannot be concluded by just analysing the linguistic features of utterance (3) only.

A probabilistic model is completely specified by a *joint probability distribution* (jpd), which assigns a number between 0 and 1 to every instantiation of the RVs. However, the number of probabilities to be assessed in order to specify the jpd increases exponentially with the number of variables. This problem can be overcome by identifying *conditional independencies* between the RVs, reducing the number of probabilities needed for specifying the jpd. For example, we may indicate that if we know the value of CQ and CR, then learning the value of PCQ gives us no information on QM: PCQ and QM are called *conditionally independent*, given CQ and CR.

These conditional independencies can be specified by means of a *Bayesian network*. A Bayesian network (Pearl, 1988) is a DAG (Directed Acyclic Graph) in which the nodes represent RVs and the arcs reflect the informational dependencies between these variables. In Figure 18.2, a Bayesian Network is depicted containing the RVs given above. It reflects a number of conditional independencies, including the one indicated above.

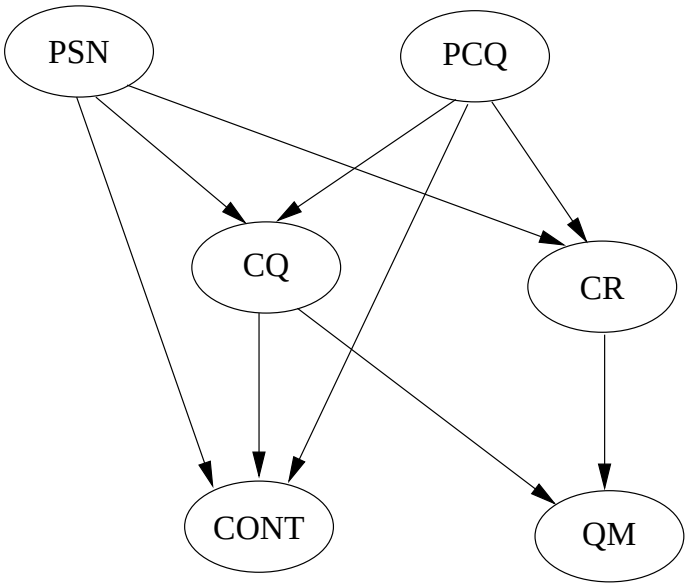


Figure 18.2: Simple Bayesian Network for utterance interpretation.

Associated with each RV is a *conditional probability distribution* (cpd) given its parents in the network. This means that for CQ a cpd is specified, given its parents PSN and PCQ. In the cpd's we should have numbers which reflect the qualitative relationships between the RVs involved. In Table 18.2, some of the chosen distributions are given. From the distribution of CQ, one can see that if PSN=*true* and PCQ=*true*, then CQ is more probably true (0.7) than false (0.3). For this example, the numbers have been assessed by using 'expert' knowledge; for more accurate and realistic models, statistical analysis on data using an annotated dialogue corpus is needed for the assessment. For learning Bayesian networks using data, see for example Heckerman (1995).

PSN	
true	0.75
false	0.25

PCQ	
true	0.8
false	0.2

CQ				
PSN	true		false	
PCQ	true	false	true	false
true	0.7	0.2	0.2	0.5
false	0.3	0.8	0.8	0.5

Table 18.2: Some of the probability distributions specifying the Bayesian network of Figure 18.2.

The RVs concerning the Belief State, PSN and PCQ, have no parents, so their cpd's are not conditional, but *prior* distributions (also given in Table 18.2). Although we have specified all prior distributions in the network as fixed, the distributions of RVs like these actually depend on the Belief State at the time-step in which the previous utterance was processed, and is therefore subject to updating. However, in order to keep our story within limits, we will not go into this dynamic extension of our model, but have based the prior distributions intuitively on the course of the dialogue passage given before. Here, the previous act by S, performed in utterance (2), was probably a NEG-ANSWER (probability 0.75) and the previous act of C, performed in utterance (1) a REF-QUERY (probability 0.8).

This network can now be used for *probabilistic inferences*: we can determine the probability that e.g. a REF-QUERY was performed, given the *partial* information that e.g. C's utterance contained a question mark, like in (3) of the dialogue passage given before. This process of determining the *posterior probability distribution* is called *belief updating*. In this process, the formula for the joint probability distribution (jpd) over all RVs in the network plays a central role. By making use of the conditional independencies implicitly given by the network structure, the jpd is given by the product of the specified cpd's:

$$P(PSN, PCQ, CQ, CR, CONT, QM) = P(PSN) \cdot P(PCQ) \cdot P(CQ|PSN, PCQ) \cdot P(CR|PSN, PCQ) \cdot P(CONT|PSN, PCQ, CQ) \cdot P(QM|CQ, CR) \tag{18.1}$$

Suppose the server gets to know that there was a question mark in the utterance. He will be interested in the updated probability that the dialogue act performed is a REF-QUERY, given this new information. This probability can be calculated by first applying the definition of conditional probability:

$$P(CQ = true|QM = true) = \frac{P(CQ = true, QM = true)}{P(QM = true)} \tag{18.2}$$

Both numerator and denominator can now be obtained from the jpd (18.1) by summing over all possible configurations of the other RVs in the network. Let $S = s$ and $T = t$ denote instantiations of the RVs in $\{PSN, PCQ, CR, CONT\}$ and $\{PSN, PCQ, CR, CONT, CQ\}$ respectively. Then we get our posterior probability from 18.3 and 18.4.

$$P(CQ = true, QM = true) = \sum_s P(CQ = true, QM = true, S = s) \tag{18.3}$$

$$P(QM = true) = \sum_t P(QM = true, T = t) \tag{18.4}$$

Finally, we show some results of probabilistic inferences in the example network. For three different cases of particular information (which is called the 'evidence' $\mathcal{E}$), we have calculated the posterior probability distribution of CQ and CR, given that information. This has been done with two different choices for the prior distributions of PSN and PCQ. From the results in Table 18.3, one can observe that with the original priors (upper row of the table), the probability that REF-QUERY was performed changes with the given evidence. Especially the case where S gets to know that the utterance shows a continuation pattern (starting with the word "en"), clearly reflects the correctness of classifying the utterance as a REF-QUERY (with probability 0.804) and not REQUEST (with probability 0.313).

In the case of 'uniform' priors (lower row of Table 18.3), i.e., the probabilities of both values *true* and $false$ are 0.5 for both PSN and PCQ, one can observe that there is much more indifference between CQ and CR than before. This illustrates how the role of beliefs concerning the course of the dialogue (as part of the Belief State) in dialogue act classification can be taken into account.

Prior distributions	$\mathcal{E}$	$P(CQ = true\|\mathcal{E})$	$P(CR = true\|\mathcal{E})$
$P(PSN) = 0.75$ and	(none)	0.515	0.298
$P(PCQ) = 0.8$	(QM=true)	0.562	0.380
	(QM=true, CONT=true)	0.804	0.313
$P(PSN) = 0.5$ and	(none)	0.400	0.413
$P(PCQ) = 0.5$	(QM=true)	0.521	0.537
	(QM=true, CONT=true)	0.624	0.485

Table 18.3: Results of Probabilistic Inferences.

4 Related Work

Other work on dialogue modelling which is based on some notion of dialogue acts includes e.g. Dynamic Interpretation Theory (DIT) and the Information State model (Poesio and Traum, 1998). In DIT (Bunt, 1995), dialogue acts are defined as functional units used by the speaker to change the context. They consist of a *semantic content* and a *communicative function*, so a dialogue act changes the context in a way that is given by the communicative function, using the semantic content as a parameter. According to the Information State model, both of the dialogue participants keep track of the *Conversational Information State* (CIS), in which *grounded conversational acts* and also *ungrounded contributions* are recorded. A CIS is characterised by a feature structure, containing embedded feature structures for both dialogue participants.

Concerning the classification of communicative actions, various research has been done. In plan-based approaches (Perrault and Allen, 1980), communicative acts (speechacts) are predicted on the basis of recognition of plans that the speaker has. Therefore, the interpretation of utterances is extended from identifying direct speechacts from linguistic features, to indirect speech acts, taking into account the course of the dialogue in terms of a speaker's plans. This rule-based approach may however lead to difficulties when dealing with uncertainty.

In other approaches, statistical methods are used to model dialogue, see for example Nagata and Morimoto (1994) and Stolcke et al. (2000). In the latter for example, a probabilistic model obtained from statistical analyses of a dialogue corpus is presented. Dialogues are modelled in a Hidden Markov Model, with states corresponding to dialogue acts and observations corresponding to utterances (in terms of word sequences, acoustic evidence and prosodic features). The transition probabilities are obtained from n-gram analysis of dialogue acts and the observation probabilities are given by local utterance-based likelihoods.

Also the use of Bayesian networks for interpreting utterances in a dialogue has been proposed before. Pulman (1996) proposes a framework for classifying communicative actions (in his approach, *conversational moves*) using Bayesian networks, that is very similar to our approach. An interesting difference with our approach, is in the structure of the Bayesian network used. While we have chosen

for arcs from nodes representing dialogue acts to nodes representing linguistic features (like the arc from CQ to QM), Pulman has chosen arc in the opposite direction. One could say that in our approach a model of the speaker's behaviour is given, which is used to derive the most probable dialogue act he performed. Pulman's network however, models the hearer, using various sources of information as 'inputs' to derive the most probable conversational move made by the speaker.

Other work on the use of Bayesian networks in dialogue systems includes research, where the emphasis is on the *user modelling* part within a specific (task-)domain (Akiba and Tanaka, 1994), in stead of the aspect of dealing with partial linguistic information in understanding communicative behaviour.

5 Conclusion

In this paper we have shown a probabilistic approach to interpreting natural language utterances in a dialogue. We have described how Bayesian networks can be used to interpret partial information from natural language analysis in terms of dialogue acts. Not only the linguistic information from utterances is taken into account, but also knowledge about the course of the dialogue and about the mental state of the speaker. Using Bayesian networks, these various sources of information can be integrated into one probabilistic model, where the most probable dialogue act (given any evidence w.r.t. the linguistic content of the utterance) can be found through belief updating (Section 3).

Current research is concerned with the construction of a Bayesian network, or perhaps various Bayesian networks, for dialogue act classification. This means that we have to find the appropriate random variables, representing all relevant aspects of identifying a dialogue act type, identify a sufficient number of conditional independencies among these variables (i.e. find the network structure), and assess the conditional probability distributions associated with the network. One way of doing that is using data in the form of an annotated dialogue corpus to train a Bayesian network. Ongoing work in this respect concerns the annotation of the SCHISMA corpus, using the dialogue act annotation scheme, mentioned in Section 2, where also relevant linguistic features need to be included in the annotation, e.g. the sentence type, indication of punctuation or the occurrence of certain keywords.

Where this paper merely concerned the classification of dialogue acts as part of the interpretation of an utterance, we are more generally interested in using of Bayesian networks for modelling a conversational agent that has to reason under uncertainty.

Bibliography

Akiba, T. and Tanaka, H. (1994). A Bayesian approach for user modeling in dialogue systems. Technical Report TR94-0018, Tokyo Institute of Technology.

Allen, J. and Core, M. (1997). Draft of DAMSL: Dialog Act Markup in Several Layers. Dagstuhl Workshop.

Austin, J. L. (1962). *How to Do Things with Words.* Harvard University Press.

Bunt, H. C. (1995). Dynamic interpretation and dialogue theory. In Taylor, M., Bouwhuis, D. G., and Neel, F., editors, *The Structure of Multimodal Dialogue*, volume 2. John Benjamins, Amsterdam.

Heckerman, D. (1995). A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research.

Nagata, M. and Morimoto, T. (1994). First steps towards statistical modeling of dialogue to predict the speech act type of the next utterance. *Speech Communication*, 15:193–203.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann.

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

Poesio, M. and Traum, D. (1998). Towards an Axiomatization of Dialogue Acts. In Hulstijn, J. and Nijholt, A., editors, *TwenDial'98: Formal Semantics and Pragmatics of Dialogue*, number 13 in TWLT.

Pulman, S. G. (1996). Conversational games, belief revision and Bayesian networks. In Jan Landsbergen, Jan Odijk, K. v. D. and van Zanten, G. V., editors, *Computational Linguistics in the Netherlands*. SRI Technical Report CRC-071.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Ess-Dykema, C. V., Ries, K., Shriberg, E., Jurafsky, D., Martin, R., and Meteer, M. (2000). Dialogue act modelling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–374.

Traum, D. R. (1999). Speech acts for dialogue agents. In Wooldridge, M. and Rao, A., editors, *Foundations of Rational Agency*, pages 169–201. Kluwer.

Paper 19

An HPSG-based representation model for illocutionary acts in crisis talk

Claudia Sassen
Universität Bielefeld, Fakultät für Linguistik und Literaturwissenschaft
Germany
claudia.sassen@uni-bielefeld.de
http://www.uni-bielefeld.de/~csassen

Abstract

The present paper addresses the extension of an HPSG formalism to the description of illocutionary acts from spontaneous speech[1]. According to Searle the syntactic and semantic features of an utterance provide essential information about its illocutionary force (see e.g. (Searle and Vanderveken, 1985)). In order to make linguistic features of this sort explicit an HPSG formalism is employed since this allows a detailed description of the syntactic structure of sentences and also their semantic treatment (see (Pollard and Sag, 1987, 1994)). However, the scope of Pollard & Sag is narrower than what a description of utterances would require. And although they extend their formalism by the context-attribute CONX (Pollard and Sag, 1994), which permits the integration of features such as pragmatic agreement and background conditions, further extensions are necessary. The formalism is applied here to a variety of *crisis talk*, i.e. an excerpt from a *cockpit voice recording* (CVR) transcript.

1 A modified formalism

Different illocutionary forces are constituted by different features. Searle's *conditions of success* are indicators which help to disambiguate the type of illocutionary force that underlies an utterance. Furthermore, the conditions of success are useful in order to make claims about whether an illocutionary act of an utterance has successfully or unsuccessfully been performed. In addition to these conditions, which are based on an extended semantics, there are conditions of syntactic relevance and those that relate to surface features of language expressions. They jointly function with the conditions of success as *illocutionary force indication devices*. *Head Driven Phrase Structure Grammar* (HPSG) supplies a formalism that allows a detailed description of the syntactic structure of sentences and also their semantic interpretation (Pollard and Sag, 1987, 1994). However, the HPSG rules and principles do not go beyond the structure of the sentence, let alone a dialogue, i.e linguistic signs produced by more than one speaker, and even though Pollard and Sag (1994) introduce the CONX (=context) attribute[2]

[1]Thanks to Dafydd Gibbon, Peter Kühnlein and John Walmsley for discussion of the manuscript and to Peter Ladkin for pointing out to me the problem of cockpit tower communication.

[2]The CONX attribute allows the integration of features such as pragmatic agreement ((Pollard and Sag, 1994, 92-95)) and the background conditions that provide linguistically relevant information about the states of affairs of an utterance, e.g. SPEAKER, ADDRESSEE and UTTERANCE-LOCATION (Pollard and Sag, 1994, 332).

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

it does not serve an adequate linguistic description. What is more, natural language expressions can hardly be modeled by HSPG, since it is oriented towards the ideal speaker/hearer according to the Chomskyan paradigm and not tuned to imperfect beings. Consequently, the HPSG formalism needs to be extended. At least with regard to the syntactic features they call *parts of speech* Pollard & Sag point out that their list of sorts *is not intended to be exhaustive* and that they *leave open the question of the precise inventory* (Pollard and Sag, 1994, 22). In this way they would allow an extension of their formalism at least on a syntactic level. In this paper the HPSG methodology is used conservatively with regard to current usage (Pollard and Sag, 1987) and the HPSG-application non-conservatively, i.e. the formalism is employed in large parts freed from its original interpretation. Hence, the approach is HPSG-*based* and deviates from the traditional conventions in the following respects:

1. HPSG rules are applied to tokens of spontaneous speech instead of the traditionally analysed abstract sentences (see (Searle and Vanderveken, 1985)). Thus, the HPSG-based structure proposes a solution in what way Searle's *f(p)*, i.e. natural language expressions (see section 3), can be translated into a logical form *F(P)*. This is something that Searle & Vanderveken consider, but do not pursue since they limit their research to idealised data. The ensuing context-dependence treated in the current paper is captured by an extended set of types of HPSG-entities. Different substructures are added to the semantic attribute which have been adopted from Searle's conditions of success (Searle and Vanderveken, 1985).

   In the resulting HPSG-based entry *illocutionary force* and *proposition* jointly function as a semantic attribute of a complex sign which has a four-dimensional structure ⟨SYN, DTRS, SURF, SEM⟩ with two *compositional* and two *interpretative* dimensions. The compositional dimensions refer to the syntactic features of the sign such as its distribution in the immediate linguistic context (SYN) and to the internal components of which it is constituted (DTRS). The interpretative dimensions stand for its surface representation (SURF) including aspects of orthography and word order (also its phonetic and perhaps gestural realisation) and for its semantic (SEM) features that include contextual properties (cf. (Gibbon and Sassen, 1997)).

2. The *head-feature principle*, which is conventionally applied to phrasal syntax, is extended to the semantic attribute of the type *illocutionary act* which has been motivated by Searle, who argues that propositions are bound to the performance of illocutionary acts:

   > In the performance of an act of the form F(P) the illocutionary point is distinct from the propositional content, but it is achieved only as part of a total speechact in which the propositional content is expressed with the illocutionary point. We will say therefore that the illocutionary point is achieved on the propositional content. (Searle and Vanderveken, 1985, 15)

   On the evidence of some illocutions that may occur without a proposition (e.g. *Hooray for the Raiders!*) and since the proposition is derived as an abstract entity from the illocution, the illocutionary component is interpreted as a head in relation to the propositional component. The principle will be useful in order to model propositions that are distributed over contributions, possibly of different speakers (see (Rieser and Skuplik, 2000)). Head and sister are in a dependency relation.

The HPSG-based formalism further elaborates on the idea that the conditions of success and other parameters can be construed as input to a rule whose output makes statements about the success or failure of the performance of a speechact. The following sections will show in which way the formalism is developed and applied to a token from a *crisis talk* scenario.

2 Crisis talk and application of the formalism

The formalism developed is applied to a token of *crisis talk*, i.e. a type of dialogue that occurs in difficult and serious situations, which require quick decision and unorthodox strategies in order to be

solved. Unlike classical spoken language scenarios, such as service encounters or instances of small talk, crisis talk is non-Gricean, i.e. it violates most of Grice's maxims, it is usually emotional and has a high frequeny of uptake securing techniques and within-turn repetitions. Crisis talk is typically at play in dialogues like negotiations with criminals, summit meetings of politicians or in private conflicts.

In the context of this paper crisis talk relates to an instance from the scenario of aviation accidents. The token selected is the utterance *disconnect the autopilot*, which has a directive illocutionary force and lacks an explicit performative verb. It is orthographically reduced to its corresponding sentence. The token comes from a cockpit voice recording transcript of an air disaster that took place at Puerto Plata, Dominican Republic in 1996 (see (Aviation Safety Network, 2000)). The transcript documents the crew's communication before their airplane crashes into the sea. Here is an extract from the transcript where the token under linguistic discussion is marked by an arrow:

(1) Birgen Air B757 Accident Intra-Cockpit Communication 6 Feb, 1996

HCP085:	0346:25	disconnect the autopilot,	⟵
		is autopilot disconnected?	
HCO086:	0346:25	already disconnected, disconnected sir	
HFE087:	0346:31	...	

Topic of the conversation between Captain (HCP) and Copilot (HCO) is the autopilot. HCP commands HCO to disconnect it. In the same turn he asks whether his command has been completed while HCO confirms this overlapping with parts of HCP's turn. The extract includes a time code whose scale is specific to the cockpit voice recorder. The numeric extensions after the speaker codes mark turn numbers.

3 Conditions and Rules and Their Relation to the HPSG-based model

3.1 Conditions

Conditions in the given analysis are the conditions of success and the context of utterance which form the illocutionary force indication devices. They are integrated into the HPSG-based feature-structure as displayed in figure 19.1, cf. p. 227. Here the conditions of success formulate the parameters necessary for the successful and non-defective performance of a speech act from which the rules are generated. They are necessary in order to be able to unequivocally identify the type of illocutionary force that is expressed by a particular utterance. The conditions of success form a septuple of elements, which is non-arbitrary. It consists of the *illocutionary point, mode of achievement of the illocutionary point, degree of strength of this illocutionary point, propositional content conditions, preparatory conditions, sincerityconditions* and the *degree of strength of the sincerity conditions.* Unlike the other elements the *input-output*-condition does not figure in (Searle and Vanderveken, 1985), but has been deliberately taken over from Searle's earlier work (Searle, 1969, 1979). The input-output-condition pertains to the uptake relation of the communicative channel between speaker and hearer (see also (Austin, 1962)), while in the latter work Searle puts most emphasis on the speaker.

3.2 Rules

Rules are derived from the conditions which are integrated into the HPSG-based feature-structure (see figure 19.1, p. 227). Unlike the conditions, the rules are stated externally to the HPSG-based feature-structure. In order to validate the attribute-value matrix the argument-slots of the rules are filled with the parameters of the feature-structure. The totality of the rules' output pertains to the illocutionary force of the token at issue. The combination of conditions and rules result in a structural description of the utterance.

The idea to identify the illocutionary force of an utterance and to determine its success or failure in the performance of a particular speechact through a rule is expressed by the definition below. It decides about the appropriate logical form of utterances in context. In other words, it assigns each utterance in context its relevant logical form.

(19.1) $$\boxed{R(\langle i, f(p)\rangle, \langle F(P)\rangle) = 1}$$

Description of the rule definition:

- The rule R is constituted by one or more elements of a *context of utterance* i, the natural language expression (or token) *f(p)*, and its formal description *F(P)*, i.e. the illocutionary force indication devices which include the conditions of success of every type of illocutionary force. F and f stand for the illocutionary force indication devices and P and p for the propositional content of an utterance.
- *Context of utterance* refers to a set of contingent features here applying to the contextual features of *f(p)*, i.e. speaker, hearer, time, location and framing utterances.
- *i, f(p)* and *F(P)* mark the input of the rule. If *f(p)* together with its *i* matches a particular set of templates (rules generated from *F(P)*) the output applies to the successful performance of a speechact, hence the value *1*. If at least one component of *i* or *f(p)* does not match the templates the output of some other value indicates failure in the performance of a speech act.

The rules which are derived from the conditions are listed below for the illocutionary force of the type *directive*. For spatial reasons the present paper will not deal with the other types of force (*assertives, declaratives, commissives* and *expressives*), but will limit itself to brief rule definitions.

Semantic Rules

The semantic rules are derived from Searle's conditions of success. Since the inventory of Searle's and Vanderveken's propositional logic is too limited, some signs had to be added in order to formulate the rules for the attribute value matrix (see figure 19.1, p. 227).

Illocutionary Point Rule A speaker a_i succeeds in achieving the directive illocutionary point (Π_3) on a proposition P in a context i (for short: $i\Pi_3 P$, where the index marks the directive) iff in that context in an utterance he makes an attempt to get the hearer b_i to carry out the future course of action represented by P (Searle and Vanderveken, 1985, 39). The second part of the rule can be re-written for this context as an action/attempt (A) by the speaker (a_i) to elicit (elicitation = E) an action from the hearer (b_i), hence

(19.2) $$\boxed{\mathrm{i}\Pi_3\mathrm{P} \text{ iff } \mathrm{A}(\mathrm{a}_i)\ \mathrm{E}(\mathrm{b}_i,\ \mathrm{P})}$$

Mode of achievement rule A speaker a_i in the context i achieves the directive illocutionary point on P by invoking his position of authority over the hearer b_i. Hence

(19.3) $$\boxed{\mathrm{mode}(||command||)(\mathrm{i},\ \mathrm{P}) = 1}$$

Since the mode of achievement of a command restricts the conditions of achievement to its illocutionary point it is a *special mode of achievement* (Searle and Vanderveken, 1985, 40).

Degree of strength of the illocutionary point A speaker a_i in the context i achieves the illocutionary point Π on the proposition P with the degree of strength k: $\boxed{i\Pi^k P}$ with $\mathrm{k} \in \mathrm{Z}$ (Searle and Vanderveken, 1985, 41). In the AVM-structure of the current paper, however, k obtains the value $||command||$, since no comparative value is part of the present discussion, hence

(19.4) $$\boxed{i\Pi_3^{||command||} P}$$

Propositional content rule Some illocutionary forces like directives place restrictions on propositional contents. Searle & Vanderveken introduce the function Θ_{fut}, which pertains to temporal relations and associates with each possible context of utterance i a set of all propositions that are future with respect to the moment of time t_i (Searle and Vanderveken, 1985, 43). From this results the temporal relation between the utterance time (t_i) and denotation time (t_{denot}). For this paper they are defined as

t_i: the time interval during which an utterance is produced. It is expressed by information from the time line of the CVR transcript.

t_{denot}: the time interval or point of time during which something that is referred to is the case.

(19.5) $$\boxed{\Theta_{fut},\ \mathrm{t}_i \prec \mathrm{t}_{denot} \Rightarrow \mathrm{Prop}_{||command||}}$$

Preparatory Rule The preparatory rule specifies for each context of utterance i and proposition P, which states of affairs the speaker a_i must presuppose to obtain in the world of the utterance w_i if he performs the illocution *F(P)* in i (Searle and Vanderveken, 1985, 43). The issuance of a command requires three rules:

1. The speaker a_i be in a position of institutional authority (Aut) over the hearer b_i: $\Sigma_{||command||}$(i, P) = [the proposition that a_i at time t_i is in a position of authority over b_i as regards $P \cup \Sigma_!$(i, P)] (Searle and Vanderveken, 1985, 201) rewritten as:

   (19.6) $$\boxed{\Sigma_{||command||}(\mathrm{i},\ \mathrm{P}) = \mathrm{Aut}(\mathrm{a}_i,\ \mathrm{b}_i,\ \mathrm{t}_i, P)}$$

2. The hearer b_i is capable (Cap) of carrying out the future course of action (A_{fut}) represented by P:

   (19.7) $$\boxed{\mathrm{Cap}(\mathrm{b}_i)\ \mathrm{A}(\mathrm{b}_i,\ \mathrm{P}) \Rightarrow \mathrm{DIR},\ \mathrm{command}}$$

3. It is not obvious, i.e. common knowledge (C), to both, speaker a_i and hearer b_i that b_i will perform the action at t_i without being commanded:

   (19.8) $$\boxed{\neg\ \mathrm{C}(a_i, b_i)\ \mathrm{A}(b_i, P) \Rightarrow \mathrm{DIR},\ \mathrm{command}}$$

Sincerity Rule The sincerity rules of an illocutionary force F are defined by specifying for each context of utterance i and proposition P which psychological states the speaker a_i expresses in the performance of *F(P)* in i. A speaker who commands a hearer to do something is sincere iff he desires (Des) him to do it (Searle and Vanderveken, 1985, 45):

(19.9) $$\boxed{\Psi_{||command||}(\mathrm{i},\ \mathrm{P}) = [\mathrm{W(P)}]}$$

Degree of strength rule of the sincerityconditions/rules Depending on the type of illocutionary force psychological states are expressed in speechacts with greater or lesser strength (η). For most illocutionary forces F, their degree of strength of illocutionary point and of sincerityconditions are identical, however, in the case of commands this may be different (see (Searle and Vanderveken, 1985, 45):).

(19.10) $$\boxed{\mathrm{degree\ (F)} > \eta}$$

Surface Rules

Word Order Rule (WOR) Searle mentions word order as another illocutionary force indication device. The word order rule focusses on the position of the verb within the utterance in question. In commands such as the token at hand it appears in first position of the utterance:

(19.11) $$\boxed{\mathrm{VF}_{utt} \Rightarrow \mathrm{DIR}}$$

4 Description of the model

This section will only comment on the most important features of the HPSG-based representation model (see figure 19.1, p. 227).

4.1 General structure

Illocutionary force and proposition jointly function as a semantic attribute of a complex sign.

The composite entry for the lemma *token*, which pertains to the *whole illocutionary act* (see (Searle and Vanderveken, 1985, 8)) consists of two parts: first, an item of type F with *head*-features, second an item of type P with *complement*-features. In the context of type *F(P)* under the SYN (syntax) attribute, the operator is interpreted as combination, e.g. A⌢B. Under the SEM (semantics) attribute, the operator is interpreted as unification, e.g. A⊔B.

4.2 Particular structures of the item of type *F* for a directive

The item of type *illocutionary act* has the attributes SYN and SEM

The SYN-attribute

Within the framework of the present token *disconnect the autopilot* the utterance does not have an explicit performative verb which would have been an essential key to its illocutionary force. For this reason the SYN-attribute of F does not have a value.

It might be argued whether the illocutionary component F can have a SYN-attribute at all. In case this view is favoured it may be a means in order to model the utterance *I order you to disconnect the autopilot*, which contains the explicit performative verb *order*. From the perspective of the illocutionary force the utterance would be treated as a directive, from the perspective of the propositional content it would be an assertion.

The SEM-attribute

The semantic attribute contains the result of the association of the output of the individual rules derived from the conditions. The result pertains to the *illocutionary force* (FORCE) of the token (Π_3 = directive) and its illocutionary point (*command*). The SEM-attribute further includes

- the I/O-attribute that marks the input-output condition and applies to the uptake between speaker and hearer. Its value is *noise*.
- the POINT attribute of the illocutionary point condition
- the MODE_{POINT}-attribute that pertains to the mode of achievement of this illocutionary point
- the $\mathrm{STRENGTH}_{POINT}$-attribute of the degree of strength of the illocutionary point
- the PREPI–PREPIII-attributes of the preparatory conditions
- the SINC-attribute of the sincerity condition
- the $\mathrm{STRENGTH}_{SINCERITY}$-attribute of the degree of strength of the sincerity condition.

Searle elaborates on an extended semantics that comprises features which could be termed as having a pragmatic nature. However, no distinction is made between these two dimensions, i.e. the formalism is not extended by a pragmatic attribute but instead by further substructures of the semantic attribute. Thus, with regard to the formal description of contextual features the following attributes and substructures are included:

- the CONX-attribute (context) that breaks down into the substructures PARTIC-attribute (participant) and DISCREL-attribute (discourse relations). The former has the attributes SPEAKER/SUPERORD and HEARER/SUBORD with the values *pilot* and *copilot*, respectively. The latter refers with its THEME-attribute to the preceding utterance (value: *emergency*) and relates with the RHEME-attribute to the current token (value: *disconnect the autopilot*). A third substructure of the CONX attribute is the SETTINGS attribute with its value *cockpit*.

Within the framework of a fine-grained differentiation of illocutionary forces it might be wise to include a PERLOC-attribute which refers to the perlocutionary effect of an utterance. This, however, goes beyond the scope of this paper and will be treated elsewhere.

In order to arrive at an analysis of the propositional content it is separated from the utterance context. Since this analysis requires surface information the syntactic analysis of the utterance's constituent structure is provided in the proposition-entry.

4.3 Particular structure of the item of type *P*

The type *P* is constituted by the attributes SYN, SEM and SURF.

The SYN-attribute

The SYN-attribute breaks down into the attributes HEAD and SUBCAT according to the syntactic *head feature principle*. They have substructures of a conventional syntactic analysis as in (Pollard and Sag, 1987).

The SEM-attribute

The SEM-attribute consists of the CONT-attribute (propositional content) and the TREL-attribute (temporal relations). The former attribute breaks down into reference (=REF) with the value *the autopilot* and predication (=PRED) *disconnect*. The TREL-attribute displays as value a formula that indicates a future act.

The SURF-attribute

The type *P* bears the surface attributes PHON (phonology), PUNC for punctuation, WORD ORDER, and ORTH for orthography, which has as value a list of all lexical components of the token.

