Proceedings of

GÖTALOG 2000

Fourth Workshop on the
Semantics and Pragmatics of Dialogue

Göteborg University
15 – 17 June 2000

Gothenburg Papers in
Computational Linguistics 00–5

GOTHENBURG PAPERS IN COMPUTATIONAL LINGUISTICS

00–5
Massimo Poesio & David Traum (eds.)
Proceedings of GÖTALOG 2000
June 15 – 17, 2000

This paper can be found on
http://www.ling.gu.se/publications/GPCL.html
or ordered from
Department of Linguistics, Göteborg University,
Box 200, SE-405 30 Göteborg, Sweden

Previous workshops:

I	1997	MunDial'97	Universität München
II	1998	Twendial'98	Universiteit Twente
III	1999	Amstelogue'99	Amsterdam University

The organisers would like to acknowledge support from

The European Commission – Language Engineering Project LE4-8314:
Task Oriented Instructional Dialogue (TRINDI)
`http://www.ling.gu.se/research/projects/trindi`

The Bank of Sweden Tercentenary Foundation:
Information Exchange in Dialogue
`http://www.ling.gu.se/research/projects/indi`

NUTEK and HSFR joint Language Technology Program:
Swedish Dialogue Systems
`http://www.ida.liu.se/~nlplab/sds/`

Proceedings of GÖTALOG 2000,
the Fourth Workshop on the Semantics and Pragmatics of Dialogue,
held at Göteborg University, June 15 – 17, 2000.

Department of Linguistics, Göteborg University
Box 200
SE – 405 30 GÖTEBORG

Printed by Göteborg University, Göteborg 2000

Preface

This volume contains the proceedings of the Fourth Workshop on the Semantics and Pragmatics of Dialogues, held at Göteborg University on June 15-17, 2000. This workshop is part of a series of workshops whose aim is to bring together researchers working on the semantics and pragmatics of dialogues in fields such as artificial intelligence, formal semantics and pragmatics, computational linguistics, and psychology. The first of these workshops, MunDial 97, was organized by Anton Benz and Gerhard Jäger and took place in Münich in March 1997; the two subsequent workshops took place in Twente and Amsterdam, respectively. The increasing number of submissions (54 papers were submitted to this year's workshop) suggests that these workshops provide a stimulating forum for this kind of work.

Part of what makes these workshops so attractive is their success in attracting stimulating invited speakers. We are extremely grateful to this year's speakers - Jens Allwood, Herbert Clark, Paul Dekker, and Ronnie Smith - for accepting our invitation and contributing their insights to this year's meeting.

While receiving a large number of submissions is encouraging and ensures that this year's workshop will be yet again of high quality, it also makes the selection task very difficult. We wish to thank the members of the Program Committee, who ended up reviewing more papers than originally expected, and in a shorter time, managing nevertheless to produce highly informative reviews:

> Cecile Balkanski (LIMSI-CNRS, France); Johan Bos (U Saarlandes, Germany); Jonathan Ginzburg (King's College London, UK); John Gurney (Army Research Lab, US); Masato Ishizaki (JAIST, Japan); Gerhard Jaeger (ZAS Berlin, Germany); Yasuhiro Katagiri (ATR, Japan); Jan van Kuppevelt (IMS Stuttgart, Germany); Ian Lewin (SRI, UK); Diane Litman (AT&T, USA); Johanna Moore (U Edinburgh, UK); Joakim Nivre (Gothenburg University); Hannes Rieser (U Bielefeld, Germany); David Sadek (France Telecom); Len Schubert (U Rochester, USA); and Henk Zeevat (U Amsterdam, Netherlands)

Thanks also to Philippe Bretier (France Telecom), Mark Core (U Edinburgh), and F. Panaget (France Telecom) for additional reviews. As in previous editions of the workshop, we decided to accept fewer papers so as to give the authors more time for presentation; thus, only 22 papers were accepted. However, because of the large number of high quality submissions, we also decided to accept a few more papers as posters. The review process was blind, to increase the fairness of reviewing.

This year's workshop is organized in collaboration with the TRINDI project, LE4-8314. We also wish to thank Göteborg University for the use of their facility, and especially the local organizers - Robert Andersson, Maria Björnberg, Robin Cooper, Elisabet Engdahl, Staffan Larsson and Katarina Magnusson - for doing most of the hard work involved in organizing such an event.

Massimo Poesio and David Traum (Program Chairs)

Contents

Invited Papers

Contributed Papers

Short Papers and System Descriptions

Parameters of Dialog analysis

Jens Allwood

Department of Linguistics, Göteborg University
Box 200, SE–405 30 Göteborg, Sweden
jens@ling.gu.se

Abstract

The notion of "dialog" as well as the related notions of "dialog analysis" and "dialog theory" have become increasingly popular in recent years. The popularity of the notion of "dialog" includes several disciplines such as philosophy, linguistics, computer science, psychology, sociology, anthropology, literary studies etc. The goal of this paper is to give an overview of some possible parameters of dialog analysis. These parameters can then be used to locate different approaches to dialog in relation to each other. The parameters will first be brielfly characterized and then some of them will be commented on more in depth.

Uptake and its Role in Conversation

Herbert Clark

Stanford University
Stanford CA 94305, USA
herb@psych.stanford.edu

Abstract

Illocutionary acts have been viewed in two main ways. In the first, which originated with Searle, they are acts that speakers perform autonomously. In the second, proposed by Austin, they are acts performed by speakers in coordination with their addressees. In my terminology, they are participatory acts, which are one individual's parts of joint acts. By now Searle's view is standard, and Austin's is almost forgotten. The essential difference between them is this: In Austin's view the content of an illocutionary act is determined in part by the addressee's uptake, whereas in Searle's view it is not. I review a range of phenomena from spontaneous conversation as evidence for Austin's view and against Searle's. What finally counts, I argue, is not what speakers mean, but what they are jointly taken to mean.

Trying to Understand Misunderstanding: How Robust Can Spoken Natural Language Dialogue Systems Be?

Ronnie Smith

Department of Computer Science, East Carolina University
Greenville, NC 27858, USA
rws@cs.ecu.edu

Abstract

A ubiquitous problem with AI systems is the difficulty in "knowing that you don't know." Spoken natural language dialog systems are not exempt from this dilemma. This talk will overview several different studies that have been undertaken in order to make dialog systems behave more robustly in the presence of miscommunication. The studies use data collected via experimental interaction with the Circuit Fix-It Shop, a dialog system originally constructed to test the validity of a model for integrated dialog processing that is continuing to be used as a testbed for evaluating the effectiveness of techniques for the prevention, detection, and repair of miscommunication in human-computer dialog.

Support for Update Semantics (House Version)

Paul Dekker

ILLC/Department of Philosophy
Faculty of Humanities
University of Amsterdam
dekker@hum.uva.nl

Abstract

In this paper classical systems of update semantics are studied from the wider perspective of information exchange. We present independent compositional statements of the content of, update with, and support for first order expressible sentences. It is shown that a proper update with the contents of supported utterances is safe, in the sense that it does not corrupt the information distributed over the interlocutors. The pragmatic outlook on update and support also allows us to escape from some of the objections that has been raised against first order analyses of natural languages connectives, notably that of conditionals as material implication. The adopted outlook furthermore provides inspiration for a plausible analysis of functional dependencies and of certain cases of what has been called quantificational and modal subordination.

1. Introduction

Dynamic Semantics (DS) is a branch of linguistics originating from the model-theoretic and referentially-based work of Montague and his followers. The characteristic feature of dynamic semantics is that it systematically takes into account certain pragmatic aspects of interpretation. In the original versions of dynamics semantics, the object of study is the (dynamic) logic of the effects which the utterance of indicative sentences may bring about in an utterance situation (Discourse Representation Theory, File Change Semantics, Dynamic Predicate Logic, Update Semantics); in other versions also the logic of the satisfaction-conditions and pre-conditions are systematically studied (Situation Semantics, Epistemic Semantics).

In this paper we adopt a more systematic outlook on the semantics and pragmatics of assertions in natural language, one that takes into account both update effects on the side of the hearer, as well as the information which a speaker can be said to be committed to in support for her assertions. The main inspiration for this paper comes from (Grice, 1975; Grice, 1978; Stalnaker, 1978; Stalnaker, 1998). Technically, it elaborates upon the work of (Kamp, 1981; Heim, 1982; Groenendijk and Stokhof, 1991; Veltman, 1996).

We will proceed as follows. In the next section we introduce crucial notions and issues (semantic as well as pragmatic) in the context of a simple fragment of natural language which has the expressive power of propositional logic. In section 3 we lift the discussion to the first order level, and show how the relatively trivial results and observations from the propositional logic can be preserved, in a non-trivial way, at the first order level. Section 4 discusses a conceptual issue which automatically arises at the first order level. Maybe surprisingly, this doesn't have to do with the dynamics of first order interpretation—which has been given a natural explanation in section 3—, but with the absence of it in certain contexts. The insights gained in this section constitute motivation for a natural extension of the empirical scope of the first order system with functional dependencies, which is sketched in section 5. In the concluding section we summarize the results.

2. The Exchange of Propositions

In this section we introduce the basic notions which this paper focuses on, that of the 'content of', 'update with' and 'support for' assertions. The concepts are introduced for a fragment of natural language which can be analyzed in terms of that of propositional logic.

2.1. Propositional Semantics

Let us assume we have a language of proposition logic built up from a set of proposition letters by means of negation and conjunction. Interpretation is stated relative to a model $M = \langle W, V \rangle$, consisting of a set of possible worlds W and an initial valuation function V, which assigns sets of worlds to proposition letters.

Satisfaction is defined relative to such a model M and a world $w \in W$:

Definition 1 (Propositional Satisfaction)

- $w \models_M p$ iff $w \in V(p)$
  $w \models_M \neg\phi$ iff $w, \not\models_M \phi$
  $w \models_M \phi \wedge \psi$ iff $w \models_M \phi$ and $w \models_M \psi$

Obviously, these are the standard satisfaction conditions.

In terms of the satisfaction conditions we can spell out the contents of our sentences as the sets of possible worlds in which the sentences are true:

Definition 2 (Propositional Contents)

- the content of ϕ in M, $[\![\phi]\!]_M$, is $\{w \mid w \models_M \phi\}$

It is completely straightforward to see that the contents of our sentences could as well have been spelled out independently and in a compositional fashion, for:

Observation 1 (Composition of Contents)

- $[\![\neg\phi]\!]_M = W \setminus [\![\phi]\!]_M$
  $[\![\phi \wedge \psi]\!]_M = [\![\phi]\!] \cap [\![\psi]\!]$

Clearly, this little propositional system has a classical, Boolean semantics.

2.2. Pragmatics of Assertion

Let us now see how our propositional system language can be used to characterize aspects of propositional information *exchange*. Agents are assumed to exchange information about the (actual state of the) world, and the information they have can be modeled by means of the sets of possible worlds which might be the actual one for as far as the agents know.

So what can we say that an agent a may pick up when somebody else asserts a particular sentence ϕ (and if a accepts the assertion)? Obviously, this is the update of the information t which he already had with the information which he gets:

Definition 3 (Propositional Updates)

- the update of t with ϕ in M, $[\![\phi]\!]_M(t)$, is $t \cap [\![\phi]\!]_M$

After accepting ϕ, an agent is taken to believe the world to be as he thought it was before the update, and also as it is said to be by the utterance of ϕ.

Let us now take a look at the other party in the game of information exchange, and see how the speaker's information relates, or must relate, to what she says to be the case. One of the main goals of the game of information exchange is that agents exchange true information, so one of the major principles is that the speaker's information state s support the contents of her utterances:

Definition 4 (Propositional Support)

- s supports ϕ in M, $s \models_M \phi$, iff $s \subseteq [\![\phi]\!]_M$

If one says that the actual world is a ϕ-world, then one ought not to conceive it possible that the actual world is not a ϕ-world.[1]

2.3. Coherent Information Exchange

The present propositional notions of content, update and support are closely and coherently related. Update and support can also be specified independently, and in a compositional fashion:

Observation 2 (Compositional Updates)

- $[\![\neg\phi]\!]_M(t) = t \setminus [\![\phi]\!]_M(t)$
  $[\![\phi \wedge \psi]\!]_M(t) = [\![\psi]\!]_M([\![\phi]\!]_M(t))$

Observation 3 (Compositional Support)

- $s \models_M \neg\phi$ if for no $\emptyset \subset s' \subseteq s$: $s' \models_M \phi$
  $s \models \phi \wedge \psi$ iff $s \models_M \phi$ and $s \models_M \psi$

This is appealing, because it shows that we can study the logic of update and support separately and in a formally transparent fashion, without having to worry about the implications of this for the philosophically well-motivated notion of linguistic content.

It is also easily seen that the update with a supported assertion is safe. The information which an agent has after the update with a supported assertion is contained in the joint information which he and the speaker had before the assertion:

Observation 4 (Supported Updates)

- if $s \models_M \phi$ then $s \cap t \subseteq [\![\phi]\!]_M(t)$

This fact guarantees that a proper exchange of propositional information does not corrupt the information exchanged. Thus, if the interlocutors start out with correct information, and only exchange supported information, then each of the player's information after the exchange is correct, too. Surely, this is a key property of a sound system of information exchange.

2.4. Pragmatics of Conditionals

In this paper implication is, as usual, defined in terms of negation and conjunction: $\phi \rightarrow \psi =_{\mathrm{df}} \neg(\phi \wedge \neg\psi)$. Thus:

Observation 5 (Implication satisfaction)

- $w \models_M \phi \rightarrow \psi$ iff, if $w \models_M \phi$ then $w \models_M \psi$

Several objections have been raised against such a material implication analysis of conditional sentences, one of these being that it is too weak. An implication is already satisfied by w if the antecedent clause is not satisfied. When we take pragmatic matters into account, however, such an objection loses its bite.

Consider the support a speaker may have. If it is true that if $s \models_M \phi$ then $s \models_M \psi$, it does not yet follow that $s \models_M \phi \rightarrow \psi$. More is required to make the latter point true, viz., that $\forall w \in s$: if $w \models_M \phi$ then $w \models_M \psi$. This is a stronger requirement with an intensional flavor. State s supports $\phi \rightarrow \psi$ if s has evidence for some trans-world dependency between ϕ and ψ.

This point can be appreciated further when we consider the following two related facts:

Observation 6 (Role Switches)

- $\sigma \models \neg\phi$ iff $[\![\phi]\!](\sigma) = \emptyset$
- $\sigma \models \phi \rightarrow \psi$ iff $[\![\phi]\!](\sigma) \models \psi$

Support for a negation $\neg\phi$ consists in evidence against an attempted *update with* ϕ. Similarly, support for an implication consists in support for the consequent which is functionally dependent on an update with the antecedent. Notice that this functional dependence should not be a trivial one. Grice's quantity maxims exclude evidence for $\neg\phi$ or ψ to support an utterance of $\phi \rightarrow \psi$.

When we consider a fragment of natural language which can be modeled by a language of propositional logic, the notions of content, update and support are seen to be coherently related. Content of, update with, and support for the utterance of a propositional sentence can be defined separately, and any one of these notions can be defined in terms of another. Besides, exchange of information can be defined in a safe way. What about a fragment of natural language with the expressive power of first order predicate

[1] This notion of support may serve to implement Grice's maxims of quality. It can also be used to specify Hamblin's notion of commitment: a speaker can be said to be committed to having the information supporting what she says.

logic? The next section discusses what are appropriate notions of content, update and support when we are concerned with the utterance of sentences with indefinite elements (in particular pronouns, and indefinite noun phrases)?

3. First Order Exchange

In this section we sketch the semantics and pragmatics of a first order language, built upon three assumptions, relatively well-motivated in the philosophical literature (cf., e.g., (Stalnaker, 1998)). First, indefinite noun phrase (modeled by means of existentially quantified phrases) are generally used with referential intentions; second, anaphoric pronouns may refer back to the individual which a preceding indefinite was intended to refer to; third, both kinds of terms are linearly ordered in discourse. The set up of this section mirrors that of the previous one. We first specify a satisfaction semantics for a first order language (which includes pronouns), and then we develop suitable notions of content, update and support. (The observations, definitions and results of this section are extensively discussed in (Dekker, 2000).)