5 Conclusion

The present paper offers an HPSG-based analysis which integrates propositional and illocutionary information from spontaneous speech. It elaborates on the idea that *illocutionary force* and *proposition* jointly function as semantic attributes of a complex sign which consists of nested feature structures and shows that the inventory of traditional HPSG does not suffice for an adequate representation model. Consequently, the formalism has been extended to include additional attributes and substructures and a rule which applies the head feature principle to the SEM attribute. The modified head-feature principle will be useful in order to model propositions that are distributed over contributions, possibly of different speakers. Rieser & Skuplik address this problem (Rieser and Skuplik, 2000) and concentrate on tokens such as

(2) a. `A: jetzt nimmst du`

b. `B: eine Schraube`

c. `A: eine orange mit einem Schlitz`

which they interpret as one turn whose propositional content *eine orange Schraube mit einem Schlitz nehmen* is spread over the contributions of speakers A and B. Each part of the proposition is dependent on the illocutionary force of the class *directive*. The force is e.g. indicated by the imperative mood of the verb *nehmen*. As displayed in the HPSG-based feature structure head and sister form a dependency relation. The example above is illustrated in figure 19.2.

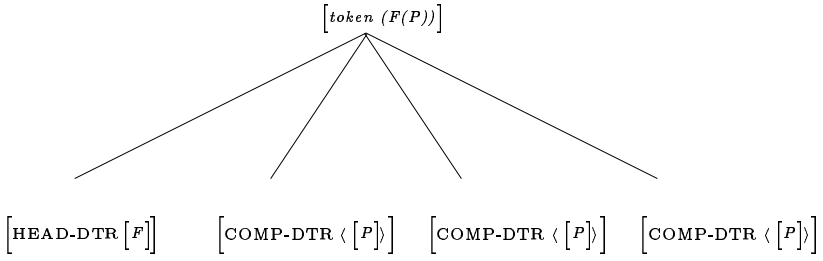


Figure 19.2: An underspecified tree structure of a proposition that is distributed over several speaker contributions.

Ongoing work on this topic is aiming at the description of a sign that consists of more than one utterance, i.e. the integration of dialogue threads in an HPSG-notation.

Bibliography

Austin, J. (1962). *How to do things with words.* London: Oxford University Press.

Gibbon, D. and Sassen, C. (1997). Prosody particle-pairs as discourse control signs. In Kokkinatis, G., Fakotakis, N., and Darmatas, E., editors, *EUROSPEECH 97, Proceedings of the 5th European conference on Speech Communication and Technology*, volume I. University of Patras, Greece.

Pollard, C. and Sag, I. (1987). *Information-based syntax and semantics, Vol.I: Fundamentals.* Stanford: Center for the Study of Language and Information.

Pollard, C. and Sag, I. (1994). *Head-Driven Phrase Structure Grammar.* Center for the Study of Language and Information Stanford. Chicago & London: The University of Chicago Press.

Rieser, H. and Skuplik, K. (2000). Multi-speaker utterances and coordination in task-oriented dialogue. In *Gothenburg Papers in Computational Linguistics 00-5.* http://www.ling.gu.se/gotalog/FinalP/rieser2.ps.

Aviation Safety Network, Ranter, H., and Lujan, F. (2000). CVR and ATC transcripts. http://aviation-safety.net/cvr/transcripts.htm.

Searle, J. (1969). *Speech acts.* Cambridge University Press.

Searle, J. (1979). *Expression and Meaning.* Cambridge: Cambridge University Press.

Searle, J. and Vanderveken, D. (1985). *Foundations of illocutionary logic.* Cambridge: Cambridge University Press.

Figure 19.1: An HPSG-based structure for the token *disconnect the autopilot*.

Paper 20

(Non-)Temporal Concepts Conveyed By *before*, *after* and *then* in Dialogue

Thora Tenbrink and Frank Schilder
Department for Informatics, Hamburg University
{tenbrink|schilder}@informatik.uni-hamburg.de

Abstract

In this paper, we analyse *before*, *after*, and *then* in dialogue contexts where they denote temporal order. Based on a corpora investigation, we present four refined options of expressed temporal precedence relations and outline how our results can be combined in one consistent formalisation that integrates various separate insights of former approaches. Especially, we point out the interrelation between (temporal) proximality and (semantic) presuppositional effects. The four different options for temporally ordering non-overlapping events are: 1. *Sometime earlier*: Pure temporal precedence; general order irrespective of the absolute times of the events or the situation context; 2. *Within a specific time frame*: Proximal time frame; relative order within a limited time span determined by the discourse or situation context independent of intermediate events on the same granularity level; 3. *Next event*: Immediate succession at a specific level of granularity derivable from the discourse context; 4. *Specific time*: The reference times are explicitly given.

1 Introduction

Studies in the field of formal semantics tend to capture only single characteristics of temporal connectives such as *before* and *after* by focusing on either temporal constraints or presuppositional effects. So far, the various approaches have not been integrated into one consistent formalisation, nor have they been extensively evaluated with data drawn from natural language corpora. In this paper, we work out both temporal and non-temporal concepts conveyed by *before*, *after*, and *then* in dialogue contexts. We show how the different connectives can be used to convey semantic interconnections between events as well as the conceptualisation of an immediate successor in time. Our outlined formalisation, which is based on diverse former approaches, captures these insights in one consistent approach.

Before and *after* have traditionally been viewed as the prototypical linguistic expressions denoting temporal order. Consequently, their usage has been studied extensively to infer information on psychological processes and cognitive, e.g. developmental, effects concerning temporal phenomena (see e.g. (Coker, 1975); (Wrobel, 1994)). However, in such studies the existence of *then* is largely ignored. Moreover, the question of which additional non-temporal phenomena *before* and *after* are capable of expressing is rarely addressed. Presuppositional effects, for example, as described by Lascarides and Oberlander (1993) or Lagerwerf (1998) are usually not considered. In

(1) Jane went to England after she won the scholarship.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

the fact that Jane won the scholarship is presupposed and still holds even when the sentence is negated. Moreover, Schilder (2001) points out that it is conceivable (but not necessarily true in all meaningful contexts) that Jane went to England *because* of the scholarship such that a causal relation is implied. Further scrutiny of the linguistic context of this sentence would decide on the latter point by providing background knowledge.

Thus, *before* and *after* are not, as previously assumed, purely markers of temporal relationships, but presuppose more complicated relationships between events. On the one hand, this calls for a closer analysis of the non-temporal implications conveyed by *before* and *after*. On the other hand, the question arises what linguistic means speakers employ in order to express a purely temporal connection. Moreover, the nature of the temporal relationship between the events involved also needs to be specified. One example by Herweg (1991) will illustrate this point:

(2) a. Peter entered the flat after Mary opened the door.

b. Peter did not enter the flat after Mary opened the door.

The temporal occurrence of the described events is dependent on the situational context indicated by the sentence. Intuitively, in (2a), Peter enters the flat within an appropriate (according to conventional standards) period of time after Mary opened the door. (2b) does not imply that Peter never again entered the flat after Mary had opened the door. The intended reading is that Peter did not react to the opening of the door by entering the flat.

Habel et al. (1993) propose the existence of a conceptual *immediate successor*. Using expressions like *immediately afterwards* or *and next*, it is possible to determine the next event in time without specifying the exact temporal relationship between the events. Two conceptions are crucial to the notion of an immediate successor in time: first, as Avrahami and Kareev (1994) point out, contextually embedded events are cognitively packaged as independent entities which may be separated from each other ("Cut Hypothesis"): "A sub-sequence of stimuli is cut out of a sequence to become a cognitive entity if it has been experienced many times in different contexts." Without this effect, the world would be experienced as a continuous stream of events and states. Second, events can be conceptualised, and linguistically described, on different levels of granularity Habel et al. (1993). Thus, one event is conceived of as the immediate successor of another event only at a specific level of granularity. If a speaker switches to a different level, further events can be introduced as occurring between exactly those events which were just described as directly succeeding each other Habel (1995).

In order to communicate, it is not necessary that speakers are informed about their listeners' subjective assessments of granularity levels. Communication is already successful if the listener is capable of interpreting the speaker's utterance such that the speaker intends a specific level of granularity. Thus, in order to analyse the impact of the concept of granularity levels on the interpretation of temporal expressions, the linguistic strategies speakers employ in order to convey the intended information need to be determined rather than defining criteria for the assessment of granularity levels.

In the remainder of this paper, we analyse natural language data of two different styles of speech in order to specify how speakers order non-overlapping events in time using *before*, *after*, and *then*. Additionally, we specify the interrelation between (temporal) proximality and (semantic) presuppositional effects. We present dialogue examples that demonstrate context-dependent temporal constraints on the time of events, and the conceptualisation of an immediate successor at a specific level of granularity. Where these phenomena occur, we find a fairly clearcut division between either a semantic interconnection between the two described events (discernible by the linguistic or nonlinguistic context or by background knowledge), or a conceptualisation of an immediate successor in time. Finally, we outline a formalisation of our insights gained through the corpora, combining former approaches Herweg (1991); Sánchez Valencia et al. (1994); Schilder (2001).

2 Corpora analysis

Spoken language corpora analysis offers the opportunity to point to specific, and real, usages of linguistic expressions which are hard to detect through intuitive reflection. By discovering actual

strategies speakers employ to convey temporal order, insights will be gained concerning the linguistic options speakers have at their disposal to communicate abstract relationships to their interaction partners. Crucial to the success of such communication are shared background knowledge and the situational context.

Data from two different corpora are analysed. The first corpus is taken from the CHILDES collection MacWhinney (2000) containing 26 20-minute sessions of interaction between the 2- to 3-year-old Trevor and his father Demetras (1989). These data were videotaped in 1985-1987, transcribed and made available for computer analysis. Trevor and his father are native speakers of American English. Most uses of the expressions under analysis stem from the father. No attempt is made here to analyse the effect of his language input on Trevor's language development. The second corpus is a sample of the Corpus of Spoken Professional American-English (CPSA; http://www.athel.com/corpdes.html), containing press conference transcripts from the White House, and a record of faculty meetings at UNC and Committee Meetings held at various locations around the country to discuss the creation of different kinds of national tests.

These two corpora were chosen in order to determine whether the identified concepts are reflected in different kinds of context, and whether they occur independent of speech style. As our research question addresses the occurrence of linguistic phenomena rather than their distribution, quantitative analyses are not carried out. In these corpora, the following usages of *before*, *after*, and *then* were found:

2.1 *Sometime earlier*

Before may be used to indicate that one event happened at some unspecified time before the other event, or that it never happened at any time before. In these cases, no non-temporal interconnections are implied, and there are no indications that the first event (if it happened at all) happened in temporal proximity to the second. Of the three expressions we analysed, only *before* seems to be capable of expressing a non-proximal temporal relationship.

Sentence-final *before*, where used in the sense of *before now*, and *before* used to indicate that something did not happen previously, may indicate this kind of general order, as in:

(3) a. FATHER: you changed his clothes and put (th)em in the corner. oh we did, yeah a couple times. you mean when we made a tape before?

b. STRICKLAND: I would like to welcome two people who haven't been with us before.

As adding *now* at the end of these sentences does not change their meaning, the reference time is the present moment. Here, the time of the prior event (or non-event) is unspecified: an expression like "ever" may be added without changing the intended meaning. However, situation knowledge may rule out this standard interpretation of sentence-final *before*:

{enumsentenceSTRICKLAND: Eunice had her hand up before.

Here, proximity of the present time and the event is involved, as it is clear from the situation context that Eunice had her hand up during the time of the meeting. In contrast to the previous examples, it is not possible to add *now* here without changing the sentence's meaning. Instead, the sentence may be interpreted as an elliptical version of a sentence such as:

(4) Eunice had her hand up before Peter had his hand up.

yielding a temporally ordered list of speaker rights based on the times of their signals during the present meeting.

2.2 *Within a specific time frame*

Wherever it is inferable from the context - and possibly reinforced by pointers such as *shortly* - that the mentioned events happen in close temporal proximity, and where it is not indicated that the events succeed each other immediately, there is some non-temporal interconnection between the events that

is either explicitly mentioned in the linguistic context, or that is derivable from situation or world knowledge.

As Knott and Dale (1994), for example, point out with regard to discourse relations in general, different systems of categorisation may be descriptively adequate for the analysis of texts. Thus, implying no claims for objectivity or universal adequacy wrt. our categorisation of discourse semantic interconnections, we point out that there are, in fact, discernible interconnections which can be roughly structured into a descriptive categorisation system. In order to capture the kinds of interconnections that may occur, we start from Lascarides' and Asher's (1993) proposal (chosen because they are specifically concerned with temporal phenomena). Two of their five kinds of discourse relations, *Elaboration* and *Background*, involve overlap between the events. As the three temporal connectives under analysis here do not allow temporal overlap (except for *then* in specific discourse contexts, see below), these two relations do not occur. Two other relations concern causality (*Explanation* and *Result*), depending on which event causes which. Only sentences involving *after* and *then* can express this relation. All three connectives, however, may express *Narration* (cases in which one event is a consequence of, but not strictly speaking caused by, the other event). Where this discourse relation is found, we specify the clauses with respect to their informational status, i.e., whether their content is new to the hearer or given in the dialogue context. Moreover, we focus on whether there is some kind of dependency between the events, or whether they are unrelated.[1]

Explanation & Result: Causality. As pointed out by Schilder (2001), causal relations between the clauses may not occur with *before*, but they may with *after*. Our corpora investigation confirms this claim and shows that *then* may also be used to convey causality. In:

(5) MANDEL: And we can make sure that it gets distributed in a broad range of communities. Then, we get a complete diversity of responses in the review process.

the later event is clearly caused by the prior one. The sentence may not be modified to include *before* instead:

(6) #We can make sure that it gets distributed in a broad range of communities before we get a complete diversity of responses in the review process.

Narration: 1. Insertion. Where future actions are expressed, speakers may insert one event, or a time span, in between the present moment and another event, using *before* or *then*. In that case, the event described by the temporal clause is presupposed, and the inserted one is new to the hearer leading to an up-date of the common ground that is shared by the speakers, as in:

(7) STRICKLAND: Just one thing going back to the issue of time before you speak, Eunice.

A special case of this are *rules*: statements which concern not only the event currently under focus, but express a general regularity, as in:

(8) CHILD: now can we knock it down? FATHER: you're supposed to ask before you knock it down.

Thus, this kind of *Narration* relation is distinct in that the consequence is known beforehand, while the event which it is the consequence of is new to the hearer.

Narration: 2. Regulation. Speakers use *before* or *after* to regulate the order of future actions. In this case, both events are known from the discourse context, and the utterance's new information concerns the order itself, as in:

[1] Although it is also possible to contrast events with each other in a discourse involving the three connectives under analysis, temporal connectives alone do not seem to be capable of expressing this relationship, but need further expressions to point it out, as in *but then*. Therefore, such cases are left out of the analysis.

(9) FATHER: can I drink it now?
CHILD: no. after dinner.
FATHER: after dinner. okay.

Narration: 3. Dependency. The later event (E2) may be based (non-causally) on the earlier one (E1). This can occur with all tenses, and all three expressions can be used for it. In these cases, E1 can be viewed as a precondition of E2, in the sense that the second event would be pointless or could not even occur (see Schilder (2001)) without the earlier one (without, of course, being caused by it). This kind of discourse relation is not, in our view, related to specific kinds of information status.

(10) a. VOICE: So are you saying that there was a meeting; it was for the White House to inform them of the announcement [E2] after it was made [E1] and to brief them in advance, not to tell them please don't go out and [...]

b. JACKSON: [...] we even send out some surveys for courses that were taught the previous semester [E1] then attempt to get the information back [E2].

Thus, our corpora analysis of discourse segments where two events are represented as occurring in close temporal proximity, but not necessarily directly after one another, reveals that if there is no *causal* relationship, there must be a different discourse semantic interconnection between the events. One such possibility is non-causal dependency; in other cases, the interconnection is specified by the information status of the hearer. In order to gain a meaningful interpretation of the discourse segment, the hearer is required to derive the intended interconnection between the events.

2.3 *Next event*

Often, speakers conceptualise events as one following immediately after another at one specific level of granularity Habel et al. (1993). Our corpora analysis revealed that, in such cases, no further semantic interconnections are implied or required. Discourse contexts that involve a temporal ordering of temporally close events therefore allow for either immediate succession or semantic interconnection.[2]

Where immediate succession is involved, no further events on the same level of granularity (= no comparable events) may happen between the two mentioned events. We propose to test this by adding a clause expressing one such event using either *or rather*, or *more exactly*. *Or rather* establishes a contrast to the previous clause, indicating that, originally, immediate succession of the two events was intended. *More exactly*, however, indicates a specification of the preceding representation rather than a contrastive relation. Thus, the events are intended to be understood as being temporally proximate but not necessarily immediate successors. The following example illustrates a case of temporal proximity without implying immediate succession:

(11) a. FATHER: so after dinner I get pie?
CHILD: yeah.
FATHER: what do I get now?

b. – after dinner I get pie - *more exactly*, I get a cup of coffee, and then I get pie

c. ?after dinner I get pie - *or rather*, I get a cup of coffee, and then I get pie

In this example - a case of *regulation* - the father does not intend to find out what he will do right after dinner; instead, he inquires whether he gets pie (E2) after having finished the regular meal (E1), contrasting the time of E1 with *now* in the following conversational turn. In contrast, another example indicates that *after* may in certain contexts also be used to express immediate succession:

[2] In Lascarides' and Asher's terms, clause sequences involving immediate succession would probably be classified as cases of *Narration* as well, to point out that the narrative is carried forward. Such clauses do not, however, exhibit the additional semantic characteristics described in section 2.2.

(12) a. CHILD: lookit dis air plane.
FATHER: where's he going?
CHILD: um right dere.
FATHER: oh. where's he goin after that?
CHILD: um it's not goin to duh zoo.

b. ?After that, he's going to the zoo. *More exactly*, he's going to the station, and then to the zoo.

c. – After that, he's going to the zoo. *Or rather*, he's going to the station, and then to the zoo.

The appropriateness of *or rather* indicates that the succession of events is interrupted in order to insert the further event. The intermediate event is thus interpreted as a contrast to the previous representation. However, if one switches to a different level of granularity, the contrast disappears:

(13) – After that, he's going to the zoo. *More exactly*, he travels through the air, and then he's going to the zoo.

Our corpora contained only very few examples of immediate succession involving *before* and *after*. Preferably, *then* is used to express the concept that one event immediately succeeds another. Although *before* and *after* are also capable of expressing this concept, they are only seldom used for it, and if so, the linguistic context often indicates that the usual interpretation involving a semantic interconnection is overruled, as in:

(14) FATHER: then I'll get you a new piece later.
CHILD: yeah.
FATHER: (o)kay.
CHILD: no get new piece after dis one.

Here, *after* is contrasted with *later*, thus it is clear that the child wants to get the piece directly after the first one in contrast to some later time.

Immediate succession may occur on either the agent, the patient, or the narrative level:

(15) MYERS: It's more of a periodic view, and then he will be briefed privately out of the view of your eyes and ears. Then he will have lunch with the Vice President.

In this example, the steps the agent (the President) will take are described one after the other. Obviously, this indicates a listing of events ordered by their temporal occurrence implying no semantic interconnections, focusing on the agent. The following example, on the other hand, focuses on the events that happened to the patient (i.e., an object found by the father):

(16) FATHER: I found it on the windshield of a car.
CHILD: yeah.
FATHER: then I uh pulled it out and I showed it to Mommy. and then I brought it home .

This example illustrates that it is possible to represent a succession of events with a focus on the patient even if the patient is not in syntactic subject position, as the example is narrated by the agent in active voice.

Finally, narrations may involve a focus on neither agents nor patients but rather on "what happens next", i.e. the narrative level itself, as in:

(17) CHILD: they have to go upstairs.
FATHER: there they come.
CHILD: weal high up stair.
FATHER: way up.
CHILD: and then they. it's way up. and then this guy jumps in.
FATHER: he's in the racquet club pool.

2.4 *Specific time*

Neither immediate succession nor semantic interconnections are implied if reference times are explicitly given, as in:

(18) VOICE: It's now a month after the IEA suggested it had to know within weeks.

With *then*, explicit reference times (ERTs) evoke simultaneity where *then* occurs in sentence-final position (cf. Glasbey (1993)), as in:

(19) BAYNE: (...)or do we have to wait until 2056. (...) BROWN: Great. And I'll be dead by then. BAYNE: And you wouldn't be counted then.

Glasbey distinguishes between sentence-final occurrences of *then* together with an ERT, in which cases *then* is used anaphorically, and other sentence-final occurrences which convey a *Background* or *Elaboration* relation between the events. In our corpora, sentence-final *then* occurs only with ERTs or in questions, which Glasbey does not deal with. In questions, we identify the immediate successor relation, as in:

(20) FATHER: an you jumped on the bed. [...] what'd you do then? CHILD: go go sweep. FATHER: no! you didn't go to sleep!

However, the default interpretation of immediate succession for sentence-initial or mid-sentence *then* can be overruled by the existence of an ERT, as in:

(21) MYERS: [...] Saturday, he will give his radio address live at 10:06 a.m. and then leave that night at roughly 11:00 p.m., maybe a little bit before, for Brussels.

3 Formal analysis

In this section, we sketch a formal analysis of *before*, *after* and *then* by combining insights from former approaches. Starting with a standard semantics for *before* and *after* by Sánchez Valencia et al. (1994) and Glasbey's analysis for *then* Glasbey (1993), two further accounts are added. First, a proximal time frame is specified following Herweg (1991) to cover the data described in section 2.2. Second, the presuppositional features of the connectors *before* and *after* are described by further specifying the account by Schilder (2001). The semantics for *then* does not require the latter constraints. Instead, the core meaning of temporal *then* is the immediate successor relation. Finally, the influence of the reference time explicitly mentioned by the sentence is captured by the formalisation.

3.1 Standard semantics for *before, after* and *then*

Traditionally, the connectors *before, after* and *then* are viewed as only expressing a temporal precedence relation between two described situations (e.g. Landman (1991)). As a starting point, we adopt the formalisation presented by Sánchez Valencia et al. (1994) within a Davidsonian framework employing event variables e_p and e_q that hold for the respective propositions p and q Davidson (1967).[3]

$$pAq : \exists e_p \exists e_q [\tau(e_q) < \tau(e_p) \wedge p(e_p) \wedge q(e_q)]$$

$$pBq : \exists e_p [p(e_p) \wedge \forall e_q [q(e_q) \rightarrow \tau(e_p) < \tau(e_q)]]$$

Note that the formalisation for *before* is non-veridical for the temporal clause that describes the proposition q, because the situation described by this clause does not have to occur. Sánchez Valencia et al. (1994) and others note that *before* allows such non-veridical readings in certain contexts.

(22) Mary left the party before she punched anyone.

[3] τ is the run time function for events that gives back a time interval.

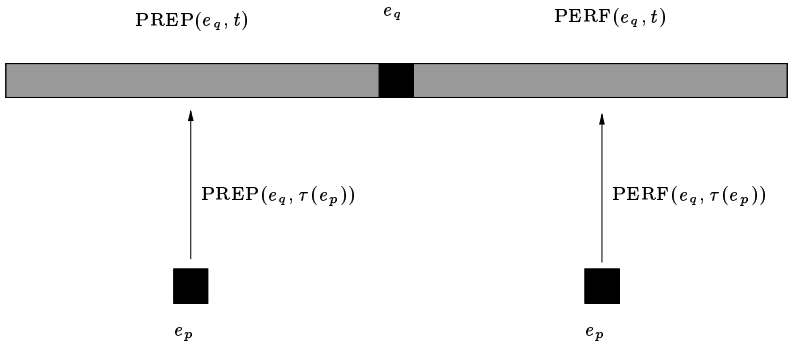


Figure 20.1: The proximal time span for a given event e_q contains a PREP and a PERF phase

The semantics for *then* normally only requires a temporal precedence relation holding between the described situations:

$$pTq : \exists e_p \exists e_q [\tau(e_p) < \tau(e_q) \wedge p(e_p) \wedge q(e_q)]$$

The obvious shortcoming of these semantics for temporal connectives is the lacking of any temporal restriction regarding the temporal precedence relations. Only the sequences listed in section 2.1 are captured by this formal description. Sánchez Valencia et al. (1994) themselves already note that the precedence relation ($<$) is too weak for a correct semantics of *before* and *after*, since the situation e_p occurs within a contextually dependent time frame before or after the second situation e_q.

3.2 Proximality and presupposition

The phenomenon of a proximal time frame is best studied by Herweg (1991). He defines *proximality* as a relation between events and times.[4] With respect to the semantics of *before* and *after*, this contextually dependent time frame ensures that the situation e_p is located close to the situation e_q (cf. (8)). More specifically, the predicate PREP(e, t) holds for a limited time period t before the event e, whereas PERF(e, t) holds for a time period t after the event e. We incorporate the notion of *proximality* into the semantics given earlier as follows (cf. Figure 20.1):

$$pAq : \exists e_p \exists e_q [p(e_p) \wedge q(e_q) \wedge \text{PERF}(e_q, \tau(e_p))]$$

$$pBq : \exists e_p [p(e_p) \wedge \forall e_q [q(e_q) \rightarrow [\text{PREP}(e_q, \tau(e_p))]]]$$

The improved definitions for *before* and *after* cover all the cases where e_p is temporally close to the situation e_q. Sequences containing a sentence-final *before* and a negated proposition in the main clause, as in (3b), are not excluded by this more restrictive semantics. This elliptic construction has to be interpreted as *before now*. Note that the negation of an event e_p is interpreted as a state, where the situation e_p does not occur. This state coincides with the proximal time span for *now* ('n') (i.e. PREP(n, t)), which is the set of all time points preceding n.

In all other cases, the situation e_q described by the temporal clause delimits the time t for PROX(e_q, t). However, it is not clear from the definition of PROX how this proximal time interval is to be derived. In order to flesh out this definition we demand discourse semantic relations connecting main and temporal clause. These discourse relations are in fact indispensable for a correct semantics of temporal connectives, as indicated by work on the presuppositional effects of discourse connectives Lascarides and Oberlander (1993); Schilder (2001). Our corpora study supports this claim by providing evidence for several different discourse semantic connections that link the clauses of temporal sentences. The presuppositional characteristics of the temporal connectives *before* and *after* can be formalised within van der Sandt's framework van der Sandt (1992), as shown in Schilder (2001).

[4]See (Herweg, 1991, p. 64) for axioms characterising this relation.

In the following, we will briefly outline how this presuppositional approach can be incorporated into the semantic formalisation given earlier. Generally speaking, a temporal clause presupposes a situation e_q that has to be (locally or globally) accommodated within the context via a linking relation.[5] We assume that the linking relations are derived similarly to a discourse relation connecting discourse segments. Moreover, section 2.2 lists the discourse relations we found in the data investigated. While the relations *Elaboration* and *Background* are blocked, the relation *Narration* is derivable for all temporal connectives. Additionally, only *after* and *then* allow the relations *Explanation* and *Result*, whereas *before* is not capable of expressing a causal relationship between the described situations (cf. (6)). Moreover, we also found that a more fine-grained set of discourse relations was necessary, in particular with respect to the usages of the temporal connectors in dialogue. There are three relations that can be subsumed by *Narration*.[6]

Insertion. The speaker wants to add another situation e_q to the common ground shared by all participants of the conversation while the situation e_p is in the attentional focus of the hearer(s). The special case of a *rule* that is expressed by a temporal clause is triggered by the presence of a modal verb such as *supposed to* or *must*.

Regulation. In contrast to a narrative where the predominant time is the past tense, dialogues contain quite often planning and scheduling utterances where the temporal sequence of situation has to be determined. Here, the formal description needs to specify an intention for scheduling events for speaker and hearer.

Dependency. A stronger connection than a simple temporal sequence is expressed by this discourse relation. This relation allows the inference for pAq that if situation e_q had not been then situation e_p could not have occurred. This inference holds for pBq, respectively.

3.3 Immediate successor

The semantics of sentence-initial *then* is further specified as follows. First, the constraints of the proximal time span apply to *then* as they do to *before* and *after* (cf. Ehrich (1992)). Second, *then* expresses an immediate successor relation NEXT[7] wrt. a given granularity level. In contrast to *before* and *after* this relation suffices for the usage of *then*.[8]

$$pTq : \exists e_p \exists e_q [p(e_p) \wedge q(e_q) \wedge \mathrm{PREP}(e_q, \tau(e_p)) \wedge \mathrm{NEXT}(e_p, e_q)]$$

The semantics for sentence-final *then* as presented by Glasbey (1993) can easily be incorporated into van der Sandt's modification of DRT. The existence of an ERT and the derivation of an *Elaboration* or *Background* relation would be presupposed by the usage of a sentence-final *then*.

3.4 Reference time

Specifying a reference time by the temporal clause binds all presuppositions that are triggered by the connective. Only *Narration* as a default is derived. The reference time also overwrites the NEXT constraint for *then*, as in example (21).

[5] In Schilder (2001) relations that hold between the clauses of the temporal sentence are called sentential (linking) relations in contrast to discourse linking relations. The discourse linking relations are required for the placement of a temporal sentence in a wider discourse context and hence these relations are out of the scope of the present paper.

[6] Note that *Narration*(e_1, e_2) is derived as a default for a narrative sequence with the only constraint that e_1 and e_2 have to share a common topic. The informal description of *Insertion*, *Regulation* and *Dependency* can be seen as a first attempt to further specify how a common topic could be derived.

[7] We adopt the formal definition for NEXT that Herweg (1991) gives wrt. *sobald* ('as soon as').

[8] *Then* may be used in a context that allows one to infer such a link, but it is not a necessary requirement for the semantics of this temporal connector.

4 Conclusions

Between our two corpora, we found a fairly balanced distribution of the differing usages of the three temporal connectives we analysed. Speakers addressing both professional or less skilled speakers such as children use *before*, *after*, and *then* to convey differing temporal relationships between events. Most often, the events described are either in close proximity and semantically linked, or they occur immediately after one another. A significant preference for *then* to express immediate succession indicates its suitability for packaging temporal events as independent and separable entities succeeding each other. *Before* and *after* are used to express a more general (usually non-immediate) temporal relationship between events that are in some non-temporal way interconnected semantically. These findings are reflected by a formalisation that combines other accounts for the semantics of temporal connectives. We integrate the semantic impact of discourse relations together with the notion of temporal proximality into the standard semantics for *before*, *after* and *then* and outline how more presuppositional information can be integrated into the formalisation.

Bibliography

Avrahami, J. and Kareev, Y. (1994). The emergence of events. *Cognition*, 53:239–261.

Coker, P. (1975). On the acquisition of temporal terms: before and after. *Papers and Reports on Child Language Development*, 10:166–177. Committee on Linguistics, Stanford University.

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.

Demetras, M. (1989). Working parents conversational responses to their two-year-old sons. Technical report, University of Arizona.

Ehrich, V. (1992). Wann ist *jetzt* – Anmerkungen zum adverbialen Zeitlexikon des Deutschen. *Kognitionswissenschaft*, 2(3/4):117–135.

Glasbey, S. (1993). Distinguishing between events and times: some evidence from the semantics from *then*. *Natural Language Semantics*, 1:285–312.

Habel, C. (1995). Representing space and time: Discrete, dense or continuous? - is that the question? In Eschenbach, C. and Heydrich, W., editors, *Parts and Wholes. Integrity and Granularity*, pages 97–107. Graduiertenkolleg Kognitionswissenschaft, Hamburg. Report Nr. 49.

Habel, C., Herweg, M., and Pribbenow, S. (1993). Wissen über Raum und Zeit. In Görz, G., editor, *Einführung in die künstliche Intelligenz*, chapter 1.4, pages 139–204. Addison-Wesley, Bonn.

Herweg, M. (1991). Temporale Konjunktionen und Aspekt. Der sprachliche Ausdruck von Zeitrelationen zwischen Situationen. *Kognitionswissenschaft*, 2(2):51–90.

Knott, A. and Dale, R. (1994). Using linguistic phenomena to motivate a set of rhetorical relations. *Discourse Processes*, 18(1):35–62.

Lagerwerf, L. (1998). *Causal Connectives have Presuppositions.* PhD thesis, Catholic University of Brabant.

Landman, F. (1991). *Structures for Semantics.* Kluwer, Dordrecht.

Lascarides, A. and Asher, N. (1993). Temporal interpretation, discourse structure and commonsense entailment. *Linguistics and Philosophy*, 16:437–493.

Lascarides, A. and Oberlander, J. (1993). Temporal connectives in a discourse context. In *Proc. of the 7 th EACL*, pages 260–268, Dublin, Ireland. Association for Computational Linguistics.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, Hillsdale, NJ, 3 rd edition.

Sánchez Valencia, V., van der Wouden, T., and Zwarts, F. (1994). Polarity, veridicality, and temporal connectives. In Paul Dekker, M. S., editor, *Proc. of the 9 th Amsterdam Colloquium*.