3.1. Satisfaction of First Order Assertions

Let us assume a language built up from variables, pronouns ($\mathrm{p}_1, \mathrm{p}_2, \ldots$) and relational constants by means of negation, existential quantification and conjunction. Variables are dealt with in the usual way by means of variable assignments and pronouns look back in the discourse for preceding antecedents. A pronoun p_i is interpreted as coreferential with the i-th existential found when going back in the discourse from the place where the pronoun occurs. Relative to a sequence e, a pronoun p_i selects the i-th individual e_i from that sequence, thus indicating that it is coreferential with the i-th potential antecedent in preceding discourse.

Formulas are interpreted relative to sequences of individuals satisfying both their existentials and pronouns. (For a start, we neglect 'worldly' information.) Given that existentials (associated with indefinites) are used with referential intentions, and since these occur in a linear order, the 'length' $n(\phi)$ of a formula ϕ is that the number of existential quantifiers in ϕ not in the scope of a negation. Satisfaction is defined, furthermore, relative to a first order model $M = \langle D, E\rangle$ consisting of a domain of individuals D and an interpretation E for the non-logical constants. It is defined as follows:

Definition 5 (First Order Satisfaction)

- $e \models_{M,g} Rt \ldots t'$ iff $\langle [t]_{g,e}, \ldots, [t']_{g,e}\rangle \in E(R)$
  $e \models_{M,g} \exists x\phi$ iff $e{-}1 \models_{M,g[x/e_1]} \phi$
  $e \models_{M,g} \neg\phi$ iff $\neg\exists c \in D^{n(\phi)}$: $ce \models_{M,g} \phi$
  $e \models_{M,g} \phi \wedge \psi$ iff $e \models_{M,g} \psi$ and $e{-}n(\psi) \models_{M,g} \phi$
  where $e{-}m$ is the sequence $e_{m+1}, e_{m+2}, \ldots$

This satisfaction semantics is dynamic in that it accounts for inter-sentential anaphoric binding. Due to its treatment of existentials, pronouns, and its dynamic notion of conjunction, the following two formulas are equivalent:

Observation 7 (Dynamic Conjunction)

- $\exists x(Dx \wedge \exists y(Py \wedge Fxy)) \wedge \exists z(Tz \wedge S\mathrm{p}_1\mathrm{p}_2 z) \Leftrightarrow$
  $\exists z\exists x(Dx \wedge \exists y(Py \wedge Fxy) \wedge (Tz \wedge Sx\mathrm{p}_1 z)) \Leftrightarrow$
  $\exists z\exists x\exists y(Dx \wedge Py \wedge Fxy \wedge Tz \wedge Sxyz)$

This mirrors the natural language equivalence between:

(1) A diver found a pearl. She sold it to a tourist.
(2) A diver sold a pearl she found to a tourist.

Our satisfaction semantics also accounts for a worn-out example like the famous 'donkey-sentence':

(3) If a farmer owns a donkey he beats it.
(4) Every farmer beats every donkey he owns.

since, e.g.,

Observation 8 (Dynamic Implication)

- $(\exists x\phi(x) \rightarrow \psi(\mathrm{p}_1)) \Leftrightarrow \forall x(\phi(x) \rightarrow \psi(x))$

The interested reader is referred to (Dekker, 2000) for a further exposition of this system.

3.2. First Order Contents and Updates

In the propositional system of section 2 formulas were defined to be satisfied by certain possible worlds, and in the first order system by certain sequences of individuals, in an extensional manner. The two definitions can be combined by resorting to intensional models $\mathcal{M} = \langle W, D, I\rangle$, which consist of a set of worlds W, a domain of individuals D, and an interpretation function I such that for all $w \in W$: $\mathcal{M}_w = \langle D, I_w\rangle$ is an (extensional) model:

Definition 6 (First Order Contents)

- $[\![\phi]\!]_{\mathcal{M},g} = \{we \mid e \models_{\mathcal{M}_w,g} \phi\}$

The first order contents of a formula ϕ are given by (Heim, 1982)'s satisfaction sets, sets of worlds and sequences of individuals which jointly satisfy the conditions imposed upon them by ϕ.

Although our first order notion conjunction from section 3.1 has been seen to be dynamic, it can be understood to derive from a (Boolean) form of intersection:

Observation 9 (Conjunction is Boolean)

- $[\![\phi \wedge \psi]\!]_{\mathcal{M},g} = [\![\psi]\!]_{\mathcal{M},g} \cap [+n(\psi)][\![\phi]\!]_{\mathcal{M},g}$

In this observation $[+n(\psi)]\tau$ indicates the *update* of the satisfaction set τ with the fact that $+n(\psi)$ more terms have been used after the satisfaction set $\tau = [\![\phi]\!]_{\mathcal{M},g}$ has been established, i.e., after the assertion of ϕ.

With the notion of conjoining contents at hand, first order updates can be specified as follows:

Definition 7 (First Order Updates)

- $[\![\phi]\!]_{\mathcal{M},g}(\tau) = [\![\phi]\!]_{\mathcal{M},g} \cap [+n(\phi)]\tau$

Under this definition, the update notion of conjunction turns out to be a form of composition:

Observation 10 (Conjunction as Composition)

- $[\![\phi \wedge \psi]\!]_{\mathcal{M},g}(\tau) = [\![\psi]\!]_{\mathcal{M},g}([\![\phi]\!]_{\mathcal{M},g}(\tau))$

Interestingly, the dynamics of conjunction (and that of interpretation more in general) thus can be seen to reside in the temporal order of updating with the successive conjuncts. Conversely, the dynamic notion of conjunction (composition) is seen to be based on an underlying (Boolean) notion of intersection.

Our notions of content and update are not merely derived notions, for

Observation 11 (First Order Compositionality)

- both the contents of and the dynamic update with a first order formula ϕ can be defined independently and compositionally

The definitions are given in (Dekker, 2000). As a matter of fact, it appears to be immaterial whether one adopts a classical or an update notion of meaning as the basic one if one sets out to deal with intersentential anaphoric relationships

3.3. Compositional Support and Coherence

Like we said in the introduction to this section, a speaker is assumed to use indefinites (and other terms, for that matter) with referential intentions. When we turn to the support for first order utterances, these intentions have to be dealt with explicitly.

Indefinite terms (like definite ones) are assumed to be supported by specific subjects in the speaker's state of information. These states of information which an agent a has about sequences of individuals are modeled by sets of worlds and sequences of individuals, viz., the worlds w which a thinks might be the actual one, with the sequences of individuals e as the individuals which she has information about (her subjects). A speaker's information state, thus, can also be modeled as a Heimian satisfaction set.

In order to indicate that a speaker has a certain subject in mind when uttering an indefinite or pronoun, we use links l which relate the terms she utters to the subjects of her information state. Given that the (uttered) terms are linearly ordered, these links can be specified as sequences of subjects of the speaker. Thus, e.g., a link ij for an utterance of a formula ϕ indicates that the last term in ϕ is supposed to be supported by the i-th subject of the speaker, and the one but last term by her j-th subject.

First order support can now be defined as follows:

Definition 8 (First Order Support)

- $\sigma \models_{\mathcal{M},g,l} \phi$ iff $\sigma \subseteq [l][\![\phi]\!]_{\mathcal{M},g}$

In this definition, $[l]\tau$ translates the information $\tau = [\![\phi]\!]_{\mathcal{M},g}$ has about $n(\phi)$ subjects into information about the corresponding subjects of σ (given by l), and that information is supposed to be supported by σ's subjects. (See (Dekker, 2000) for more discussion.)

Under the present definition, support for a dynamic conjunction corresponds to its dynamic satisfaction, for:

Observation 12 (First Order Conjunction Support)

- $\sigma \models_{\mathcal{M},g,l} \phi \wedge \psi$ iff $\sigma \models_{\mathcal{M},g,l} \psi$ and $\sigma \models_{\mathcal{M},g,l-n(\psi)} \phi$

The linking functions l can be seen to mediate between satisfying sequences of individuals and supporting subjects. Support for two anaphorically related terms actually consists in support for the two terms by one and the same subject so that

Observation 13 (Anaphoric Support)

- $\sigma \models_{\mathcal{M},g,l} \exists x \phi \wedge F\mathrm{p}_1$ iff $\sigma \models_{\mathcal{M},g,l} \exists x(\phi \wedge Fx)$

Like the notions of content and update, that of support can be stated independently:

Observation 14 (Compositional Support)

- the support for a first order formula ϕ can be defined independently and compositionally

(The definition is spelled out in (Dekker, 2000).) Actually, it turns out to be immaterial whether one adopts a support or another (satisfaction, update) notion of meaning as basic.

Now that we have defined (related) notions of first order update and support, we may turn back to the issue of safe information exchange, this time at the first order level. Its characterization is somewhat more involved:

Observation 15 (Supported First Order Updates)

- if $\sigma \models_{\mathcal{M},g,l} \phi$, then $(\sigma \cap [m]\tau) \subseteq [l'][\![\phi]\!]_{\mathcal{M},g}(\tau)$

provided that $l' = l \cup (m \circ -n(\phi))$ be a function

There is no room, here, to discuss the details of the side conditions, but, basically, it amounts to this. An update of τ with an utterance of ϕ is safe, if the utterance is supported by σ under a link l, and if pronouns unresolved in ϕ are justly matched with subjects in σ. Notice that if ϕ itself is resolved, that is, if ϕ contains no unresolved pronouns, the situation is entirely similar to that of the propositional setting. If ϕ is not resolved however, then the hearer should, of course, take care in resolving pronouns in the right way. We, again, refer to (Dekker, 2000) for further discussion.

3.4. Support for Dynamic Implications

In section 2.4 we have seen that the support for an implication consists in more than the simple truth or satisfaction of a material implication, even though it is eventually based on a material satisfaction analysis. Thus one argument against the material implication semantics of conditionals was made harmless. At the first order level, one more argument against this analysis is countered, as our notion of support also solves what has been known as "Peirce' Puzzle" (Read, 1992).

The following two formulas are equivalent in ordinary predicate logic:

(5) $\exists x(\phi(x) \leftarrow \psi(x))$
(6) $\exists x \phi(x) \leftarrow \forall x \psi(x)$

Peirce considered the following sentence:

(7) There is some married woman who will commit suicide in case her husband fails in business.

Although (7) is of the form (5), it seems to express a much stronger statement than (8), which is of the, deemed equivalent, form (6):

(8) Some married woman will commit suicide if all married men fail in business.

Peirce therefore proposes an alternative analysis, as to which (7) states that there is some married woman who under all possible courses of events would commit suicide if her husband would fail, again, with a stronger interpretation of the embedded conditional.

However, upon a material implication analysis of conditionals, the support (not satisfaction-) conditions of (7) are the same as those Peirce argues for. For (7), under an analysis as in (5), to be supported, the speaker must have a woman in mind such that upon all courses events which the speaker conceives possible, that woman commits suicide in case her husband fails. That is, again under Gricean assumptions, (7) requires a speaker to acknowledge a dependence between failure and suicide, relative to a woman which the speaker has in mind. (Likewise, (8) also expresses a dependence, but a different one, and not relative to some specific woman.)

4. Dynamics of Dialogue

4.1. Blocking Issues

Under the pragmatic perspective adopted in this paper, it is no mystery why indefinites and pronouns interact in the dynamic way they do. Like definite noun phrases, indefinites noun phrases are used with referential intentions, and anaphoric pronouns simply pick up the referents intended to be associated with their antecedents. This has been characterized using sequences of individuals in a Tarskian fashion. The dynamics of conjunction has been seen to reside in the (assumed) linear construction of discourse. Basically, conjunction is intersection, but dynamic conjunction in addition acknowledges the order of terms (the order of referential intentions) in a discourse. These principles are independently motivated on pragmatic grounds.

Thinking of it, it is not so much the dynamics of indefinite reference and anaphoric co-reference that is conceptually puzzling, but the lack of it when indefinites figure in certain constructions: under a negation, in the antecedent of conditionals, in the restriction of quantifiers, and also in interrogatives and imperatives. If, as we think, anaphoric potential derives from referential intentions, then why do these referential intentions vanish in these negative (and other) contexts?[2]

In this section we attribute the blocking effects of the mentioned constructions to their typical role in discourse and dialogue. As we will see in the next section, a proper acknowledgment of the typical role of these constructions in discourse and dialogue also gives us a natural handle on the type of dependencies which indefinites in these constructions do give rise to.

4.2. Negation as a Role Switcher

We think the blocking effects of negations and other constructions can be understood well when we take into account their typical role in actual dialogues and the thematic structure of such dialogues. A negation *Not S* may serve to answer the issue—raised explicitly or implicitly—whether *S* is true. For instance, consider an utterance of (9):

(9) Farly doesn't run a sushi bar.

Typically, such an utterance does not serve to state of Farly and of some sushi bar that the first doesn't run the last. It is much more likely that it states—possibly in answer to the question whether Farly runs a sushi bar—that he doesn't, that is, that there is no such bar which Farly runs.[3] Generally, then, a speaker need not have a particular sushi bar in mind when uttering (9), and the reason may be that, intuitively, the existence of such a sushi bar is not part of what the speaker claims to have evidence for. Rather, the existence of such a bar appears to be part of the issue which the speaker addresses—negatively in example (9)—, or even part of what the hearer might have claimed just before. So actually, when somebody utters (9), she is normally not coming up with a sushi bar herself, but she is claiming to have evidence against the existence of such a bar, were anybody else thinking of the possibility of there being one, or even of thinking of claiming there actually to be one.

As a matter of fact, this intuition is already implicitly spelled out in our definition of negation support. For it turns out that the first observation in (6) also holds for our first order system:

Observation 16 (First Order Role Switch)

- $\sigma \models \neg\phi$ iff $[\![\phi]\!](\sigma) = \bot$[4]

State σ supports $\neg\phi$ iff the update with ϕ is absurd, that is, iff σ has evidence against ϕ. Notice that the support for $\neg\phi$ can be thus, spelled out in terms of an update with ϕ, and that updates with indefinites do not presuppose referential intentions with the updater.[5]

4.3. Implication as a Double Role Switcher

The conception of negation as a role switching device has a further pay off when we consider the pragmatics of conditionals $\phi \rightarrow \psi$, which are defined in terms of negation (and conjunction). For instance, the evidence which a speaker may bring to bear upon her assertion of a conditional sentence can be seen to consist in her evidence for the consequent of the conditional, were she to accept anybody else's evidence for the antecedent.

Also this observation is implicitly accounted for, as, evidently, the second observation in (6) holds as well in our first order system:

[2] Surely this 'generalization' requires some qualification. Indefinite noun phrases can be used 'specifically' in all of the mentioned contexts—under a negation, in the restriction or scope of adnominal and adverbial quantifiers, etc. However, we think that the majority of indefinites in these contexts is not 'specific' in this sense.

[3] One has to be careful with these types of sentences, because alternative interpretations are easily made available by emphasizing, e.g., *Farly*, or *run*. We here assume the utterance to carry what may be called a 'neutral' intonation.

[4] We here assume ϕ to be resolved.

[5] The indicated role switch is of course reminiscent of the one adopted in systems of game-theoretical semantics (*GTS*, cf., e.g., (Hintikka and Sandu, 1997)). In *GTS*, the truth of ϕ is defined in terms of the existence of a winning strategy for a 'verifier'. A verifier is supposed to come up with evidence for ϕ and be able to supply witnesses for (indefinite) terms in ϕ. However, when it comes to a negation $\neg\phi$, the verifier gets the role of the 'falsifier' of ϕ, who has to refute any attempt to verify ϕ by somebody else.

Observation 17 (First Order Double Role Switch)

- $\sigma \models (\phi \rightarrow \psi)$ iff $\exists l$: $[\![\phi]\!](\sigma) \models_l \psi$[6]

Not only does this provide motivation for the often attested existential closure over the indefinites in the antecedent of a conditional, but it also suggests the speaker's evidence for the consequent clause to be functionally dependent on the possible witnesses for these indefinites. As we will see below, the functional type of support which the speaker can be required to have for indefinites in the consequent clause can be cashed out by subsequent pronouns, if these are also read functionally.

4.4. Background and Focus in Discourse

In the preceding discussion the typical impact of negations (and implications) has been spelled out in terms of the different roles of the speaker and a hearer. In *GTS*, these roles have been qualified as that of an (initial) verifier and that of an (initial) verifier, but we think the pragmatic division of labour at issue is more general than that.

Coherent discourses and dialogues generally consist of assertions which have an (explicit or implicit) 'background' or 'topic' part, and an (explicit) 'focus'.[7] Typically—that is, if context or intonation have no interfering effects—one could say that e.g., the contents of negated sentences, the antecedents of implications, and the restrictions on quantifiers constitute a background or topic, which the speaker is not automatically supposed to support, but which she is supposed to react upon. It is the focus part of her utterance which she can be required to have support for, possibly in functional dependence on such a background.