Schilder, F. (2001). Presupposition triggered by temporal connectives. In Bras-Grivart, M. and Vieu, L., editors, *Semantic and Pragmatic Issues in Discourse and Dialogue: Experimenting with Current Theories*, the Elsevier Science series Current Research in the Semantics/Pragmatics Interface. Elsevier. forthcoming.

van der Sandt, R. A. (1992). Presuppositon projection as anaphora resolution. *Journal of Semantics*, 9:333–377.

Wrobel, H. (1994). *Sprachverstehen als kognitiver Prozeß: zur Rezeption komplexer Temporalsätze*. Westdeutscher Verlag, Opladen.

Part IV

Computational Perspectives

Paper 21

A Basic System for Multimodal Robot Instruction

Alois Knoll and Ingo Glöckner
Technische Fakultät, Universität Bielefeld
{knoll|ingo}@techfak.uni-bielefeld.de

1 Introduction

Due to recent developments in enabling technologies Brooks and Stein (1994) (processing power, mechatronics, walking machines, articulated vision heads and more) but also due to findings and developments in other fields (e.g. studies of the human brain, linguistics, psychology), we currently observe a shift in the view of what artificial intelligence is and how it can be put to work in operational autonomous systems. This sets the stage for putting perceptive, cognitive, communicative and manipulatory abilities together to create truly interactive robot systems.

In the past, there have been a number of attempts to teach robots by showing them a task to be performed. We note, however, that such systems for "teaching by demonstration" or skill transfer have not met with much success. We identify three main reasons for this failure: **(i)** Instruction input is monomodal, mostly through a fixed camera. This precludes the system from constructing cross-modal associations by evaluating clues from more than one modality. It also prevents the instructor from giving additional explanations in "natural" modalities, e.g. teaching movements of the hand supplemented by instructive speech statements. **(ii)** Partly due to monomodality the instruction is not in the form of a dialogue between the instructor and the robot. Dialogue-oriented interaction may be the source of additional information in "normal" instruction mode, but it becomes indispensable in the case of error conditions.

2 Human-Humanoid Interaction

In view of the aforementioned needs and deficiencies we present some of our theoretical work involving methodology form linguistics and robotics. We intend to show how future robot systems will be able to carry on dialogues in several modalities over selected domains. Endowing a humanoid robot with the ability to carry a goal-directed multimodal dialogue (vision, natural language (NL), speech, gesture, face expressions, force, ...) for performing non-trivial tasks is a demanding challenge not only from a robotics and a computer science perspective: it cannot be tackled without a deeper understanding of linguistics and human psychology Grangle and Suppes (1994). There are two conceptually different approaches to designing an architecture for incorporating NL input into a robotic system: the Front-End and the Communicator approach.

The "Front-End" Approach. The robot system receives instructions in NL that completely specify a – possibly very complex – task the instructor wants to be performed. Examples are Restaino and Meinicoff (1985); Kawamura and Iskarous (1994); Laengle et al. (1995). The input is analysed and in a subsequent separate step the necessary actions are taken. Upon completion of the task, i.e. after having carried out a script invoked by the instruction fully autonomously, the system is

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

Technische Fakultät, Universität Bielefeld
{knoll|ingo}@techfak.uni-bielefeld.de.

ready for accepting new input. This approach is ideal for systems that have to deal only with a limited set and scope of tasks, which do not vary much over time either. It much less lends itself to tasks that presuppose a high degree of flexibility during their processing. Inadvertent changes of the environment resulting from the robot's actions, which would require a re-formulation of the problem, cannot be considered. Such situations cannot be dealt with unless the whole decisionmaking competence is transferred to the robotic system. For non-trivial tasks this is currently impossible; it is questionable whether it is at all desirable to try not to make use of the instructor's sensory system and intelligence (see the discussion of rationales for the introduction of sensor-based manipulation primitives in Hirzinger et al. (1994)). Neither is it possible to make specific references to objects (and/or their attributes) that are relevant only to certain transient system states because the instructor cannot foresee all of these states (cf. the well-known AI "frame problem"). These references, however, are often indispensable for the system to work correctly, i.e. as intended by the instructor. With this approach the system cannot produce requests for specific and more detailed instructions because those, too, may arise only during the sequence of actions.

Communicator or Incremental Approach. If the nature of tasks cannot be fully predicted, it becomes inevitable to decompose them into (a set of) more elementary actions. Ideally, the actions specified are atomic in such a way that they always refer to only one step in the assembly of objects or aggregates, i.e. they refer to only one object that is to be assembled with another object or collection thereof (aggregates). The entirety of a system that transforms suitable instructions into such actions is called an *artificial communicator* (AC). It consists of cognitive NL processing, sensor subsystem and the robotic actors. From the instructor's point of view the AC should resemble a human communicator (HC) as closely as possible Moratz et al. (1995). This implies several important properties of AC behaviour: **(i)** All modules of the AC must contribute to an event-driven *incremental* behaviour: as soon as sufficient NL input information becomes available the AC must react. Response times must be on the order of human reaction delays. **(ii)** One of the most difficult problems is the disambiguation of instructor's references to objects. This may require the use of sensor measurements or NL input resulting from an AC request for more detailed information. **(iii)** In order to make the system's response seem "natural", some rules of speechact theory should be observed. The sequence of actions must follow a "principle of least astonishment", i.e. the AC should take the actions that the instructor would expect it to take. Furthermore, sensor measurements (and their abstractions) that are to be communicated about must be transformed into a human comprehensible form. **(iv)** It must be possible for the instructor to communicate with the AC about both scene or object properties (e.g. object position, orientation, type) and about the AC system itself. Examples of the latter are meta-conversations about the configuration of the robot arms or about actions taken by the AC. **(v)** The instructor must have a view of the same objects in the scene as the AC's (optical) sensors. **(vi)** The AC must exhibit *robust* behaviour, i.e. all system states, even those triggered by contradictory or incomplete sensor readings as well as nonsensical NL input must lead to sensible actions being taken.

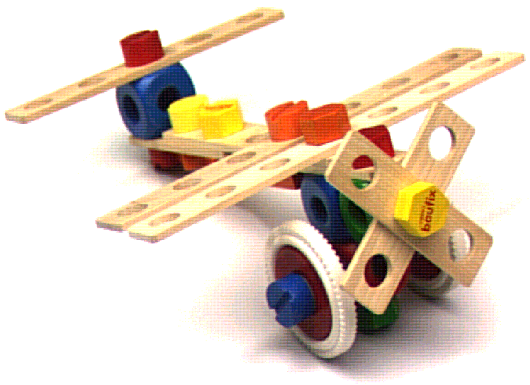


Figure 21.1: The fully assembled "aircraft".

In other words: The AC must be seamlessly *integrated* into the handling/manipulation process. More importantly, it must be *situated*, which means that the situational context (i.e. the state of the AC and its environment) of a certain NL (and further modalities) input is always considered for its interpretation. The process of interpretation, in turn, may depend on the history of utterances up to a certain point in the conversation. It may be helpful, for example, to clearly state the goal of the assembly before proceeding with a description of the atomic actions. There are, however, situations in which such a "stepwise refinement" is counterproductive, e.g. if the final goal cannot be easily described. Studies based on observations of children performing assembly tasks have proven to be useful in developing possible interpretation control flows. From an engineering perspective the two approaches can be likened to *open loop control* (Front-End Approach) and *closed loop control* (Incremental Approach) with the human instructor being part of the closed loop.

3 Scenario for Practical Evaluation

For studying situated goal-directed multimodal assembly dialogues, a prototypical scenario was chosen carefully. In this scenario a human instructor and an AC cooperate in building aggregates from elements of a toy construction set intended for children of the age of 4 years and up. The elements are made of wood (with little precision); their size is well suited to the parallel jaw grippers of our robots. The goal pursued in the sample conversations is the construction of the "aircraft" shown in fig. 21.1.

Due to several mechanical constraints its complete construction is difficult for children. As observed during some of the experiments even some adults had problems assembling the aircraft although they were provided with the exploded view of the assembly. It remains to be shown that this can be done with robots using no specialised tools. In principle, however, it may one day become possible to replace the HC with an AC and to achieve the same goals through the same dialogue.

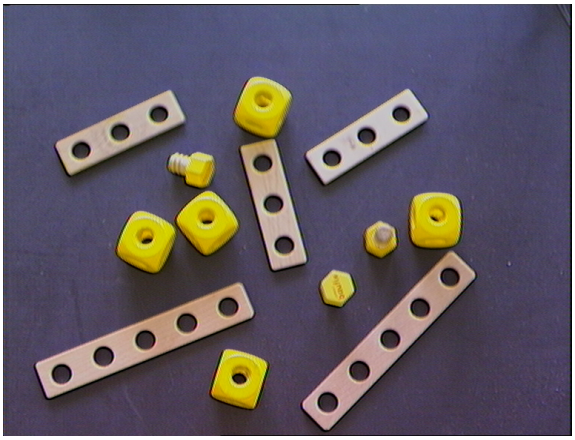


Figure 21.2: Randomly positioned construction elements: Cubes, Slats, Bolts.

To illustrate *only one individual problem* occuring from a linguistic point of view, we briefly turn to the question of *object naming* in this scenario. In an assembly dialogue between HCs each object of the scenario may be referenced using a variety of different names. Before a sensible dialogue between HC and AC may take place, however, an unambiguous binding between an object and its reference name must be established. This binding must be identical on both the HC and AC side. Since there is no common naming convention in natural language that is precise enough, a straightforward way of generating (initial) bindings is negotiation. Before entering the assembly, object names are assigned in an opening phase. The AC might, for example, point at one of the objects of fig. 21.2 (e.g. by highlighting it on a monitor) and ask the HC "What do we call / What do you want to call this object?" The HC's answer is then used as the name for the remainder of the assembly session.

While acceptable for testing purposes, such a procedure is obviously too inconvenient, time consuming and hence impractical in real-world applications involving dozens of objects. This is the reason, therefore, that the AC must possess the ability to react in a flexible manner to all (most) of the conceivable object names. It would be both difficult, cumbersome and intractable in the general case to compile all possible names for all possible objects in all possible situations. Fortunately, linguistic experiments have shown that rules may be postulated that HCs obey in assembly-type dialogues. These rules can be used to reduce the "name space" the AC must consider. Some of them follow: **(i)** Even with simple items like the cube in fig. 21.2, HCs frequently switch between names. Apart from *cube* the object ist called *die, dice or block*. **(ii)** An object may be referenced not by its generic name but by its function in the situational context: the slat is named as such but also as *wing*, the cube may be called *nut* when used as the counterpart of the bolt. **(iii)** Particularly in this scenario objects are named after their geometrical shape where frequently a projection from three into two dimensions can be observed, e.g. the cube becomes a *square*.

The AC must recognise and cope with the principles and conditions under which these transformations occur Heydrich and Rieser (1995).

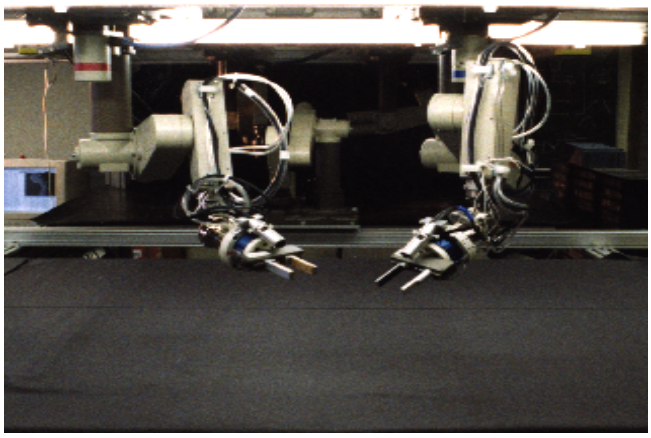


Figure 21.3: A view of the set-up for assembly.

4 Dialogue Control in Action

Even the construction of an aggregate of only a few elements may consist of a great number of elementary actions. Every assembly step resulting from an instruction comprises three distinct phases:

- The recording of the scene content using the sensory system;
- the processing of what is seen/sensed and the development of a plan for achieving a set (sub-)goal;
- the assembly of available elements with the actors.

In other words: Every assembly step is composed of perceptive, cognitive and manipulative actions. Each of these may be atomic or complex and mirror (i.e. is the consequence of) specific instructions given by the HC.

While system architectures are conceivable that implement a temporally interleaved processing of perception, cognition and action, our system currently works strictly sequentially. At the beginning of each individual assembly step the scene is analysed visually. The objects are detected and their locations are computed. A geometrical model of the scene is generated. Once this model is available, the AC requests an instruction from the HC (the instructor). These instructions can be of the type

- Assembly (Construction): "Take the red screw";

- Scene Control: "The screw is/should be located on the left hand side of the bar";
- Meta-Level-Control: "Move the elbow up a little" or "Turn this camera a little clockwise";

where the latter type of meta-level instructions very rarely occurs in human construction dialogues. The instructions are analysed linguistically and interpreted according to a hypotheses model of the scene and the history of the construction process, e.g. taking into account that a robot that has already grasped an object cannot grasp another one. As part of the cognitive phase a simple planner transforms complex into atomic actions.

Unlike standard motion sequence planners, this planner must also draw on knowledge obtained from cognitive linguistic observations. For example, an HC does not necessarily give all the instructions required for fulfilling the preconditions of a certain manipulative action. In some sense the problem is underdetermined; the planner must provide a solution within the given degrees of freedom. A simple example: The HC would not instruct the AC to grasp a screw (let alone a specific screw if more than one is available), before giving an instruction involving a screw. The reasoning about what the HC may have meant and the necessary inferences are left to the AC's planner with no help other than the cognitive knowledge mentioned above. Currently, in such a situation, our system selects the object that the robot can grasp most easily (following a *principle of economy*). In the future this will be extended in such a way as to make an attention control possible, i.e those objects are chosen that are in the *focus of the discourse*.

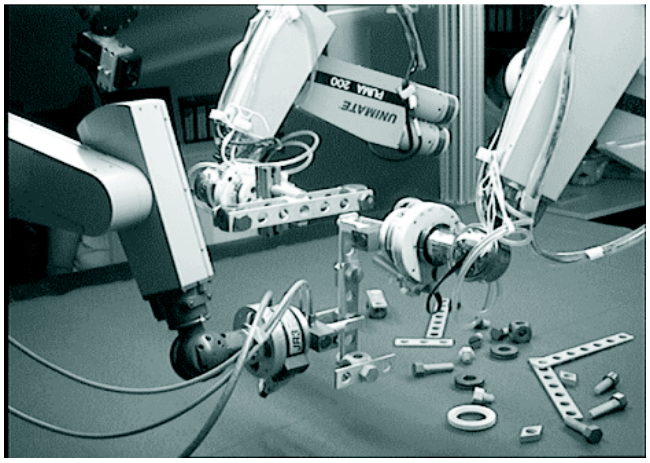


Figure 21.4: A view of the flexible assembly cell in action.

Meta-level instructions/statements are necessary for interrupting the dialogue whenever the HC wants to guide the AC to a better sequence of actions than the latter is able to find autonomously. This is in contrast to most meta-level utterances in human dialogues, which normally deal with the (format of) the dialogue itself ("What are you doing there?", "Be more polite!").

Another important application of these instructions is error handling: imagine a situation in which the robot arm has run into a singularity while following move instructions by the HC. The typical HC, of course, has no comprehension of this problem. In such a case the AC must explain the (imminent) error, and a dialogue must be conducted about possible (consensual) ways leading out of the error situation. Sometimes errors pertaining to the actuators may be anticipated. If in such a case proper use is made of the NL-production facility of the AC, errors may even be *prevented*. A further source of errors are utterances by the HC that the AC does not understand correctly. If the AC fails to comprehend the meaning of a statement, the HC must recognise the AC's problem and act accordingly. For this reason the linguistic components were so designed as to provide transparent messages whenever an error occurs. There are three classes of errors: lexical, syntactical and semantical. The reason for a lexical error is a certain (uncommon) word missing in the system's lexicon or a word having been misspelled. A syntactic error is reported when the parser cannot combine the individual words, i.e. it

cannot compile a sensible syntactical structure. A semantic error occurs if the action required by the HC cannot be taken. This normally happens when the preconditions of the action are not met (and the necessary steps cannot be inferred); in particular if the necessary objects are not present in the scene. After completion of the perception-cognition-manipulation sequence for a single assembly step, this cycle is repeated until the aggregate is finished.

4.1 Experimental Setup

To complement the AC's cognitive component a manipulation unit or *cell* was built using standard robots that cooperate and come as close as possible to the geometry of humans and their hand/arm. The similarity of the geometry often makes it easier for an HC giving instructions to the AC to image himself into the problems arising from the AC's point of view. It also enables the immediate transfer to a humanoid (torso). The cell mimics the situation of an assembly carried out by an HC sitting at a table (and possibly being given instructions). In such a setting the construction elements are placed on the table, the HC's arms/hands cooperate from above the table.

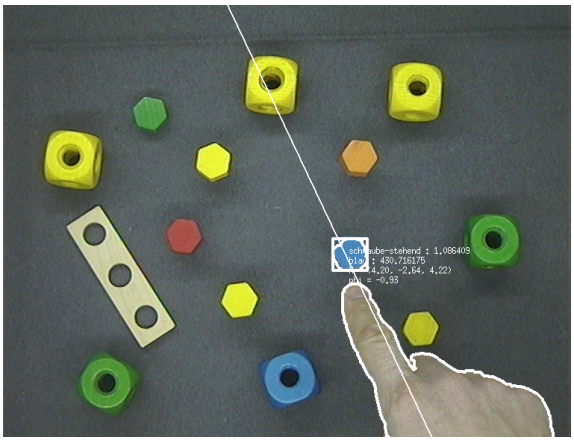


Figure 21.5: Recognition of simple gestures for identifying NL reference to a certain object

Up to now the following assembly skills have been implemented on both manipulators:

Pick-up: Most elements of the construction set can be picked up from any location on the table. The approach of the end effector's tip to the desired grasping point is controlled in real time using "self-viewing visual servoing"Meinicke and Zhang (1996).

Put-down: Elements or aggregates can be put on the table or on other objects. Prior to releasing the gripper controlled forces and torques may be applied to the object.

Peg-in-hole: Most combinations of objects that can be passed through one another can be handled. If necessary, a reflex can be activated that lets one of the robots find the center of the hole by following a spiral path under force control.

Screwing: This is by far the most complex operation available. It requires sensitive force/motion control. It involves the **(i)** approach phase in which the true thread position is determined; detection of the contact angle between screw and start of the thread; **(ii)** re-grasping of the bolt head after completing one revolution; **(iii)** application of the tightening torque. The latter is particularly difficult because the wooden screws tend to block. Special types of adaptive fuzzy controllers for force control have proven to be superior in performance to standard PID controllers Zhang et al. (1997).

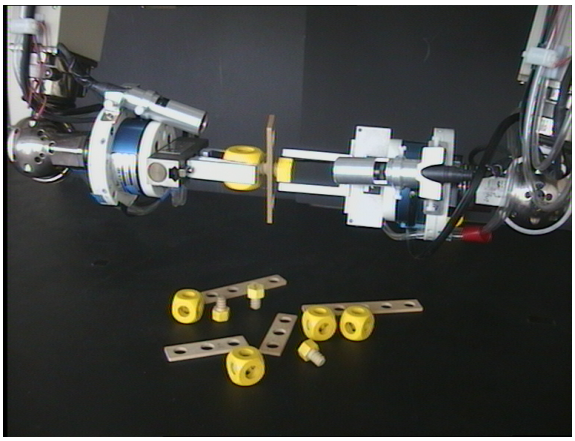


Figure 21.6: Screwing by cooperating robots.

4.2 Sample Dialogue and Results

Table 21.1 shows the beginning of a sample dialogue which was carried out in order to build the "elevator control" aggregate of the aircraft (fig. 21.1) out of three elementary objects. The objects were laid out on the table in a way similar to fig. 21.2 (i.e. there were many more objects positioned in arbitrary order on the table than necessary). The instructor had a complete image in his mind of what the assembly sequence should be. Alternatively, he could have used the assembly drawings in the construction kit's instructions and translated them into NL.

Initialisation			
(2) Constructor: *Yes, let's get started* Confirmation Initialisation	(i) No, not today (ii) One moment, please! [need to have my hands free] ...	Initialise hardware/software	
(3) **Today, we want to build a [Baufix-] aircraft [together] [, we'll start with the elevator control]!** Problem Specification		Activate domain knowledge	– Only sensible, if knowledge about object domain has been acquired – "Baufix" as opposed to "Lockheed" specifies domain (properties) – "Build..." focuses on target object, "Build together..." focuses on cooperation
(4) *All right!* Confirmation problem specification	(i) I know nothing about these aircrafts (ii) Again!	(Enter learning mode)	Teaching: "An aircraft is ..." or: Discussion about assembly plan
(5) *I'll check whether there are all necessary objects available.* Preconditions of action	I need another screw [, cube ,...].	Object recognition in scene Update scene model Problem formulation output	Precondition: Common field of view Ins/Cons Cons has knowledge about necessary objects
(6) *What would you call the [rectangular] object [in the upper right corner, to your left hand side, to my left, ...]?* Negotiation of object naming (conventions)		Learning/update naming in knowledge base (baptising act)	Precondition: – Common field of view Ins/Cons – Negotiations about naming and locations only sensible if an abstract object model is available to Cons
(7) **This is a cube!** <**points to it**> Object naming		Focus to hand Recognition of gesture	
(8) *OK, that's what we'll call it!* Accept bject naming			Gender and further properties can only be determined if there is an entry in the computer lexicon
(9) **Take a screw!** Instruction	[First,] you need a screw!	Find reference object in scene	– Object indefinite, Cons selects on its own – Alternative is an indirect instruction that needs not be exectued immediately [but before any others]
(10) *I am taking one!* Commenting action	(i) ...+ with my left hand (ii) I canot see a screw	Motor and Sensorimotor control	(i) Cons expresses selection (ii) Cons signals that it knows about the importance of the hand and indicates its orientation
(11) **Now, take the three-hole slat!** Instruction	...+ with three holes!	Infer that we need another arm	Definite object naming, works only if there is only one slat in the scene.
(12) *I see more than one [such] slats.* Cons' Identification of Contradictions or Ambiguities	(i) ...+ I would [rather] take the one on top. (ii) ...+ which {one | of the ones I see } do you want me to take?	Detect ambiguities	(i) Cons makes full use of its autonomy (ii) Cons produces two utterances: problem statement and request for information (object spec) it needs. How precise must it be?
(13) **Take this one!** <**points to one**> Instructor's resolution of contradictions	(i) Take the one I am pointing to! (ii) Take the one to {my, your} left! (iii) Take the one you want <and/or suitable gesture> (iv) Take the right one		(i) Makes sure Cons/Ins refer to the same object (ii) Needs reference frame (and info about Ins' location) (iii) E.g. nodding (iv) Location instead of colour/shape
(14) *I have got it.* Action Confirmation	And now the screwing?		Anticipation of the most probable follow-up action
(15) **Screw the bolt through the slat!** (Erroneous) Instruction	(i) Insert the screw {into | through} the slat! (ii) Put the slat on the screw! (iii) Insert the screw through the center hole!	Inference processes about the functions of the objects	Roles and object functionality (do not) match (i) Syntactic structure matches the roles of the objects (ii) Correction of the roles (iii) Instruction to avoid Cons' info request
...	...	...	...

Table 21.1: An excerpt form a sample dialogue, as partly implemented on the set-up.

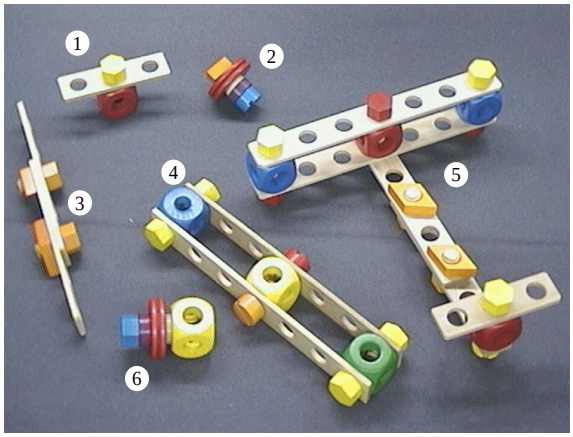


Figure 21.7: Finished aggregates that can currently be built in multimodal dialogues.

Lines input by the HC are typeset in **bold face**; AC output is in *italics*. The first AC input request in *Line 1* is output after it checked all modules of the setup are working properly. The necessary classification and subsequent steps are based on the colour image obtained from the overhead colour camera.

After the AC finding out if all objects are present and after going through an optional object naming procedure (*Lines 6...8*) the HC input in *Line 9* first triggers the action planner, which decides which object to grasp and which robot to use. Since the HC did not specify either of these parameters, both are selected according to the principle of economy. In this case, they are so chosen as to minimise robot motion. The motion planner then computes a trajectory, which is passed to the RCCL subsystem. Since there are enough bolts available, the AC issues its standard request for input once the bolt is picked up.

HC input *Line 11* results in the other robot picking up the slat. Before this may happen, however, it has to be cleared up, which slat to take (*Lines 12...14*). This involves the incorporation of the gesture recogniser (fig. 21.5). In *Line 15* the screwing is triggered, involving the peg-in-hole module mentioned above followed by the screwing module. The screwing is shown in fig. 21.6. Many uncertain parameters have to be taken into account; in particular the bolt axis is never in line with the effector's z-axis. Using the adaptive force control mentioned above, however, angles between the two axes of up to 15 degrees can be accomodated without blocking (if the thread of the bolt is not excessively worn out). For reasons of space the subsequent steps of the dialogue have to omitted here; they show how error handling and many other operations can be performed – most of which humans are not aware of when they expect machines do do "what I mean". Fig. 21.7 shows typical objaect that can – in principle – be built with the setup as developed up to now.

5 Conclusions

We introduced a scenario and a robot system that experimentally show the way humans may communicate with robot systems (and future humanoid robots) in a very natural way using all modalities. The scenario consists of only a limited set of construction elements but offers a rich variety of different tasks. It may serve equally well as the basis for construction experiments in cognitive linguistics (between HCs) and for benchmarking the perceptive, cognitive and manipulative skills of a real-world humanoid robotic system.

Bibliography

Brooks, R. and Stein, L. (1994). Building brains for bodies. *Autonomous Robots*, 1(1):7 – 25.

Grangle, C. and Suppes, P. (1994). *Language and Learning for Robots.* CSLI Publications, Stanford, Ca.

Heydrich, W. and Rieser, H. (1995). Public information and mutual error. Technical report, SFB 360, 95/11, Universität Bielefeld.

Hirzinger, G., Brunner, B., Dietrich, J., and Heindl, J. (1994). Rotex - the first remotely controlled robot in space. In *Proc. IEEE Conference on Robotics and Automation.* IEEE Comp. Soc. Press.

Kawamura, K. and Iskarous, M. (1994). Trends in service robots for the disabled and the elderly. In *Proc. IROS '94 – IEEE/RSJ/GI Int. Conf. on Intell. Robots and Systems.* IEEE Press.

Laengle, T., Lueth, T., Stopp, E., Herzog, G., and Kamstrup, G. (1995). KANTRA - a natural language interface for intelligent robots. Technical report, SFB 314 (VITRA) - Bericht Nr. 114, Universität des Saarlandes.

Meinicke, P. and Zhang, J. (1996). Calibration of a "self-viewing" eye-on-hand configuration. In *Proc. IMACS Multiconf. on Comp. Eng. in Syst. Appl.*, Lille, France.

Moratz, R., Eikmeyer, H., Hildebrandt, B., Knoll, A., Kummert, F., Rickheit, G., and Sagerer, G. (1995). Selective visual perception driven by cues from speech processing. In *Proc. EPIA 95, Workshop on Appl. of AI to Rob. and Vision Syst.*, TransTech Publications.

Restaino, P. and Meinicoff, R. (1985). The listeners: Intelligent machines with voice technology. *Robotics Age.*

Zhang, J., Collani, Y., and Knoll, A. (1997). On-line learning of B-spline fuzzy controller to acquire sensor-based assembly skills. In *Proc. IEEE Int. Conf. on Robotics and Automation*, Albuquerque.

Paper 22

VoiceXML generator of slot-filling transaction dialogue

Masahiro Araki, Tasuku Ono, Takuya Nishimoto, and Yasuhisa Niimi
araki@dj.kit.ac.jp
http://www-vox.dj.kit.ac.jp/araki/

Abstract

We demonstrate a semi-automatic dialogue system generator from the Internet information contents. This generator translates XML documents to VoiceXML, which controls a conversation between user and computer system. The translation is made by dialogue library according to three task class: slot-filling, database search, and explanation. In this paper, we describe an algorithm of generating a slot-filling transaction dialogue pattern.

1 Introduction

The World Wide Web contains a wealth of information. Many people get and send information using the Web by GUI (Graphical User Interface) based browsers. If we could access web content via voice channels, we could get and send such information anytime and anywhere.

If we can translate HTML (Hyper Text Markup Language) based interaction pattern to the dialogue pattern for such voice-based dialogue systems, we can achieve good computer and telephony integration. However, because many Web pages use HTML as only a page layouting method, it is difficult from such pages to extract structured information which can be used for constructing dialogue pattern.

On the other hand, XML (eXtensible Markup Language) is one of the most promising languages for knowledge representation and information exchange on the Internet. The strong points of an XML document are: (1) it is well structured and (2) the tag name has a semantic information about the contents. Therefore, we select XML as a contents description language.

In XML-based interactive Web pages, most interaction methods are written in XSL (eXtensible Stylesheet Language). However, we assume, in our prototype version, that dialogue structure can be extracted from the XML documents.

Concerning about dialogue description language, we selected VoiceXML (Voice eXtensible Markup Language) VoiceXMLForum (2000) as an output. VoiceXML is expected to be a new standard used when making the Internet content and information accessible via voice and phone. Mixed-initiative, cooperative dialogue can be realized by specifying the pattern of dialogue in VoiceXML.

Therefore, we set our goal as to make a converter from XML documents to VoiceXML. In interactive Web pages, one of popular pages is a slot-filling type transaction page such as reservation of train, airplane, hotel, etc. As a result, we concentrate on implementing slot filling type dialogue generator for starters.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

araki@dj.kit.ac.jp
http://www-vox.dj.kit.ac.jp/araki/.

Section 2 reports an XML to VoiceXML converter of slot-filling transaction dialogue using a dialogue library. Section 3 explains grammar adaptation method in this task. Section 4 shows our implementation. Section 5 includes a conclusion and our future projects.

2 XML-to-VoiceXML Converter

There are many interactive task domains which can be implemented by a spoken dialogue system; *e.g.* telephone shopping, on-line trade, on-line banking, etc. These task domains can be classified by considering the direction of information flow (see Table 22.1) Araki et al. (1999).

Table 22.1: Classes of task domains according to the direction of information flow.

Information flow	class	example domain
user $\rightarrow$ system	slot-filling	telephone shopping
user $\leftrightarrow$ system	database search	bibliography search
user $\leftarrow$ system	explanation	route direction

A slot-filling class is the simplest task of dialogue system. Typical domains of this class are telephone shopping, on-line banking, and on-line trade, etc. The user has a purpose and has almost all the information in order to achieve the goal. Most subdialogue for these tasks are devoted to filling the values of the slots. The back-end application of this task is a slot-filler. Each slot has a slot name and may have constraints to its value. The task is completed when all the necessary slots are filled by appropriate values.

We prepared three levels of dialogue library for slot-filling class. Figure 22.1 shows a top level library. This level provides the structure of VoiceXML documents.

```
(Opening)
(Slot-filling)
{Submitting data}
if status=="OK" then (Confirmation)
                else (Apologize)
endif
(Closing)
```

( ): subdialogue, { }: system function

Figure 22.1: Top level library of the slot-filling class.

In slot-filling class, the construction of dialogue state, which is defined by top level dialogue library, is shown in Figure 22.2.