Given this it is no mystery that, by default, indefinites in the background part of an assertion do not introduce possible referents for pronouns used in subsequent utterances. Since the speaker need not be required to support that part of her utterance, these indefinites fall beyond her 'pragmatic jurisdiction' so to speak, and they are not assumed to be used with referential intentions. Indefinites in focus, however, do require speaker's support, and generally are associated with referential intentions. However, since the focus may be functionally dependent on a background, the referential intentions associated with indefinites in focus may be functional, too.

5. Functional Dependencies

In the preceding section we already indicated that support for, and satisfaction of, the consequent of an implication can be functionally dependent on that of its antecedent. However, so far, the possible witnesses of indefinites in the consequent, or, rather, the possible witness-*functions*, did not appear in the satisfying sequences or supporting states themselves. In this section we show that a more principled account of the dependencies is not only possible, but also desirable. We here restrict ourselves to stating the relevant satisfaction conditions. Corresponding notions of update and support can be obtained by suitable generalizations.

5.1. Transparent Universal Quantification

Consider the following well-known example:

(10) Harvey courts a girl at every convention. She usually comes to the banquet with him. (after Karttunen)

Surely this sentence may serve to state something about a particular girl which Harvey courts at every convention, but—knowing Harvey—it is probably not about one particular girl. Thus, (10) may serve to state that, at every convention, there is a girl Harvey courts, and which he takes to the banquet with him. Which girl that is is of course functionally dependent on which convention we consider, and this is where functional witnesses come in.

An utterance of the first sentence of (10) can be taken to involve a reference to possible functions g, which associate girls which Harvey courts with the conventions he visits. Therefore, the utterance of the second sentence can be taken to refer back to these and to state that for most c, if c is a convention which Harvey visits, then $g(c)$ accompanies Harvey to the banquet of c.

Such readings can be derived compositionally by combining the techniques from, e.g., (Jacobson, 1999) with the account of anaphoric relationships sketched in this paper. Consider the following notion of universal quantification:

Definition 9 (Transparent Universal Quantification)

- $wfe \models_g \forall x\phi$ iff $\forall d \in D$: $wf(d)e \models_{g[x/d]} \phi$
  (with $f : D \rightarrow D^{n(\phi)}$)

This definition of universal quantification is transparent as it it doesn't invoke any existential closure over the indefinites in ϕ, their witnesses being given by the function f in the satisfying sequence, relative to the possible values of x. Upon this definition (and assuming an extension of our language with function variables), we can state the following equivalence:

Observation 18 (Dynamic Skolem Quantifiers)

- $\forall x \exists y \phi(x,y) \Leftrightarrow \exists f \forall y \phi(x, f(x))$

This Skolem equivalence is special, since it is dynamic: the witness function f is accessible for subsequent pronouns. For example (10) this means that the pronoun "she" can be associated with a suitable antecedent.

5.2. Transparent Implication

As has already been indicated, conditional sentences also license functional anaphoric dependencies. Consider:

(11) If a book is printed with Kluwer it has an index. It can always be found at the end. (after Heim)

Support for an utterance of the first sentence of (11) may consist of a witness function f, assigning indices to books printed with Kluwer. If the speaker has such a function in mind, then she may refer back to it with a pronoun when she subsequently utters the second sentence. The second

[6] The whole implication is assumed to be resolved.

[7] Cf., e.g., (Jackendoff, 1972; Karttunen and Peters, 1979; von Stechow, 1991; Rooth, 1992, Ch. 6) for a number of formally quite different analyses of this distinction, which we, however, think are really close in spirit to the one we have in mind. Notice that, when we use the term 'focus', we do not mean 'contrastive focus' here.

utterance then is assumed to be about books printed with Kluwer, too, and expresses that, always, if b is a book printed with Kluwer, then $f(b)$ can be found at the end of b.

The satisfaction of implications can be adjusted in the following principled, and pragmatically motivated way:

Definition 10 (Transparent Implication Satisfaction)

- $wfe \models (\phi \rightarrow \psi)$ iff $\forall c \in D^{n(\phi)}$: if $wce \models \phi$ then $wf(c)ce \models \psi$ (with $f : D^{n(\phi)} \rightarrow D^{n(\psi)}$)

Also with this definition, we escape existential closure over indefinites in the scope (consequent) of the implication. The indefinites are supposed to be satisfied by the values of the satisfying function f, relative to the possible values of the indefinites c in the antecedent.

That this notion of implication is really transparent can be seen from the following equivalence:

Observation 19 (Dynamics Skolem Implication)

- $(\exists x\phi(x) \rightarrow \exists y\psi(y)) \Leftrightarrow \exists f(\exists x\phi(x) \rightarrow \psi(f(\mathrm{p}_1)))$

Again, the witnesses supporting indefinites in the consequent are available for subsequent anaphoric pronouns, as long as these can be conceived to be functionally dependent upon possible witnesses for indefinites in the antecedent.

The interpretation of implications in our framework is a strong one (as it is in most systems of dynamic semantics) in the sense that it amounts to universal quantification over the possible value of indefinites in the antecedent clause. A sophisticated use of the witness functions f, however, allows us to generate weak and asymmetric readings as well, both in a transparent fashion.

5.3. Transparent Beliefs

Techniques similar to the ones discussed above can be used to approach the interpretation of indefinites in modal and belief contexts. Consider two typical examples of so-called modal subordination:

(12) Mary thinks there is a burglar in the house. She thinks he came in through the chimney.

(13) A wolf might come in. He would eat you first.

When asserting (12) the speaker can be taken to refer to what constitutes Mary's representation of possible witnesses of a possible burglar in her belief state; similarly, the witness for a wolf can be seen to be dependent on the possibilities in which a wolf comes in. In either case, the witness is available as a referent for a subsequent pronoun, if that pronoun can be interpreted as functionally dependent upon the same parameters, viz., Mary's belief state, or the possibility that a wolf comes in.

We here refrain from spelling out the exact details of a suitable definition of the satisfaction of these modal statements, as it involves some technical complications. We refer to (Dekker and van Rooy, 1998) for further discussion.

5.4. Specific Indefinites

A last issue which we want to touch upon here is the phenomenon of specific indefinites, cf., e.g., (Abusch, 1994; Reinhart, 1997; Kratzer, 1998). As we already said above, indefinites sometimes do escape from contexts which normally constitute a background, and which systematically forbid genuine quantifiers to be raised from there. Here are a couple of representative examples:

(14) If a certain linguist shows up, we are supposed to be particularly polite, but do you remember who? (Reinhart)

(15) Max did not consider the possibility that some politician is corrupt. (Kratzer)

(16) If three relatives of mine die, I will inherit a house. (Ruys)

(17) If each student improves into two subjects, then noone will fail the exam. (Schlenker)

We will not discuss, here, how such indefinites exactly manage to escape these contexts. (Apparently, it has to do with information structure.) Rather, we are concerned with the pragmatic or semantic impact of the readings obtained.

As Kratzer observes, the specific readings of the above examples are quite a bit pragmatically infected, but we think they are infected more than has generally been acknowledged. For both upon a wide scope indefinite interpretation, and upon a choice function analysis, the above sentences seem to suffer from the same problem as the one addressed by Peirce. Consider, for instance, example (14). Upon both analyses the example can be seen to be satisfied as long as we can pick up any linguist who doesn't show up. Intuitively, this is not correct. Upon a specific interpretation, an assertion of (14) ought to be made with a particular linguist in mind, relative to whom our expected politeness is functionally dependent upon the possibility of her showing up. As the reader may remember from the discussion in section 3.4, no such problem will arise within our pragmatically oriented framework. Our notion of support requires the speaker to dispose of precisely that kind of information.

Conclusion

In this paper we have studied meaning and interpretation from a pragmatic point of view. We have sketched a systematic and coherent account of the content of, update with, and support for sentences of a first order logic. Our general perspective on information exchange has allowed us to deal with intersentential anaphoric binding, from all three perspectives.

We have shown that our notions of content, update and support each can be defined independently, and in terms of one another. We have also shown under which conditions first order information exchange is safe, crucially relying upon updates with supported assertions. We have argued that a pragmatic notion of support indeed undermines some of the objections which has been leveled against the semantic account of conditionals as material implications. The pragmatic outlook upon the use of indefinites has finally inspired a more principled analysis of the support of indefinites in background focus structures, one which enabled a straightforward extension of our empirical scope.

With regard to the last issue more has to be said though. In order for pronouns to be interpreted functionally, the background of the antecedent indefinites has to be recovered, and we have said nothing about how this can be achieved. The theses (Geurts, 1995; Frank, 1997; van

Rooy, 1997; Stone, 1998) provide some recent, dynamic, analyses of this type of reconstruction.

6. References

Dorit Abusch. 1994. The scope of indefinites. *Natural Language Semantics*, 2:83–135.

Paul Dekker and Robert van Rooy. 1998. Intentional identity and information exchange. In Jerry Seligman and Patrick Blackburn, editors, *Proceedings of ITALLC'98*, Chiayi, Taiwan. National Chung Cheng University.

Paul Dekker. 2000. Meaning and use of indefinite expressions. manuscript, University of Amsterdam, submitted for publication in the *Journal of Logic, Language and Computation*.

Anette Frank. 1997. *Context Dependence in Modal Constructions*. Ph.D. thesis, Universität Stuttgart, Stuttgart.

Bart Geurts. 1995. *Presupposing*. Ph.D. thesis, Universität Stuttgart, Stuttgart.

H.P. Grice. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics, Vol III: Speech Acts*. Academic Press, New York.

H.P. Grice. 1978. Further notes on logic and conversation. In P. Cole, editor, *Syntax and Semantics, Vol IX: Radical Pragmatics*. Academic Press, New York.

Jeroen Groenendijk and Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy*, 14(1):39–100.

Irene Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts, Amherst. published in 1988 by Garland, New York.

Jaakko Hintikka and Gabriel Sandu. 1997. Game-theoretical semantics. In Johan Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 361–410. Elsevier, Dordrecht.

Ray Jackendoff. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, Massachusetts.

Pauline Jacobson. 1999. Towards a variable-free semantics. *Linguistics and Philosophy*, 22:117–84.

Hans Kamp. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, editors, *Formal Methods in the Study of Language*. Mathematical Centre, Amsterdam. Reprinted in J. Groenendijk, T. Janssen, and M. Stokhof (eds.), *Truth, Interpretation and Information*. Foris, Dordrecht, 1984.

Lauri Karttunen and Stanley Peters. 1979. Conventional implicature. In Choon-Kyu Oh and David A. Dinneen, editors, *Syntax and Semantics 11 – Presupposition*, pages 1–56. Academic Press, New York.

Angelika Kratzer. 1998. Scope or pseudoscope? are there wide-scope indefinites? In Susan Rothstein, editor, *Events in Grammar*, pages 163–196. Kluwer, Dordrecht.

Stephen Read. 1992. Conditionals are not truth-functional: an argument from Peirce. *Analysis*, 52:5–12.

Tanya Reinhart. 1997. Quantifier scope: How labor is divided between qr and choice functions. *Linguistics and Philosophy*, 20:335–397.

Mats Rooth. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116.

Robert Stalnaker. 1978. Assertion. In Peter Cole, editor, *Syntax and Semantics 9 – Pragmatics*, pages 315–332. Academic Press, New York.

Robert Stalnaker. 1998. On the representation of context. *Journal of Logic, Language and Information*, 7:3–19.

Matthew Stone. 1998. *Modality in Dialogue: Planning, Pragmatics and Computation*. Ph.D. thesis, IMS, University of Pennsylvania.

Robert van Rooy. 1997. *Attitudes and Changing Contexts*. Ph.D. thesis, IMS, Stuttgart.

Frank Veltman. 1996. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261.

Arnim von Stechow. 1991. Focussing and background operators. In Werner Abraham, editor, *Discourse Particles*. John Benjamins, Amsterdam.

Lifelong Discourse Representation Structure

Gábor Alberti[1]

Pécs University
H-7624 Pécs, Ifjúság 6, Hungary
albi@btk.jpte.hu, aalberti@mail.matav.hu

Abstract

We argue that the key to the solution of both the theoretical problem of working out a realistic picture of the "hearer's (permanently changing) information state" (abbr. HIS) within the framework of Discourse Representation Theory (DRT; van Eijck and Kamp, 1997) and the empirical problem of a wide range of classical formal-semantic puzzles (concerning the creation (or retrieval?) of referents for pronouns and definite descriptions in universal and belief contexts and in other special cases) lies in one and the same discovery: HIS is essentially to be regarded as a discourse representation structure, a gigantic "lifelong" DRS furnished with a partially ordered set of *worlds*, a (multiple) *cursor* (pointing to temporal, spatial and rhetorical reference points) and a *meaning function.* The structure we propose is based on three denumerably infinite set of *pegs* (in Landman's (1986) sense), those of *referents*, *predicate names* and *worlds*, whose inner structures and the rich system of connections among them are to be defined by simultaneous recursion (see the Appendix after References).

1. Dynamic perspective on semantics

Our starting-point is DRT (e.g. Kamp, 1981; Heim, 1983; van Eijck & Kamp, 1997), which we consider to be a successful attempt to extend the sentence-level Montagovian model-theoretic semantics (Dowty *et al.,* 1981), which had not only failed to exceed this level but had also been unsuccessful in the treatment of certain types of anaphoric relations, to the discourse level. Its essence lies in the discovery that the failure of the immediate interpretation of sentences / discourses in the static Montagovian world model is to be attributed to the fact that the discourse under interpretation is permanently becoming part of the world in which it is being interpreted; thus a level of discourse representation *must* be inserted in between the language to be interpreted and the world model serving as the context of interpretation. This *dynamic perspective* of DRT can be captured by regarding the content of DRSs as a "(partial) function mapping information states to information states" (Zeevat, 1991: 17).

Nevertheless, the picture of HIS in DRT and related dynamic theories (e.g. Groenendijk & Stokhof, 1990, 1991) is oversimplified and in certain areas simply counter-intuitive: practically (total) models are used as information states. In this approach atomic statements (e.g. 'Peter loves Mary') are to be held to be *tests,* which means that an assertion heard is supposed to be either corroborated or rejected. The typical case is excluded: to regard an assertion as a new piece of information for the hearer. (Another basic problem of DRT —the one concerning the compositional transition between syntax and DRS— is discussed in Alberti (1998, 1999))

Hearers, counter to the oversimplified picture sketched above, practically always have a partial knowledge. And nothing other than DRS serves the purpose of representing partial knowledge.

2. Lifelong DRS

The hearer's permanently changing information state can be defined by simultaneous recursion (see the table in Appendix).

We essentially regard this definition as a generalization of the (also simultaneously recursive) definition of DRSs given in van Eijck & Kamp (1997); or this original definition may be regarded as one suitable for discourses with an empty mutual background knowledge shared by speaker and hearer. This stipulation on background knowledge mentioned may often serve (or have served so far) as a useful working hypothesis — especially when DRT is compared to the Montagovian model-theoretic semantics— but it is undoubtedly far from being a realistic picture of discourses. Now it is high time, hence, to turn to the general situation on the basis of experiences and results obtained by studying the restricted area in the last two decades.

First of all, three denumerably infinite sets of *pegs* should be assumed to be at the hearer's disposal: those of *referents* (R), *predicate names* (P) and *worlds* (W). They are 'pegs' (Landman, 1986) in the sense that before use they contain no information, they are only carriers of information. The inner structure of these sets and the rich system of connections among them are due to six (partial) functions/relations:

[1] The expression "Lifelong DRS" is due to an anonymous reviewer of an earlier article of mine on the topic (Alberti 1996/1999), who was selected by Anna Szabolcsi, the guest-editor of *Acta Linguistica Hungarica.* I would like to thank both of them for the expression.
The Hungarian Széchenyi Professorial Scholarship (1999-2002) has made it possible for me to write this article. I am also grateful to the Netherlands Institute for Advanced Studies in the Humanities and Social Sciences (NIAS, Wassenaar) and the Hungarian National Scientific Research Fund (grant no. OTKA F026658) for their contribution to my (travel) costs.

i. The *extension of predicates* is a partial function ext : P → Pow(R*) from predicates to the powerset of referent sequences (* denotes the Kleene star).[2]

ii. Another partial function ref : R → Pow(P×R*) *(referent function)* assigns each referent in its domain a set of sequences consisting of a predicate name and referents; an element of Pow(P×R*) is essentially a basic kind of DRS.

iii. Relation prc *(precedes* or <) is a partial ordering in W×W with a least element, denoted by v (the *basic world).*

iv. wrl : (P ∪ R) → W is also a partial function *(world function);* it assigns a predicate name or a referent a world.

v. There is a *cursor,* a partial function cur : {W,R} → W∪R*, which chooses an *active* world (cur(W)) and a sequence of referents playing distinguished roles in different respects in the current state of the hearer's (permanently changing) information state: cur(R) = ⟨$\mathrm{cur}_{temporal}$(R), $\mathrm{cur}_{spatial}$(R), $\mathrm{cur}_{rhetorical}$(R), ...⟩.

vi. There is also a *meaning function:* a partial function mea : P → Pow(P×R*). It maps a predicate name to a DRS (meaning postulates).