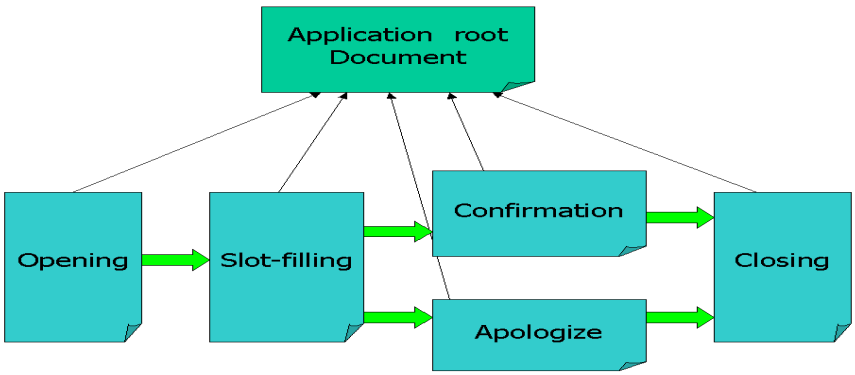


Figure 22.2: Dialogue state of slot-filling class.

Figure 22.3 shows a middle level library which provides exchange pattern of dialogue mapping on to a VoiceXML document. Bottom level library is a build-in dialogue of VoiceXML, e.g. acquiring phone number, name, etc.

```
Opening{
    Message=value-of(first-element)
    output(block, "This is"+Message+"site")
}

Slot-filling{
    foreach Element (contents)
        library-checkup(tag-of(Element), Library)
        if has_options(Element) == true
            then make-grammar(Element, Grammar)
                 output(field, Grammar)
            else if Library != NULL
                    then output(field, Library)
                    else error
                 endif
        endif
    end
}

Confirmation{
    foreach Variable (variables)
        output(YN-question, Variable)
        output(if,call(Variable))
    end
}

Closing{
    Message=value-of(first-element)
    output(block, "Thank you for visiting"+Message+"site")
    output(block, "Good bye")
}
```

Figure 22.3: Middle level library of slot-filling class.

In generating VoiceXML files, the dialogue library maps onto the overall structure of the dialogue and exchange pattern of each subdialogue, instantiating the variables according to XML contents.

For the slot-filling class task, the XML of this class consists of some slots (with or without options), a submit procedure and reset procedure.

Figure 22.4 is an example of an XML file on the hotel reservation task. The output VoiceXML files should be constructed as prompting the page name, gathering values for each slot and confirming these values. A system initiative dialogue is suitable in this class. Figure 22.5 is an output of our converter. For simplicity, opening VoiceXML and slot-filling VoiceXML are merged into one file.

```
<?xml version="1.0"?>
<!-- slot-filling -->
<element>
<title> ABC hotel reservation</title>
<contents>
  <check-in-date/>
  <check-out-date/>
  <roomtype>
    <room>single room</room>
    <room>twin room</room>
    <room>double room</room>
  </roomtype>
  <name>
    <first name/>
    <last name/>
  </name>
  <phone/>
</contents>
</element>
```

Figure 22.4: XML file of a hotel reservation system.

```
<?xml version="1.0"?>
<vxml version="1.0">
 <form>
  <block> This is ABC hotel reservation site </block>

  <field name="check-in-date" type="date">
   <prompt> What day is your check in date? </prompt>
   <help> Please say day and month. </help>
  </field>

  <field name="check-out-date" type="date">
   <prompt> What day is your check out date? </prompt>
   <help> Please say day and month. </help>
  </field>

  <field name="roomtype">
   <prompt> Which room type do you prefer? </prompt>
   <help> Please say single or twin or double. </help>
   <grammar>
      single [room] {single} | twin [room] {twin} |
      double [room] {double}
   </grammar>
  </field>

  <subdialog name="name" src="name.vxml">
   <filled>
    <assign name="firstname" expr="name.first" />
    <assign name="lastname" expr="name.last" />
   </filled>
  </subdialogue>

  <subdialog name="phone" src="phone.vxml"/>

  <block>
    <submit next="reservation"
            namelist="check-in-date check-out-date ...">
  </block>
 </form>
</vxml>
```

Figure 22.5: Output VoiceXML in a hotel reservation system.

3 Grammar Adaptation

Each field of VoiceXML file must have a grammar which defines user's input utterance. In our converter, we prepare three patterns of grammar adaptation.

1. The tag with options:

   If XML tag has its options, these options are directly consists input grammar in filed tag. In case all the options include same word (e.g. *room* in Figure 22.4), the word can be omitted.

2. The tag whose name is the same with prepared grammar:

   If tag name is the same with existing grammar name (i.e. build-in grammar or pre-defined grammar), we use field tag with type attribute (e.g. acquiring phone number in Figure 22.4), or subdialog tag with src attribute which specifies the grammar.

3. otherwise:

   First, the tag name is decomposed with word. It is because many XML tag names are compound noun. Then, search the nearest concept which has a input grammar in the thesaurus and adapt the grammar to the field tag.

4 System Descriptions

We have implemented XML-to-VoiceXML converter. We used Microsoft Visual Basic and MSXML library for implementation. Our converter generates a set of VoiceXML files from XML of slot filling class (Figure 5). The generated VoiceXML works on a Japanese VoiceXML interpreter. Our interpreter was implemented on the basis of extended specification of VoiceXML on Japanese.

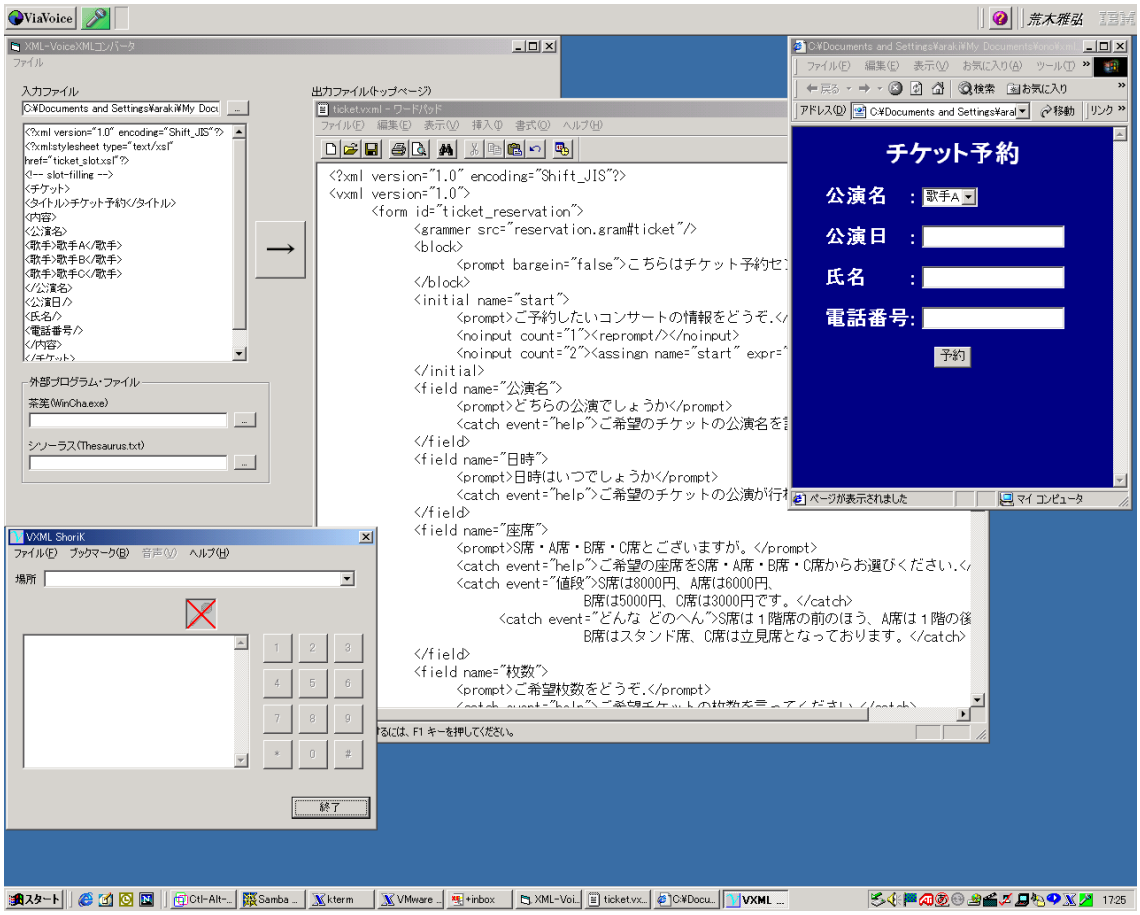


Figure 22.6: XML-to-VoiceXML Converter and Japanese VoiceXML interpreter.

5 Conclusion

This paper described the aim and current status of our project. We have implemented a XML-to-VoiceXML converter based on a dialogue library. The merit of this approach is that contents providers need not to have technical knowledge about how to construct dialogue pattern. All they have to do is to make the contents in XML and post-edit the output VoiceXML files, if necessary. On the other hand, the demerit of our approach is the limitation of XML pattern. In the present system, we subject a severe limitation on the original XML pattern.

Our future works are:

- taking away strict limitations on XML by surveying the structure of several Web sites.
- trying to convert more complex task such as database search and explanation

Bibliography

Araki, M., Komatani, K., Hirata, T., and Doshita, S. (1999). A dialogue library for task-oriented spoken dialogue systems. In *Proc. IJCAI Workshop on Kowledge and Reasoning in Practical Dialogue Systems*, pages 1–7.

VoiceXMLForum (2000). Voice extensible markup language voicexml.

Paper 23

A Java Toolkit for Dialogue Evaluation

Jean-Baptiste Berthelin and Yann Girard
jbb@limsi.fr, y.girard@kub.nl
http://www.limsi.fr/Individu/jbb
http://cwis.kub.nl/~fdl/general/people/girard.htm

Abstract

This paper presents two steps in the design of future NL dialogue systems. Step one is the constitution of three sets of corpora, corresponding to different roles of the computer in the conversation. Step two is the Java programming of observers and inspectors, which are dialogue specialists. From a global evaluation of several NL dialogues, we plan to extract principles for artificial conversation.

1 Introduction

1.1 A two-step approach

We are currently exploring human conversational skills from a computational point of view. This implies two separate developments: first, we construct a set of relevant corpora, and secondly, we create a "toolkit for dialogue evaluation", which means a set of software tools that can manage various aspects of conversation transcripts, especially the problematic ones. Simpler tools are for semi-automatic processing, namely, annotation and the establishment of dialogue trees, as described in Berthelin et al. (1999). Advanced tools, like our recent and future Java "observers" and "inspectors", correspond to more ambitious goals: they contribute to the development of a system for the automatic evaluation of natural dialogues and conversations. We deal with design issues for such a system in Berthelin et al. (2000). A more general perspective is given at `http://www.limsi.fr/Individu/jbb/research.html`.

1.2 Global dialogue processing

Initial works in the same field insisted on rules with a local span, cf Litman and Allen (1990), Pollack (1990) (both in Cohen et al. (1990)) and also Lambert and Carberry (1991).

We try, contrariwise, to process each dialogue transcript *as a whole*, dedicating different specialized agents to different aspects of the linguistic interaction. It is made possible by having many different, very specialised agents, each of them looking for a particular kind of words, or patterns, in the dialogue transcript. As output, they produce quantitative evaluations, which, in turn, are processed by so-called Inspectors, in search of global dialogue patterns.

1.3 Destination of the system

The main interest of our approach is that such a system, once it reaches a suitable level of development, will act as a guide in artificial conversation, so that typical repair strategies will apply when needed.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

jbb@limsi.fr, y.girard@kub.nl
http://www.limsi.fr/Individu/jbb
http://cwis.kub.nl/~fdl/general/people/girard.htm.

The main difficulty is that we shall first have to "teach" the system, by rewarding Inspectors when their judgement is relevant, and secondly, to create a sentence generation capacity in future versions.

2 Corpora and methodology

2.1 Three kinds of corpora

We consider three families of corpora: "natural", "Oz" and "artificial". They correspond to steps in the evolution of human conversation in presence of the computer. The machine is given a growing role within the exchange, from mere transcription medium to simulated conversationist.

2.2 Real-world conversations

We are first interested in transcripts of ordinary exchanges. They range from idle chat (see the [Conversations] website[1]) to highly specialized dialogues. We particularly examined a corpus of lunchtime exchanges, where misunderstandings occur frequently, and are quickly repaired. We also found interesting structural properties in purposeful dialogues, as is the case e.g. when users buy railway tickets over the phone. Even these purposeful interactions can be problematic, as shown in the [bookseller] website[2], where a polite and incompetent bookseller entertains a long series of exchanges with a customer, knowing all the time that he cannot meet his requirements.

Most of our examples are either in English or in French, but in some cases, language varies within one conversation, as can be daily observed at the [Boz] website[3], a good place for observing tangled interactions.

2.3 Oz dialogues

Oz dialogues are those where a human being impersonates a talking computer. They take place when software designers want to know what the user would do with an automatic system. In spite of their heuristic relevance, they are not easy to evaluate, since the "human chatterbot" in them is often placed in awkward positions (I must act as a machine with a capacity to act "almost" like a human being, sometimes one gets lost at it). Nevertheless, we did rely on them, especially in pedagogical contexts. They are especially relevant in the design of expert systems, in which rule conditions have to be derived from actual input by users.

2.4 Artificial conversations and dialogues

Artificial conversations and dialogues are exchanges with a fully automatic system. Their situation is paradoxical.

The terms of the paradox are the following:

- The real thing is hard to get: due to the multiplicity of the constraints involved in "serious" exchanges, attempting to design an (even restricted) Natural Language interface with decent conversational abilities ends up with incredibly complex models of the interaction, having to combine linguistic knowledge with situational and interpersonal information.
- Trivial solutions exist: when provided with some approximation of a conversational system, people find a use to it, and tend to compensate for the feeble relevance of the system's contribution, by "bringing their own meanings" into the exchange. So, while the best design is still insufficient, cheaper systems actually find an occasional audience. This was once true of Jason Hutchen's MegaHal, and, presently, of Rollo Carpenter's Jabberwacky. MegaHal transcripts

[1] `http://www.lutecium.org/stp/conversations.html`
[2] `http://www.limsi.fr/Individu/jbb/bookseller-libraire.html`
[3] `http://www.geocities.com/Athens/9448/chat.html`

show that the system has a capacity of combining input by different users, yielding unexpected sentences. But as it lacks a representational dimension, it cannot make an essential contribution to dialogue modelling.

One could dismiss conversation simulators as mere practical jokes, but we shall not do so. They provide us with something precious, namely, human reactions to an absurd conversational situation. Here we find that repair strategies can go very far. Metacommunication is often at work, especially in "wrong language" situations. Our challenge is to capture them, and incorporate them in our dialogue evaluation toolkit.

3 A Java Toolkit

3.1 Numeric Descriptions for Dialogues

Since our strategy calls for the automatic evaluation of natural dialogues, we are currently developing software tools in Java for this purpose. The main idea is that a global representation of a dialogue can be obtained by combining a number of elementary observations. Therefore, the dialogue is described by a two-dimensional array of numbers, one dimension being discrete time (first to last utterance), while the other dimensions represent the diversity of elementary observers. Each observer is looking for one special feature of the dialogue, e.g. bad words, theme shifts, speechacts, emphasis, style discrepancies, and so on. The array they collectively construct is named SoD-vector, meaning "state of Dialogue".

3.2 Pattern Recognition

What we then look for is patterns in the SoD-vector. This task is accomplished by subjective inspectors, acting as filters for particular dialogue patterns. To take a very simple case, developed in Berthelin et al. (2000), there is an inspector which tries to check if a conversation from Castaing (1993) took place over the phone. It may find that observers of that field have marked almost all sentences "telephone, 100 percent", while one isolated sentence has only "telephone, 90 percent". The inspector then decides that it is globally 100 percent, and modifies the "90" mark accordingly. This example is oversimplified. We are currently in the process of examining our various corpora, in search of possible tasks for sophisticated inspectors.

3.3 A Measure of Resemblance Between Dialogues

This approach leads naturally to a classification of dialogues. When two transcripts satisfy exactly the same inspectors, they belong to a class. Problems occuring in a dialogue of a given class can be tentatively solved by strategies which have worked in another dialogue of that class. Sometimes, this strategy will reveal crucial differences which otherwise would have gone unnoticed.

4 Conclusion and perspectives

The results of evaluation are relevant to the task of informing future systems of which attitude is best suited when a conversation has a given set of problematic properties. This cannot be completely disconnected from the actual context of the dialogue. Joking, for instance, is OK when you are teaching, less so when dealing with legal cases. Therefore, much remains to be done in the field of dialogue engineering. Nevertheless, we consider that a global approach with a numeric component is quite promising.

Bibliography

Berthelin, J.-B., Girard, Y., Grau, B., and Vilnat, A. (2000). In the beginning was the "END": Evaluation of natural dialogues as a step towards improving artificial ones. In *Third Workshop on Human-Computer Conversation, Bellagio, Italy, 3-4-5 juillet 2000.*

Berthelin, J.-B., Grau, B., Robba, I., and Vilnat, A. (1999). A cross-corpus vision of conversation and dialogues. In *Proceedings of the thirty-second meeting of the Societas Linguistica Europea in Ljubljana.* University of Ljubljana.

Castaing, M. F. (1993). Corpus de dialogues dans un standard. Technical report, LIMSI-CNRS, France.

Cohen, Morgan, and Pollack, editors (1990). *Intention in Communication.* MIT Press.

Lambert, L. and Carberry, S. (1991). A tripartite plan-based modelof dialogue. In *Proc 29th ACL.*

Litman, D. J. and Allen, J. F. (1990). *Intention in Communication*, chapter Discourse Processing and Commonsense plans. MIT Press.

Pollack, M. (1990). *Intention in Communication*, chapter Plans as complex mentals attitudes. MIT Press.

Paper 24

GoDiS: flexible dialogue in multiple domains

Staffan Larsson, Robin Cooper, Stina Ericsson
{cooper, stinae, sl}@ling.gu.se

Abstract

This paper introduces the GoDiS system (Bohlin et al. (1999)). GoDiS is a prototype dialogue system for information-exchange and action-oriented dialogue, capable of accommodating questions and tasks to enable the user to present information in any desired order, without explicitly naming the dialogue task. In the demo we show how GoDiS has been adapted to various domains. GoDiS has also been extended to handle spoken dialogue in addition to written dialogue.

1 Introduction

GoDiS is an experimental dialogue system based on Ginzburg's concept of Questions Under Discussion (QUD)[1]. GoDiS is implemented using the TrindiKit (Larsson and Traum (2000); Larsson et al. (2000)), a toolkit for implementing dialogue move engines and dialogue systems based on the Information State approach. While originally built for fairly simple information exchange dialogue, it is being extended to handle action-oriented and negotiative dialogue. One of the goals of the information state approach is to encourage reusability and plug-and-play; to demonstrate this, GoDiS has been adapted to several different dialogue types and domains, including information exchange dialogue in travel agency and autoroute domains, and action-oriented dialogue when acting as an interface to a mobile phone or computerized agenda. GoDiS has also been enhanced with speech input and output.

2 GoDiS architecture

GoDiS [2] is implemented using the TrindiKit software package developed in the TRINDI project[3]. The TrindiKit is a toolkit for building and experimenting with *dialogue move engines* and *information states* (IS), We use the term *information state* to mean, roughly, the information stored internally by an agent, in this case a dialogue system. A *dialogue move engine* (DME) updates the information state on the basis of observed dialogue moves and selects appropriate moves to be performed. The GoDiS architecture shown below is an instantiation of the general TrindiKit architecture.

[1]Work on this paper was supported by SIRIDUS (Specification, Interaction and Reconfiguration in Dialogue Understanding Systems), EC Project IST-1999-10516, and D'Homme (Dialogues in the Home Machine Environment), EC Project IST-2000-26280. The first author also wishes to thank STINT (The Swedish Foundation for International Cooperation in Research and Higher Education). Parts of this paper has previously been published in Bohlin et al. (1999); Larsson et al. (2000c); Larsson and Cooper (2000); Larsson et al. (2000b)

[2]The acronym stands for Gothenburg Dialogue System.

[3]www.ling.gu.se/research/projects/trindi/

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

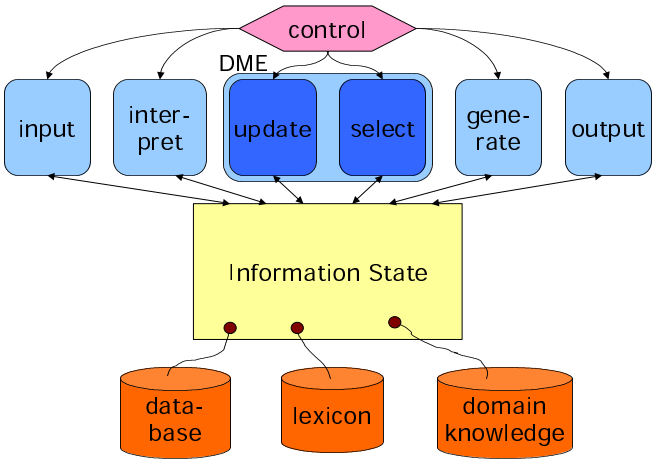


Figure 24.1: GoDiS architecture

The components of the architecture are the following:

- the Information State (IS)
- domain-independent modules, operating according to module algorithms
- the domain-independent Dialogue Move Engine (DME), consisting of two modules (Update and Select); the DME is responsible for updating the IS based on observed moves, and selecting moves to be performed by the system.
- a controller, wiring together the other modules, either in sequence or through an asynchronous mechanism.
- three domain-dependent resources: Database, Lexicon, and Domain Knowledge

The TIS is accessed by modules through conditions and operations. The types of the various components of the TIS determine which conditions and operations are available.

The modules in the GoDiS system are the following:

- Input: Receives input utterances from the user and stores it in the input interface variable.
- Interpretation: Takes utterances (stored in input) and gives interpretations in terms of dialogue moves (including semantic content). The interpretation is stored in the interface variable latest_moves.
- Generation: Generates output moves based on the contents of next_moves and passes these on to the output interface variable.
- Output: Produces system utterances based on the contents of the output variable

In GoDiS, a number of resources are hooked up to the information state. There are three resources in GoDiS: a lexicon, a database and a domain resource containing (among other things) domain-specific dialogue plans.Currently, there are (in some cases incomplete) GoDiS resources for several domains, including the following:

- travel agency

- autoroute
- Xerox device manual
- mobile phone interface
- handheld computer agenda interface

3 Information State

The question about what should be included in the information state is central to any theory of dialogue management. The notion of information state we are putting forward here is basically a simplified version of the dialogue game board which has been proposed by Ginzburg. We are attempting to use as simple a version as possible in order to have a more or less practical system to experiment with.

The main division in the information state is between information which is private to the agent and that which is (assumed to be) shared between the dialogue participants. What we mean by shared information here is that which has been established (i.e. grounded) during the conversation, akin to what Lewis in Lewis (1979) called the "conversational scoreboard". We represent information states of a dialogue participant as a record of the type shown in figure 24.2.

$$\begin{bmatrix} \text{private} : & \begin{bmatrix} \textsc{agenda} : \textsc{Stack}(\textsc{Action}) \\ \textsc{plan} : \textsc{StackSet}(\textsc{Action}) \\ \textsc{bel} : \textsc{Set}(\textsc{Prop}) \\ \textsc{tmp} : \text{(same as } \textsc{shared}) \end{bmatrix} \\ \text{shared} : & \begin{bmatrix} \textsc{bel} : \textsc{Set}(\textsc{Prop}) \\ \textsc{qud} : \textsc{Stack}(\textsc{Question}) \\ \textsc{lu} : \begin{bmatrix} \textsc{speaker} : \textsc{Participant} \\ \textsc{moves} : \textsc{assocSet}(\textsc{Move},\textsc{Bool}) \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 24.2: The type of information state we are assuming

The private part of the information state includes a set of beliefs and a dialogue *plan*, i.e. is a list of dialogue actions that the agent wishes to carry out. The plan can be changed during the course of the conversation. For example, if a travel agent discovers that his customer wishes to get information about a flight he will adopt a plan to ask her where she wants to go, when she wants to go, what price class she wants and so on. The *agenda*, on the other hand, contains the short term goals or obligations that the agent has, i.e. what the agent is going to do next. For example, if the other dialogue participant raises a question, then the agent will normally put an action on the agenda to respond to the question. This action may or may not be in the agent's plan.

The private part of the IS also includes "temporary" shared information that saves the previously shared information until the latest utterance is grounded, i.e. confirmed as having been understood by the other dialogue participant[4]. In this way it is easy to retract the "optimistic" assumption that the information was understood if it should turn out that the other dialogue participant does not understand or accept it. If the agent pursues a cautious rather than an optimistic strategy then information will at first only be placed in the "temporary" slot until it has been acknowledged by the other dialogue participant whereupon it can be moved to the appropriate shared field.

The (supposedly) shared part of the IS consists of three subparts. One is a set of propositions which the agent assumes for the sake of the conversation and which are established during the dialogue. The second is a stack of questions under discussion (QUD). These are questions that have been raised and are currently under discussion in the dialogue. The third contains information about the latest utterance (speaker, moves and integration status).

[4]In discussing grounding we will assume that there is just one other dialogue participant.

4 Accommodation in GoDiS

In this section, we introduce the concepts of question and task accommodation, and discuss how such behaviour has been implemented in GoDiS.

4.1 Accommodating a question onto QUD

Dialogue participants can address questions that have not been explicitly raised in the dialogue. However, it is important that a question be available to the agent who is to interpret it because the utterance may be elliptical. Here is an example from a travel agency dialogue[5]:

```
$J: what month do you want to go
$P: well around 3rd 4th april / some time there
$P: as cheap as possible
```

The strategy we adopt for interpreting elliptical utterances is to think of them as short answers (in the sense of Ginzburg Ginzburg (1998)) to questions on QUD. A suitable question here is *What kind of price does P want for the ticket?*. This question is not under discussion at the point when P says "as cheap as possible". But it can be figured out since J knows that this is a relevant question. In fact it will be a question which J has as an action in his plan to raise. On our analysis it is this fact which enables A to interpret the ellipsis. He finds the matching question on his plan, accommodates by placing it on QUD and then continues with the integration of the information expressed by *as cheap as possible* as normal. Note that if such a question is not available then the ellipsis cannot be interpreted as in the dialogue below.

A. What time are you coming to pick up Maria?
B. Around 6 p.m. As cheap as possible.

This dialogue is incoherent if what is being discussed is when the child Maria is going to be picked up from her friend's house (at least under standard dialogue plans that we might have for such a conversation).

Question accommodation has been implemented in GoDiS using a single information state update rule **accommodateQuestion**, seen below in (1)[6]. When interpreting the latest utterance by the other participant, the system makes the assumption that it was an **answer** move with content A. This assumption requires accommodating some question Q such that A is a relevant answer to Q. The check operator "answer-to(A, Q)" is true if A is a relevant answer to Q given the current information state, according to a (domain-dependent) definition of question-answer relevance.

(1) RULE: **accommodateQuestion**
CLASS: **accommodate**
PRE:
- val(SHARED.LU.SPEAKER, usr)
- in(SHARED.LU.MOVES, **answer(**A**)**)
- not (lexicon :: **yn_answer(**A**)**)
- assoc(SHARED.LU.MOVES, **answer(**A**)**, false)
- in(PRIVATE.PLAN, findout(Q))
- domain :: answer-to(Q, A)

EFF: { push(SHARED.QUD, Q)

4.2 Accommodating the dialogue task

After an initial exchange for establishing contact the first thing that P says to the travel agent in our dialogue is "flights to paris". This is again an ellipsis which on our analysis has to be interpreted as

[5]This dialogue has been collected by the University of Lund as part of the SDS project. We quote a translation of the transcription done in Göteborg as part of the same project.

[6]These rules have been simplified slightly for readability.

the answer to a question (two questions, actually) in order to be understandable and relevant. As no questions have been raised yet in the dialogue (apart from whether the participants have each other's attention) the travel agent cannot find the appropriate question on his plan. Furthermore, as this is the first indication of what the customer wants, the travel agent cannot have a plan with detailed questions. We assume that the travel agent has various plan types in his domain knowledge determining what kind of conversations he is able to have. Each plan is associated with a task. E.g. he is able to book trips by various modes of travel, he is able to handle complaints, book hotels, rental cars etc. What he needs to do is take the customer's utterance and try to match it against questions in his plan types in his domain knowledge. When he finds a suitable match he will accommodate the corresponding task, thereby providing a plan to ask relevant question for flights, e.g. when to travel?, what date? etc. Once he has accommodated this task and retrieved the plan he can proceed as in the previous example. That is, he can accommodate the QUD with the relevant question and proceed with the interpretation of ellipsis in the normal fashion.

This example is interesting for a couple of reasons. It provides us with an example of "recursive" accommodation. The QUD needs to be accommodated, but in order to do this the dialogue task needs to be accommodated and the plan retrieved. The other interesting aspect of this is that accommodating the dialogue task in this way actually serves to drive the dialogue forward. That is, the mechanism by which the agent interprets this ellipsis, gives him a plan for a substantial part of the rest of the dialogue. This is a way of capturing the intuition that saying *flights to Paris* to a travel agent immediately makes a number of questions become relevant.

Task accommodation and plan retrieval is taken care of by two rules: **accommodateTask** and **retrievePlan**, respectively. The first matches the move(s) in SHARED.LU.MOVES with a task to which it is relevant and stores this task in SHARED.BEL, and the second puts the plan to achieve that task in the PRIVATE.PLAN field.

(2) RULE: **accommodateTask**
CLASS: **accommodate**
PRE: { val(SHARED.LU.SPEAKER, usr)
in(SHARED.LU.MOVES, $Move$)
domain :: relevant_to_tasks($Move$, $\{Task\}$) }
EFF: { add(SHARED.BEL, task(Task)

(3) RULE: **retrievePlan**
CLASS: **managePlan**
PRE: { emptyRec(PRIVATE.PLAN)
in(SHARED.BEL, task($Task$))
domain :: plan($Task$, $Plan$) }
EFF: { set(PRIVATE.PLAN, $Plan$)

5 Sample dialogues

We demonstrate the use of question and task accommodation in different domains, including "travel agency" and "mobile phone interface". The dialogue below shows how GoDiS is able to accommodate the task and integrate information in the order that the user gives it.

```
 $S> Welcome to the travel agency!
$U> flights to paris
$S> What city do you want to go from?
$U> gothenburg, in april, as cheap as possible
$S> Do you want a return ticket?
$U> no
$S> It will cost 50 crowns
```

```
$U> thanks, bye!
$S> Thank you for your visit!
```

We also show how the concept of question accommodation can be used in conjunction with dialogue plans generated from menu structures (Larsson et al. (2001). If information presented by the user is relevant to a specific branch of the menu system then the system will be able to jump directly to that branch without requiring the user to step through all the intervening nodes.

```
 $S> What do you want to do?
$U> add new number to phonebook for John
$S> What is the number?
$U> 0123 45678
$S> OK, new number added for John.
```

If the user does not know what to do the system can present the options. This means that users can use their own conception of the knowledge space and not be locked to that of the designer of the system.

```
 $S> What do you want to do?
$U> John
$S> Do you want to add a new number, change a number, delete a number, give a
    ringing tone, or search?
$U> add a new number
$S> What is the number?
$U> 0123 45678
$S> OK, new number added for John.
```

Bibliography

Allwood, J. (1995). An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.

Bohlin, P., Cooper, R., Engdahl, E., and Larsson, S. (1999). Information states and dialogue move engines. In Alexandersson, J., editor, *IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

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

Ginzburg, J. (1998). Clarifying utterances. In Hulstijn, J. and Niholt, A., editors, *Proc. of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*, pages 11–30, Enschede. Universiteit Twente, Faculteit Informatica.

Larsson, S., Berman, A., Bos, J., Grönqvist, L., Ljunglöf, P., and Traum, D. (2000). TrindiKit 2.0 manual. Technical Report Deliverable D5.3 - Manual, Trindi.

Larsson, S. and Cooper, R. (2000). An information state approach to natural interactive dialogue. In *Proceedings of LREC2000 Workshop on Natural Interactive Dialogue*.

Larsson, S., Cooper, R., and Engdahl, E. (2000b). Question accommodation and information states in dialogue. In *Proceedings of the Third Wokshop in Human-Computer Conversation*.

Larsson, S., Cooper, R., and Ericsson, S. (2001). menu2dialog. In *Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. To appear.

Larsson, S., Ljunglöf, P., Cooper, R., Engdahl, E., and Ericsson, S. (2000c). GoDiS - an accommodating dialogue system. In *Proceedings of ANLP/NAACL-2000 Workshop on Conversational System*.

Larsson, S. and Traum, D. (2000). Information state and dialogue management in the trindi dialogue move engine toolkit. *NLE Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering*.

Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359.

Searle, J. R. (1969). *Speech Acts*. Cambridge University Press, New York.

Sidner, C. L. (1994). An artificial discourse language for collaborative negotiation. In *Proceedings of the forteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, pages 814–819.

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.

Paper 25

An information state update approach to collaborative negotiation

Robin Cooper, Stina Ericsson, Staffan Larsson, Ian Lewin
{cooper, stinae, sl}@ling.gu.se, ian.lewin@netdecisions.co.uk

Abstract

Using the information state approach to dialogue analysis, we sketch an account of negotiative dialogue starting from Sidner's artificial negotiation language. This account is adapted to the Questions under Discussion (QUD)-based information state used by the GoDiS system. Some problems with this account are pointed out, and we attempt to analyse why these problems arise and how they might be resolved. Finally, an alternative account to negotiative dialogue is outlined.

1 Introduction

In the TRINDI project (see e.g. Poesio et al. (1999); Larsson and Traum (2000)), Larsson and Traum (2000); Bos et al. (1999)) an information state update approach to dialogue analysis was developed which treats utterances in terms of their update effects on the information state of the dialogue[1]. One of the aims of the SIRIDUS project is to explore ways of extending this work to handle *negotiative dialogue*. The aim of this paper is to explore the nature of negotiation in dialogue.

We first attempt to characterise the concept of negotiation and make some relevant distinctions. Using the information state approach, we then sketch an account of negotiative dialogue starting from Sidner's artificial negotiation language. This account is adapted to the Questions under Discussion (QUD)-based information state (Ginzburg (1998)) used by the GoDiS system (Bohlin et al. (1999)). Some problems with this account are pointed out. Finally, the concept of an Issue Under Negotiation is introduced to model the fact that in negotiation, several alternative solutions (answers) to an issue can be discussed and compared before a solution is finally settled on.

1.1 The concept of negotiation

Negotiation can occur at many different levels in a dialogue. In particular, we may identify: negotiation in the task domain, negotiation over dialogue strategy and negotiation over meaning. In a shopping domain, customers may negotiate with salesmen over quantity, price and product features. They may also negotiate over dialogue strategy. A salesman may wish to focus on product features first in order to encourage a customer to buy a higher value product; whereas the customer may wish to focus on

[1]Work on this paper was supported by SIRIDUS (Specification, Interaction and Reconfiguration in Dialogue Understanding Systems), EC Project IST-1999-10516, and D'Homme (Dialogues in the Home Machine Environment), EC Project IST-2000-26280. The third author also wishes to thank STINT (The Swedish Foundation for International Cooperation in Research and Higher Education).

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

price first. They may also negotiate over meaning, for example, over the precise meaning of "Palm-compatible" or whether "faulty product return" includes cash reimbursement or just replacement by a similar item.

There are a couple of different kinds of negotiation that can be isolated. Negotiation may be *collaborative*. DPs[2] may be negotiating how to achieve a common goal and may find that they do not have any conflicting opinions. That is, negotiation does not necessarily imply conflicting goals or interests.[3]

Related to collaborativity is *argumentation*. DPs may argue for some proposals and against other proposals. To handle argumentation, a dialogue system would have to deal with *argumentation acts* (Traum and Hinkelman (1992)). Usually, noncollaborative negotiation will require argumentation, but in the case of collaborative negotiation it is not always needed (especially concerning issues where one of the DPs has the right to decide on a solution on her own, see below).

As our initial domain for exploring negotiative dialogue, we have chosen a travel agency setting, partly because this is a well-known domain for research on dialogue systems with many examples in the literature (most importantly, in the work of Sidner reported below), and partly because it offers an example of a relatively simple type of negotiation. In a travel agency setting, the customer and the travel agent may negotiate which flight the customer should take. This may involve discussing and comparing several different flights, but it is usually a collaborative, non-argumentative (at least in most cases) type of dialogue.

2 Sidner's artificial negotiation language

In Sidner (1994a), Candace Sidner defines an artificial discourse language for collaborative negotiation. Her aim is to understand dialogues in which agents recognize a shared goal, and then plan and act *jointly*. Negotiation, for Sidner, is "the interactive process of attempting to agree on the goals, actions and beliefs that comprise the planning and acting decisions of the collaboration". In Sidner (1994b) she discusses the application of her scheme to an example from the American Express Travel Dialogue corpus (Kowtko and Price (1989)).

2.1 Negotiation language constructs

Sidner identifies eleven constructs for her artificial language. The constructs are presented here using a and b as variables over agents, and p and q as variables over propositions. Sidner states that `propose + accept` and `propose + reject` are the "most typical characteristics of negotiation in discourse". Consequently, the central moves she defines are

PFA(a,b,p)	ProposeForAccept	a proposes p to b
AP(a,b,p)	AcceptProposal	a accepts b's proposal of p
RJ(a,b,p)	Reject	a rejects b's proposal of p

In addition, she defines a further five moves.

RP(a,b,p)	RetractProposal	a retracts his proposal of p (to b)
CO(a,b,p,q)	Counter	a counters b's proposal of q with p
AOP(a,b,p,q)	AcceptOtherProposal	a accepts b's proposal of q and retracts his own p
PR(a,b,p,q)	ProposeReplace	a proposes p and rejects b's proposal of q
PA(a,b,d,c)	ProposeAct	a proposes action *act*

Of these five moves, only *RP* is actually a new primitive. The two moves *AOP* and *PR* are simple constructions out of other moves. *AOP* is equivalent to *RP* + *AP*, that is, first retracting one's own

[2] dialogue participants

[3] This view may not correspond perfectly to the everyday use of the word "negotiation". It is, however, common practice in other fields dealing with negotiation (e.g. game theory, economy) to include collaborative negotiation (cf. Lewin et al. (2000)).

proposal and then accepting your partner's. *PR* is equivalent to *RJ + PFA*, that is, first rejecting your partner's proposal and then proposing your own. *CO*, a counter, is also a construction over two instances of *PFA*. A counter consists of making one new proposal p and another (complex) proposal that p is a reason for thinking q is false. *PA(a,b,d,c)* (ProposeAct) is a proposal in which an action d (in context c) is proposed rather than a belief.

There are also three acknowledgment moves:

AR(a,b,p)	AckReceipt	a acknowledges b's proposal of p
ARJ(a,b,p)	AckReject	a acknowledges b's rejection of p
ARP(a,b,p)	AckRetractedProposal	a acknowledges b's retraction of p

Sidner's work is particularly interesting from the Information State Update perspective because she also defines a semantics for these moves. For each move, a postcondition on the beliefs and intentions of the dialogue participants is defined. Postconditions may state that certain propositions are mutually believed by the dialogue participants. In addition, some moves are associated with operations on two stacks: a stack of `open beliefs` and a stack of `rejected beliefs`. Sidner states that the stacks capture part of the attentional state of the discourse. That is, they represent what the negotiation is currently *about*, namely that some proposal p is currently under discussion. Sidner uses the term 'the state of communication' to refer to her postconditions. The conditions describe who believes and intends what and who believes what has been communicated.

For example, figure 25.1 gives the postcondition for a ProposeForAccept, PFA(a,b,p)[4]. The idea is that after a proposal (to believe p) has been made, the state of communication is that a believes p, a intends that b should believe it and that a believes he has communicated p to b.

1 believes(a,p)
2 intends(a, achieve(a, believes(b, p)))
3 believes(a, communicated(a,b,p))

Figure 25.1: Postconditions on ProposeForAccept(a,b,p)

Sidner has no need for a specific move for asking questions. Utterances containing questions are analysed as proposing the action that the other DP provides some piece of information, as exemplified by the analysis of the Y/N-question "Did John come?" as

(PA agt1 (Should-Do agt2 (Tellif agt2 '(john did come))))

Consequently, Sidner's project appears to turn into an analysis of all dialogue in terms of negotiation. We return to this point in 4.

2.2 Application of Sidner's theory to real dialogue

In Sidner (1994b), Sidner discusses in general terms the application of her scheme to a dialogue between a travel agent (TA) and a customer (BC). The dialogue from the AMEX travel planning corpus (Kowtko and Price (1989)) is illustrated in figure 25.2.

[4]This type of formalisation is similar to previous work on dialogue in the BDI tradition, e.g. Allen and Perrault (1980)

0	**BC**	My name is B C and I would like to plan a trip
1	**TA**	and the date you need to leave?
...		
19	**TA**	there is one on United that leaves Oakland at eleven thirty p.m. and arrives Chicago five twenty five a.m.
20	**BC**	so that's a two hour hold there
21	**TA**	yes
22	**BC**	waiting for that flight ok any others?
23	**TA**	uh not from Oakland. departing from San Francisco it's about the same actually American has an eleven forty one flight from San Francisco that arrives Chicago five fifty four (and
24	**BC**	that's) and hour and a half. so that's that's a a wash
25	**TA**	yeah or wait just one moment. or United has a twelve oh one a.m. departure that arrives at Chicago five fifty two a.m.
26	**BC**	oh that sounds good

Figure 25.2: Excerpt from Amex Transcript

Perhaps surprisingly, the very first utterance labelled as a proposal is utterance 0, namely *My name is BC*. Sidner claims that this apparently simple assertion is a proposal about a belief, albeit a 'mundane one', which BC wishes to share. The claim is counterintuitive simply because *My name is BC* is not negotiable. One can negotiate what one's name will be (perhaps as part of a marriage contract) but not what it is. One can only argue about what it actually is.

When analysing utterances 19 through 25, Sidner points out there are a number of alternative proposals on offer throughout this section of the dialogue and that there is more going on than a 'simple linear format of making statements or asking questions, followed by responses'. This does indeed seem an important part of negotiation: there may be several proposals or offers on the table at once, and they may be evaluated and compared. Other offers may be solicited. Sidner analyses utterance 19 as a proposal and 23 and 25 as *counterproposals*, even though they are all generated by TA. She herself remarks that counterproposals are usually brought by a collaborator in response to a proposal. It is unclear why she makes this analysis of 19, 23 and 25. It appears perfectly possible and very natural to analyse them as a simple sequence of proposals. We will return to this point in 4.

3 Analysing Sidner's language using the information state update approach

In this section we discuss the reformulation of some of Sidner's rules in terms of the kind of update rules that were used in the GoDiS system in the TRINDI project (Traum et al. (1999)). The full set of Sidner's rules and corresponding GoDiS update rules can be found in (Lewin et al. (2000)).

3.1 The GoDiS information state

In the GoDiS approach a variant of Ginzburg's notion of QUD (questions under discussion), Ginzburg (1996), roughly corresponds to Sidner's openStack but contains questions instead of propositions. However, each agent has their own view of what the QUD might be, allowing for misunderstandings concerning what proposals are actually being considered in the dialogue.

Sidner's second stack is the rejectedStack where proposals that have been rejected are stored. We do not currently have a field corresponding exactly to this. Instead of placing a proposal (proposition) p on a rejected stack we add $\neg p$ to the shared commitments (beliefs) field. (Again this has to be done

separately for each agent first in the move and then in its integration counterpart for the other agent.) At the moment we see this as being sufficient to achieve the effects of a rejected stack, although if we were to discover the need for a separate field it would be straightforward to add it. [5]

The type of records we are assuming for our information states is shown below.

The information state is divided into a private and a shared part. The PLAN and AGENDA fields in the private part contain the dialogue plan and the short term goals, typically the next action to be performed, respectively. These two fields correspond roughly to Sidner's postconditions concerning intentions. The PRIVATE field also includes a set of propositions representing the agent's (private) beliefs. These private beliefs influence the agent's 'negotiative' behaviour. For example, in the descriptions below of the negotiative moves and update rules, a proposition p's being among an agent's private beliefs is required for an agent to put p forward as a proposal.

The SHARED field is divided into three subfields. The first of these is a set of shared beliefs – the beliefs an agent assumes to be shared by the dialogue participants. The next field, the QUD, is a stack of questions under discussion, and the third field, LU, is a record containing information about the latest utterance. The MOVES subfield is an association set, where each move is associated with a boolean indicating whether the move has been integrated or not.

3.2 Towards an implementation of Sidner's language in GoDiS

Although Sidner only states a postcondition for her moves, one can quite easily extract an information update from it. Only the third condition in figure 25.1 – postconditions on proposeForAccept – naturally arises as a *result* of undertaking the PFA action itself. Presumably, a already believed p and intended that b should believe it before undertaking the action.

In fact if we extract all the update effects of Sidner's central primitive moves (*PFA*,*RJ*,*AP* and *RP*), then we can obtain the very simple list of additive updates shown in figure 25.3, and the same can be done for the other moves.

PFA(a,b,p)	believes(a, communicated(a,b,p))
AP(a,b,p)	mutually-believe(a, b, p))
RJ(a,b,p)	believes(a, communicated(a, not(believes(a,p)),b))
RP(a,b,p)	believes(a, communicated(a, not(believes(a,p)),b))

Figure 25.3: Information State Additions of Sidner's Negotiative Moves (1)

Sidner's update effects on the stacks of open and rejected are summarised in figure 25.4. This table reveals an odd asymmetry in Sidner's account: proposals (*PFA*) do not have any effect on the stacks until acknowledged (*AR*) but rejection and retraction have immediate effect on the stacks, with no need for acknowledgement (*ARJ* and *ARP*, respectively). The same goes for acceptances, which don't even have any acknowledgement-message defined.

[5] In fact, it is not clear how items that are put on the Rejected stack are ever to be used. If it merely makes a historical record of items that were discussed but rejected it is surprising there is no similar record of things that were accepted.

PFA(a,b,p)	
AP(a,b,p)	pop(OpenStack)
RJ(a,b,p)	pop(OpenStack), push(p,RejectStack)
RP(a,b,p)	pop(OpenStack)
AR(a,b,p)	push(p,OpenStack)
ARJ(a,b,p)	
ARP(a,b,p)	

Figure 25.4: Stack Operations of Sidner's Negotiative Moves

Thus, when reformulating Sidner's postconditions in an information state update framework, it becomes clear that conditions and effects must be separated out from the postconditions. Taking the postconditions of *PFA* and converting them into GoDiS style conditions and effects, we then find the following[6]:

PFA proposeForAccept(a,b,p)
Conditions
$p \in$ a.pr.bel (Sidner line 1)
Effects
(a, propose(p)) $\in$ a.sh.lu (Sidner line 3)

Sidner's line 2 in figure 25.1 can be derived by the rule of inference :

$(\mathrm{dp}(a,b) \wedge (a, \mathrm{propose}(p)) \in \mathrm{a.sh.lu}) \rightarrow \mathrm{intend}(a, \mathrm{achieve}(a, \mathrm{bel}(b,p)))$

where $\mathrm{dp}(a,b)$ means that b is agent a's dialogue partner. Such rules of inference could be seen as a bridge between the simple information states we use and the more general BDI approach to reasoning (Allen and Perrault (1980)). While BDI information is not always directly represented in our information states as such we believe that it can often be inferred and that this could be exploited if more general BDI reasoning is required. We currently do not have such reasoning in our implemented systems, however.

The first condition, line 1 in figure 25.1, is essentially a selection condition, i.e. something that has to hold for a in order for a to select a propose-move (provided a is honest). Obviously, an agent's believing p will not be the *only* condition for its proposing p, but it can be regarded as a necessary one. Having selected a propose move, and produced a linguistic utterance corresponding to this move, the effects of *PFA* formulated above will similarly be necessary but maybe not sufficient effects of the prose move.

The effect of pushing p on the openStack is not included in the PFA formulation. Instead, this is an effect of the *AR* ("acknowledge receipt") move. AR is used by an agent to indicate that a proposal made by the dialogue partner has been received, that is, heard and understood (to the best of the receiver's knowledge). The postconditions for AR are shown in figure 25.5. An AR move does not commit the speaker to the proposed proposition in any way. However, the dialogue participant who made the proposal is committed to believing p, and also to the intention of making the other dialogue participant believe p.[7] Sidner gives this move in terms of three mutual beliefs. An AR move also pushes the proposition p onto openStack, indicating that p is now open for discussion.

[6]We will use an abbreviatory notation in reformulating Sidner's rules which we illustrate here by example. 'p $\in$ a.pr.bel' will stand for a condition that proposition p is a member of the set in PRIVATE:BEL in a's information state. '$p?$ $\in$ b.sh.qud' will stand for a condition that the question whether p is on the stack SHARED:QUD in b's information state.

[7]Note that 'commitment' as used here corresponds to a shared belief.

1 mutually_believe(a, b, believe(b,p))
2 mutually_believe(a, b, intend(b, achieve(b, believe(a,p))))
3 mutually_believe(a, b, communicated(b,p,a))

Figure 25.5: Postconditions on AcknowledgeReceipt(a,b,p)

We break Sidner's *AR* down into two update rules, one for the agent who is acknowledging receipt and one for the agent who is integrating this acknowledgement.

AR acknowledgeReceipt(a,b,p)
Conditions
(b, propose(p)) ∈ a.sh.lu (part of Sidner's line 3)
Effects
believe(b,p) ∈ a.sh.bel (part of Sidner's line 1)
push(p?, a.sh.qud) (Sidner's call to openStack)

Sidner's line 2 is derived by the rule of inference:
(dp(a,b) ∧ (b, propose(p)) ∈ a.sh.lu) → intend(b, achieve(b, bel(a, p)))
in addition to the inference rule used in PFA, as the intention is now a mutual belief.

integAR integrateAcknowledgeReceipt(a,p,b)
Conditions
(b, acknowledgeReceipt(p)) ∈ a.sh.lu (part of Sidner's line 3)
Effects
believe(a,p) ∈ a.sh.bel (part of Sidner's line 1)
push(p?, a.sh.qud) (Sidner's call to openStack)

A difference between Sidner's rules and GoDiS in its current version is that the SHARED.BEL in the GoDiS information state only records beliefs about propositions, not beliefs about *who* is holding a particular belief. In order to model this, two new fields can be added to the information state: SHARED.BEL.SYS and SHARED.BEL.USR, which, in addition to the separate SHARED.BEL field, keep track of shared beliefs concerning who believes what.

4 Discussion

In this section, we use the insights gained from analysing Sidner's account using the information state approach (and from Sidner's own application of the theory to real dialogue), and point out some problems with Sidner's analysis. We also suggest an analysis which attempts to solve these problems.

4.1 Negotiation of uptake vs. negotiation of alternatives

An immediate consequence of applying the information state update approach to Sidner's artificial language is that it requires reformulating postconditions as information state updates. That is, instead of only saying what is true after a move has been performed, we have to provide *conditions* on the performance of that move, and *effects* of performing the move. Conditions place restrictions on what the information state must look like in order to apply an update, and effects specify how the information state is to be updated. This also requires specifying exactly when the effects of a move are integrated into the information state.

As we saw in the previous section, it turns out that Sidner's account is asymmetric in regard to when the effects of different moves are integrated; some (*PFA*) require an acknowledgement to affect the openStack, while others (*RJ, RP, AP*) do not need acknowledgment. The strategy resulting from the assumption that the effects of a move can be integrated before an acknowledgment has been

received can be called *optimistic*, The absence of such an assumption leads to a *pessimistic* strategy, where the effects of a move cannot be integrated until an acknowledgment has been received.

A third difference between Sidner's model and ours, is that GoDiS has a single general acknowledge move which is why the TMP field is needed for pessimistic grounding and uptake. GoDiS could however be reformulated so that a number of different acknowledge moves are used, just as in Sidner's language, in which case the TMP field would no longer be needed.[8]

Both Clark (1996) and Allwood (1995) distinguish four levels of action involved in communication (S is the speaker, H is the hearer):

- Acceptance/uptake: whether H accepts (the content of) S's utterance
- Understanding: whether H understands S's utterance
- Perception: whether H perceives S's utterance
- Contact: whether H and S have contact, i.e. if they have established a channel of communication

These levels of action are involved in all dialogue, and to the extent that understanding and uptake can be said to be negotiated, all dialogue has an element of negotiation built in. This is reflected in Sidner's account, when she views as proposal utterances which would be analysed as answers or assertions in the GoDiS framework.

When recasting Sidner's formal analysis in the QUD-based GoDiS framework as done in section 3, it becomes clear that Sidner assumes a pessimistic approach to both grounding and uptake of proposal-moves (but not e.g. to reject- and accept-moves, where an optimistic approach to uptake is apparently used). Utterances which would be analysed as answers or assertions in the GoDiS framework are seen as proposals which need to be explicitly acknowledged and accepted to achieve full effect. By contrast, in GoDiS we have previously assumed an optimistic approach to both grounding and uptake.

Whichever approach is chosen, it is clear that strategies for grounding and acceptance are involved in all utterances in a dialogue. But as we noted in section 2.2, some dialogue is negotiative in a different sense: there may be several proposals on offer at once, and they may be evaluated and compared before a final decision is made. Or in a slightly different terminology: there may be several potential solutions (or answers) to a problem (or issue) on the table at once. This feature is not present in all dialogue; for example, in simple information exchange dialogues, questions are usually answered directly without any discussion of possible alternatives.

We believe that Sidner's account fails to make a distinction between negotiation of understanding and uptake (which is a feature of all dialogue) and negotiation of different alternative solutions to an issue (which is not a feature of all dialogue). This may explain why "My name is BC" in utterance 0 is analysed as a proposal; it is, after all, subject to the same process of grounding and uptake as any other utterance. The failure to make this distinction may also account for the asymmetry observed in the discussion of figure 25.4 in section 2, regarding the treatment of acknowledgments for proposals, rejections, and acceptances.

When negotiation is regarded as negotiation of alternatives, it becomes natural to view proposal-moves as those moves which add new alternative solutions to some issue under negotiation. This gives proposal moves a different status than in Sidner's account, and allows proposal-moves to coexist with ask- and answer-moves. On this view, proposal-moves are regarded as "core speechacts" in the sense of Traum and Hinkelman (1992), and as such they are subject to the same mechanisms of grounding and uptake as any other core speech acts. This also means that proposal-moves may optimistically assumed to be grounded and taken up, in the same way as ask- and answer-moves. The full effect of a proposal-move with content c (which is achieved when it is grounded and taken up) is, on this analysis, that c is added as an alternative answer to an issue under negotiation.

[8] From a dialogue interpretation point of view it is however questionable whether different types of acknowledgement can be used, in particular if they are all, as in Sidner's examples, linguistically realised as "uh-huh". The recognition of a particular acknowledgement would in that case need to take the dialogue history into consideration.

4.2 Proposals and counterproposals

As an account of this latter type of negotiation, however, Sidner's account has some drawbacks. In section 2.2, we noted that Sidner analysed utterance 19 as a proposal, and utterances 23 and 25 the Amex transcript as counterproposals, event though they are all uttered by **TA**. So, why are they not all analysed as proposals?

As stated in section 2, a counterproposals (*CO(a,b,p,q)*) is a construction over two instances of *PFA*. A counter consists of making one new proposal p and another (complex) proposal that p is a reason for thinking q is false. Note that this establishes a connection between p and q which would not have been present if p had been merely proposed. We believe that this is, in fact, the reason that utterances 23 and 25 are analysed as counterproposals. If they were seen as proposals, there would be no place in the analysis for the fact that they connected. Unfortunately, this analysis forces connected proposals to be in conflict with each other. Consequently, this analysis excludes cases where alternatives are not mutually exclusive, which is the case e.g. when buying a CD. This points to the need for a way to represent the fact that proposals are connected which does not entail that they are in conflict.

A related point is that a proposal of p, once it is accepted, commits the speaker to intending to achieve that the hearer believes p. But in fact, in many cases (including travel agencies) it seems that the agent may often be quite indifferent to which proposed alternative the user selects. So we also need an account of proposal-moves which does not require the speaker to intend the addressee to accept that particular proposal.

4.3 Negotiable and non-negotiable issues

In section 2.2, we noted one problem with Sidner's analysis of an Amex transcript is that seems strange to see "My name is BC" in utterance 0 as a proposal. This is especially true if one is thinking of proposals in the sense of proposing an alternative answer to an issue. We argue that the reason for this is, simply, that the issue of BC's name is not a negotiable issue. A straightforward way of resolving this problem is to make a distinction between negotiable and non-negotiable issues in a dialogue. The notion of negotiability is an activity-dependent one; an issue which is negotiable issue in one activity may not be so in another. Also, issues which are not originally assumed to be negotiable may become negotiable if a DP opens it for negotiation, e.g. by questioning a previously accepted proposal.

4.4 Issues Under Negotiation

To resolve the problems raised here, we need to do three things:

- make a distinction between utterance-related negotiation (grounding, uptake) and negotiation of alternatives
- provide an account of negotiation of alternatives which makes it possible to represent alternatives as regarding the same issue, regardless of whether the alternatives are in conflict or not
- make a distinction between negotiable and non-negotiable issues in an activity

As has been hinted above, negotiation can be thought of as the process of providing a solution (an answer) to an issue (a question). *Issues Under Negotiation* (IUN) can be thought of as questions (often *wh*-questions) e.g. *which flight do you want to take?*. Proposals are suggestions of answers to these questions. Often, proposals can only be understood in the context of an IUN; for example, "there's a flight at 07:45" in the context of the IUN "what flight should the user take" amounts to proposing that the user take a flight at 07:45, although this was not explicitly stated. This is similar to the way elliptical utterances can be interpreted using QUD. This account will be further developed in the GoDiS framework in Larsson (2001).

Bibliography

Allen, J. F. and Perrault, C. (1980). Analyzing intention in utterances. *AIJ*, 15(3):143–178.

Allwood, J. (1995). An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.

Bohlin, P., Cooper, R., Engdahl, E., and Larsson, S. (1999). Information states and dialogue move engines. In Alexandersson, J., editor, *IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

Bos, J., Bohlin, P., Larsson, S., Lewin, I., and Matheson, C. (1999). Dialogue dynamics in restricted dialogue systems. Technical Report Deliverable D3.2, Trindi.

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

Ginzburg, J. (1996). Interrogatives: Questions, facts and dialogue. In *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.

Ginzburg, J. (1998). Clarifying utterances. In Hulstijn, J. and Niholt, A., editors, *Proc. of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*, pages 11–30, Enschede. Universiteit Twente, Faculteit Informatica.

Kowtko, J. and Price, P. (1989). Data collection and analysis in the air travel planning domain. In *Proceedings of DARPA Speech and Natural Language Workshop, October*. freely available at http://www.ai.sri.com/~communic/amex.

Larsson, S. (2001). Issues under negotiation. in progress.

Larsson, S. and Traum, D. (2000). Information state and dialogue management in the trindi dialogue move engine toolkit. *NLE Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering*.

Lewin, I., Cooper, R., Ericsson, S., and Rupp, C. (2000). Dialogue moves in negotiative dialogues. Project deliverable 1.2, SIRIDUS.

Poesio, M., Cooper, R., Matheson, C., and Traum, D. (1999). Annotating conversations for information state updates. In *Proceedings of Amstelogue'99 Workshop on the Semantics and Pragmatics of Dialogue*.

Sidner, C. L. (1994a). An artificial discourse language for collaborative negotiation. In *Proceedings of the forteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, pages 814–819.

Sidner, C. L. (1994b). Negotiation in collaborative activity: A discourse analysis. *Knowledge-Based Systems*.

Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999). A model of dialogue moves and information state revision. Technical Report Deliverable D2.1, Trindi.

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.

Paper 26

An XML architecture for the HCRC Map Task Corpus

Amy Isard
amy.isard@ed.ac.uk
http://www.ltg.ed.ac.uk/~amyi

Abstract

This paper describes an XML architecture for dialogue annotation, which represents multiple overlapping data streams. Different annotation levels are stored in separate files and linked to a common base level, ensuring that the annotations are maintainable and that changes to one level have minimal effects on another. Some tools and techniques which take advantage of this architecture to allow the annotations to be presented in flexible user-friendly formats are also described.

1 Introduction

Researchers at the Human Communication Research Centre (HCRC) at the University of Edinburgh have a long history of basic dialogue research using vertical analysis on recorded and transcribed data, on which different phenomena are annotated and their relationships determined empirically. For this work, our main resource has been the HCRC Map Task corpus (Anderson et al. (1991)), which consists of 128 task-oriented dialogues with an average length of around 6 minutes. The two speakers face each other across a table, and each has a map in front of them with pictures on it known as landmarks. One speaker, known as the information giver, also has a route marked on the map, and the task is for the information giver to describe this route to the other speaker (the information follower) who tries to reproduce the route on his/her own map. The speakers are prevented from seeing one another's maps by a screen in the middle of the table, and in half the dialogues the screen also prevents the participants from seeing each other's faces.

This corpus has been annotated with a wide range of phenomena including dialogue moves, games and transactions (Carletta et al. (1997)), disfluencies (Lickley and Bard (1998); Branigan et al. (1999)), syntax (McKelvie (1998)), gaze - which records each time that each participant looks up (usually at the other person) or down (usually at the map) - (Boyle et al. (1994)) and landmark references - where a participant refers to a picture of a landmark which appears on the map - (Bard et al. (2000)).

In pursuing this work, we have developed a way of representing multiple streams of data with a common time line and many overlapping hierarchies, while ensuring that the annotations are maintainable, so that changes to one level of annotation have the minimum possible impact on any other. In this paper, we describe our data representation and how we work with it.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

amy.isard@ed.ac.uk
http://www.ltg.ed.ac.uk/~amyi.

2 Design Criteria

Our research methodology has clear advantages in terms of providing a structured framework on which to build, but it raises some serious technical challenges for data representation. The representation must be clear and machine parsable using standard techniques, to minimize programming effort. It must represent tree structures because we want to make the hierarchical relationships between elements explicit. For example, a dialogue move, a sentence and a dialogue game may start at the same time, but we want it to be clear that the move is structurally part of the game, whereas the sentence belongs to a different annotation hierarchy. It must allow multiple data streams (e.g. overlapping speakers, gaze, external noises) and if each annotation level is viewed as a tree, it must allow overlapping branches across different levels (e.g. syntax, discourse, and gaze). The representation must also allow for extensions in unforeseen directions, be maintainable, and allow concurrent editing. It must also be possible to display the data in ways tailored to the individual research task, as it is impossible to make sense of all the annotation levels simultaneously.

3 The Base Technology

A number of partial solutions existed in SGML (Goldfarb (1990); W3C (1999)) and XML (Ducharme (1999); W3C (2000a)) for the problems we were trying to solve. The Corpus Encoding Standard (CES) (Ide and Priest-Dorman (2000)) describes guidelines for encoding standards for natural language corpora, based on the Text Encoding Initiative (TEI) (Sperberg-McQueen and Burnard (1994)). The TEI provides methods for allowing overlapping in SGML, but these create a data structure in the form of a chart, and we preferred to work with a data representation based on a set of trees, so we we chose to take a slightly different approach, described in more detail in section 4. Meanwhile, a number of recent developments in the SGML/XML area provided a new approach to creating the functionalities described in section 2. In our structure, each data stream and each level of annotation is stored as a separate XML file, and linking between files is done using a mechanism known as hyperlinking, which allows elements in one file to point to one or more elements in another file. Our implementation of this is very similar to the XLink and XPointer (W3C (2000c)) proposals from the World Wide Web Consortium (W3C), which are close to becoming accepted standards, and when they are established as standards, we will convert our data to conform to them. There are several tools, described in more detail in section 5 which take advantage of this design, including "knit" which makes it possible to expand the hyperlinks within one XML hierarchy to explicitly include the linked elements within a document, or to replace an original element with the links. Stylesheets of various types including XSLT (W3C (2000b)) can be used to transform or display the data. A stylesheet is a document which contains a declarative set of rules for converting one XML document into another XML document, so it can for example be used to create HTML which can then be displayed in a web browser.

4 Structure of the Corpus Annotation

As described above, the corpus annotations are stored in a number of linked XML files. In the Map Task corpus, the lowest level of structure is the timed transcription unit and we have a separate base level transcript for each speaker. Each of these base files contains a sequence of word items, other noises, and silences, which have start and end times which refer to the speech signal. All other annotation is held in separate files, and hyperlinks are used to "point" between files. Timings are generally only included in the base level, but it is a simple matter to "inherit" times upwards through a tree to other annotation levels.