The starting-point of the simultaneously recursive definition of HIS as a sextuple <ext, ref, prc, wrl, cur, mea> is fixing a one-member base: <∅, ∅, ∅, ∅, cur(W)=v, ∅> where ∅ denotes the empty set. We propose seven kinds of recursive steps (Appendix); their names are intended to refer to their operation: *expansion (of extensions) of predicates* (EXP), *introduction of a new predicate* (INP), *cursor move* (CUM), *introduction of a new referent into the active world* (IREA) and *a new world* (IREN), *referent assignment to a (generalized) DRS* (RED). Observe the first four components of HIS (the LDRS) define a DRS: the usual box structure corresponds to the tree of worlds, and function wrl is responsible for linking referents to boxes. The recursive steps are to capture different linguistic and extralinguistic ways of gathering information at the hearer's disposal. The linguistic ways will be sketched below, and will be described as (certain combinations of) special cases of the above listed recursive steps.

3. Where is the referent?

Subsections 3.1-5. provide an informal sketch of the treatment of a couple of famous semantic puzzles in the approach based on defining HIS as a Lifelong DRS. The first subsection demonstrates the use of RED and the cursor-moving operation (CUM).

3.1. Referents, predicates, and then referents again

(1) shows the bidirectional connection between referents and predicates. On the one hand, we describe properties ('r2 ∈ pretty'), classes ('r1 ∈ boy') and relations ('⟨r1, r2⟩ ∈ ext(love)') of referents by means of predicates. On the other, we can also refer to "products" of this linguistic activity (RED) by means of referents ('admitted(r1,r3,r2)' where r3 is assigned to the situation expressed by the first sentence; and 'surprise(r4,r5)' where 'friend-of(r5,r1)' and ref(r4) essentially corresponds to the DRS expressing the information of the second sentence (according to a preferred reading)).

(1) The boy loves a pretty Dutch girl. He admitted IT to her. His friend was surprised by IT.

The sentence in (2a) below describes a strange custom whose (Davidsonian) referent r_{sc} belongs to, say, the basic world (v), whereas the farmer (r_f) and the donkey (r_d), and then the merchant (r_m) belong to "later" worlds: wrl(r_{sc})=v, wrl(r_f)=w1, wrl(r_d)=w1, wrl(r_m)=w2, where v<w1<w2, according to partial ordering prc. The alternative continuations (2b-d) can "activate" different worlds (v, w1 (+w3), w2, respectively) due to conjunctions, adverbs etc., which determines accessibility of referents (EXP, CUM, IREA, IREN and RED are concerned).

(2) a. If a farmer owns a donkey, HE sells IT to a merchant.
b. ... Mary is surprised at THIS STRANGE CUSTOM.
c. ... Or HE hires IT out to A FOREIGNER.
d. ... Although HE usually gets little money from HIM.

The participants of sentence (2b) belong to world v, including the referent of 'this strange custom' identical with the Davidsonian argument r_{sc} of sentence (2a). Sentence (2c), with 'or' as its first word, "accepts" participants of w1 (the farmer and the donkey) but, instead of world w2, it evokes a world w3, which the merchant (r_m) does not belong to but in which there is a 'foreigner': v < w1 < w3 (where w2 and w3 are incommensurable elements according to partial ordering prc), and wrl($r_{foreigner}$)=w3. Continuation (2d) makes the cursor choose world w2 (cur(W)=w2) due to the adverbial expression 'usually' referring to the generalizing / generic content of the conditional sentence in (2a); and it is in this world w2 that the merchant can be retrieved (by the pronoun 'him').

The most important law illustrated by these examples is that a referent r belonging to a world w can be referred to in worlds w' such that w'≥w. The farmer, for instance, but not the merchant, can be referred to in world w3 because v < w3, but w2 and w3 are incommensurable worlds.

(3a-c) show the same strategy in different areas: the first sentences "activate" a world (distinct from the basic world v), which remains active due to temporal/aspectual/etc. tools. These tools "retain" us in the world of the game (3a), of the typical convention where Harvey is present (3b), and of the speaker's wishes (3c).

(3) a. Every player chooses a pawn. HE puts IT on square one. (canonical scenario)
b. Harvey courts a girl at every convention. SHE is usually very pretty. (univ. quantifier)
c. I wish Mary had a car. Peter could drive IT / THE CAR too. (modal expression)

There may be even explicit references to worlds to be activated, as is shown below in (4) (CUM, IREN). 'The

[2] The mapping ext(love) = {⟨r_P,r_M⟩, ⟨r_M,r_P⟩, ⟨r_P,r_A⟩}, for instance, can be understood as follows: there is a HIS in which love ∈ P, and the r's are referents (corresponding to, say, persons named Peter, Mary and Ann), and the given hearer is assumed to know that Peter and Mary love each other, and Peter lo ves Ann, too. As HIS expresses partial knowledge, the hearer is not assumed to be sure that other people do not love each other.

first case' refers to a potential world with a secretary while 'the second case' makes another potential world active, which the gardener belongs to (the task of making coffee referred to by 'it' is assumed to belong to v). These two alternative (incommensurable) worlds are introduced due to the disjunctive structure in the first sentence.

(4) We ought to employ either a secretary or a gardener. In the first case THE SECRETARY would make coffee from now on, whereas in the second case IT would be Mary's task but THE GARDENER should sweep the yard.

3.2. Where is the referent that does not exist?

Due to RED, as in the case of (1), HIS after working up sentence (5a) below may contain the referent that 'this victory' in continuation (5b) is intended to retrieve. In continuations (5c-e), 'this victory' refers to different —underspecified— situations.

In (5c) 'the victory of our team A over the Spanish team' *—with no temporal anchor—* is referred to. Sentence (5d) is about 'the victory of *one of our teams* over the Spanish team.' Finally, 'this victory' in (5e) refers to 'an arbitrary victory of one of our teams over anybody.'

(5) a. Yesterday our team A won a victory over the Spanish team.
b. ... THIS VICTORY is marvelous.
c. ... Did not THIS VICTORY happen the day before yesterday?
d. ... I wish our team B could replicate THIS VICTORY today.
e. ... I wish our team B could replicate THIS VICTORY over the English today.

Operation RED enables us to create all the appropriate referents by means of the "generalization function" G mentioned in Appendix: G is to replace certain referents with "variables." More precisely, referents not used up earlier can be used as "variables," because such referents have not been individualized by any kind of information so predicating something of them is no more than making existential statements about them, and these existential statements are to be regarded as logical entailments of sentence (5a). Thus no new information is applied in the course of the replacement of referents carried out by partial function G — in harmony with the fact that no new information is at the hearer's disposal relative to his/her information state just after working up sentence (5a).

In the case of continuation (5d), for instance, where 'the victory of *one of our teams* over the Spanish team' is referred to, partial function G is to replace the referent belonging to team A with a non-individualized referent, which is a legitimate operation because statement (5a) implies that 'a team (of ours) has won a victory over the Spanish team.' The relevant entailment of (5a) in the case of continuation (5e) is that 'a team (of ours) has won a victory over another team,' and concomitantly partial function G is to substitute variable-like referents for both the referent of team A and that of the Spanish team.

Case (5c) shows that temporal referents must not have been ignored. Here the role of G amounts to substituting a variable-like referent for the referent belonging to the particular point of time when team A is assumed by the speaker to defeat the Spanish team.[3]

By now the case illustrated by continuation (5b) has become "extreme." Here partial function G required to calculate the reference of 'this victory' on the basis of the Davidsonian referent belonging to statement (5a) is to choose to be an empty function (whose domain is empty).[4]

HIS after working up the first sentence ((6a) below) in the following discourses contains no referent for a priest / dog. We claim, however, that the priest's referent can be created — by extending the mentioned stage of HIS without exploiting (really) new information, due to regarding HIS as a Lifelong DRS. That is what makes case (6) similar to the phenomenon illustrated by (5); but now not logically derivable existential entailments will be applied.

Here our starting-point is that it is a plausible assumption that a marriage is associated with a potential priest, but not with a dog, at least in Christian cultures. (6) warns us that this "association" does not amount to a logical implication but a *licensed* piece of cultural/encyclopedic knowledge (Kálmán, 1990), since it is not claimed in (6a) that the marriage was a religious one.

(6) a. Joe got married yesterday.
b. ... THE PRIEST spoke very harshly.
c. ... *THE DOG barked very loudly.

Operation SPED enables us to create the priest's referent, by applying it to the following pair of associated DRSs: <"x gets married," "y organizes x's marriage where y is a priest"> ∈ ASS, where ASS is one of the elements of the set P of predicates. This formula roughly means that if somebody gets married, it is typical that a priest organizes the ceremony. Note, however, that we do not regard it as being excluded that the same hearer's information state contains the following statement simultaneously: if somebody gets married, it is typical that a registrar organizes the ceremony.

As for technical details, here we need a "specifying function" S (see Appendix), which is to substitute Joe's particular referent for the "general" x (here variable-like referents are used again). This operation is permitted because the first member of the associated pair of DRSs mentioned above is considered to be true by the hearer with Joe's referent in the role of x (Joe got married indeed); and the crucial element of the result of the operation is as follows: the hearer (already) *thinks* that there is a priest who organized Joe's marriage in the precise sense of 'thinking' that this priest has a referent belonging to the basic world of HIS (or at least to a world preceding the fictive world that referent x belongs to).

One might think that (6c) can serve as a well-formed continuation of (6a) in an appropriate context. Suppose, for instance, that (the hearer knows that) Joe has a dog which barks loudly whenever it feels that its owner is in danger... This piece of information is to be regarded as part of some interpersonal knowledge at the hearer's disposal. It may be formulated by means of associated pairs of non-specific DRSs in the extension of relation

[3] This temporal referent should be replaced in cases (5d,e) as well.

[4] If G(r) is not defined, referent r is not to be replaced.

ASS, too; so the referent of the dog can be produced as the priest's one above.

The only difference lies in the source / nature of "mediating" information. The exploited information belongs to a supposed interpersonal knowledge in the case of (6c), to the hearer's cultural / encyclopedic knowledge in the case of (6b), and it is worth mentioning here that logical consequences were used in the case of the phenomenon illustrated by example (5). All these three kinds of information, together with the sort of information that can be referred to as 'lexical,' should be assumed to be stored in HIS in similar format, perhaps separated from each other, but all kinds should be accessible in the course of processing a discourse from sentence to sentence. In this way a wide range of phenomena is accounted for where "non-existing" referents should be "retrieved," or rather produced; and not only a certain referent is introduced as a result of our approach but the hearer commits him-/herself to a whole story that the given referent is a participant of (Kálmán 1990).[5]

3.3. Cursor

The temporal reference point changing from sentence to sentence, whose introduction (to DRT) is proposed by Kamp & Reyle (1993: 5.2.2), is very easy to capture in LDRS. It can be defined as a "cursor" pointing to the currently active temporal referent. We are going to argue that it would be useful to use at least four cursors, or lesser cursors with multiple values.

First of all, we need a *world cursor,* a function that assigns the set W of worlds a distinguished world, which can be called to be 'active' in the current state of HIS. As was mentioned in passing in subsection 3.1., antecedents of expressions in a sentence processed are to be sought among referents belonging to the active world or preceding worlds. The movement of this cursor can be described (and restricted) according to the partial ordering prc of worlds. Certain linguistic factors seem to make the cursor remain at a world or turn to another world *adjacent* to the last active world according to the partial ordering of worlds

There are extralinguistic factors, however, that seem to make the cursor choose a world independently of its last value. When two people enter into conversation with each other, for instance, they should "activate" the worlds in their LDRSs carrying their shared interpersonal knowledge. It can be accounted for in this way why a question like "What about Peter" cannot be interpreted in certain situations whilst it is perfect in other situations without any special introductory discourse. The problem is not that the given hearer knows no Peter or more Peters; practically every hearer can be assumed to know several Peters. This fact, however, does not imply that the question discussed will always be ill-formed. The question will obviously prove to be perfect if, and only if, a single referent named Peter happens to belong to the *active* world of the conversation.

Revealing rules of movement of the world cursor requires much future research, of course. What is claimed here is that furnishing traditional DRSs with a cursor pointing to a distinguished "box" is a promising idea; and what makes this picture realistic is that DRSs with this cursor are assumed to belong to the hearer interpreting a discourse, and not to the discourse itself, which is the essence of the LDRS approach.

Let us return to the kind of cursor corresponding to Kamp & Reyle's (1993) temporal reference point. Discourse (7a) below, similar to one analyzed by the authors mentioned above (p526), serves as an illustration of our approach. Now we need a cursor whose values are temporal reference points. It can be defined as a function from the set R of referents which chooses one referring to a point of time, which can be denoted by $\mathrm{cur}_{\mathrm{temp(oral)}}(R)$.

(7) a. A man entered the White Hart. Bill served him a beer. The man paid.

After processing the first sentence, which describes an event, the cursor takes the referent r_{enter} of the point of time when the man entered the pub as its value. Then let r_{serve} denote the time of the event of Bill's serving him a beer. Kamp & Reyle's (1993) statement on discourses consisting of events can be paraphrased in the framework of LDRS as follows: the temporal referent of the new event (chronologically) follows the active temporal referent. In the case discussed, hence, $r_{\mathrm{enter}} = \mathrm{cur}_{\mathrm{temp}}(R) < r_{\mathrm{serve}}$. For the third sentence, then, r_{serve} will serve as an active temporal reference point, and we obtain on the basis of the generazation mentioned above: $r_{\mathrm{serve}} = \mathrm{cur}_{\mathrm{temp}}(R)' < r_{\mathrm{pay}}$.

Other kinds of relevant reference points can be obtained by a straightforward generalization of the cursor function demonstrated above. Let cur(R) be defined as a vector of referents with the active temporal value as its first element. Then a spatial reference point may occupy the second position, to be denoted by $\mathrm{cur}_{\mathrm{spat(ial)}}(R)$. We can formulate rules such that, after processing a sentence describing a movement, the referent of the place associated with the *goal* thematic role will serve as the spatial cursor value. In example (7a), thus, the event described by the second sentence is understood to take place in the pub named White Hart, which is the goal of the entering event described in the first sentence. That is, the beer mentioned in the second sentence is served in this pub (at least that is the default reading). The paying event that the third sentence describes is also to be understood to take place in the same pub (according to the preferred reading) because the second sentence triggers no shift in active spatial referent (suppose the man getting the beer has a beneficiary role in the serving situation, and not a goal role).

Topic shift, illustrated below in (7b) by a Hungarian example (Pléh, 1982), is a further phenomenon that can be accounted for by means of the referent cursor. This third

[5] An anonymous reviewer of an earlier version of this paper has considered this method of producing referents to be too productive (at least from a practical point of view). What I accept is that some additional tool is required but I retain that the kinds of phenomena discussed in the subsection require HIS to contain the pieces of information of different nature stored in some way or another. The additional tool can be some kind of (also permanently changing) *weighing* sensitive to the (frequency or temporal distribution of) use of associative connections in the domain of relation ASS. Thus a model of oblivion should be worked out. Continuation (6c) is well-formed, for instance, only if the hearer happens to be clearly aware of the interpersonal piece of information that Joe has a single special dog. Cultural / encyclopedic pieces of knowledge, however, belong to a more stable sphere of the content of HIS.

element of cur(R) can be called the 'active rhetorical referent' ($\mathrm{cur}_{\mathrm{rhet(orical)}}$(R)).

(7) b. Anna$_{i}$ megverte Marit$_{j}$. A következõ percben pro$_{i/*j}$ / AZ$_{j/*i}$ sírni kezdett. Aztán pro$_{?}$ hazament.
Anna hit Mari-acc. The following minute-in pro/that weep-inf began. Then home-went.
'Ann hit Mary. In the following minute she burst into tears. Then she went home.'