Figure 26.1 shows shows the structure of the annotations for one speaker in the corpus; each box represents a separate file and arrows represent links between files (described in section 4.1). There is a separate parallel set of files for the other speaker. In this example there is another annotation level apart from the base level, gaze, which also points directly to the speech signal. Speakers can look up or down while either or both participants are speaking, so it was not clear that there was a principled

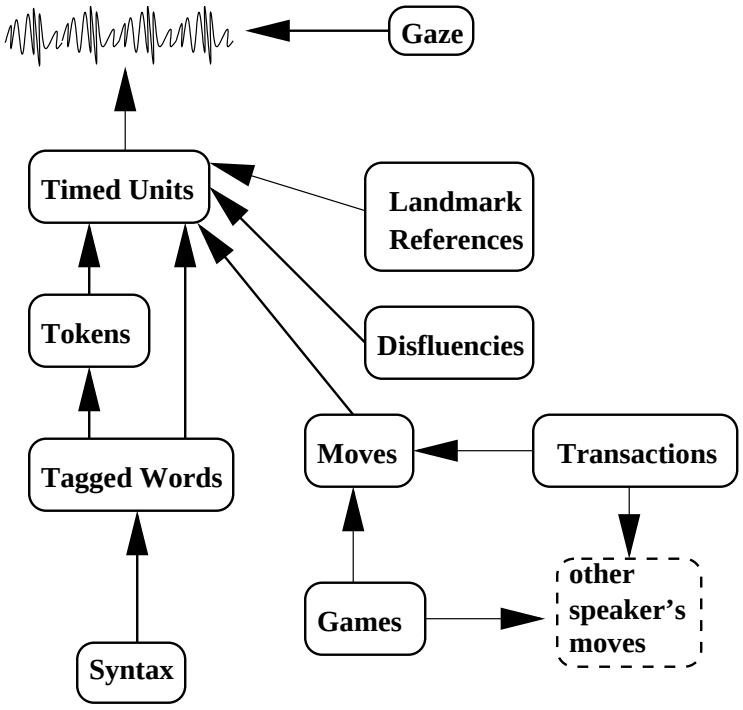


Figure 26.1: Architecture

way to link gaze to the word files. Therefore gaze elements have start and end time attributes, and when gaze is considered along with other levels of annotation, comparisons are made according to the shared timeline.

The choice of particular annotations is corpus specific - more, fewer or different annotation levels could be used without affecting the general architecture. It is also possible to use this architecture with a speech corpus which does not have timings for each word, by omitting the start and end times from the word-level elements and attaching them instead to a higher level for which timings are available, such as utterances. The architecture could also be applied to a text corpus; in this case, the word level will still be the base level and all the other levels will point to it. In this case there would only be one base level file, as there would not be separate speaker streams.

Figure 26.2 shows another view of the corpus, with a stretch of text and several overlapping annotation hierarchies which refer to it.

4.1 Links Between XML Files

In our design, a hyperlink attribute on an element points to a contiguous stretch of elements in a different file. Figure 26.3 shows a) part of a base-level timed unit file, and b) a dialogue moves file with pointers to the base file. In the base file, the elements are of type tu (timed unit) and sil (silence). Each element has three attributes: a unique identifier (ID), start time, and end time, and tu elements additionally have content, which is the transcription of the word. In the move file, each move element has four attributes: an ID, speaker, a label specifying the type of dialogue move, and an href attribute, which provides the hyperlinking. The href attribute consists of two parts, first the name of the file which the element(s) pointed to can be found in, and second a list of ID(s) of the element(s) which are pointed to. If it is a single element, just one ID is listed (e.g. id(q1ec1g.1)) but it is also possible to include a sequence of elements (e.g. id(q1ec1g.4)..id(q1ec1g.14)). In this second case all the elements between the two which are named will also be included.

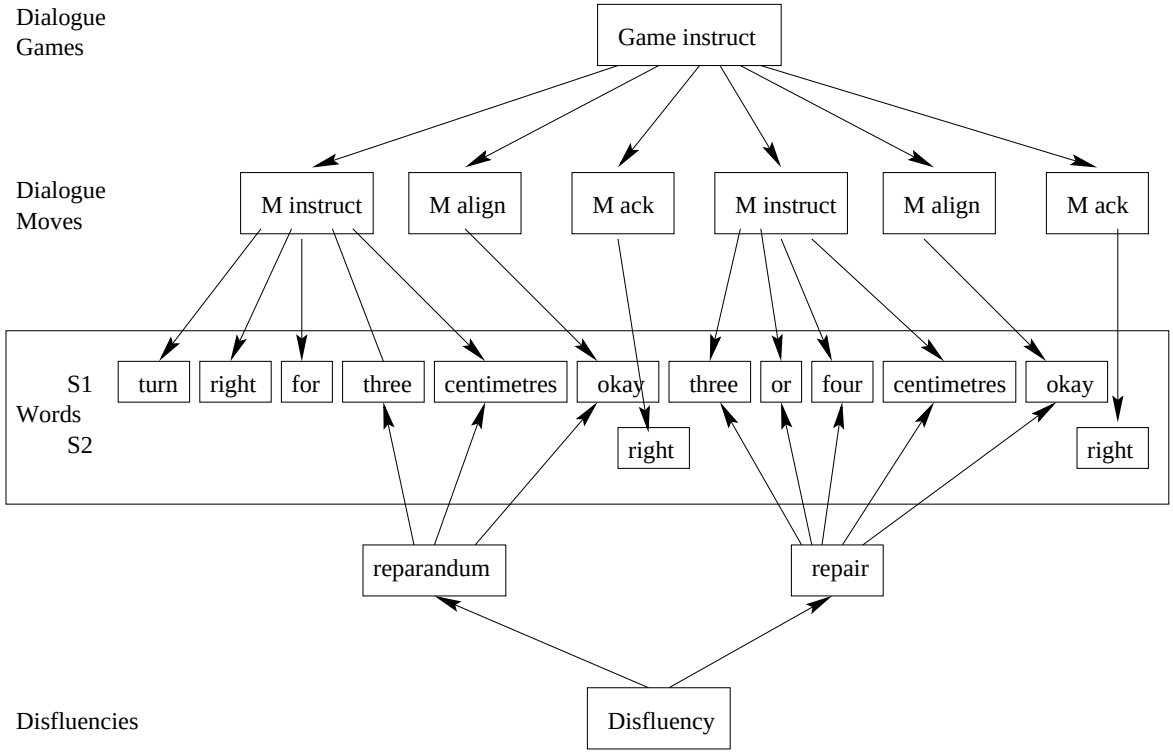


Figure 26.2: Annotation Structure

5 Working with the Data

XML coding makes the structure of a corpus explicit, and this facilitates the process of querying the corpus. There are a number of tools developed by the Edinburgh Language Technology Group (LTG (2000)) which take advantage of the hyperlink semantics used in the corpus. One example is "knit" which, depending on some user-set parameter, expands the hyperlinks either to include the linked elements within the original document, or to replace the original element with the links. This then allows queries to be performed over an entire hierarchy using one of any number or XML query languages (Marchiori (1998)). Knitting and XML querying suffice when a single annotation hierarchy is being processed, but it is more complicated to perform queries across hierarchies. The fact that our each annotation level is linked to the same base level facilitates the formulation of queries which involve more than one hierarchy. For example, one might want to find out whether a particular type of disfluency occurs more frequently during one type of dialogue move than another. The MATE workbench (McKelvie et al. (2001)) takes advantage of the hyperlink semantics, and allows queries and displays to be performed over an entire corpus. It is also possible to perform queries across more than one document using XSLT stylesheets; this is currently rather tortuous but should be simpler once XLink and XPointer become fully standardized.

The XML annotation of the corpus also allows us to provide flexible display options using stylesheets and generic software. The online Map Task Corpus Demo (Isard and Aylett (2001)) shows some flexible display solutions produced in real time using stylesheets to produce HTML, and allows the user to select a number of annotation levels and display formats.

a)

```
<!DOCTYPE timed_unit_stream SYSTEM "dtd/maptask-timed-units.dtd">
<timed_unit_stream id="tu.q1ec1.g">
<tu id="q1ec1g.1" start="0.0000" end="0.3294">okay</tu>
<tu id="q1ec1g.4" start="0.3294" end="0.8432">starting</tu>
<tu id="q1ec1g.5" start="0.8432" end="1.3702">off</tu>
<sil id="q1ec1g.6" start="1.3702" end="1.5777"/>
<tu id="q1ec1g.7" start="1.5777" end="1.8413">we</tu>
<tu id="q1ec1g.8" start="1.8414" end="2.2201">are</tu>
<sil id="q1ec1g.9" start="2.2201" end="2.3518"/>
<tu id="q1ec1g.10" start="2.3518" end="2.8722">above</tu>
<sil id="q1ec1g.11" start="2.8722" end="2.9644"/>
<tu id="q1ec1g.12" start="2.9644" end="3.0369">a</tu>
<tu id="q1ec1g.13" start="3.0369" end="3.5244">caravan</tu>
<tu id="q1ec1g.14" start="3.5244" end="3.9394">park</tu>
 ...
</timed_unit_stream>
```

b)

```
<!DOCTYPE move_stream SYSTEM  "dtd/maptask-moves.dtd" [
<!ENTITY gfile "q1ec1.g.timed-units.xml">
]>
<move_stream id="move.q1ec1.g">
<move id="q1ec1.g.move.1" who="giver" label="ready"
 href="&gfile;#id(q1ec1g.1)"/>
<move id="q1ec1.g.move.2" who="giver" label="instruct"
 href="&gfile;#id(q1ec1g.4)..id(q1ec1g.14)"/>
...
</move_stream>
```

Figure 26.3: Part of one speaker's base level timed unit transcription and corresponding dialogue move transcription

6 Discussion

Our techniques have some disadvantages, in that they require the creation of a large number of files, which are not easily human-readable as they stand, and that tools such as "knit", MATE or XSLT must be used to work with a whole hierarchy, or a set of hierarchies. However, these are outweighed by the advantages described in this paper.

There are some issues which have not yet been fully addressed; we maintain individual files using standard version control software (RCS) but we have not yet developed a coherent strategy for dealing with the knock-on changes which occur to higher-level files when a base-level file is edited. Multimodality options are also still under development. We can currently link to a speech or video using offset times from the beginning of the signal and start and end attributes on elements. However, there is for instance no software currently available which would allow us to integrate the XML files with, for example, an interface for annotating sections of a map.

No other currently available approach comes close to providing all of the functionality which we require, and because we use standards developed for other purposes, we can make use of many freely available tools, greatly reducing programming effort. We can also make use of new XML techniques as they are developed in the future. It is easy to add new levels of annotation to an existing corpus without reference to existing annotations, and editing of one annotation level has minimal effects on other levels.

Bibliography

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4):351–366.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the intelligibility of referring expressions. *Journal of Memory and Language*, 42(1):1–22.

Boyle, E., Anderson, A., and Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, 37(1):1–20.

Branigan, H., Lickley, R., and McKelvie, D. (1999). Non-linguistic influences on rates of disfluency in spontaneous speech. In *Proceedings of the 14th International Conference of Phonetic Sciences*.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson, A. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

Ducharme, B. (1999). *XML: The Annotated Specification*. Prentice Hall.

Goldfarb, C. F. (1990). *The SGML Handbook*. Oxford University Press.

Ide, N. and Priest-Dorman, G. (2000). The Corpus Encoding Standard. http://www.cs.vassar.edu/CES/.

Isard, A. and Aylett, M. (2001). The HCRC Map Task Corpus Demo. http://www.ltg.ed.ac.uk/~amyi/maptask.

Lickley, R. and Bard, E. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech*, 41(2).

LTG (2000). LTXML. http://www.ltg.ed.ac.uk/software/xml.

Marchiori, M. (1998). W3C Query Languages Workshop 98. http://www.w3.org/TandS/QL/QL98.

McKelvie, D. (1998). SDP - Spoken Dialogue Parser. Technical report, HCRC, University of Edinburgh. HCRC/RP-96, May 1998.

McKelvie, D., Isard, A., Mengel, A., Moeller, M. B., Grosse, M., and Klein, M. (2001). The MATE Workbench - an annotation tool for XML coded speech corpora. *Speech Communication*, 33(1-2):97–112. Special Issue: Speech Annotation and Corpus Tools.

Sperberg-McQueen, C. M. and Burnard, L. (1994). TEI Guidelines for Electronic Text Encoding and Interchange (P3). http://etext.lib.virginia.edu/TEI.html.

W3C (1999). Standard Generalized Markup Language (SGML). http://www.w3.org/MarkUp/SGML/.

W3C (2000a). Extensible Markup Language (XML). http://www.w3.org/xml.

W3C (2000b). Extensible Stylesheet Language Transformations (XSLT). http://www.w3.org/TR/xslt.html.

W3C (2000c). XPointer and XLink. http://www.w3.org/XML/Linking.

Paper 27

Dialogue Understanding in Dynamic Domains

Bernd Ludwig
University of Erlangen-Nuremberg
FORWISS
bdludwig@immd5-wv.uni-erlangen.de

Abstract

This paper describes an approach to dialogue understanding for dynamic applications. It shows, how from a pragmatics first perspective, dialogue situations describe multiple orthogonal dimensions of the function and content of an utterance in a task-oriented dialogue and how instantiations of such situations can be employed to control the dialogue behavior of a dialogue system as well as the analysis of individual utterances. Situations are application independent and complemented by a black box abstraction of integrated application modules. They function as a basis for reasoning about actions that are executed by an automatic system when conducting a rational dialogue. In addition, we sketch a programming language for specifying dialogue actions as a methodology to adapt a generic dialogue system to various applications.

1 Introduction

While there is a large number of dialogue systems for spoken language around, only few of them can be adapted to different domains and applications. These state-based systems rely on finite automata models or variants of them to describe the flow of dialogues the system will be able to analyze (see e.g. Aust and Oerder (1995)). This approach guarantees a system whose dialogue flow always follows a safe and a-priori known sequence of operations. In addition, it enables the developer to describe the application task (mostly an information dialogue about trains, flights, etc.) with a simple transition graph that can be compiled into a finite state machine automatically. However, as K. and D. (1997) shows, this "precompiled" dialogue structure is also of great disadvantage as it cannot react sufficiently to the requirements of a dynamically changing situation.

On the other hand, recent literature on discourse theory reveals many structural aspects that serve to describe the semantics of an utterance in terms of how information states are modified Cooper (1997). In addition, Poesio and Traum (1997) extend the classical Discourse Representation Theory (DRT) by Kamp and Reyle (1993) and incorporate so called *conversational events* as discourse referents. On the basis of this work, our paper describes a plan-based approach to define a programming language for the design of dialogue systems. The approach is similar to the GOLOG Levesque et al. (1994) implementation of the situation calculus McCarthy and Hayes (1969), and extends it with a procedural component that actually executes actions that can be inferred only in GOLOG. In this way, the dynamics of dialogues are captured formally in order to compute the reactions of an automatic dialogue system to a given natural language input. For the aim of adaptability, instead of defining application specific transition graphs, the approach tries to factor out orthogonal dimensions of language

and discourse understanding. (Dialogue) situations are used to describe the current state of affairs of a dialogue in terms of these dimensions. Reasoning about (dialogue) actions is employed to compute the system's dialogue plans.

Throughout this paper, we will refer to an example dialogue taken from the EMBASSI[1] project.

U$_1$: *Welche Filme kommen heute abend? (What films are on this evening?)*
S$_2$: *Krimis oder Spielfilme? (Thrillers or popular films?)*
U$_3$: *Ich würde gerne eine Komödie sehen. (I'd like to watch a comedy.)*
S$_4$: *Um 21:45 beginnt "Dead Man". ("Dead Man" is on at 21:45.)*
U$_5$: *Und Krimis? (and thrillers?)*
S$_6$: *"Tatort" um 20:15. ("Tatort" at 20:15).*
U$_7$: *Bis wann läuft der? (How long is that on for?)*
S$_8$: *[Er dauert] Bis 22:15. ([It is on] Till 22:15.)*
U$_9$: *Dann möchte ich den Spielfilm aufnehmen. (I'd like to have the popular film taped.)*
S$_{10}$: *Ok. Der Videorecorder ist programmiert auf Sender BR3 von 21:45 bis 23:05. (Ok. The recorder has been set to tape channel BR3 from 21:45 to 23:05.)*

2 Modeling the Application Domain

For the purpose of easy adaptability to new dialogue scenarios our approach abstracts from a certain scenario as e.g. the retrieval of train information from a data base: The data base is viewed as the application for the scenario and is described by defining the operations it is capable of. A simple data base could be described with a single operation:

$$\texttt{retrieve}(\langle\text{departure}\rangle, \langle\text{destination}\rangle, \langle\text{departure time}\rangle, \langle\text{arrival time}\rangle)$$

The parameter slots of the operation encode the relational model of the data base. More complex applications as car navigation systems, would need a much larger number of operation descriptions in order to be specified completely.

To explain how the process abstraction works, we give a formalization for the notion of "operation descriptions". The basic idea is that we know the complex data types as well as the functions (each with a set of parameters) the application uses. For the discussion here, we assume these definitions to be written in an object-oriented language, but this is not a necessary prerequisite. In the EMBASSI domain, an **AvEvent** is the data structure for entries in a electronic TV programme that can be consulted by the user or other application modules in order to retrieve information about TV programmes:

AvEvent **Info**
date: **TimeInt**
location: **AvEventLocation**
title: **Title**
genre: **Genre**

The function **Record** is responsible for getting **AvEvents** taped on a VCR:

Record **Action**
avevent: **AvEvent**
date: **TimeInt**

So, an instance is an **AvEvent** if and only if it is an **Info** (as **Info** is the super class of **AvEvent**, its *date* is a **TimeInt**, its *location* is an **AvEventLocation**, its *title* is a **Title**, and its *genre* is a **Genre**.

[1]EMBASSI is a joint project sponsored by the German Federal Ministry of Research with the goal to develop systems for multi modal assistance to operate devices. The knowledge of how to use electronic devices is delegated from the user to software, In our case, we work on speech dialogue control of an audio and video home theater.

On the other hand, **Record** is an **Action** and has got two parameters, one of type **AvEvent** and one of type **TimeInt**. The following sufficient and necessary definition holds:

$$\mathbf{Record}(a) :\Leftrightarrow \mathbf{Action}(a) \wedge \forall e : (\mathit{avevent}(a,e) \rightarrow \mathbf{AvEvent}(e) \wedge \forall t : (\mathit{date}(a,t) \rightarrow \mathbf{TimeInt}(t))$$

for all a. We allow methods as well as classes to have instances, since in natural language, references can be made to objects ("Which film did you tape?") and actions ("Did you tape the film?").

We use the above semantics for definitions of classes and methods to specify the concepts of an application domain. This approach is not restricted to object-oriented languages, but can even be applied to HTML and XML document definitions. Fortunately, there is a sublanguage of first order logic for the formulae we need in order to cover the relevant definitions of data types: Description Logics (DL; see Donini (1996)). In DL, the semantics for **Record** is written as

$$\mathbf{Record} :\Leftrightarrow \mathbf{Action} \wedge \forall \mathit{avevent}.\mathbf{AvEvent} \wedge \forall \mathit{date}.\mathbf{TimeInt}$$

The meaning of this formula is identical to the one above. DL offer decidable sound and complete algorithms for the subsumption problem. For that reason, we employ DL for representing knowledge about the application domain that delivers pragmatic constraints which are useful for the construction of the semantics of utterances.

As a consequence of abstracting from scenarios, the modules employed for computations in an application scenario are treated as black boxes by the dialogue system. They are characterized by their domain model defined as outlined above and exchange messages in KQML with the dialogue system and among each other about instances of application concepts. For example, in order to perform the user request in $\mathbf{U}_9$, the **Record** method of the responsible module m must be invoked. For this purpose, the following instances of application concepts are necessary:

Time t_s	**Time** t_e	**Date** d_s	**Date** d_e	**TimeInt** i	**AvEvent** f
t_s.min=45	t_e.min=05	d_s.year=2000	d_e.year=2000	i.startdate=d_s	f.title="Dead Man"
t_s.hour=21	t_e.hour=23	d_s.month=05	d_e.month=05	i.starttime=t_s	f.date=i
		d_s.day=22	d_e.day=22	i.enddate=d_e	f.location=BR3
				i.endtime=t_e	f.genre=comedy

The actual method call would be: m.**Record**(f). It implies the existence of a **TimeInt** i and an **AvEvent** f with the values above. All information contained in the call must be available as data for the dialogue system in order to resolve references to methods as well as to instances used as arguments. Discourse Representation Structures (DRS) – the "data structure" of DRT – do just this: They store a set of discourse referents (the names of instances) and a set of relations about them. The following DRS contains the same information as the above method call:

$$\rho := \left[\begin{array}{l} d_s\ d_e\ t_s\ t_e\ i\ f\ r\ m \\ \hline \mathbf{Date}(d_s)\ \mathbf{Date}(d_e) \\ \mathrm{year}(d_s, 2000)\ \mathrm{year}(d_e, 2000)\ \mathrm{month}(d_s, 05)\ \mathrm{month}(d_e, 05)\ \mathrm{day}(d_s, 22)\ \mathrm{day}(d_e, 22) \\ \mathbf{Time}(t_s)\ \mathbf{Time}(t_e) \\ \mathrm{min}(t_s, 45)\ \mathrm{min}(t_e, 05)\ \mathrm{hour}(t_s, 21)\ \mathrm{hour}(t_e, 23) \\ \mathbf{TimeInt}(i)\ \mathbf{AvEvent}(f) \\ \mathrm{startdate}(i, d_s)\ \mathrm{title}(f, \text{"Dead Man"})\ \mathrm{starttime}(i, t_s)\ \mathrm{date}(f, i)\ \mathrm{enddate}(i, d_e) \\ \mathrm{location}(f, \mathrm{BR3})\ \mathrm{endtime}(i, t_e)\ \mathrm{genre}(f, \mathrm{com}) \\ \mathbf{AvEventLocation}(\mathrm{BR3})\ \mathbf{Comedy}(\mathrm{com}) \\ \mathbf{Module}(m)\ \mathbf{Record}(r)\ \mathrm{agent}(m, r)\ \mathrm{avevent}(r, f) \end{array}\right]$$

This DRS is communicated to m as the content of a KQML message. The KQML performative `achieve` is used in order to make m execute the desired recording. The execution of the recording may produce different results depending on several factors, including the current state of the hardware or the availability of the requested **AvEvent**.

These factors influence the flow of the current dialogue that cannot be predetermined by a transition diagram, as this required to foresee all possible eventualities. In the example, different reactions to

$\mathbf{U}_9$ are possible depending on whether the system was able to execute the recording or not. A diagram would have to provide a transition for each possible result of the command in order to continue the dialogue in a user friendly manner. As there can be many sources of errors the diagram gets complex quickly. In the EMBASSI application, there are about 50 operations – each depending on a number of preconditions and producing different effects if the state of the application changes. For that reason – and in particular, to enable the adding of new operations when an application module extends its functionality – the modules engaged in the execution of a user request respond by giving feedback about the outcome of each performed operation. If in the example the recording was successful, the response (a **tell** message in KQML) would be:

$$\alpha_1 := \left[\frac{\rho}{\mathbf{Task}(\rho)\ \mathrm{status}(\rho, \mathrm{ok})\ \mathbf{OK}(\mathrm{ok})}\right]$$

However, if the video recorder was already occupied, the system would communicate this information to the user:

$$\alpha_2 := \left[\begin{array}{l} \dfrac{\rho\ \epsilon}{\mathbf{Task}(\rho)\ \mathrm{status}(\rho, \mathrm{failed})\ \mathbf{Error}(\mathrm{failed})\ \mathrm{failed}(\rho, \epsilon)\ \mathbf{Reply}(\epsilon)} \\ \epsilon : \left[\dfrac{v}{\mathbf{Vcr}(v)\ \mathrm{status}(v, \mathrm{occupied})\ \mathbf{Error}(\mathrm{occupied})}\right] \end{array}\right]$$

Of course, this would alternate the dialogue:

$\mathbf{S}_{10'}$**:** *The VCR is occupied from 21:45 on.*

If available, the called module may propose a different plan to still achieve the initial goal: Assuming in our example, that additionally to the (default) device **Vcr** v for recordings, there is also a hard disk that could also record **AvEvent**s, the following message could be sent to the dialogue system instead of α_2:

$$\kappa := \left[\begin{array}{l} \dfrac{\rho\ \epsilon}{\mathbf{Task}(\rho)\ \mathrm{status}(\rho, \mathrm{failed})\ \mathbf{Error}(\mathrm{failed})\ \mathrm{failed}(\rho, \epsilon)} \\ \mathbf{Task}(\pi)\ \mathrm{satisfies}(\rho, \pi)\ \mathbf{Proposal}(\pi) \\ \epsilon : \left[\begin{array}{l} v \\ \hline \mathbf{Vcr}(v)\ \mathrm{status}(v, \mathrm{occupied}) \\ \mathbf{Error}(\mathrm{occupied}) \end{array}\right] \quad \pi : \left[\begin{array}{l} hdd\ r \\ \hline \mathbf{Record}(r) \\ \mathrm{device}(r, hdd)\ \mathbf{HardDisk}(hdd) \end{array}\right] \end{array}\right]$$

To be consistent with the domain model, we have to assume here that **Record** has got an additional parameter for the **Device** to be used for the recording. Using the pragmatic information obtained from the response, the dialogue system could now continue the dialogue as follows:

$\mathbf{S}_{11}$**:** *Record on hard disk instead?*

As the above discussion indicates, a dialogue system for such a scenario has to cope with dynamically changing knowledge about the state of affairs in the application. For example, as a consequence of programming the vcr from 21:45 till 23:05, in a subsequent dialogue turn another recording in this interval would be impossible.

3 Integration of Discourse and Application

3.1 Incorporating Pragmatic Actions into Discourse Structure

The flow of a dialogue is therefore determined by the conditions that satisfy requests from the dialogue system at the current moment. In order to reason about satisfiability under a given situation, a notion of time is required. As transitions between situations are a consequence of executing actions, we have adopted the ideas from Situation Calculus (see Levesque et al. (1994)) to represent and reason about information changing in time. Situation calculus is a first-order theory of how a formally modeled world

may change as a consequence of executing actions. Situations result from actions and are represented by recursively nested terms of actions beginning with an initial situation.

A major difficulty with using situation calculus in dialogue processing is that actions have an ontological state different from objects. As a consequence, no reference to actions can be made – which is quite common in spoken dialogues. This problem of reference is addressed in DRT: Objects for whom predicates hold are stored as discourse referents and exist for the whole discourse, not only for a single formula. We exploit this advantage by introducing *events* and *situations* in the ontology for dialogues. Situations are ordered via a linear order before which has the initial situation as minimal element. *Events* are associated to situations via the function situationevent. *Events* may be of different type depending on their function. Up to now, three types have been used:

- **Task**: contains the description of an action to be executed by an application module.
- **Reply**: contains the description of the results of a computation.
- **Proposal**: contains the description of an action proposed by an application module in order to fulfill a user request.

To give an example, the DRS for the situation s_0 after ρ has been sent to module m looks like this (ρ... points to the definition of ρ above):

$$\Sigma_0 := \left[\begin{array}{l} \mathsf{s}_0\ \rho \\ \hline \mathsf{situation}(\mathsf{s}_0)\ \mathbf{Task}(\rho)\ \mathrm{status}(\rho, \mathrm{open})\ \mathsf{situationevent}(\mathsf{s}_0, \rho) \\ \rho : \ldots \end{array}\right]$$

The situation after α_2 has arrived is

$$\Sigma_1 := \left[\begin{array}{l} \mathsf{s}_0\ \mathsf{s}_1\ \rho\ \epsilon \\ \hline \mathsf{situation}(\mathsf{s}_0)\ \mathbf{Task}(\rho)\ \mathrm{status}(\rho, \mathrm{open})\ \mathsf{situationevent}(\mathsf{s}_0, \rho) \\ \mathsf{before}(\mathsf{s}_0, \mathsf{s}_1)\ \mathsf{situation}(\mathsf{s}_1)\ \mathbf{Reply}(\epsilon)\ \mathsf{situationevent}(\mathsf{s}_1, \epsilon) \\ \mathbf{Task}(\rho)\ \mathrm{status}(\rho, \mathrm{failed})\ \mathbf{Error}(\mathrm{failed})\ \mathrm{failed}(\rho, \epsilon) \\ \rho : \ldots\ \ \epsilon : \left[\begin{array}{l} v \\ \hline \mathbf{Vcr}(v)\ \mathrm{status}(v, \mathrm{occupied})\ \mathbf{Error}(\mathrm{occupied}) \end{array}\right] \end{array}\right]$$

Σ_1 describes the situation after the response from module m has been integrated. The information in Σ_1 allows to infer whether the user request has been fulfilled or is still pending:

$$\begin{array}{rcl} \text{task-complete}(T, S) & \leftarrow & \mathsf{situation}(S) \wedge \mathsf{situationevent}(S, T) \wedge \text{complete}(T) \\ \text{task-complete}(T, S) & \leftarrow & \mathsf{situation}(S) \wedge \mathsf{situationevent}(S, T) \wedge \neg\text{complete}(T) \wedge \\ & & \mathsf{before}(Y, S) \wedge \text{task-complete}(T, Y) \\ \text{complete}(X) & \leftarrow & \mathbf{Task}(X) \wedge \mathrm{status}(X, \mathrm{failed}) \\ \text{complete}(X) & \leftarrow & \mathbf{Task}(X) \wedge \mathrm{status}(X, \mathrm{ok}) \end{array}$$

By employing such rules and the information about events (in particular, their status) as shown in the example above, the flow of a dialogue can be controlled by the dialogue system. A-priori defined successor states in a transition diagram are substituted by describing the current state explicitly with the help of a DRS. Instead of transitions between states, dialogue operations are invoked, as we will discuss later. Rules as those shown above are used to evaluate preconditions for dialogue operations, while the execution of a dialogue operation creates a new situation reflecting the new dialogue state.

In the EMBASSI system feed back from an application module often arrives with a long delay. For example, when a programmed recording has been completed several hours after the user had told the system to do so. For this purpose, user requests must be stored and even be available for reference in order to define the semantics of questions like "Have I already programmed the vcr to tape the foot ball match?" The answer would be "yes" if there was a situation before the last one in which a **Task** for recording the corresponding **AvEvent** had been started.

Up to now, the presentation has addressed only the issue of invoking functions defined in the application domain without explaining how natural language input is analyzed in order to produce such function calls. We will now turn to the discussion of this point.

3.2 Basic Dialogue Operations

Our approach on dialogue understanding uses Discourse Representation Theory (DRT) as a basis for modeling dialogues. This means that a dialogue is represented as a Discourse Representation Structure (DRS) where new contributions (i.e. user and system utterances) are added incrementally during the dialogue. Three elementary operations use and eventually modify the content of the DRS (for a similar idea see Poggi and Pelachaud (2000)):

- **add information:** Update the content of the DRS with the given information if no inconsistency arises by doing so.
- **test:** Verify the satisfiability of the content.
- **act:** Test whether the preconditions for the described operation hold. If so, invoke the module that executes the operation and update the DRS according to the result returned.

On the other hand, any utterance can be declarative, interrogative or imperative. How is the function of an utterance as chosen by the speaker related to the assignment of a dialogue act by the hearer and, eventually, his reaction? While the speaker's motivations for selecting a particular form for an utterance are unavailable to the hearer, he or she can at least assume that the utterance has got the communicative goal to make him (or her) react in some way. A direct mapping between the forms an utterance can have and the operations that can be performed given a discourse representation as introduced above would be the following:

- **add information**: *state something – declarative*
- **test**: *ask something – interrogative*
- **act**: *command to do something – imperative*

So, if the hearer analyzes an utterance to be of one of the three types above, the reaction intended by the communicative act of uttering something would be the execution of the corresponding dialogue operation. However, results from conversational analysis show that there is no such direct relationship between form and function of an utterance. Our hypothesis is that for rational dialogues between cooperative and honest dialogue participants we can explain the relation between the form of an utterance chosen by the speaker and the function assigned to it by the hearer if we take information about the dialogue context and the dialogue participants into account.