The "dropped" pronoun in the second sentence is to be interpreted as referring to the person referred to by the topic / subject of the first sentence whereas the explicit demonstrative pronoun can refer to the other participant mentioned in the first sentence, yielding "topic shift." The referent of the dropped pronoun (pro) in the third sentence is unambiguously determined by the particular form of the second sentence: it is identical with the referent of the subject of the second sentence; thus there is no topic shift between the second and the third sentence.

These observations are easy to account for by means of the rhetorical cursor value. After processing the first sentence, the referent belonging to its subject / topic is to be chosen to play the role of the active rhetorical referent. As for fixing the referent of the subject of the second sentence, the following general rule seems to be valid: a dropped pronoun retrieves the active rhetorical referent whilst the task of a demonstrative pronoun is just the opposite: its referent is to differ from the active rhetorical referent, causing topic shift. This is a straightforward case of division of labor between alternative linguistic expressions. The interpretation of the third sentence corroborates the rule formulated above: there is no topic shift, i.e. the referent of its dropped pronoun is identical with the active rhetorical referent after processing the second sentence, which, however, depends on the choice between the two versions in the subject position of the second sentence.

3.4. Each other's belief referents

We follow Zeevat (1991: 20) in assuming that "...making the assumption that one can refer to private objects and that the idea of a private model can be worked out suffices for dealing with most of the classical belief puzzles. The contribution of DRT in this respect is to supply a theory about the structure of this private model and a set of rules for the evolution of this private model under the influx of new information." LDRS is obviously an attempt to work out the "private model" Zeevat is speaking about.

Let us study the following famous example of Geach's in order to illustrate one of the most stubborn kind of belief problems. The problem lies in the fact that it is difficult to account for the *coreference* between 'a witch' and 'she' in a traditional logic because neither witch exists so neither has a "normal" referent.

(8) Hob thinks that a witch poisoned his pig and Nob thinks that SHE killed his goat.

In our approach, the hearer may have a world (presumably not the basic one) where Hob's witch's referent can be found and it may be assumed to coincide with that of Nob's witch, since it is a straightforward assumption that their beliefs rely on the same rumor. Furthermore, people can be assumed to be able to link each other's corresponding referents together in conversations. The game named Quiz or Twenty Questions serves as clear evidence: two people begin to speak about something entirely unknown to one of them. Almost the only thing they know is that they are speaking about the *same* entity; that is, new pieces of information should be associated with one and the same referent, marked out at the beginning, throughout the whole game.

Let us highlight the crucial element of our approach based on regarding hearers' information states as LDRSs. As was shown in 3.1., "verbal products," created in the course of talking to each other, can be referred to as easily as perceived entities of the real world around us; and the same holds true of hearers' information states: referents in them can also be referred to. A HIS is like a painting or a map in this sense: its details can be talked about.

3.5. Meaning function and Qualia Structures

Pustejovsky's (1995, 129) *long record* is intended to illustrate the promising possibility for embedding his Qualia Structures (or something similar) —together with the explanatory capacity of this theory— in LDRSs.

(9) John bought A LONG RECORD.

What Pustejovsky questions is that the semantic content of the expression in question is to be captured by the (simplified) logical formula *long(x)∧record(x),* which a simple DRS representing the content of the expression could be based on, too. We are arguing that Pustejovsky's argumentation according to which the given meaning of *long* is not reasonable to store in a constant (core) lexicon can be reconciled with an LDRS-compatible approach in which the human sentence-to-sentence processing is assumed to "generate" temporary extended lexicons relative to which the above mentioned simple formula can be retained.

Suppose the proper meaning of *long* is not at the hearer's disposal at a certain point (HIS1) of working up sentence (9): (s)he tries to construct the appropriate DRS from pieces like 'bought(r1, r2),' 'record(r2),' 'long$_{i}$(r2),' but there is no proper index i. Hence, HIS1 has to be extended to another HIS (without exploiting new information), as in the cases discussed in section 3.2. Although now it is not a referent that is being sought / produced, firstly a referent can be found on the basis of a piece of cultural information: that of the playing time of the record, for which, say, 'long$_{17}$(r3)' is semantically well-formed, where 'playing-time-of(r3,r2).' Note that a certain property of *polysemy* helps find the predicate we have happened to mark with 17: the phonetic form of this predicate is identical with that of the (still) unknown one.

And now the hearer knows not only what is *long* but what the intended meaning of long$_{i}$ has been. (S)he may assign a new predicate peg to this predicate, denoted by, say, long$_{147}$, and (s)he may define mea(long$_{147}$) on the basis of long$_{17}$ and the connection discovered (INP).

4. Summary

It has been argued in this article that the key to the solution of both the theoretical problem of working out a realistic picture of the hearer's (permanently changing) information state within the framework of Discourse Representation Theory and the empirical problem of a wide range of classical formal-semantic puzzles lies in one and the same discovery: HIS is essentially to be regarded as a discourse representation structure, a gigantic Lifelong DRS furnished with a partially ordered set of worlds, a (multiple) cursor (pointing to temporal, spatial and rhetorical reference points) and a meaning function.

LDRS can be regarded as a generalized version of DRS and can (also) be defined by simultaneous recursion. Three denumerably infinite sets of peg-like elements are assumed to be at the hearer's disposal as a starting-point (in addition to an empty LDRS_0, corresponding to the moment of the ideal hearer's birth): those of referents, predicate names and worlds. The inner structure of these sets and the rich system of connections among them are due to six (partial) functions/relations; their definitions are based on the simultaneously recursive technique mentioned above (Section 2 and Appendix).

Section 3 has provided a sketchy review of the treatment of a couple of famous semantic puzzles in the approach based on defining HIS as a Lifelong DRS.

Subsection 3.1. has been devoted to the discussion of problems concerning +/– accessibility of referents from given (partly fictive) worlds evoked by certain parts of discourses.

3.2. has dealt with questions concerning the retrieval, or rather construction or calculation, of "non-existing" referents on the basis of logical, lexical, cultural/ encyclopedic and/or interpersonal pieces of "mediating" information, all kinds to be assumed to be stored in HIS in similar format and to be accessible in the course of processing a discourse from sentence to sentence.

3.3. has contained arguments for the introduction of a cursor pointing to the active world, distinguished temporal and spatial reference points, and an "active rhetorical referent" playing a crucial role in accounting for topic-shift phenomena.

In 3.4. we have argued that most of the classical belief puzzles can be treated in LDRS due to our approach according to a person's information state, including her/his beliefs and wishes, is like a painting or a map in the sense that its details can be talked about and referred to by referents (of no special status).

Finally, 3.5. has demonstrated the possibility for embedding a Pustejovskyan (1995) multistratal "generative lexicon" in our extended theory of discourse representation.

5. References

Alberti, G., 1996. Generative Argument Structure Grammar: A strictly compositional syntax for DRS-type representations. Published in *Acta Linguistica Hungarica* 46/1-2 (1999): 3-68. Budapest and Dordrecht: Akadémiai and Kluwer Academic Publ.

Alberti, G., 1998. GASG: The Grammar of Total Lexicalism Based on Prolog. To appear in L. Hunyadi (ed.), *Studies in Applied Linguistics.* Debrecen: KLTE.

Alberti, G., 1999. GASG: The grammar of total lexicalism. *Working Papers in the Theory of Grammar* 6/1. Theoretical Linguistics Programme, Budapest University (ELTE) and Research Institute for Linguistics, Hungarian Academy of Sciences.

Dowty, D.R., R.E. Wall, and S. Peters, 1981. *Introduction to Montague Semantics.* Dordrecht: Reidel.

Groenendijk, J., and M. Stokhof, 1990. Dynamic Montague Grammar. In L. Kálmán and L. Pólos (eds.), *Papers from the 2nd Symposium on Logic and Language.* Budapest: Akadémiai Kiadó. 3–48.

Groenendijk, J., and M. Stokhof, 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14:39-100.

Heim, I., 1983. File change semantics and the familiarity theory of definiteness. In: R. Bäuerle, Ch. Schwarze, and A. von Stechow (eds.), *Meaning, use and interpretation of language.* 164-190. Berlin: De Gruyter.

Kálmán, L., and L. Pólos, 1990. Deferred information: The semantics of commitment. In L. Kálmán and L. Pólos (eds.), *Papers from the 2nd Symposium on Logic and Language.* Budapest: Akadémiai Kiadó. 125–157.

Kamp, H., 1981. A theory of truth and semantic representation. In J. Groenendijk, T. Janssen and M. Stokhof (eds.), *Formal methods in the study of language.* Amsterdam: Mathematical Centre.

Kamp, H., and U. Reyle, 1993. *From Discourse to Logic.* Dordrecht: Kluwer Academic Publ.

Landman, F., 1986. *Towards a theory of truth and information: The status of partial object in semantics.* Dordrecht: Foris.

Pléh, Cs., 1982. Topic and subject prominence in Hungarian. In F. Kiefer (ed.), *Hungarian Linguistics (Linguistic and Literary Studies in Eastern Europe* 4.). Amsterdam: John Benjamins.

Pustejovsky, J., 1995. *The Generative Lexicon.* Cambridge: MIT Press.

van Eijck, J., and H. Kamp, 1997. Representing discourse in context. In J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language.* Amsterdam and Cambridge: Elsevier and MIT.

Zeevat, H., 1991. *Aspects of Discourse Semantics and Unification Grammar.* Ph.D. thesis, Univ. of Amsterdam.

name	EXTENSION OF PREDICATES ext: $P \to Pow(R^*)$	ASSIGNMENT OF REFERENTS ref: $R \to Pow(P\times R^*)$	TREE OF WORLDS $prc \subset W^2$ $(prc = <)$	ASSIGNMENT OF WORLDS $wrl : (P \cup R) \to W$	CURSOR $cur:\{W,R\}\to W\cup R^*$	MEANING POSTULATE $mea : P \to Pow(P\times R^*)$	comments
base ($LDRS_0$)	$ext(p) = \varnothing$ $(\forall p)$	$ref = \varnothing$	$prc = \varnothing$	$wrl = \varnothing$	$cur(W) = v$, $cur(R) = ?$	$mea = \varnothing$	ref, wrl and mea are partial orders
1. **EX**PANDING **P**REDICATE (1^{st} row: condition)	$ext'(p)\neq\varnothing$			$cur'(W) \geq wrl'(p)$ $cur'(W) \geq wrl'(r_i)$		mea'(p) exists	$pr_1r_2...r_n$ is a statement
EXP (2^{nd} row: definition)	$ext(p) =$ $ext'(p) \cup \{r_1r_2...r_n\}$				permitted : $cur(W)\neq cur'(W)$ $cur(R)\neq cur'(R)$, depending on p and its distinguished arg's	$mea'(p) \neq mea(p)$ is permitted	there is a subgroup of distinguished predicates in P: e.g. ASS, ⇒, ∈, "I-believe"
2. **IN**TRODUCTION OF A NEW **P**REDICATE	$ext'(p) = \varnothing$			$cur'(W) \geq wrl'(r_i)$, wrl'(p) does not exist		mea'(p) does not exist (mea'(q) exists)	
INP	$ext(p) = \{r_1r_2...r_n\}$			$wrl(p) = cur(W)$	permitted : $cur(W)\neq cur'(W)$ $cur(R)\neq cur'(R)$, depending on p and its distinguished arg's	mea(p) (already) exists	there are relations in P to provide information on associated phon. forms
3. **C**URSOR **M**OVE							cursor moves backwards...
CUM			cur'(W) and cur(W) are adjacent acc. to partial order prc				...or forwards due to ±linguistic factors
4. **I**NTRODUCTION OF A NEW **RE**FERENT INTO THE **A**CTIVE WORLD				wrl'(r) does not exist			
IREA			$cur(W) = cur'(W)$	$wrl(r) = cur(W)$			
5. **I**NTR. OF A NEW **RE**FERENT INTO A **N**EW WORLD			v and w are incommensurable acc. to partial order prc'	wrl'(r) does not exist, $wrl'^{-1}(W)$ is empty			
IREN			$cur'(W) < cur(W) = w$, and they are adjacent acc. to prc	$wrl(r) = cur(W)$			
6. **RE**FERENT ASSIGNMENT TO A (GENERALIZED) **D**RS	$r_{1,1}...r_{1,arg(p1)}\in ext'(p_1)$, ... $r_{k,1}...r_{k,arg(pk)}\in ext'(p_k)$			$wrl'(G(r_{ij}))$ do not exist			$G\subset R^2$ is a partial function ($G\cap id=\varnothing$): "generalizing function"
RED	$G(r_{1,1}...r_{1,arg(p1)})\in ext(p_1)$, ... $G(r_{k,1}...r_{k,arg(pk)})\in ext(p_k)$	$ref(r)=G(\{p_1r_{1,1}...r_{1,arg(p1)}, ..., p_kr_{k,1}...r_{k,arg(pk)}\})$		$wrl(G(r_{ij})) \geq cur(W)$			(where id is the identity function of set R)
7. **SPE**CIFICATION OF AN ASSOCIATED **D**RS	$<q,s>\in ext'(ASS)$, and $\exists S$: $S'(r_{q1,1}...r_{q1,arg(pq1)})\in ext'(p_{q1})$ ... $S'(r_{qk,1}...r_{qk,arg(pqk)})\in ext'(p_{qk})$	$ref'(q)=\{p_{q1}r_{q1,1}...r_{q1,arg(pq1)}, ..., p_{qk}r_{qk,1}...r_{qk,arg(pqk)}\}$, $ref'(s)=\{p_{s1}r_{s1,1}...r_{s1,arg(ps1)}, ..., p_{sm}r_{sm,1}...r_{sm,arg(psm)}\}$,		we need an "appropriate" S' (which replaces only "variables")		formula ASS(q,s) belongs to mea(p)	
SPED	$S(r_{s1,1}...r_{s1,arg(ps1)})\in ext(p_{s1})$, ..., $S(r_{sm,1}...r_{sm,arg(psm)})\in ext(p_{sm})$, where $S' \subseteq S$			"appropriate" S (which replaces certain "variables" with non-used referents in cur(W))			now $S\subset R^2$ is a "specifying" function

Appendix. Table 1: (simultaneously recursive) definition of LDRS (Lifelong Discourse Representation Structure)

Accessibility, Duration, and Modeling the Listener in Spoken Dialogue

E. G. Bard* and M. P. Aylett*†

* Human Communication Research Centre and Department of Theoretical and Applied Linguistics,
Adam Ferguson Building, University of Edinburgh, EH8 9LL, U.K.
ellen@ling.ed.ac.uk
†Centre for Speech Technology Research
2 Buccleuch Place, University of Edinburgh, Edinburgh EH8 9LW, U.K.
matthewa@cogsci.ed.ac.uk

Abstract

Referring expressions are thought to be tailored to the needs of the listener, even when those needs might be very costly to assess, but tests of this claim seldom manipulate listener's and speaker's knowledge independently. The Map Task enables us to do so. We examine two 'tailoring' changes in repeated mentions of landmark names: falling clarity of word articulation and rising accessibility of referring expression. Clarity results replicate Bard et al. (2000). Standardized word duration fell for speaker-Given listener-New items (Expt 1). Hence it was unimportant whether the listener heard an earlier mention. Reduction between mentions was no greater when it could be inferred that the listener could see the named item (Expt 2), and no less when the listener explicitly declared that they could not (Expt 3). Hence it was unimportant whether the listener could see the landmark. Reduction was unaffected by whether the repeater could see the mentioned landmark (Expt 4). Articulation thus depends only on what the speaker has heard previously. In contrast, accessibility was more sensitive both to listener (Expt 1) and speaker knowledge (Expt 4). The results conform most closely to a Dual Process model: fast, automatic, word-by-word processes let the speaker's own experience prime articulation, while computationally costly assessments of listener knowledge control influence referring expression design where competing tasks permit.

1. Introduction

Speakers are said to design their utterances to suit the needs of their listeners, insofar as those needs can be known (Ariel, 1990; Clark and Marshall, 1981; Gundel,., Hedberg, and Zacharski,, 1993; Lindblom, 1990). Certainly there is variation in form. Clarity of pronunciation varies with predictability from local context (Hunnicutt, 1985; Lieberman, 1963) and with repeated mention (Fowler and Housum, 1987). Forms of referring expression differ in elaboration with the more readily interpreted, those having more accessible antecedents, syntactically simpler (*a blacksmith's cottage* v *it*) (Ariel, 1990, Fowler, Levy, and Brown, 1997; Gundel, J.K., Hedberg, N., and Zacharski, R., 1993; Vonk, Hustinx, and. Simmons, 1992). Yet maintaining an incrementally updated model of what the listener knows, including the established common ground, and of what the listener needs to know is a considerable cognitive task. Because speaker's and listener's knowledge overlap and because it may be impossible to assess the listener's knowledge accurately, it is suggested that speakers often default to an account of their own knowledge as a proxy for the listener's (Clark and Marshall, 1981). In fact, many discussions of this topic simply assume that the two are the same: they describe or manipulate the speaker's knowledge without independently manipulating the listener's (see Keysar, 1997).

This paper presents two attempts to examine the hypothesis that referring expressions are genuinely tailored to the speaker's model of the addressee. One deals with the articulation of individual words, the other with the form of referring expressions. They have different implications for psychological models of dialogue. Current models of the production of language indicate that noun phrase structure and articulation are generated within units of different sizes, phonological phrases or tone groups on the one hand and phonological words, lexical words, or syllables on the other (Wheeldon and Lahiri, 1997; Levelt and Wheeldon, 1994). Speech appears to be produced in a cascade, with smaller units being prepared for articulation as the succeeding larger unit is being designed. Thus, incrementally updating a listener model in order to articulate each phonological word would impose a much heavier computational burden on a speaker, than updating it phrase by phrase.

We will first present hypotheses which the literature offers us for the way in which speakers manage the task of modeling listeners while planning and producing speech. Then we will report two studies which test these hypotheses on materials from a single corpus. Both make the same comparisons. Bard et al. (2000) excised a balanced sample of spontaneously uttered words, and measured their intelligibility to naïve listeners as well as their duration. The new results in the present paper report duration and accessibility of all suitable referring expressions in the corpus. Finally, we will discuss the implications of the comparison for the nature of listener modeling in on-line utterance generation.

2. Modeling listener knowledge while speaking

The literature offers several versions of the hypothesis that what we say is tailored to the needs of our listeners. They can be arranged in order of the computational demands they would impose on speakers.

Lindblom's *H-and-H Hypothesis* (1990) makes the heaviest demands. It posits that speakers adjust the articulation of spoken words to the knowledge which the listener can currently recruit to decoding the speech signal. Thus, speakers hyper-articulate when listeners lack such auxiliary information and hypo-articulate when redundancy is high. There is ample evidence that linguistic environments which provide more redundancy contain word tokens articulated with greater speed and less precision. (Bard and Anderson, 1983, 1994; Fowler and Housum, 1987; Lieberman, 1963; Samuel and Troicki, 1998). The question is whether this relationship depends on the speakers updating and consulting a model of the listener's current knowledge each time they prepare the prosodic character of a phonological word or the articulation of its syllables. Though the H-and-H view does not preclude defaulting to the speaker's own knowledge, it is framed under in terms of genuine listener knowledge. To adjust articulation on line to a non-default account of listener knowledge, speakers should observe listeners continuously for signs of misunderstanding or disagreement. Wherever speaker's knowledge and listener's knowledge differ, listener's knowledge should take precedence. In effect the H-and-H Hypothesis corresponds to a *Negligible Defaulting Hypothesis.*

The second alternative arose from a consideration of how speakers might manage the many tasks involved in generating appropriate utterances in dialogue. Brown and Dell (1987) propose a modular division between the initial formulation of utterances, a process based on speaker knowledge, and the monitoring and revision of output, processes based on a model of listener knowledge, or more precisely, of common ground. Called the *Monitoring and Adjustment Hypothesis* by Horton and Keysar (1996), this model defaults first and pays later – if necessary. Because responsibility for tailoring utterances to the listener's needs is shared by the interlocutors (Carletta and Mellish, 1996), the speaker's attention can initially be devoted to utterance planing rather than to listener modeling. Faultless utterances, those for which speaker- and listener-knowledge are alike, are produced quickly and accurately. Poorly designed utterances can be revised in response to explicit requests from the listener, which are received well after the initial planning of the faulty utterance is complete. If the Monitoring and Adjustment Hypothesis holds, post-feedback utterances should reflect any aspects of listener-knowledge which the feedback has conveyed. Otherwise, listener-knowledge should be irrelevant to production.

The third proposal deals with co-presence, middle- or long-term characteristics of listeners which affect likely overlap with speakers' own knowledge. Various kinds of 'co-presence' in social or regional background (Isaacs & Clark, 1987; Fussell & Krauss, 1992), physical location during the interaction (Schober, 1993), or recent experiences (Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992; Brennan & Clark, 1996) are taken into account. Although this work is usually interpreted as showing that the 'initial design' (Horton & Keysar, 1996) of conversational speech is sensitive to listeners' needs, it does not directly address on-line processes. Since most discussions of this notion focus on lasting characteristics of listeners, we assume that it is also intended to reduce the number of occasions in a dialogue when a speaker must update a model of the listener. If so, speakers should attend to evidence for and against co-presence, and defaults could hold for some undefined time after positive evidence. We will call this the *Co-presence Default Hypothesis.*

Finally, Bard et al. (2000) develop a suggestion of Brown and Dell (1987) which we will call the *Dual Process Model.* It proposes a division between fast and automatic processes, which have no computational cost, and slower, more costly processes requiring inference or attention. The former include priming, an unconscious process that allows the performance of an activity or the recognition of a stimulus to reduce the reaction time for or the duration of a behaviour. (Balota, Boland, and Shields, 1989; McKoon and. Ratcliff, 1980; Mitchell and Brown, 1988; O'Seaghdha, 1997.). Only the speaker's own experience is effective in priming. The latter include all those complex forms of reasoning usually implicated in the ability to construct a model of the listener. In competition with this set are the computations which underlie the ability to plan a dialogue or keep track of a shared task. When there is competition for time and attention, the second set of processes may suffer (Horton and Keysar, 1996), leaving the speaker with only cost-free defaults in the form of his own knowledge.

Of these four hypotheses, the second and third make roughly the same predictions for speakers' ability to tailor form of referring expression and word articulation to the listener's needs. Where speaker and listener have different pertinent knowledge which the speaker might access, Monitoring and Adjustment would predict that both form of expression and articulation will reflect the speaker's own knowledge until some corrective feedback points out the discrepancy. Co-presence tells us that long- or mid-term information is available for the whole language production process.

The other two hypotheses might distinguish between the two measures. . H-and-H, the Negligible Defaulting Hypothesis, makes no comment on units larger than words. What is essential is that an account of listener needs is available for each lexical or phonological word. This could be provided in two ways. In the more complex, speakers must conduct two parallel series of updates on the listener model: word-by-word while uttering one phrase and simultaneously, as if that phrase were complete, while constructing the next. Thus different states of listener knowledge would have to be modeled at the same time. This alternative seems so demanding that to preserve the essential predictions of H-and-H, word-by-word operations would have to take precedence, leaving phrase-by-phrase operations either impoverished or non-existent. Thus word intelligibility and duration should be the more sensitive to listener needs. In the simpler arrangement, the redundancy of each word would be assessed as part of the design process preceding the construction of their phrase. Thus clarity and accessibility should be equally sensitive to the listener's knowledge, because they are designed around the same reasoning about that knowledge.

The Dual Process Hypothesis makes a straightforward prediction. Here, the critical issue is the scale or duration of the process and not the stage when it occurs. Under the time pressures imposed by real conversations, smaller scale processes involved in articulatory design of phonological words should seldom allow scope for costly interaction with the listener model, and would have to be controlled by speaker knowledge. Larger-scale processes, like planning an NP, could cycle slowly enough to permit updating the listener model, drawing inferences from it, and the like. Accordingly, form of referring expression could be more sensitive than duration to any records of listener knowledge which speakers maintain. This hypothesis does not predict uniformly good tailoring of referring expressions to listener knowledge, however, because the task of updating the listener model may have an inherently low priority.

3. Studies of intelligibility and accessibility

3.1. Given-ness and referring expressions

To test the foregoing hypotheses, we made use of the effects of Given status. Word tokens in expressions introducing New items are longer and clearer than those referring to Given items (Fowler and Housum, 1987). Forms of referring expression are known to differ in elaboration so that changes with repeated mention are usually abbreviations (*a blacksmith's cottage....it*) which can be assigned a place in a scale of referential accessibility (Ariel, 1990, Gundel., Hedberg, and Zacharski, 1993) To compare the two systems, we used a corpus of spontaneous speech designed to vary what each interlocutor could see, coded to what each had mentioned or heard mentioned, and to what feedback each had given the other. Thus it was possible to select items which were Given to one or both interlocutors on the basis of what each saw, said or heard in the dialogue. Table X summarizes the comparisons which formed the basis of 4 experiments.

3.2. Method

3.2.1. Materials

All materials were drawn from the HCRC Map Task Corpus (Anderson et al., 1991), 128 unscripted dialogues from 64 pairs of *Glasgow* University undergraduates communicating routes defined by labeled cartoon landmarks on schematic maps of imaginary locations. Instruction Giver's and Follower's maps for any dialogue matched only in alternate landmarks. Participants knew that their maps differed but not where or how. In no case could either player see the other's map. The corpus was balanced for familiarity of participants and for ability to see the interlocutor's face. Each participant served as Instruction Giver for the same route to two different Followers and as Instruction Follower for two different routes.

Digital stereo recordings with one channel per speaker were segmented at word boundaries. All the words of any expression referring to a landmark were coded for the appropriate landmark, tagged for part-of-speech, and parsed.

Bard et al. (2000) excerpted individual words from references to the labeled schematic landmarks around which the route is defined in cases where both mentions used at least some of the same open class words (*the rift valley...the rift valley*; *the rift valley...the valley*). Except where the design of the experiment dictated otherwise, items were restricted to the Instruction Giver's initial encounter with a map and were balanced for familiarity of interlocutors and for the availability of a visual channel. Items forming part of disfluencies or interrupted by overlapping speech were excluded.

The present study examined all expressions which refer to landmarks that were mentioned more than once within a dialogue, with the exception of those which were ambiguous as to accessibility. Note that the items for which duration measures are suitable, like those assessed for intelligibility, must include the same words in both mentions. Repetitions using different words in different mentions (*the rift valley...it*), may only be assessed via accessibility.

Score	Definition		Examples
0	numeral + indef art +	noun sequence	*one mountain* *a mountain*
1	def article + possessive +	nominal	*the mountain* *my one*
2	 deictic adj +	possess pron deictic pron nominal	*mine* *that* *this mountain*
3		other pron	*it*

Table 1. Accessibility scale for referring expressions

3.2.2. Dependent variables

Intelligibility loss. Individual open class items from matching repeated mentions were excerpted from context, as were control tokens of the same landmark names read in lists by the original speakers. A standard set of phonetic conventions was used to determine the positions of word boundaries (Laver et al., 1989). All words were overlaid with noise and presented to panels of 9 to 15 naïve informants for identification. The tokens of a word were distributed among informants by Latin square. Intelligibility is the percentage of listeners identifying a word. Intelligibility loss is the difference between the intelligibility of the clearer read token and of the more reduced running speech token of the same word. (See Bard et al. for further details.)

Duration loss. Both studies used normalized duration (Campbell and Isard, 1991). The normalization makes use of the distributions of lengths typical of each phoneme and assigns to each word token a value *k* representing its position in the expected log length distribution for words of its dictionary phoneme composition and stress pattern. The *k*-score makes it possible to compare length-relative-to-expected-length for words of quite different composition. All comparisons were based on the difference between the *k*-durations of a read control form and the corresponding item in running speech.

Accessibility. The 27 items with relative clauses in their first mentions were excluded because of a conflict in coding schemes. All other first and second mentions of landmarks (N = 1136) were classed by accessibility on the scale displayed in Table 1.

3.3. Experiment 1: Listener identity

3.3.1. Design

Experiment 1 examines introductory mentions of the same shared landmarks in Givers' two trials with the same map. In the first trial, the landmark was New to the discourse for both players. In the second, it was Given for the speaker, an Instruction Giver who had mentioned it before, but New to each successive listener (hence the value 'no' in each of the 'how Given status is achieved – by listener" cells in Table 6). The identity of the listener and the state of progress through the map route were both route pre-printed on their maps. Thus, if the Negligible

Measure	Introduction	
	1	2
Word articulation:		
Intelligibility loss	0.072	0.182
k-duration loss	0.498	0.558
Referring expression form:		
Accessibility	0.466	0.552

Table 2. Effects of re-introductions to new listeners on articulatory clarity (intelligibility or length loss relative to citation form) and on form of referring expression (accessibility).

Measure	Repetition	
	Self	Other
Word articulation:		
Intelligibility loss	0.081	0.081
k-duration loss	0.127	0.192
Referring expression form:		
Accessibility	0.878	0.745

Table 3. Effects of self- v other-repetition on change in articulatory clarity (intelligibility and duration loss relative to citation form) and in form of referring expression (accessibility) with repeated mention

expressions are tailored to the listener's needs even when these differ from the speaker's, then introductory mentions of the same landmark should not differ in clarity: no Given-ness effect should be warranted because the named entities are not Given for the listener. Similarly, accessibility of referring expressions should not change. The Monitoring and Adjustment prediction is for changes in both, because the listener in the second trial has had no opportunity for feedback. The Dual Process prediction is that intelligibility will be insensitive to the listener's knowledge and fall, because it depends on the speaker's previous mention. Only accessibility ought to reflect the listener's knowledge and remain unchanged.

3.3.2. Results

Table 2 begins with the intelligibility results of Bard et al.: second introductions show significant loss of intelligibility relative to first introductions (i.e., a greater difference between the carefully pronounced form read in a list and the token produced in running speech). The present study also finds changes in articulation: second introductions are significantly shorter than first (i.e., increasingly different from citation forms) for 239 pairs of words on repeated introductions, (F_2(1, 238) = 12.48; $p <$ 0.0005). In contrast, accessibility does not increase on average over 116 pairs of introductory mentions (F_2(1,115) < 1). Thus, duration appears to reflect the Given status of the item for the speaker, while form of referring expression reflects the fact that the freshly introduced landmark is New for each listener. Greater sensitivity in form of referring expression is predicted only by the Dual Process Hypothesis.

3.4. Repeater identity: inferred listener knowledge

3.4.1. Design

Experiment 2 compared repeated mentions of shared landmarks within and between speakers. As Table 6 shows, in self-repetitions the second token refers to a landmark which is Given to the repeating speaker because he or she has seen the item, and both uttered and heard the original mention. The landmark's status vis-à-vis the listener is less certain. In other-repetitions, the second token is Given to the current speaker only by virtue of having been heard before, but Given to current listener who has mentioned the item, heard it mentioned, and must have been able to see the landmark to introduce it at all.

Negligible Defaulting should promise that either articulation or both articulation and form of referring expression will be show greater change in such cross-speaker repetitions, because an updated account of the listener's knowledge should include the inference that the item was Given to that player in those three ways. In contrast, Monitoring and Adjustment predicts no effect of original speaker on any measure, because no corrective feedback is involved. Co-presence will make the same prediction if we assume that it is satisfied by common experience of the discourse (a 'yes' in both 'Heard' columns in Table 6) without inferring what the listener can currently see. Dual Process predicts that any effect will be found in accessibility, which is designed over intervals long enough to permit inferences to be made.

3.4.2. Results

Table 3 shows that changes in articulatory clarity were the same in self- and other-repetition. Like the intelligibility results of earlier experiments, *k*-duration fell relative to citation form controls with repeated mention (mention: F_2(1,691) = 63.75, $p < 0.0001$) but showed no difference between the 263 other-repetitions and the 430 self-repetitions. (mention x prior speaker: *n.s.*). Accessibility for 90 other-repetitions and the 430 self-repetitions. behaved in the same way (F_2(1,269) = 177.12, $p < 0.0001$; mention x prior speaker: *n.s.*). Once more the listener's experience was not the critical factor, and repetitions of any mention which the speaker has heard are treated alike..

3.5. Feedback: signalled listener knowledge

3.5.1. Design

Experiment 3 provides a more direct test of the effects of listener knowledge. When one speaker introduces an unshared landmark, the listener, who lacks it, may provide corrective feedback indicating the discrepancy between the players' maps. Sometimes, however, that listener fails to find or signal the discrepancy. To test for the effects of feedback on second mentions, we use repeated mentions by the same speaker with and without accurate intervening feedback from the listener.