Consider $\mathbf{U}_9$ in the example. Although it is a declarative utterance, it is certainly intended as a request to the hearer to tape the film, and not as a pure information about the intention. However, as $\mathbf{U}_9$ has got this declarative aspect, it can be interpreted as

add information(*I'd like to have the film taped.*)

As an effect of this **add** operation the updated discourse representation contains a new information: it is the hearer's intention to tape the film "Dead Man" at 21:45. The reaction depends on the communicative behavior of the hearer: if cooperative, he would try to satisfy the intention by

act(*tape the film*)

This requires some pragmatic action to be taken whose results would be integrated into the updated discourse representation, as shown in the DRS α_1, α_2, and κ above. So, the declarative utterance has actually been interpreted as the imperative "Tape the film!".

If the hearer did not intend to satisfy requests recognized during a conversation or did not behave cooperatively, possible reactions could be "Are you sure this is the right kind of film for you.", "I prefer sports events", or "Then you should program your vcr correctly." Similarly, the utterance $\mathbf{U}_1$ is directly analyzed as an instruction to **test**. Even in this case, the answer depends on the pragmatic capabilities of the hearer. Normally, a TV programme has to be consulted in order to give an answer like $\mathbf{S}_4$ or $\mathbf{S}_6$.

The effect of **test** would be an update of the discourse structure: information about the state of the application (and therefore about what the dialogue is about) is incorporated and used as content

for further system utterances. So, computations carried out by application modules are related to their corresponding dialogue modules via the mechanism outlined above.

3.3 Complex Dialogue Operations

Dialogues have got a normally quite complex coherence structure that has to be reconstructed correctly by a dialogue system. As the previous section showed, knowledge of the attentional structure is important even in order to construct correct semantic interpretations for utterances. In our approach, the state of dialogue goals is controlled by means of

- the notion of **obligation** as introduced by Traum and Allen (1994) in order to keep track of requests initiated by the user.

  As explained above, $\mathbf{U}_9$ is first processed as a declarative utterance updating the current discourse representation in a way that it is now **obliged** to do something. The system executes an **act** which itself updates the discourse representation.

- the incorporation of pragmatic actions into the discourse structure.

  In Sect. 3.1, we showed how the dialogue system is able to determine the state of tasks it had delegated to application modules. Reasoning about the updated discourse representation, the system may infer whether a user request has been completed or is still pending.

This procedural approach to the automatic analysis of the function(s) of utterances in dialogues allows to define the semantics for a number of modalities and performatives. Analogously to **want** (that, as shown, is defined as **oblige to act**) **ask** is understood as **oblige to test** and **state** as **oblige to add information**.

Oblige is not the only notion necessary for understanding dialogues. Additionally, one needs definitions for modalities like *must*, *can*, and social conventions of conversation like *greet* or *thank*.

4 Preconditions for Basic Operations

The aim of analyzing natural language input is to construct a unique description of an operation to be executed by the dialogue system itself or some application module. For this purpose, during processing a word lattice, DRS are constructed as representations for operation descriptions. When ambiguities induce different descriptions, the utterance is marked accordingly preventing basic operations from being executable. To mark an utterance, information about it is incorporated in the DRS for the current dialogue. In this way, the parsing module and the dialogue module share common data about the utterance parsed. When no basic operation is executable, the utterance cannot be considered as *grounded* and a clarification dialogue must be started as explained later. By establishing this "break point" between discourse and application pragmatics, we distinguish clearly discourse from task structure which are known to be non-isomorphic. On the other hand, the common data structure is a basis for abstracting from discourse and application functionality and by reasoning about the content of such a DRS one can define application independent dialogue strategies for handling misunderstandings.

4.1 Syntax of Utterances

Parsing a word lattice received from the speech recognizer involves analyzing the categories of words and phrases as well as the syntactic relations between phrases. They are used to find semantic dependencies between phrases (see Abney (1991)). Consider $\mathbf{U}_1$ in the example. The chart parser segments it into the following chain of chunks:

[*What films*] [*are*] [*on*] [*this evening*]?

This segmentation is unique with respect to the used chunk grammar. The DRS for the dialogue is updated as follows:

$$\Sigma_0 := \left[\begin{array}{l} \mathsf{s}_0\ \alpha \\ \hline \mathsf{situation}(\mathsf{s}_0)\ \mathbf{ConversationalEvent}(\alpha)\ \mathsf{situationevent}(\mathsf{s}_0, \alpha) \\ \mathsf{has\text{-}chunk}(\alpha, C_1)\ \mathsf{has\text{-}chunk}(\alpha, C_2)\ \mathsf{has\text{-}chunk}(\alpha, C_3)\ \mathsf{has\text{-}chunk}(\alpha, C_4) \\ \mathbf{Chunk}(C_1)\ \mathbf{Chunk}(C_2)\ \mathbf{Chunk}(C_3)\ \mathbf{Chunk}(C_4) \\ \mathsf{syntacticlevel}(C_1, \mathrm{unique})\ \mathsf{syntacticlevel}(C_2, \mathrm{unique})\ \mathsf{syntacticlevel}(C_3, \mathrm{unique}) \\ \mathsf{syntacticlevel}(C_4, \mathrm{unique}) \end{array}\right]$$

4.2 Intension

For the semantics of an utterance, we have to distinguish between its meaning in terms of a terminology (*intension*) and the enumeration (i.e. *extension*) of all objects satisfying the intension. In $\mathbf{U}_1$, there are two readings for the chunk *What films* on the basis of the terminology for the application. This results in two different DRS for the meaning of the utterance:

$$\beta_1 := \lambda f.\left[\begin{array}{l} a\ f\ \mathrm{film} \\ \hline \mathbf{TakePlace}(a)\ \mathrm{date}(\mathsf{a}, \mathrm{today}) \\ \mathbf{PartOfDay}(a, \mathrm{ev}) \\ \mathrm{agent}(a, f)\ \mathbf{AvEvent}(f) \\ \mathrm{genre}(f, \mathrm{film})\ \mathbf{Feature}(\mathrm{film}) \end{array}\right] \qquad \beta_2 := \lambda f.\left[\begin{array}{l} a\ f\ \mathrm{film} \\ \hline \mathbf{TakePlace}(a)\ \mathrm{date}(\mathsf{a}, \mathrm{today}) \\ \mathbf{PartOfDay}(a, \mathrm{ev}) \\ \mathrm{agent}(a, f)\ \mathbf{AvEvent}(f) \\ \mathrm{genre}(f, \mathrm{film})\ \mathbf{Thriller}(\mathrm{film}) \end{array}\right]$$

The reason is: The word *film* has an ambiguous meaning: $\mathsf{avevent} \cap \forall \mathsf{genre.feature}$ and $\mathsf{avevent} \cap \forall \mathsf{genre.thriller}$. This fact is reflected in the following entries for the DRS of the dialogue (added to Σ_0):

$$\Sigma_1 := \left[\begin{array}{l} \mathsf{s}_0\ \alpha \\ \hline \mathsf{syntacticlevel}(C_3, \mathrm{unique})\ \mathsf{intensionallevel}(C_1, \mathrm{ambiguous})\ \mathsf{intensionallevel}(C_2, \mathrm{unique}) \\ \mathsf{intensionallevel}(C_3, \mathrm{unique})\ \mathsf{intensionallevel}(C_4, \mathrm{unique}) \\ \mathsf{content}(C_1, \phi_1)\ \mathsf{content}(C_1, \phi_2) \\ \phi_1 : \lambda \mathsf{f}.\left[\begin{array}{l} f\ \mathrm{film} \\ \hline \mathbf{AvEvent}(f)\ \mathrm{genre}(f, \mathrm{film}) \\ \mathbf{Feature}(\mathrm{film}) \end{array}\right] \quad \phi_2 : \lambda \mathsf{f}.\left[\begin{array}{l} f\ \mathrm{film} \\ \hline \mathbf{AvEvent}(f)\ \mathrm{genre}(f, \mathrm{film}) \\ \mathbf{Thriller}(\mathrm{film}) \end{array}\right] \end{array}\right]$$

4.3 Extension

When an intension has been found, the discourse referents satisfying this description must be determined. This amounts to finding the referents in the dialogue that satisfy the constraints given by the semantics of the current utterance. Often, referential expressions are a source for ambiguity in this process. For example, in $\mathbf{U}_9$, the noun phrase *the popular film*, can be satisfied only by inferring that among others a popular film is a comedy; for this genre a discourse referent can be found: the one referring to "Dead Man" introduced in $\mathbf{S}_4$. These facts are described by the DRS ρ. According to Kamp and Reyle (1993) its meaning is in first order logic:

$$\begin{aligned} \exists i, f, r: \quad & \mathbf{AvEvent}(f) \cap \mathbf{TimeInt}(i) \cap \mathrm{date}(f, i) \\ \cap \quad & \mathbf{AvEventLocation}(\mathrm{BR3}) \cap \mathrm{location}(f, \mathrm{BR3}) \\ \cap \quad & \mathrm{title}(f, \text{"Dead Man"}) \cap \mathbf{Comedy}(\mathrm{com}) \cap \mathrm{genre}(f, \mathrm{com}) \\ \cap \quad & \mathbf{Record}(r) \cap \mathrm{avevent}(r, f) \end{aligned}$$

I.e. i, f, r are in the extension of the concept description for

$$\begin{aligned} & \mathbf{Record} \cap \exists \mathrm{avevent}.(\mathbf{AvEvent} \cap \exists \mathrm{date}.\mathbf{TimeInt} \\ \cap \quad & \exists \mathrm{location}.\mathbf{AvEventLocation} \cap \exists \mathrm{title}.\mathbf{Title} \cap \exists \mathrm{genre}.\mathbf{Comedy}) \end{aligned}$$

which is subsumed by the intension of the utterance:

$$\begin{aligned} & \mathbf{Record} \cap \forall \text{avevent}.(\mathbf{AvEvent} \\ \cap \quad & \forall \text{date}.\mathbf{TimeInt} \cap \forall \text{location}.\mathbf{AvEventLocation} \\ \cap \quad & \forall \text{title}.\mathbf{Title} \cap \forall \text{genre}.\mathbf{Comedy}) \end{aligned}$$

In Sect. 2, this expression has been used to define the semantics of **Record** and **AvEvent**. Reasoning about satisfiability of concept descriptions as outlined here is used for the pragmatics driven construction of semantics for natural language phrase. Again, information about the satisfiability of the constructed meaning is stored in the DRS for the current dialogue.

4.4 Coherence of Utterances

For a correct understanding of an utterance in a dialogue, its coherence to previous utterances on the functional as well as the semantic level must be analyzed properly in order to *ground* the new contribution (see Traum and Hinkelman (1991)). In Ludwig et al. (2000), we have sketched a computational model for *grounding* that uses a first order logic of partial information. The central idea is that the DRS for an utterance describes a partial model for an operation description. This model may have various extensions (that are "less partial") depending on the information added in later dialogue turns. An utterance is coherent to a previous conversational event if it extends the model for the associated operation description. In this way, the information state reflected by the DRS for the current dialogue is updated continuously. By proving its consistency with the information state about the pragmatic situation maintained by the employed application modules, the dialogue system detects contradictions between the content of user utterances and the current state of the application.

4.5 Complex Operations Control the Dialogue Strategy

If the preconditions for the basic dialogue operations are fulfilled, an application module is invoked via an appropriate KQML message. In this way, evaluating the satisfiability of the user utterance is delegated to the responsible pragmatic component. The result of this evaluation is incorporated in the dialogue as described above. But what happens, if the facts in the DRS for the current dialogue block the invocation? In this case, task and discourse structure are not isomorphic. In our approach, such "exceptions" are handled as tasks for the dialogue system which is seen as an always present application (for discourse). As shown above, various submodules like the parser add facts about their computations to the dialogue description as application modules do. Complex dialogue operations define the communicative behavior of the dialogue system. As an example, let us discuss how utterance $\mathbf{S}_2$ is the result of a complex dialogue operation: In Sect. 4.2, the DRS Σ_1 shows the intensional ambiguity of *film*. ϕ_1 and ϕ_2 are the possible interpretations leading to a unique semantic representation for the whole utterance. Therefore, a **Proposal** is generated as in DRS κ which is uttered as $\mathbf{S}_2$. This utterance is intended as a request to repair due to the conditions just sketched. So, the focus is shifted to repairing the ambiguity. Subsequent utterances are expected to be coherent with the new focus. From $\mathbf{U}_3$ it can be inferred that the user interprets *film* in the sense of ϕ_1 thereby repairing the ambiguity. Just now the preconditions for a **test** operation are satisfied. Its result contains the information which is uttered in $\mathbf{S}_4$.

5 Conclusions and Future Work

There are a number of approaches on modeling rational interaction in dialogues (e.g. Sadek (1999); Carberry and Lambert (1999); Hulstijn (2000). Most of them base on modal logics to describe modalities and (auto)epistemic operations of dialogue participants. Sometimes, the computational tractability of the employed logics remains unclear. Following the ideas of GOLOG we showed how in a hybrid approach that mixes reasoning about actions and situations with procedurally executing them, we can

still express fundamental principles of rational interaction without using modal operators extensively. What we lose is that not the system, but only the "dialogue programmer" can reason about interaction. This loss of expressivity does not prevent key requirements on a dialogue system for natural language in real applications as there is no need to completely change the communicative behavior. The power of our approach still lies in its ability to react on dynamically changing situations and therefore to conduct flexible dialogues that are "sensitive" to the current state of affairs with the application. Often flexibility is obtained by losing robustness. When giving up the idea of finite state dialogue models, as a substitute to classify the function of utterances for a dialogue, one needs a model of dialogue acts. Stolcke (2000); Warnke (1997, 1999) show that models with a large set of dialogue acts make it difficult to assign dialogue acts to an utterance due to a statistical distribution. Only a few of them occur frequently. This has got a negative impact on the average recognition rate when tagging a corpus with dialogue acts. As a consequence, often the important acts are misrecognized. Our approach relies on a minimal set of dialogue acts (that we called basic dialogue operations) and describes the dialogue situation by a number of orthogonal features that may be computed from the information about the current dialogue. Recognition of this set of dialogue acts can be supported much more effectively by the analysis of frequencies than it would be possible for a larger set. Additionally, the dynamic computation of features allows for the definition of dialogue processing procedures that may be used to guide and modify the communicative behavior of the dialogue system.

As next steps, we plan to incorporate user profiles into the generation of **Proposals** when a large number of alternatives has to be presented to the user. Additionally, we are working on the semantics of function words in order to be able to process more than a single (pragmatic) action in an utterance.

Bibliography

(1995). *ESCA Workshop on Spoken Dialogue Systems*, Aalborg, Denmark.

Abney, S. (1991). Parsing by chunks. in: Berwick et al. (1991).

Aust, H. and Oerder, M. (1995). Dialogue control in automatic inquiry systems. in: lud (1995).

Berwick, R., Abney, S., and Tenny, C., editors (1991). *Principle-based Parsing.* Kluwer.

Brewka, G., editor (1996). *Foundations of Knowledge Representation.* CSLI Publications.

Carberry, S. and Lambert, L. (1999). A process model for recognizing communicative acts and modeling negotiation subdialogues. *Comp. Ling.*, 25(1):1–54.

Cassell, J. e., editor (2000). *Embodied Conversational Agents.* MIT Press.

Cooper, R. (1997). Information states, attitudes and dialogue. In *Proceedings of the Second Tbilisi Symposium on Language, Logic and Computation.*

Donini, F. e. (1996). Reasoning in description logics. in: Brewka (1996).

Hulstijn, J. (2000). *Dialogue Models for Inquiry and Transaction.* PhD thesis, University Twente.

K., W. and D., N. (1997). Integrating multiple cues for spoken language understanding. In *Proceedings of the CHI'95*, pages 131–5.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic.* Kluwer.

Levesque, H., Reiter, R., Y., L., Fangzhen, L., and Scherl, R. (1994). Golog: A logic programming language for dynamic domains. *Journal of Logic Programming*, 19(20):59–84.

Ludwig, B., G., G., and Niemann, H. (2000). An inference-based approach to the interpretation of discourse. *Language and Computation*, 1(2):261–76.

McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artifical intelligence. in: Meltner and Michie (1969).

Meltner, B. and Michie, D., editors (1969). *Machine Intelligence 4*. Edinburgh UP.

Poesio, M. and Traum, D. (1997). Towards an axiomatisation of dialogue acts. In Hulstijn, J. and Nijholt, A., editors, *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*.

Poggi, I. and Pelachaud, C. (2000). Performative facial expressions in animated faces. in: Cassell (2000).

Sadek, D. (1999). Design considerations on dialogue systems: From theory to technology — the case of artimis. In *ESCA Workshop Interactive Dialogue in Multi-modal Systems*.

Stolcke, A. e. (2000). Dialogue act modelling for automatic tagging and recognition of conversational speech. *Comp. Ling.*, 26(3):339–74.

Traum, D. and Allen, J. (1994). Discourse obligations in dialogue processing. In *Proceedings of ACL 94*, pages 1–8.

Traum, D. and Hinkelman, E. (1991). Conversation acts in task-oriented spoken dialog. *Computational Intelligence*, 8(3):575–99.

Warnke, V. e. (1997). Integrated dialog act segmentation and classification using prosodic features and language models. In Kokkinakis, G., Fakotakis, N., and Dermatas, E., editors, *Proceedings of the 5th European Conference on Speech Communication and Technology*, volume 1, pages 207–10.

Warnke, V. e. (1999). Discriminative estimation of interpolation parameters for language model classifiers. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 525–28.

Part V

Mental States & Dialogue

Paper 28

Toward a mechanistic psychology of dialogue: The interactive alignment model

Simon Garrod & Martin Pickering
Universities of Glasgow and Edinburgh

Researchers in language are increasingly coming to recognise that dialogue represents the most basic form of language use and so deserves special attention (see e.g. Clark (1996); Linnell (1998)). Yet, mechanistic psychological accounts of language processing concerned with such things as word production or recognition, syntactic parsing, lexical or syntactic ambiguity and so on are almost exclusively based on the study of monologue. In Clark's terms such accounts deal only with 'language as product'. In contrast, psychological studies of dialogue follow what Clark calls the 'language as action' tradition. To the extent that they address mechanistic questions these tend to be directed at dialogue strategies, such as how and when we infer common ground, rather than basic language processes, such as lexical selection or parsing.

In this talk we present a mechanistic account of dialogue processing which contrast clearly with the standard monologue based accounts. Whereas monologue based accounts treat language production and language comprehension as distinct autonomous processes, the interactive alignment model assumes that they are closely coupled to each other in dialogue. One of the basic claims of the model is that dialogue participants align their linguistic representations at all levels, the lexical, syntactic and semantic as well as at the level of the situation model underlying the dialogue. We will first consider the evidence in support of the interactive alignment account and then consider some of the implications of this account for language production and comprehension processes.

We then highlight three aspects of dialogue processing that the account addresses:

(1) The production and comprehension of what we shall call dialogue routines, Aijmer (1996),

(2) The notion of operational common ground and how it can be established through a simple dialogue inference mechanism, and,

(3) The origins of inner loop self-monitoring in language production (Postma, 2000).

All of these aspects of language processing in dialogue present challenges for traditional processing accounts. Yet, they are explained in a relatively straightforward way by the interactive alignment model.

The interactive alignment model is motivated by the idea that monologue and dialogue processing are interestingly different. The final part of the talk will report experiments which indicate that the two kinds of processing can be observed in group discussions. Fay et al. (2000) showed that in small groups (about 5 members) dialogue processing predominates, whereas in large groups (over 10 members) discussion is more like serial monologue. The experiments also demonstrated that the two kinds of language processing have interestingly different consequences for the alignment of the representations among group members. In small groups members are influenced by those they directly interact with, whereas in large groups they are influenced by dominant speakers. This result is in line with the basic predictions of the alignment model for small group language processing and the autonomous information transfer account for the large group language processing.

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

Bibliography

Aijmer, K. (1996). *Conversational routines in English: Convention and creativity.* Longman.

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

Fay, N., Garrod, S., and Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, 11(6).

Linnell, P. (1998). *Approaching Dialogue: Talk interaction and contexts in a dialogue perspective*, volume 3. John Benjamins.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, 77:97–131.

Paper 29

An Exploration of the Complex Structure and Process of Grounding in two communicative contexts, face-to-face and videoconferencing

Alison Newlands
University of Strathclyde
Department of Psychology

Abstract

This paper will discuss some of the factors that can effect the process of establishing shared understanding in task-oriented dialogues. This process is examined in terms of the complex structure of Conversational Games employed by participants in face-to-face and desktop video-conferencing. Two versions of a desk-top video conferencing system where examined, they varied in the type of audio channel available to users (full duplex vs half duplex audio channels). The theoretical framework for this research is the Collaborative model of communication proposed by Clark and colleagues (see for example, Clark and Wilkes-Gibbs (1986); Clark and Schaefer (1987, 1989); Isaacs and Clark (1987)). The results of Conversational Games Analysis showed that context did have an effect on the distribution of Conversational Games; the frequency of instructions, alignments and open-ended questions did vary with communicative context. However, these changes did not significantly alter the degree of complexity of the structure of embedded Games when comparing dialogues from face-to-face with video conferencing contexts. The analysis highlights the fact that Conversational Games are very often nested, or embedded, within other Games. Complex structures of embedding occur more frequently than Games without embedded Games. The findings raise several issues which need to be taken into consideration when discussing models of communication.

1 Introduction

1.1 The Collaborative Model of Communication and the Process of Grounding

The Collaborative Model is based upon the premise that communication is a joint activity, "a collective activity of the first order" (Clark and Brennan, 1991, p. 128). As in all collaborative activities, participants need to coordinate both the content of the activity and the process of the activity. Both forms of collaboration are required during communication if participants are to reach mutual understanding of each other's utterances. Participants need to coordinate the process of communication (for example, ensuring smooth transition of speaker turns) whilst at the same time collaborating on the content of the conversation, "working together in regular ways to produce evidence of a shared understanding." (Wilkes-Gibbs, 1995, p. 241). This is the basis for the development of shared understanding, or 'common ground', and is achieved through the process of 'grounding' (Clark and Wilkes-Gibbs, 1986).

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).

During conversation, speakers and addressees ensure that they have similar conceptions of the meaning of an utterance before they proceed with the conversation (Clark and Wilkes-Gibbs, 1986). Participants in a conversation need to establish that they have attained some level of 'shared knowledge'. The concept of shared knowledge has been defined in various ways. For example, it has been referred to as 'common knowledge' (Lewis, 1969), or as 'mutual knowledge' Schiffer (1972). Schiffer defines mutual knowledge in the following manner:

A and B mutually know that p $=_{def}$

(1) A knows that p

(1') B knows that p

(2) A knows that B knows that p

(2') B knows that A knows that p

(3) A knows that B knows that A knows that p

(3') B knows that A knows that B know that p

Etc., ad infinitum.

If Schiffer's definition of mutual knowledge was applied to everyday conversations, then participants would have an infinity of statements to check before they could be assured that they had understood each other. Clark and Marshall (1981) argue that this level of mutual understanding is not required in conversations; one-sided definitions of mutual knowledge will suffice. This can be represented as follows: when two people (A and B) are conversing then mutual knowledge would be established for A if "A knows that A and B mutually know that p" (Clark and Marshall, 1981, p. 18). However, Clark and Marshall (1981) suggest that each participant only requires half of the statements; A only requires the statements without the primes, whilst B just needs the statements with the primes. The Collaborative model accepts that perfect mutual understanding can never be fully achieved (Clark (1985); Clark and Brennan (1991)); instead, participants establish a level of mutual understanding to a 'criterion sufficient for current purposes' (Clark and Wilkes-Gibbs (1986); Clark and Schaefer (1989)).

The process of grounding occurs over a sequence of turns involving two phases, a presentation phase and an acceptance phase. These two stages are described as "A presenting an action for B to consider, and of B accepting that action as having been understood" (Clark and Schaefer, 1989, p. 151, original emphasis). If both of these phases are completed correctly, then A and B will both believe that they have reached the mutual belief that B has understood what A meant by the initial utterance. Both presentation and acceptance phases must be completed for A to have contributed to the discourse. Completion of the two phases depends on addressees providing evidence that they have understood the utterance. However, the evidence provided by the addressee is also a presentation which needs to be accepted (Clark and Schaefer, 1989). Therefore the acceptance process is recursive in nature. At what point does the cycle of presentation and acceptance of a contribution stop? Clark and Schaefer suggest that the strength of evidence required for accepting a presentation will be reduced for each recursive cycle of the two phases involved in contributing to a discourse. Eventually, usually after two or three cycles (Clark and Schaefer, 1989), the addressee will consider that the contribution has been sufficiently grounded, he will offer one of the weakest forms of evidence (e.g. showing continued attention) and the conversation can move on.

The number of recursive cycles required to establish mutual understanding may also be determined by a range of variables. (Clark and Schaefer, 1989) suggest that task related conversations may require stronger evidence of understanding than social dialogues. This view is supported by research into referential conversation by Cohen (1984) and Clark and Wilkes-Gibbs (1986). In addition, Clark and Brennan (1991) propose that the process of grounding changes with communicative context. This occurs because contexts vary in the number of channels of communication they support, and hence the range of 'grounding constraints' (ways of constraining the many possible interpretations of utterances or messages) afforded by the communicative context. This view is upheld by recent research (for example, Newlands et al. (2000); Doherty-Sneddon et al. (1997)). A further complication is that the

process of grounding can also involve the use of insertion sequences or repair techniques; these are sets of exchanges that are embedded within the ongoing exchanges (Jefferson (1972); Schegloff (1972)). This can make the structure of the dialogue quite complex, and is one of the issues currently under investigation.

1.2 Previous Empirical Research

This research follows on from studies which have examined the process of collaboration and communication in desktop-conferencing, and differences in the structure of face-to-face and mediated interactions (for example Newlands et al. (2000); Doherty-Sneddon et al. (1997); Anderson et al. (1997)). The application of Conversational Games Analysis (Kowtko et al., 1992) has shown that the range and distribution of Conversational Games differs in computer-mediated contexts (such as, email and desktop video-conferencing) from more familiar forms of communication, such as face-to-face interactions. The frequency of use of some pragmatic functions related to the process of establishing common ground varies significantly with context. For example, people 'check' their own understanding of previous utterances more frequently when they are unable to see the person with whom they are interacting (Doherty-Sneddon et al., 1997), or when the quality of the visual signals is low Newlands et al. (2000). Speakers also spend more time ensuring that their listener has understood what has been said, has completed some part of the task, before carrying on with the dialogue; this is a form of alignment (an Align conversational Game). This would indicate that the process of grounding differs in some contexts, and confirms the predictions made by Clark and Brennan (1991). It also suggests that the structure of dialogues in computer mediated contexts may be more complex than in other contexts, as they may contain a greater number of nested cycles of presentation and acceptance phases or a greater number of embedded Conversational Games.

2 Goal of the Paper

The current paper examines the way in which contributions are grounded in a range of communicative contexts in terms of conversational games. The analysis is not based on adjacent pairs of utterances, but on the way in which the Conversational Games are inter-leafed and embedded within each other. This form of analysis should highlight the structure and nature of the process of grounding, and give an indication of whether the amount of collaborative effort to establish common ground differs between face-to-face and video-conferencing interactions. This can be determined by seeing whether the number of inserted sequences or embedded Conversational Games varies with communicative context, and by determining the depth of embedded Games (that is, how many games are embedded with an already initiated game).

3 Design and Procedure

This study compares the structure of dialogues recorded whilst participants completed a collaborative problem solving task (The Map Task, Brown et al. (1984)) in two communication context, desk-top video conferencing (DVC) and face-to-face. The data for the DVC context was collected as part of a study which compares the effects of two different DVC systems on collaborative interactions Newlands (1998). The DVC system consisted of video (VIC) and audio (VAT) conferencing tools, which were publicly and freely available over the internet. VIC provides full colour JPEG encoded video at 5 - 6 frames per second, a relatively low level of temporal resolution (this was a common feature of publicly available DVC systems at the time). The video channel was identical in both DVC conditions, but the type of audio channel provided was varied between the two groups of DVC users. One group was provided with a full duplex audio channel, whilst the other group used a half duplex ('click to speak) audio set-up. In the latter context, participants had to click the mouse in an 'audio box' presented on the monitor, and hold down the mouse button whilst they talked. A between groups design was used, with eight pairs completing one Map Task in each of the DVC contexts. Eight dialogues were chosen

from the Human Communication Research Centre Map Task Corpus; these were all from the face-to-face context and were used as a comparison group for the video-mediated dialogues. The dialogues (in all contexts) represent the participants' first attempt at the Map Task, and involved participants who were familiar with their partner. The participants were all volunteers from the University population.

3.1 Method of analysis

Conversational Games Analysis (Kowtko et al., 1992) provides a framework for looking at the communicative functions (conversational goals and sub-goals) that speakers attempt to convey in their contributions. Conversational Games Analysis (CGA) is derived from artificial intelligence models of communication, specifically from the work by Power (1979), Houghton (1986) and Houghton and Isard (1987). The analysis involves coding every utterance in terms of what the speaker is attempting to achieve, and is based upon the function of the utterance rather than its linguistic form or content. In this way, patterns of pragmatic functions in the dialogues can be observed. The distribution of the Conversational Games and Moves can highlight the ways in which grounding may differ in a variety of contexts (for example, Newlands et al. (2000, 1996)). Because the analysis allows the embedding of one Conversational Game within another, CGA can also be used to illuminate the recursive nature of the grounding and the way in which complex patterns occur when side sequences or conversational repairs enter the dialogue. Eight dialogues from each of the video conferencing corpus were coded; the coder was naïve to the intended analysis of the embedded Games. The sample of dialogues taken from the HCRC corpus was already coded. The types of Conversational Games used in the analysis are defined in Table 1.

INSTRUCT: Communicates a direct or indirect request for action or instruction.

CHECK: Listener checks their own understanding of a previous message or instruction from their conversational partner, by requesting confirmation that the interpretation is correct.

QUERY-YN: Yes-No question. A request for affirmation or negation regarding new or unmentioned information about some part of the task.

QUERY-W: An open-answer Wh-question. Requests more than affirmation or negation regarding new information about some part of the task.

EXPLAIN: Freely offered information regarding the task, not elicited by coparticipant.

ALIGN: Speaker confirms the listener's understanding of a message or accomplishment of some task, also checks attention, agreement or readiness.

Table 29.1: Six Types of Games Found Necessary and Sufficient to Capture the Speaker's Communicative Intents in Coding Map Task Dialogues

4 Results

The results will be presented in two sections. First the results of the Conversational Games Analysis will be reported to demonstrate the effects of communicative context on the distribution of Games in each context. These results will be reported briefly, to confirm that differences do occur between contexts. This will be followed by analysis of the structure Games in the dialogues and the amount of embedding that occurred in each context.

4.1 Conversational Games Analysis

The frequency with which each type of Conversational Game occurred in the face-to-face and two DVC contexts (full duplex audio channel and half duplex audio channel) was calculated. The standardised frequency scores (per 100 Games) were obtained to allow for any differences in the length of dialogues in the three contexts. The mean standardised frequency of each Game in each context, by Instruction Giver (IG) and Instruction Follower (IF) are presented in below. Standard deviations are given in brackets.

Game	Role	Face-to-face	Full duplex DVC	Half duplex DVC
Instruct	IG	9.70 (3.85)	4.92 (0.55)	8.54 (2.23)
	IF	0.26 (0.37)	0.24 (0.58)	0.00 (0.00)
Explain	IG	4.13 (1.28)	3.98 (1.88)	4.03 (1.78)
	IF	6.12 (1.62)	4.62 (2.28)	5.09 (3.34)
Query-yn	IG	5.29 (2.00)	3.37 (1.57)	4.94 (1.55)
	IF	3.21 (2.58)	2.90 (1.18)	2.37 (1.31)
Query-w	IG	1.55 (1.16)	0.71 (0.85)	3.62 (2.26)
	IF	2.79 (1.32)	4.33 (1.19)	4.83 (2.97)
Align	IG	5.74 (4.19)	6.78 (4.13)	10.26 (3.81)
	IF	0.44 (0.61)	1.18 (0.72)	1.05 (0.67)
Check	IG	1.41 (1.18)	0.95 (0.63)	1.42 (2.08)
	IF	8.15 (2.12)	7.68 (1.55)	5.94 (1.72)

Table 29.2: Mean standardised frequency of Conversational Games in face-to-face and DVC contexts, by role of participant

The data presented above shows that some Conversational Games appear to be used more frequently in some contexts than in others, and an apparent effect of role of the participants. For example, Instruction Givers initiated more Instruct Games than Instruction Followers, but the number or Instruct Games varies across the three communicative contexts. To determine whether any apparent differences were significant, separate analyses of variance (2 way mixed ANOVA) were computed for each category of Conversational Game. Communicative context (face-to-face, full duplex DVC, half duplex DVC) was treated as a between group factor, and the role of the participant (Instruction Giver or Instruction Follower) as a within dialogue repeated measure.