It is difficult to see how a cooperative speaker, in the usual sense, could ignore such overt evidence. We assume that Negligible Defaulting and Co-presence joint Monitoring and Adjustment predicting that feedback will make a difference to the nature of subsequent mentions. In

Measure	Visibility to listener	
	Not denied	Denied
Word articulation:		
Intelligibility loss	-0.080	0.080
k-duration loss	0.070	0.140
Referring expression form:		
Accessibility	0.470	0.410

Table 4. Effects of feedback about listener's ability to see an entity on changes in articulatory clarity (intelligibility and duration loss relative to a citation form) and in form of referring expression (accessibility) with repeated mention.

Measure	Visibility to speaker	
	Seen	Unseen
Word articulation:		
Intelligibility loss	0.151	0.181
k-duration loss	0.114	0.183
Referring expression form:		
Accessibility	0.745	0.240

Table 5. Effects of speaker's ability to see named entity on change in articulatory clarity (intelligibility and duration loss relative to a citation form) and in form of referring expression (accessibility) with repeated mention.

fact, the repetitions with feedback are the only ones where Monitoring and Adjustment does predict an effect of listener knowledge. In all these cases, cooperative behaviour would yield a more restricted effect of repetition where the listener has denied ability to find the object, -- that is, less change in intelligibility or accessibility across repetitions. Only Dual Process, which holds that feedback may be unimportant, could account for failure to mitigate of the effects of repetition on form and articulation.

3.5.2. Results

Table 4 begins with results from Bard et al. Which require further comment. Intelligibility loss vis-à-vis a clear control form should have increased more where listeners offered no negative feedback and less where they denied having the named object on their maps. In fact, the reverse was true, with a significant interaction of mention and feedback because of increased intelligibility with no-denial repetitions and decreased with denial. However, the root of the difference lay in the first mentions, not the second, whose absolute intelligibility scores were indistinguishable. In the present study, no such complication is found. For the 73 repeated words with intervening denial and the 122 without, *k*-duration loss increased with repetition significantly and equally (mention: $F_2(1, 193) = 9.45, p = .0024$; mention x denial n.s.). Form of referring expression showed the same pattern: the change toward more accessible referring expressions on second mention was no more limited for the 44 cases with intervening denials than for the 86 without ($F_2(1,128) = 18.49, p <.0001$; mention x denial: *n.s.*). Feedback that should block defaulting does not do so. Only what the repeater has seen, heard, and said seems to play a role.

3.6. Repeater knowledge

3.6.1. Design

What the repeater knows is the subject of Experiment 4, in particular what the speaker can see. Here only cross-speaker repetitions were used, but now the landmark in question might be shared by both speakers or absent from the repeater's map. As Table 6 shows, the original introducer, the listener at the point of second mention, can see the item, has mentioned it, and has heard it mentioned. The repeater has also heard it mentioned. Experiments 2 and 3 have already shown that the original introducer's ability to see the named item does not bear on the manner of repeated mention. What we ask here is whether the speaker's visual knowledge of the named entity is also unimportant or whether articulation and form or referring expression are influenced by this kind of knowledge. If what the repeater can see is an important addition to speaker-Given status, then intelligibility loss across repeated mentions will be greater for shared landmarks, where the speaker has more knowledge of the entity than for unshared.

The Negligible Default Hypothesis predicts no effect of what the speaker can see, because the more important listener knowledge is constant across conditions. Co-presence would seem to make the same prediction. Monitoring and Adjustment allows for speaker knowledge having direct effects on articulation or referring expression design. Dual Process makes the assumption that articulation is keyed to speaker knowledge by fast priming processes. It is not clear whether visual stimuli prime word duration. Thus far, illustrations have all been via perceiving or producing the repeated word. Duea process does allow for slower, costly access to additional information, and so would allow for effects of speaker knowledge on accessibility.

3.6.2. Results

Table 5 shows the effects of repetition. Bard et al. found a robust effect of repetition on intelligibility loss vis-à-vis citation forms, but no tendency toward greater change where the repeater could see the landmark. The present results have the same interpretation: *k*-duration falls with repeated mention (mention: $F_2(1, 224) = 12.37$, $p < .0005$) but there is no significant difference between the outcome for the 144 shared repetitions and the 82 unshared (mention x introducer: *n.s.*).

In contrast, form of referring expression shows the speaker-centric result. Second mentions are made in more accessible forms in both cases (mention: F_2 $(1,138) = 24.67$, $p < .0001$), but the increase is greater for the 90 cases where the repeater can see the landmark than for the 50 where he or she cannot. (mention x sharing: $F_2(1,138) = 6.48, p < .02$). This outcome is certainly not indicative of careful adjustment to listeners alone. Nor does it indicate overall attention to speaker knowledge as a proxy for listener knowledge. It conforms best to the notion that different processes design the form and articulation of referring expressions with the former sensitive to a wider range of information.

4. Discussion

Table 6 summarizes the results reported here and in Bard et al. (2000). Each of the experiments tests for an effect on repeated mentions of some aspect of speaker or listener knowledge. Experiment 1 pitted the speaker's experience in having seen the mentioned landmark, mentioned it, and heard it mentioned against the new listener's ignorance of the item as the landmark was introduced in a second trial with a map. Experiment 2 pitted the speaker's own experience in seeing and hearing against the listener's under two conditions, when those listeners to the repetition had produced the original mention so that it might be inferred that they could see the landmark, and when they had not. Experiment 3 pitted the speaker's experience of seeing, saying, and hearing against the listener's declared inability to see the item in question. Experiment 4 kept the listener's knowledge constant as well as the speaker's experience in hearing a prior mention, but manipulated the speaker's ability to see the landmark.

In all these cases, as the shaded cells of Table 6 show, the repeating speaker had heard the original mention. In all cases the measures of word articulation were sensitive only to what the speaker had heard. These are exactly the results found by Bard et al. (2000) for a balanced but restricted sample of materials and with intelligibility as dependent variable. Thus, reductions in articulatory detail with repeated mention are conditioned by what the repeaters have heard mentioned. There is no indication that models of the listener are consulted except insofar as they conform exactly to the speaker's memory for what he or she has heard.

Form of referring expression showed a different pattern. It behaved like articulation in being insensitive to information which should have been of use in updating a model of the listener: either an indication that the listener could see the landmark under discussion or a direct statement to the effect that he or she did not (Experiments 2 and 3). Yet it did show two effects which articulation did not. In Experiment 1 accessibility of referring expression did not increase with re-introductions to new listeners. In this case, form of referring expression was tailored to the listener's needs. In Experiment 4, accessibility was enhanced more for repeated mentions of landmarks which the speaker could see than for repeated mentions of items which the speaker had only heard mentioned. Thus, accessibility is more sensitive than articulation but not in a way which support claims for the tailoring of referring expressions to listeners' needs.

Why should accessibility have these characteristics? The current results indicate that form or referring expression does not respond on-line to changes in co-presence, whether via feedback or inference. Nor does accessibility, which seems to be designed before articulation, show the characteristics that Monitoring and Adjustment would predict for initial design. In Dell and Brown's account, early processes like design of referring expressions should, if anything, be less sensitive to listener knowledge than later processes like articulation This certainly is not the case here: referring expressions patterned like duration when the two should have differed.

Experiment	Effects on repeated mention (by dependent variable)		How Given status achieved					
			By speaker			By listener		
	Word articulation	Referring expression form	Said	Sees	Heard	Said	Sees	Heard
1:different listeners	Speaker	Listener	yes	yes	yes	no	no	no
2: same/ different speakers	Speaker	Speaker	no / yes	yes	yes	yes / no	yes (inferred) /?	yes
3: +/– negative feedback	Speaker	Speaker	yes	yes	yes	no	no (declared) / yes (inferred)	yes
4.speaker +/– sees	Speaker	Speaker (additional)	no	no / yes	yes	yes	yes (inferred)	yes

Table 2. Speaker-knowledge and listener-knowledge effects on repeated mentions of landmark names. Word articulation results in terms of intelligibility (Bard et al, 2000) agree with current results of standardized word duration (*k*). Shaded cells indicate conditions in common across all experiments where repeated mentions lost clarity. Form of referring expression in terms of accessibility shows additional sensitivity to conditions in the doubly boxed cells. (yes = condition holds; no = condition does not hold; / = contrast manipulated in experiment).

Furthermore, referring expressions patterned differently from duration where the two should have been alike in reflecting the speaker's knowledge.

We would argue that Map Task speakers demonstrated the effects of competing demands on their attention, as the Dual Process Hypothesis predicts. Unlike the fast automatic processes which affect articulation and are keyed to speaker memory, slower processes compete for attention with the task in hand. Consequently only the factors grossly related to that task -- who is participating and what is on the speaker's own map -- have a noticeable effect.

We have argued elsewhere (Bard et al., 2000) that the difficulty of the communicative task may well influence the degree to which speakers appear to be modeling their listeners. We noted the Map Task is more difficult than other tasks where more cooperative behaviour is reported. For example, the tangram task involves a fixed set of shapes and players usually know that the match between their shapes will be complete and that none will have to be re-used. Hence the problem becomes easier with every trial. In contrast, the Map Task does not make it clear at the outset how many landmarks will determine each route, how many are on the map but irrelevant, how many match between players' maps, how many are duplicated on a single map, and how many have to be revisited as the task advances. If listener modeling competes for attention with task management, we might well expect the Map Task and the more complex of everyday communicative tasks to show little tendency toward tailoring for the listener. It remains to be seen whether direct manipulation of extended communicative tasks will change speakers' priorities (see Horton and Keysar, 1996, for a simple example). It also remains to be seen whether speakers will be more sensitive to fine differences in listener knowledge in any task if some kind of external record-keeping eases the computational burden. The Dual Process Hypothesis predicts that both task and memory load should have effects on the design of referring expressions, but that neither should affect the articulation of individual words.

5. Acknowledgment

This work was supported by ESRC (UK) main grant to the Human Communication Research Centre.

6. References

Anderson, A. Bader, M., Bard, E.G., Boyle, E., Doherty, G., et al., 1991. The H.C.R.C. Map Task Corpus. *Language and Speech*, 34:351-366.

Ariel, M., 1990. *Accessing Noun-Phrase Antecedents.* London : Routledge/Croom Helm.

Balota, D. A., Boland, J. E., and L. W. Shields, 1989. Priming in pronunciation: Beyond pattern-recognition and onset latency. *Journal of Memory and Language,* 28:14-36.

Bard, E. G., and A. Anderson, 1983. The unintelligibility of speech to children. *Journal of Child Lan*guage, 10:265-292.

Bard, E. G., and A. Anderson, 1994. The unintelligibility of speech to children: Effects of referent availability. *Journal of Child Language*, 21: 623-648.

Bard, E. G., Anderson, A., Sotillo, C., Aylett, M. Doherty-Sneddon, G., and A. Newlands, 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language,* 42:1-22.

Brennan, S., and H. H. Clark, 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning Memory and Cognition,* 22:1482-1493.

Brown, P., and Dell, G., 1987. Adapting production to comprehension -- the explicit mention of instruments. *Cognitive Psychology*, 19:441-472.

Campbell, W. N., and S.D. Isard, 1991. Segment durations in a syllable frame. *Journal of Phonetics*, 19:37-47.

Carletta, J., and C. Mellish, 1996. Risk taking and recovery in task-oriented dialogue. *Journal of Pragmatics,* 26:71-107.

Clark, H. H., and C. R. Marshall, 1981. Definite reference and mutual knowledge. In A.K. Joshi, B. Webber, and I. Sag (eds.), *Elements Of Discourse Understanding.* Cambridge: Cambridge University Press.

Dell, G., and P. Brown, 1991. Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In D. J. Napoli and J. A. Kegl (eds.), *Bridges Between Psychology And Linguistics.* (pp. 111-222). Hillsdale: Erlbaum.

Fowler, C., and J. Housum, 1987. Talkers' signalling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language,* 26:489-504.

Fowler, C., Levy, E. and J. Brown, 1997. Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, 37:24-40.

Fussell, S., and R. Krauss, 1992. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psycholog,* 62: 378-391.

Gundel, J.K., Hedberg, N., and R. Zacharski, 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69: 274-307.

Horton, W., and B. Keysar, 1996. When do speakers take into account common ground? *Cognition,* 59:91-117.

Hunnicutt, S., 1985. Intelligibility vs. redundancy—conditions of dependency. *Language and Speech*, 28:47-56.

Isaacs, E., and H. H. Clark, 1987. References in conversation between experts and novices. *Journal of Experimental Psychology: General,* 116:26-37.

Laver, J., Hieronymus, J., Bennett, C., Cohin, I., Dalby, J., Davies, D., Hiller, S., McAllister, M., and E. Purves, 1989. *The ATR/CSTR Speech Database Project (Report 3).* Edinburgh: Centre for Speech Technology Research, Edinburgh University.

Levelt, W., and L. Wheeldon, 1994. Do speakers have access to a mental syllabary? *Cognition*, 50:239-269.

Lieberman, P., 1963. Some effects of the semantic and grammatical context on the production and perception of speech.. *Language and Speech*, 6:172–175.

Lindblom, B., 1990. Explaining variation: a sketch of the H and H theory. In W. Hardcastle and A. Marchal (eds.), *Speech Production And Speech Modelling* (pp. 403-439). Dordrecht, Netherlands: Kluwer Academic Publishers.

Keysar, B., 1997. Unconfounding common ground. *Discourse Processes*, 24:253-270.

O'Seaghdha, P. G., 1997. Conjoint and dissociable effects of syntactic and semantic context. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 23:807-828.

McKoon, G., and R. Ratcliff, 1980. The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior,* 19:668-682.

Mitchell, D. B., and A. S. Brown, 1988. Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition,* 14:213-222

Schober, M. 1993. Spatial perspective taking n conversation. *Cognition, 47: 1-24.*

Schober, M., and H. H. Clark, 1989. Understanding by addressees and overhearers. *Cognitive Psychology,* 21:211-232.

Samuel, A., and M. Troicki, 1998. Articulation quality is invrsely related to redundancy when children or adults have verbal control. *Journal of Memory and Language,* 39:175-194.

Vonk, W., Hustinx, L., and W. Simmons, 1992. The use of referential expressions in structuring discourse. *Language and Cognitive Processes,* 7:301-33.

Wheeldon, L., and A. Lahiri, 1997. Prosodic units in speech production. *Journal of Memory and Language,* 37:356-81

Wilkes-Gibbs, D., and H.H. Clark, 1992. Coordinating beliefs in conversation.. *Journal of Memory and Language,* 31:183-194

Modality Convergence in a Multimodal Dialogue System

Linda Bell[1], Johan Boye[2], Joakim Gustafson[1] and Mats Wirén[2]

[1]Centre for Speech Technology, KTH
Drottning Kristinas väg 31, S-100 44 Stockholm, Sweden
bell@speech.kth.se, jocke@speech.kth.se

[2]Telia Research
S-123 86 Farsta, Sweden
johan.boye@trab.se, mats.wiren@trab.se

Abstract

When designing multimodal dialogue systems allowing speech as well as graphical operations, it is important to understand not only how people make use of the different modalities in their utterances, but also how the system might influence a user's choice of modality by its own behavior. This paper describes an experiment in which subjects interacted with two versions of a simulated multimodal dialogue system. One version used predominantly graphical means when referring to specific objects; the other used predominantly verbal referential expressions. The purpose of the study was to find out what effect, if any, the system's referential strategy had on the user's behavior. The results provided limited support for the hypothesis that the system can influence users to adopt another modality for the purpose of referring.

1. Introduction

1.1. The problem

When participants in a dialogue refer to specific objects on successive occasions, they typically converge towards using the same terms in their referential expressions (Brennan and Clark 1996). Such *lexical convergence* in human–human interaction has a counterpart in human–computer interaction in the sense that human dialogue participants tend to adopt the terms of the system when referring to various concepts (Brennan 1996).

In this paper, we set out to investigate whether there is a more general form of convergence in human–computer interaction in multimodal dialogue systems. In the systems that will be of interest to us here, both the user and the system have the option of using either graphical operations or verbal expressions (or both) as they refer to specific objects in the dialogue. Given that users can choose to communicate by using speech or by using a pointing device to select objects on the screen, the question was to what extent they would be affected by the system's behavior as they constructed references.

1.2. Motivation

Apart from being a problem which is interesting in its own right, we believe that the results obtained from such an investigation will have important practical consequences for the design of multimodal human–computer dialogue systems. In order to create a system that performs well, it is crucial to have a good understanding of how the system should behave, so as to increase the chances of correctly interpreting the user's input. In particular, if we can find a systematic correspondence between the feedback strategy of the system on the one hand, and the user's choice of modality in her utterances on the other (i.e. what the user expresses in words and what she expresses by means of graphical operations), a lot can be gained. The present study is a step towards pursuing this goal.