For the purposes of this paper, the results of the effect of communicative context are of greatest importance. A change in the frequency of use of certain Games (for, examples Align Games or open-questions) could indicate that participants are using a greater number of recursive cycles of acceptance and presentation phases.

The results of the analyses showed that there were significant main effects of context for some of the Games (Instruct, Query-w, Align Games) but not for Explain or Query-yn or Check Games. The role of the participant was significant across all types of Games. A significant interaction was observed for only for the Instruct Games. Details of the significant effects are outlined below.

Instruct Games: The main effect of context was significant [$F(2.21) = 7.49, p < 0.005$]. Post hoc tests (Tukey HSD) showed that Instruct Games occurred more frequently in the face-to-face context than in full duplex DVC ($p < 0.01$). These games were more frequently initiated by the IG rather than the IF [$F(1.21) = 313.40, p < 0.001$].

Query-w Games: The main effect of context was significant [$F(2.21) = 11.40, p < 0.001$]. Post hoc tests showed that Query-w Games occurred more frequently in the half duplex DVC context than in the face-to-face. These games were more frequently initiated by the IF rather than the IG [$F(1.21) = 13.43, \ p < 0.001$].

Align Games: The main effect of context was significant [$F(2.21) = 3.33, p < 0.05$]. Post hoc tests showed that Align Games occurred more frequently in the half duplex DVC context than in face-to-face Align ($p < 0.05$). These Games were more frequently initiated by the IG rather than the IF [$F(1.21) = 78.67, p < 0.001$].

Check Games were not affected by communication context ($p > 0.1$), but they were more frequently initiated by the IF than the IG [$F(1.21) = 133.47, p < 0.001$].

Explain Games were not effected by communicative context ($p > 0.1$), but they were more frequently initiated by the IF than the IG [$F(1.21) = 4.55, p < 0.05$].

Query-yn Games were not effected by communicative context ($p > 0.1$), but they were more frequently initiated by IG than the IF [$F(1.21) = 15.00, p < 0.01$].

The results replicate some of the findings reported by in the literature on the effects of mediated communication. Of special interest are the differences noted in the use of Align, since these Games are involved in the process of establishing mutual knowledge; they are used to ascertain whether a previous utterance has been understood sufficiently, or in the intended manner. The increased number of Align Games in the half duplex replicates the findings reported earlier by Newlands et al. (2000), (2000) and supports the results reported by Doherty-Sneddon et al. (1997). When participants cannot see who they are talking to, or the quality of the video channel is low, they tend to use a more cautious style of interaction, shown by an increased amount of aligning. Changes in the use of Align Games in the full duplex DVC dialogues did not reach significance when compared to the other two contexts, but the means are in the predicted direction. The increased use of questions (Align and Query-w questions) could indicate that the structure of the dialogues in the DVC contexts differs, or is more complex, than the structure of Games in the face-to-face context. The next part of the results examines this possibility.

4.2 Structure of Games and Embedded Games

Several types of patterns of contributions and embedded structures of Conversational Games were noted. Straight-forward examples of Games following an adjacancy pair structure do occur, as can be seen in Extract 1 where three Games occur in a row without any embedded Games. In the following extracts IG and IF stand for Instruction Giver and Instruction Follower (the roles of the participants), the type of Conversational Game is shown above each turn and brackets indicate the start and end of each game.

Extract 1. Sequence of Games with no embedded Games (zero level of embedding)

□ Game 1	IG Query-yn Game Starts
×	IG: Do you have farmed land?
×	IF: Nope
□ Game 1	Ends
□ Game 2	IG Query-yn Starts
×	IG: Do you have a dead tree?
×	IF: Yes
□ Game 2	Ends
□ Game 3	IG Check Starts
×	IG: In like where two streams meet you have a dead tree?
×	IF: uh huh
□ Game 3	Ends

However, other interactions can involve a series of side sequences, or embedded Games, each one being resolved before another Conversational Game was initiated (see Extract 2) below. Examples of more complex interactions frequently occurred. One game being embedded within another one is a common occurrence, and on occasions a third or fourth game can be inserted within the other two (see Extract 3). For instance, an Instruct Game could be initiated by one participant and within this Game a

series of other Games may be initiated in an attempt to resolve a misunderstanding (or request more information) before the Instruct Game could be completed.

Extract 2. Sequence of single embedded games

□ Game 1	IG Instruct starts	
×	IG: Then the part curves up to the right if you were	
×	walking along it, and then it goes up straight again	
×		
×	□ Game 2	IF Check embedded starts
×	×	IF: Straight up, yeah?
×	×	IG: Straight up
×	□ Game 2	Ends
×		
×	□ Game 3	IF Query-W embedded starts
×	×	IF: How Far?
×	×	IG: Ehm, one, two centimetres, about 3 centimetres
×	□ Game 3	Ends
×		
×	□ Game 4	IG Align embedded starts
×	×	IG: Have you got that?
×	×	IF: Yeah, yeah I'm trying to draw the straight line
×	□ Game 4	Ends
□ Game 1	Ends	

These extracts show some of the different types structures of embedded Games and the process of grounding in the dialogues analysed to date. A variety of different forms of structure occur throughout the dialogues, some involving overlapping embedded sequences which can produce very complicated exchanges. Conversational Games were found to occur at five levels of embeddedness, from no level (zero) of embeddedness (a Game with no Games embedded within it) through to four levels of Games within Games. The frequency of each type (or level) of embedding was obtained. The mean number of Games per dialogue at each level of embedding is shown in Table 3 below, for each of the three communicative contexts. (Standard deviations are shown in brackets).

Extract 3. Sequence of complex embedded games (up to 3 levels of embedded Games)

□ Game 1	IG Instruct starts				
×	IG: And then draw a little corner and you want to go straight				
×	up about one centimetre to the left of				
×	the collapsed shelter.				
×					
×	IF: to the left, okay				
×					
×	□ Game 2	IF Check embedded starts (level 1)			
×	×	IF: In which direction? Straight up?			
×	×	IG: Yeah			
×	□ Game 2	Ends			
×					
×	□ Game 3	IF Query-W embedded starts (level 1)			
×	×	IF: How far do I go?			
×	×	IG: oh my goodness, quite far, let me see			
×	×				
×	×	□ Game 4	IG Query-yn embedded starts (level 2)		
×	×	×	IG: Have you got a Saxon Barn?		
×	×	×	IF: Ehm yeah		
×	×	×			
×	×	×	□ Game 5	IF Check embedded starts (level 3)	
×	×	×	×	IF: That's the one on the left hand side?	
×	×	×	×	IG: Yes	
×	×	×	×	IF: okay	
×	×	×	□ Game 5	Ends	
×	×	×			
×	×	×	IG: right, so you have the Saxon Barn, cool		
×	×	□ Game 4	Ends		
×	×				
×	×	IG: So draw up to the Saxon Barn, about 2 centimetres higher			
×	□ Game 3	Ends			
×					
□	... sometime later Game 1 Ends				

Levels of Embedded	Face-to-face	Full duplex DVC	Half duplex DVC
zero	32.23 (23.98)	20.63 (8.43)	23.75 (9.22)
one	49.00 (12.30)	60.37 (19.94)	47.38 (18.29)
two	13.50 (5.07)	17.75 (12.95)	20.25 (17.99)
three	1.75 (1.58)	3.13 (3.75)	2.37 (3.33)
four	0.25 (.0.46)	0.37 (0.74)	0.38 (1.06)

Table 29.3: Mean number of Conversational Games per dialogue, which occur at each level of embeddedness in three communicative contexts

The data displayed in Table 3 indicates that some levels of embeddedness occurred more frequently than others. For example, Games occurring at one level of embeddedness (a Game within another Game) appears to be the most frequently occurring pattern. However, the standard deviations are large, demonstrating the fact that there was a lot of difference in the complexity of the structure of Games between pairs of participants. To determine if any observed differences were significant, the data was entered into a mixed ANOVA, with communicative context as a between groups factor, and levels of embeddedness (zero, one, two, three and four levels) as a within dialogue repeated measures.

The results of the ANOVA showed that there was no main effect of communicative context ($p > 0.1$), and no significant interaction between context and levels of embedded Games ($p > 0.1$). There was however a significant main effect of levels of embeddedness [$F(4.84) = 94.88, p < 0.001$]. Post hoc tests (Tukey HSD) showed that one Game being embedded within one other Game was the most frequently observed pattern of embeddedness, accounting for over 50% of the total number of Games. This is the pattern of embedded Games illustrated above in Extract 2. The second most common patterns were for Games to contain no embedded Games, as in Extract 1; this type of structure accounted for 25% of all Games. However, Games with 2 Games inserted within them were just as likely to occur as Games with no embedding. There was no significant difference between the occurrence of these two levels or zero levels of embedded Games, though both occurred significantly less often than one level of embeddedness ($p < 0.001$).

Several interesting phenomena have been noted during the analysis. For example, the resolution of a complex exchange of Games sometimes has an interesting effect on the process of communication. A sequence of embedded Games can build up over a length of time, and then all of the Games are completed (or resolved) when the last game has been successfully negotiated. The response to the final game seemingly resolving all of the other, embedded games. The reasons for these complex structures of embedded Games is also being investigated, to try to determine what factors would predict this form of dialogue structure, and which types of Games tend to occur at the various levels of embeddedness. It is already notable that complex structures of embedded Games are not just a consequence of participants encountering a mis-match in map landmarks. Equally complex embedding of Games can occur at other times. Deep levels of embedded Games (up to four levels of embedding) have also been observed in dialogues taken from another study, during which researchers met over a video conferencing system to discuss grant applications. Therefore the complexity of the structures observed in this study are not simply an artefact of the Map Task.

5 Conclusion

The analysis has shown that there was no effect of communicative context, perhaps a disappointing result from a theoretical point of view. However for users of video mediated conferencing systems the results are positive, as it appears that the structure of the Games in these dialogues are no more complex than those encountered in face-to-face interactions. The increased use of Align Games, Instruct Games, and open-ended questions does not make the structure of the dialogues more complex. The results also highlight the fact that one Game being initiated within another Game is a common occurrence, whereas simple question-answer sequences occur less frequently. This would indicate that adjacency pairs are not the most appropriate way in which to analyse the structure of interactions, as they do not account for a sufficient amount of the data. A different way of modelling what is going on in the dialogues is required, with the emphasis still on an 'language as action' approach (Clark, 1996), taking into account the context in which participants interact.

Bibliography

Anderson, A., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A., and Van der Velden, J. (1997). The impact of vmc on collaborative problem solving: an analysis of task performance, communicative process, and user satisfaction. in: Finn et al. (1997).

Brown, G., Anderson, A., Yule, G., and Shillcock, R. (1984). *Teaching Talk*. Cambridge UP.

Clark, H. and Brennan, B. (1991). Grounding in communication. in: Resnick et al. (1991).

Clark, H. and Marshall, C. (1981). Definite reference and mutual knowledge. in: Joshi et al. (1981).

Clark, H. and Schaefer, E. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2(1):19–41.

Clark, H. and Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*, 13:259–94.

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

Clark, H. H. (1985). Language use and language users. in: Lindsay and Aronson (1985).

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

Connolly, J. and Pemberton, L., editors (1996). *Linguistic Concepts and Methods in CSCW*. Springer-Verlag.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., , and Bruce, V. (1997). Face-to-face and video mediated communication: a comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, 3(2):1–21.

Finn, K., Sellen, A., and Wilbur, S., editors (1997). *Video-Mediated Communication*. Lawrence Erlbaum Associates.

Houghton, G. (1986). *The Production of Language in Dialogue: A Computational Model*. PhD thesis, University of Sussex.

Houghton, G. and Isard, S. (1987). Why to speak, what to say and how to say it: Modelling language production in discourse. in: Morris (1987).

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

Jefferson, G. (1972). Side sequences. in: Sudnow (1972).

Joshi, A., Webber, B., and Sag, I., editors (1981). *Elements of discourse understanding*. Cambridge UP.

Kowtko, J., Isard, S., and Doherty-Sneddon, G. (1992). Conversational games within dialogue. Research Paper HCRC/RP-31, Human Communications Research Centre, University of Edinburgh.

Lewis, D. (1969). *Convention: A philosophical study*. Harvard UP.

Lindsay, G. and Aronson, E., editors (1985). *Handbook of Social Psychology*, volume II. Random House.

Morris, P., editor (1987). *Modelling Cognition*. John Wiley.

Newlands, A. (1998). *The Effects of Computer Mediated Communication on the processes of communication and Collaboration*. PhD thesis, Glasgow University.

Newlands, A., Anderson, A., and Mullin, J. (1996). Dialog structure and cooperatiave task performance in two cscw environments. in: Connolly and Pemberton (1996).

Newlands, A., Anderson, A. H., Mullin, J., and Fleming, A.-M. (2000). Processes of collaboration and communication in desktop videoconferencing: Do they differ from face-to-face interactions? In *Proceedings of GOTALOG 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue*. Goteborg University.

Resnick, L., Levine, J., and Teasley, S., editors (1991). *Perspectives on socially shared cognition*. APA.

Schegloff, E. (1972). Notes on conversational practise: formulating place. in: Sudnow (1972).

Schiffer, S. (1972). *Meaning*. Clarendon.

Sudnow, D., editor (1972). *Studies in Social Interaction*. Free Press.

Paper 30

How much Common Ground Do we Need for Speaking?

Kerstin Fischer
kerstinf@uni-bremen.de
http://www.fb10.uni-bremen.de/anglistik/homepages/fischer.htm

1 Introduction

The question addressed in this paper is to which types of common ground speakers attend in dialogical interactions. The procedure is to investigate a particular kind of interaction in which common ground is at stake, i.e. in which speakers are uncertain about the common ground they can assume. Analysing what they request for producing utterances for their communication partner reveals to which types of common ground they orient.

Most research on common ground has been carried out on what is shared between the conversational participants on the basis of the discourse record of the current situation. Building on work by Clark and collaborateurs (Clark and Marshall, 1981; Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989; Clark and Brennan, 1991), summarized in Clark (1996), much work has addressed aspects of 'grounding', the process by which individuals add information to the common ground (Traum, 1994; Ginzburg, 1998). Thus, research has concentrated on the augmentation of the propositions representing the assumed shared knowledge on the basis of what is said in a discourse situation (Larrson et al., 2000).

However, as Clark (1996, p. 92-121) points out, the shared basis for joint action speakers draw upon consists in a number of further aspects besides the discourse record, and he provides us with a list of knowledge types speakers may use as possible resources to establish common ground. Such resources include knowledge about the human nature, a common lexicon, knowledge about scripts, and knowing how. The problem addressed in this paper, what speakers really draw upon in discourse, has rarely been studied, and if so, only in natural conversations (Kreckel, 1981; Clark, 1996). While it is certainly useful to base one's investigations on natural conversations since they constitute the most basic type of communication from many points of view (Fillmore, 1981; Diewald, 1991), studying the communication between human speakers and communication partners about whom they do not know much may be particularly suited for showing what they consider the necessary common ground for their producing of utterances. For instance, common ground cannot be presupposed in the interaction between human and artificial communication partners to the same extent as in the communication among humans. In this particular type of interaction, almost all aspects of common ground may have to be negotiated. Dialogues with so little shared knowledge between the communication partners may thus reveal how much, and which types of, information is necessary for communication to work because in these cases, common ground is attended to as potentially problematic. The methodology underlying this study relies thus on the conversation analytic principle of 'deviant case analysis', based on the idea that deviant cases are not only orderly in themselves, but, as Hutchby and Wooffitt (1998, p. 95-98) argue for the analysis of sequences, if "someone displays in their conduct that they are 'noticing' the

BI-DIALOG 2001 (P. Kühnlein, H. Rieser, & H. Zeevat (eds)).
Paper 30,
kerstinf@uni-bremen.de
http://www.fb10.uni-bremen.de/anglistik/homepages/fischer.htm.

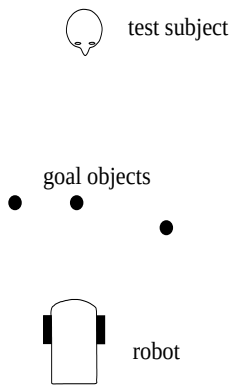


Figure 30.1: The Experimental Setup

absence of a certain type of turn from a coparticipant, then that demonstrates their own orientation to the relevance of the sequence that the analyst is aiming to describe" (Hutchby and Wooffitt, 1998, p. 98). Thus, by not analysing basically smooth and unproblematic human-to-human conversation, but by focussing instead on problematic human-to-robot (mostly mis-)communication, the types of common ground to which speakers attend may become evident. The methodology is thus to analyse the participants' displays of their noticing the absence of aspects of common ground in human-robot interaction.

The results will not only be theoretically interesting because they provide evidence for a typology of common ground, they will also have important practical consequences regarding the modelling of cognitive agents, such as robots.

2 Data

The data were elicited in an experimental setting (see Figure 30.1) for which initially, a robot (see Figure 30.2) was designed on the basis of what is known about spatial reference among humans (Moratz and Fischer, 2000). Then, a test scenario was developed in which the users' task was to make the robot move to particular locations pointed at by the leader of the experiment; pointing was used in order to avoid the prompting of verbal expressions and the use of pictures of the scene which would impose a particular perspective. Users were instructed to use natural language sentences typed into a computer to move the robot through a setting in which, for instance, a number of cubes were placed on the floor together with the robot.

Since the main aim of the experiments was to determine how participants naively approach the robot as a communication partner, the only system output was action or an unspecified error message. This avoids the 'shaping' of the users' language by the system's output (Zoltan-Ford, 1991). By being constantly forced to try out new strategies to increase the understandability of their utterances, users furthermore reveal their hypotheses about how the robot may work. The users' beliefs about the nature of the robot, about what it can perceive and do, are part of the common ground users expect, that is, to which they attend.

Fifteen different participants carried out about 40 attempts to move the robot within about 30 minutes time each. Altogether 603 instructions were elicited. The sentences were protocolled, and the users' verbal behaviour during the experiments was recorded in order to capture self-talk in which speakers anounce their strategies or their ideas about what is going wrong. After the experiments, participants were asked to put down in a questionnaire what they believed the robot could and could not understand.

Figure 30.2: The Robot

3 Types of Common Ground Attended to in the Data

According to Kreckel (1981, p. 29), the background knowledge that plays a role in interactions can be distinguished into three kinds:

1. "what remains idiosyncratic and, thus, has to be labelled experience;
2. what is based on mutual acquaintance oo knowledge about and, thus, can be considered as implicitly or explicitly shared;
3. what is based on separate acquaintance or knowledge about and, thus, may or may not be in common between communicants."

The criterion on the basis of which Kreckel distinguishes what the speakers can build upon as common is thus whether the respective knowledge is acquired jointly or separately by the interactants. Clark's (1996) analysis is more sophisticated in showing how these different types of knowledge relate to each other and by providing further, more fine-grained, distinctions (Clark, 1996, p. 54). Thus in his account, 'total common ground' comprises a 'discourse representation', which consists of a textual and a situational description, as well as the discourse record, and communal (Kreckel's common knowledge) and personal (Kreckel's shared knowledge) common ground.

Clark (1996, p. 92-96) argues that common ground is best be seen as a shared basis between the participants. This means that

> both participants have the information that b holds;
>
> b indicates to both that both have the information that b holds;
>
> b indicates to both that p.

This may give rise to reflexive knowledge about common ground:

(i) both have the information that p and that (i).

Clark (1996, p. 100-120) distinguishes two types of shared bases, communal and personal. Both of them can be distinguished into a number of subtypes:

1. communal common ground
    - human nature
    - communal lexicons
    - cultural facts, norms, procedures
    - ineffable background
    - our feeling of other's knowing
2. personal common ground
    - perceptual basis
    - actional basis
    - personal diaries
    - acquaintedness (friends and strangers)
    - personal lexicons

We can now use the evidence from our corpus of human-to-robot communication to show to which of these aspects of common ground from Clark's typology the users really attend in which ways. The different types of common ground users can be found to orient to in dialogues with the robot will be investigated on the basis of the transcripts of the human-robot interaction itself, the self-talk elicited during the experiments, and the answers participants gave in the questionnaire at the end of each dialogue. By means of this procedure, we can show which the types of common ground are of which users display that they are noticing their absence. This may also suggest which kinds of background they preferred to have for formulating their utterances for their communication partner.[1]

3.1 Communal Common Ground

Human Nature

The data do not show that speakers attend to their knowledge about the human nature, but since the robot is not a human communication partner, this is not surprising. However, for human-computer conversation in another scenario, participants could be shown to treat the computer like a human being (Fischer, 2000). Thus, in previous analyses human speakers have been found to transfer human attributes to their artificial communication partners. In contrast, in the present human-robot dialogues, the users' self-talk consists of several questions regarding the nature of the robot, for instance, regarding its orientation. In order to be able to employ an intrinsic reference system, the users requested to know where the 'front' of the robot is and what it can perceive. We can count this information need as evidence that participants orient to the robot's nature while formulating spatial instructions.

Communal Lexicons

The data show that the human users worry very much about which are the appropriate words to use, i.e. which is the common lexicon between them and the robot. Thus, during the experiments they asked questions like whether one word or another is understandable. In the questionnaire, six out of the fifteen participants hypothesized that a possible source for their miscommunication with the robot was that they could not find the right words, that they did not know the 'appropriate' lexicon.

[1] We will focus only on the human users and their beliefs about the robot, and thus spare out the perspective of what the robot may be implemented to believe about what the human users may know.

However, participants did not only have problems regarding the communal lexicon; the recordings of their utterances during the experiments as well as the questionnaire results reveal that they regarded the following linguistic aspects as problematic: orthography, formality (in particular the formal or informal way to write imperative verbs in German), but also syntax, for instance, whether relative clauses are allowed, the length and complexity of sentences, the granularity level (especially the question whether they should use natural language or metrical expressions), and, most generally, whether the language of instruction should be German, the native language of the participants, or English, 'the language computers speak.' The data thus support Clark's category, yet it has to be extended to all linguistic levels involved, not just the lexicon.

Cultural Facts, Norms, Procedures

The setting in which the experiments were carried out did not require participants to attend to cultural facts, norms, or procedures, such as scripts. However, participants are found to request one property in their communication partner that can be considered to constitute 'normal' or even 'normative' behaviour among human beings such that it can be requested of human speakers, and its lack is accountable in dialogues, namely consistency (Goffman, 1978). When the users found a hypothesis about the behaviour of the robot untrue, they complained about its lack of consistency, both verbally and in the questionnaire. Thus participants displayed attention to an instance of cultural norms, even in this restricted setting.

Ineffable Background

The example by means of which Clark (1996, p. 110) illustrates the notion of ineffable background is that a person living in San Fransisco is expected to know what Golden Gate Bridge or Coit Tower, for instance, look like. In our data, one particular problem turned up which relates best to the notion of ineffable background, but which differs from Clark's example in a particular way. This problem concerns the way people believe aspects of spatial reasoning to be related. In particular, in the experiments about half of the participants instructed the robot with a strategy which is also most common among humans, namely to name the goal object to which the robot was supposed to move. The other half, however, started off with another type of instruction, namely to describe a path along which the robot was supposed to move. Since the robot was not implemented for this kind of instruction, the system's feedback was only "error". The users' strategy now consisted in proceeding to more and more elementary strategies, up to sentences like 'move your rear wheels.' Similarly, those participants who had initially attempted a goal description but had failed because of some linguistic problem tried path descriptions later. Remarkable is that none but a single participant, who openly wondered about whether path or goal descriptions were more appropriate, returned from path descriptions or more elementary strategies to a goal description, which the robot would have understood. Even if prompted to do so, users were extremely reluctant to change their strategy. Thus, for the participants there was apparently a fixed order of simplicity vs. complexity regarding spatial instruction, which was unrelated to the robot's real implemented capabilities. For them, therefore, knowing how to move along a path constitutes the ineffable background for moving towards a goal object. To return to Clark's example, the participants behaved as if it was impossible to know Coit Tower without knowing that it is in San Francisco. Participants thus orient to ineffable background as a source of common ground in the dialogues.

Grading of Information

By grading of information, Clark (1996, p. 110-112) understands our knowledge of other people's knowing. He quotes results from experiments which show that we usually have a good idea of what our communication partners know and what they are not likely to know; that is, in general we have a good judgement of the mutuality of information. What our results show is that this is not the case with robots. The participants are uncertain about what language the robot understands, which words,

syntactic structures, formality and granularity levels are understandable to it, what it perceives (see below for a discussion of these aspects), and how it interacts with the world.

3.2 Personal Common Ground

Perceptual Basis

A joint perceptual basis constitutes the prototype for personal common ground (Clark, 1996, p. 112). In the dialogues between the human speakers and the robot investigated, the conditions for a joint perceptual basis are not given; the situation is not equally accessible to both participants, that is, the robot's perceptual capabilities are much more restricted than those of a human being. Thus, a robot may not have the information that something is the case, although for the human speaker it is 'obvious'. Accordingly, speakers were found to be much aware of the fact that their perception may differ from the robot's perception, i.e. while a fact perceived indicates to them that something holds, it may not indicate the same fact to their communication partner. Thus, the participants were uncertain about whether the scene perceived by them constitutes the same situation to the robot. The questions participants asked during the experiments were thus: 'what does it see?', 'where is its front?' and even 'does it see anything at all?'.

Actional Basis

The actional basis between the participants is constituted, according to Clark (1996, p. 114), by means of joint action, the prototype being talk. This includes the successful presentation, acceptance and acknowledgement of utterances (Clark and Schaefer, 1989). When the conversational participants in our experiment were successful in giving an instruction, the robot's resulting action can be seen as an appropriate acceptance and the user's proceeding to the next task as a verfication of this interpretation of the instruction. Users, however, were also found to change their linguistic behaviour on the basis of failed joint action, i.e. when the system answered "error" only. Usually it took the participants several attempts before they succeeded; some participants did not achieve a single joint action at all. However, once they had discovered a way to make themselves understood, they sticked to it; that is, they adapted their linguistic behaviour according to their hypotheses about common ground. Thus, users were found to attend to both successful and unsuccessful joint actions carried out in the interaction with the robot.

Personal Diaries

By personal diaries, Clark (1996, p. 114) understands the previous joint actions carried out by the participants. Here it is not entirely clear in which way the personal diary differs from previous joint actions.

Acquaintedness (Friends and Strangers)

Because of the limited interaction with the robot, participants hardly acquired acquaintedness with it. However, what the data show is that participants were constantly attempting to increase acquaintance with the robot in order to reduce their uncertainty. As results by Amalberti et al. (1993) show, users indeed adapt to machines in a way that can be described as increasing acquaintance. Thus, after three times (with breaks of at least a week in between) 60 minutes of interaction with the simulated system, participants believing to talk to a computer behaved similarly to those who had been told that they were talking to a human 'wizard'. How far acquaintedness with a robot can go, whether a private language may evolve (see also the problem of acquiring personal lexicons below), cannot be predicted on the basis of the current experiments. What the data do show, however, is that users try to increase the acquaintedness with the robot, that is, that they attend to it.

Personal Lexicons

Because of the limited interaction with the robot, participants can not be said to have acquired a personal lexicon with it, though there are interpersonal differences in their linguistic strategies (for instance, in the choice of goal- versus path-based instructions), and thus idiosyncratic communicative means may have developed. In any case, participants gave up using particular words after some time of interaction, if they suspected them to be problematic, so that speakers can be argued to attend 'negatively' to a common personal lexicon.

4 Conclusions and Prospects

The problems users have in their formulating of utterances for the robot as a communication partner point to the fact that we normally know very much about our co-participants by drawing at least on those resources mentioned in Clark's typology. The results of this study show that in the communication with an unfamiliar communication partner users indeed attend to these resources. Thus, the results indirectly support Clark's hypothesis that we build on all of those above mentioned types of information for our joint actions in human-to-human communication.

Regarding specific categories, it could be shown that the categories related to the linguistic resources have to be extended; all linguistic levels may be part of the negotiation of common ground, not just the lexicon. Furthermore, the common ground also consists of basic theories about how the world works, in this case, that moving towards a goal presupposes knowing how to move along a path and how to use the respective devices for moving (engines, wheels). Knowledge as basic as how to navigate in space is therefore also part of the category **ineffable background**. Finally, the distinction between actional basis and personal diaries was not found to be useful since the common diaries are built up on the basis of previous joint action.

What practical consequences do our results have? Clark (1996, p. 116-120) has argued that conversational participants have techniques for building up common ground, for instance, by deliberately displaying community affiliations. This may point to a way how future systems can be significantly improved: Strategies have to be found by which the artificial system can signal to its interactants what its abilities and strengths are, and thus to inform the human interlocutor about its 'nature', its linguistic and perceptual capabilities, and even its ineffable background. This has to be done subtly and implicitly since too much in-advance instruction has turned out to be unpleasant, at best (Ogden and Bernick, 1996; Fischer and Batliner, 2000). However, the results of this investigation have shown how great the need for a common ground between system and user is and how much users invest to build up hypotheses about how the robot works. These conceptualization and adaptation processes could be exploited for the improvement of future human-robot interaction.

Bibliography

Amalberti, R., Carbonell, N., and Falzon, P. (1993). User representations of computer systems in human–computer speech interaction. *International Journal of Man–Machine Studies*, 38:547–566.

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

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

Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In Resnik, L., Levine, J., and Teasley, S., editors, *Perspectives on Socially Shared Cognition*. Academic Press.

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L., and Sag, I., editors, *Elements of Discourse Understanding*. Cambridge University Press.

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

Diewald, G. (1991). *Deixis und Textsorten im Deutschen.* Number 118 in Reihe Germanistische Linguistik. Tübingen: Niemeyer.

Fillmore, C. J. (1981). Pragmatics and the description of discourse. In Cole, P., editor, *Radical Pragmatics*, pages 143–166. New York etc.: Academic Press.

Fischer, K. (2000). What is a situation? *Proceedings of Götalog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue, Göteborg University, 15-17 June 2000. Gothenburg Papers in Computational Linguistics*, 00(05):85–92.

Fischer, K. and Batliner, A. (2000). What makes speakers angry in human-computer conversation. In *Proc. of the Third Workshop on Human–Computer–Conversation*, pages 62–67, Bellagio, Italien.

Ginzburg, J. (1998). Shifting sharing and access to facts about utterances. In Heydrich, W. and Rieser, H., editors, *Proceedings of the 10th European Summer School in Logic, Language and Information – Workshop on "Mutual Knowledge, Common Ground and Public Information"*, pages 30–35.

Goffman, E. (1978). Response cries. *Language*, 54:787–815.

Hutchby, I. and Wooffitt, R. (1998). *Conversation Analysis.* Cambridge: Polity.

Kreckel, M. (1981). *Communicative Acts and Shared Knowledge in Natural Discourse.* London etc.: Academic Press.

Larrson, S., Cooper, R., and Engdahl, E. (2000). Question accomodation and information states in dialogues. In *Proceedings of the Third Workshop on Human–Computer–Conversation*, pages 93–98, Bellagio, Italy.

Moratz, R. and Fischer, K. (2000). Cognitively adequate modelling of spatial cognition in human-robot interaction. In *Proceedings of the 12th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2000*, pages 222–228, Vancouver, British Columbia, Canada.

Ogden, W. and Bernick, P. (1996). Using natural language interfaces. In Helander, M., editor, *Handbook of Human-Computer Interaction.* Elsevier Science Publishers, North Holland.

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversations.* PhD thesis, University of Rochester.

Zoltan-Ford, E. (1991). How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies*, 34:527–647.

List of Tables

List of Figures

Persons

Keywords