1.2.1. Modality switching as an error handling strategy

Errors can occur on all levels of a dialogue system, but in domains where many of the words in the recognition lexicon are similar sounding, or where there is a large morphological overlap, the problem of recognition errors may become especially difficult. Experiments by Oviatt and VanGent (1996) have shown that there is a tendency for users to switch from one modality to another when their interaction with a multimodal system becomes problematic. In these semi-simulated experiments, users were subjected to errors which required them to repeat their input up to six times. Many users went from speech to graphical input after already having repeated and rephrased their spoken input to the system several times. It appears as if people use modality switching to recover from errors after having been subjected to a series of failures in communication by a noncooperative system.

It should be interesting to examine whether it is possible for a cooperative system to promote the use of one modality rather than another without explicitly asking the user to alternate or ceasing to 'understand' the user's input. Ultimately, the goal would be to design a multimodal system with the ability to predict and prevent the occurrence of longer error sequences. A low confidence score from the speech recognizer or an error indication from another part of the system could be used by the dialogue manager as a signal to encourage a user to switch to the graphical input mode. In this way, it would perhaps be possible to avoid a succession of errors and a resulting spiral of miscommunication.

1.3. The setting

This research has been carried out within the Adapt project, whose principal aim is to study various aspects of multimodal human-computer interaction in the context of an apartment-seeking domain. The practical goal of the project is to create a multimodal dialogue system which will help users find an apartment in the city of Stockholm.

The apartment domain is highly useful for studying multimodal interaction. An apartment is a complex object that has properties suitable for graphical presentation (e.g. its location in the city), as well as properties suitable for verbal presentation (price, description of interior details, etc). Furthermore, it is not always obvious which modality is preferable for a referential construction.

For the purpose of the experiment described here, we use a simulation system where the key functionalities of the intended system are handled by a "wizard" (namely, analysis of multimodal user input, dialogue management and multimodal response generation).

2. Background

2.1. Lexical entrainment

In spontaneous human-human dialogue, participants frequently use referential expressions as a way of making the interaction efficient and concise. Clark and Wilkes-Gibbs (1986) have demonstrated that participants in a dialogue collaborate in the making of references. This collaborative effort is a sort of negotiation, where one of the interlocutors suggests a way of using a noun phrase to refer to a certain object, and the other accepts, rejects or postpones the decision. Once the participants have found a mutually acceptable way of referring to the object in question, they tend to use the term agreed on. Garrod and Anderson (1987) have established that people who repeatedly refer to the same objects in a dialogue often start using the same terms. They called this phenomenon *lexical entrainment.* Brennan and Clark (1996) have argued that lexical entrainment can be understood in terms of shared conceptualizations that are established between people engaged in conversation. After a conceptual pact has been established, speakers are sometimes overinformative in subsequent references instead of introducing a new term.

Brennan (1996) has argued that there is a phenomenon corresponding to lexical entrainment in human–computer interaction. Human dialogue participants tend to mimic the terms introduced by a spoken language system, something Brennan calls *lexical convergence.* Since computer programs generally are not constructed to negotiate about terminology, entrainment in Brennan and Clark's sense is not really possible in human–computer interaction. However, there appears to be a unidirectional influence by which the terminology of a natural language system is likely to influence the user's choice of vocabulary.

2.2. Multi-modal human-computer dialogue systems

Multimodal interfaces are potentially more flexible, powerful and effective than unimodal interfaces. Experiments in map-based simulation environments have demonstrated that a pen/voice interface can be more efficient and user-friendly than either a speech-only interface (Oviatt 1997) or a graphics-only interface (Cohen, Johnston et al. 1998). Studies of how users integrate the different input modes in multimodal dialogue systems have been previously reported in (Oviatt and Olsen 1994; Oviatt and VanGent 1996; Oviatt, DeAngeli et al. 1997). In a study where speech or pen input could be used to interact in a simulated map system (Oviatt, DeAngeli et al. 1997), it was demonstrated that people use the spoken and written modalities in a complementary way, rather than provide redundant information. Adaptable multimodal systems offer many possible advantages over unimodal interfaces, such as greater expressive power. However, if these systems are to become useful, we need to put greater efforts into studying how people use different modalities and alternate between them.

3. Method

3.1. Hypotheses

Our conjecture when embarking on this experiment was that when both system and user may choose the modality in which to construct a reference, the system will, to some extent, affect the user to enter into "modality convergence" with itself. More specifically, we were interested in testing two hypotheses with respect to modality convergence:

"Strong convergence": the user converges on the system's behavior while abandoning his previously adopted modality behavior.

"Weak convergence": the user converges on the system's behavior while retaining and integrating it with his previously adopted modality behavior.

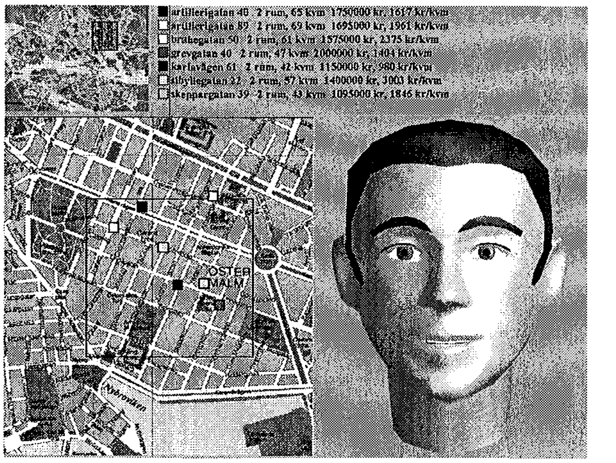
Essentially, the weak hypothesis states that the system can "entrain" the user to adopt new behaviors. The strong hypothesis additionally states that the user can be retrained and made to abandon old behaviors.

We take it that it would be possible to achieve strong convergence if the system is suitably "uncooperative", for example, if it explicitly tells the user to switch modality or if it ceases to understand a certain behavior. However, rather than trying to affect the user by putting restrictions on the system's capabilities, we were interested in investigating to what extent a cooperative system could influence the user's behavior merely by changing its own way of constructing references.

The experimental task used to test these hypotheses involves the construction of deictic references to specific apartments on a map. Subjects who referred to apartments had the option of using either graphical or verbal means, or both. The question was then to what extent the subjects' construction of deictic (and other) references would be influenced by the behavior of the system.

3.2. Simulation system

The basic vehicle for the experiment was a Wizard-of-Oz simulation tool which provided information about available apartments in downtown Stockholm. The tool included a map showing names of streets, major neighborhoods, parks, etc., an overview map allowing the user to scroll the detailed map, and an animated agent speaking with a synthesized voice (see Figure 1). For each displayed icon, limited information about the corresponding individual apartment was provided in the row of a table. Here, the apartment's address, size and listed price were displayed. Icons on the map that represented apartments at adjacent or identical positions were only allowed to overlap to a limited extent in order to keep them simultaneously visible to the user.



■ artillerigatan 40 2 rum, 65 kvm 1750000 kr, 1617 kr/kvm
□ artillerigatan 89 2 rum, 69 kvm 1695000 kr, 1961 kr/kvm
□ brahegatan 50 2 rum, 61 kvm 1575000 kr, 2375 kr/kvm
▨ grevgatan 40 2 rum, 47 kvm 2000000 kr, 1404 kr/kvm
■ karlavägen 61 2 rum, 42 kvm 1150000 kr, 980 kr/kvm
□ sibyllegatan 22 2 rum, 57 kvm 1400000 kr, 3003 kr/kvm
□ skeppargatan 39 2 rum, 43 kvm 1095000 kr, 1846 kr/kvm

Figure 1. The graphical user interface.

The user's input was sent to the wizard interface where a human operator controlled the system's response. Much care was devoted to design the wizard interface to allow rapid system response times (typically between one and two seconds), thus giving users the impression of a fully functional system. The wizard chose his answer from a button menu, where information about specific apartments from the database was included in one of a number of possible answer templates.

To investigate convergence effects, the experiment focused on two equivalent ways of forming deictic references using different modalities, namely, graphics (point-and-click) and verbal expressions. To this end, the simulations mimicked two versions of a system, called System G ("graphics-oriented") and System S ("speech-oriented"), which behaved identically except for the way the deictic references were constructed. Thus, both versions used square-formed icons to indicate apartment positions on the map. The icons were color-coded so that each displayed icon had a unique color. The sole difference between the two simulated systems was that System G, while using a deictic utterance ("This apartment has a tiled stove"), let the corresponding icon on the map "shake" in a highly perceptible way for a fixed number of seconds (1.5, to be exact). In contrast, System S constructed apartment deictic references by using a verbal expression that exploited the color-coding ("The yellow apartment has a tiled stove"), but without shaking or otherwise changing the appearance of the icon in any way. Throughout the dialogues, the two systems retained their way of referring to the individual apartments.

Because of the difficulty of verbally distinguishing a large number of colors, and in order to help focus the dialogues on a limited number of objects which could be systematically compared, both of the simulated systems displayed at most seven apartment icons at any given time. Thus, as long as the current set of apartments to match the user's constraints was larger than seven, no icons were shown on the map. The animated agent would then prompt the user to narrow down the search by saying something like, "There are too many apartments to show. Are there any particular features you'd like your apartment to have?"

To make it straightforward for the user to associate table rows with the corresponding apartment icons, each row was preceded by a color-coded icon similar to the one on the map.

3.3. Experiment

To collect the data needed to test the hypothesis, a between-subjects design was selected. 16 participants were randomly assigned to a task/system sequence and each completed two tasks. For each task order (A-B, B-A), there was a corresponding system order (G-S, S-G), resulting in four unique sequences of two tasks (AG-BS, BG-AS, AS-BG, BS-AG), aimed at counterbalancing sequence effects. Each of these sequences was completed by eight persons, and a total of 32 dialogues were thus recorded.

Each task involved finding an apartment that fulfilled certain criteria. In solving the tasks, the subjects were invited to take their time looking around, and to contrast individual apartments in order to arrive at a suitable alternative. Before subjects started an experimental session, they were asked to try the functionalities of the system. In this way, the experimenter could make sure each user knew how to carry out the various operations.

As can be seen in Figure 2, task A and B both included a map of Stockholm where different areas had been shaded. These were the designated areas in which the users were to look for an apartment in their respective scenarios. In addition, the number of rooms the apartment should have and an approximate time period for the construction of the building were indicated on scales. Pictures of interior and exterior details were also added to each task. The subjects were informed that these details (stucco and a balcony, for instance) were merely suggestions, and that they were free to ask the system about other things that might interest them.

Subjects were instructed that they could communicate with the system using an open microphone and two graphical operations with respect to the map, namely, the selection of a position by point-and-click and the selection of a rectangular area of arbitrary size. The subjects' graphical operations were echoed in the same way by the two system versions; in particular, a point-and-click on an apartment icon was echoed by highlighting the icon. 16 subjects, all volunteers, participated in the experiment. Eight of the subjects were female and eight were male, and their ages ranged from 17 to 55. The subjects were all native speakers of Swedish, and while a few of them were staff at the Department of Speech, Music and Hearing,

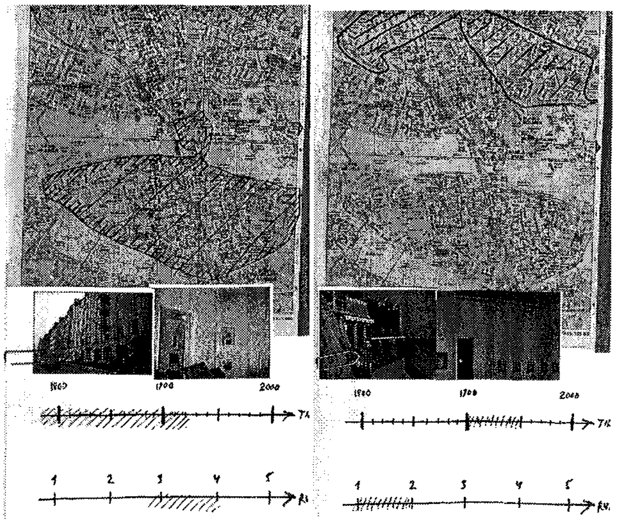


Figure 2. The scenarios, A on the left and B on the right.

none were working in the field of speech technology. All subjects reported to be familiar with computers, most of them regularly used word processing software and browsed the web, but only a couple claimed to have any significant programming skills. Each experiment session (including the introduction to the system and the post-experimental interview) lasted for approximately 30 minutes. During the post-experimental interviews, subjects were asked to give feedback on the interface and comment on their own modality choices during the dialogues. Finally, the experimenter verified that all subjects had been unaware of the fact that they had been interacting with a Wizard-of-Oz simulation rather than a real system. Those subjects who had been recruited from outside the department were rewarded with a movie voucher.

4. Data coding and analysis

4.1. Coding

Figure 3 shows an excerpt from a typical dialogue obtained during the experiment. A dialogue normally proceeds as follows: The subject begins by indicating a city area and providing a number of preferences in order for the system to come up with a set of matching apartments. The subject then asks the system about various features of the individual apartments displayed. This procedure may be iterated for a number of areas.

The dialogue excerpt shown in Figure 3 begins at the point where the system has just displayed a set of apartment icons on the map (with accompanying information in the table, as seen in Figure 1). The subject uses color references with a deictic function when he shifts focus from one apartment to another (that may or may not have been referred to previously in the dialogue).

...	
(User.44)	**Does the black apartment have a ceiling with stucco?** /subject uses mouse to select the black apartment/
(Sys.45)	The black one has stucco
(User.45)	Does the black one have a balcony?
(Sys.46)	Yes
(User.46)	Does the black one have a terrace?
(Sys.47)	I don't know
(User.47)	When was the building constructed?
(Sys.48)	The black house was built in 1928
(User.48)	When was it refurbished?
(Sys.49)	In 1994
(User.49)	Is it a quiet apartment?
(Sys.50)	I don't know
(User.50)	Which floor was that on?
(Sys.51)	The black apartment is on floor one
(User.51)	Facing the street or the back?
(Sys.52)	The street
(User.52)	**The red apartment which floor?**
(Sys 53)	The red apartment is on floor two
...	

Figure 3. A translated excerpt from one of the dialogues, using System S. Focus shifts are indicated using bold fonts.

After a focus shift, subjects typically use pronominal expressions to refer to the apartment under discussion. To a lesser extent, subjects continue to make use of constructions with a deictic function during turns subsequent to focus shifts. (Examples of this occur in Figure 3 at turns User.45 and User.46.) In these cases, however, they typically retained the particular reference construction used at the focus shift.

Our data analysis thus showed that the critical points at which modality convergence can be investigated correspond to focus shifts. The coding was therefore guided by the need to track user references to apartments made at these points. The references occurring at user turns other than focus shifts were not tagged.

References at focus shifts were tagged along two dimensions:

1. A category for each primitive type of reference construction used by the subject. We distinguished between four types:
   - g – graphical reference (that is, point-and-click);
   - c – color reference (for example, "the yellow one", "the black apartment");
   - a – address reference (a street name optionally followed by a street number, such as "Swedenborgsgatan 7");
   - m – miscellaneous (for example, "this one", "the apartment with a sauna").
2. A tag indicating whether the focus shift was initiated by the user or the system (user-init and system-init, respectively).

As an example of this, the dialogue excerpt shown in Figure 3 contains two focus shifts which are tagged as follows:

- (User.44) cg; user-first
- (User.52) c; user-first

The notation "cg" means that the subject used an integrated color and graphical reference by making a point-and-click operation in connection with a verbal utterance. Each reference categorized as "cg" was counted as one "c" and one "g", in addition to being counted as one "cg".

Focus shifts that occur initially in the dialogues, before the system has had any chance of entraining the subjects, have not been included in the count. However, we still coded them, since they could tell us something about the subjects' a priori preferences with respect to reference constructions at focus shifts.

4.2. Analysis

As previously stated, our main objective was to investigate if and how the subjects were influenced by the system in their way of referring to individual apartments. System G consistently referred to apartments using a graphical operation; system S consistently used color codes; hence we were primarily interested in the subjects' behavior in this regard, which was reflected by the values of the "g" and "c" categories. Two "g" values were calculated for each subject, one for the number of "g" references in dialogue 1 and one for the number of "g" references in dialogue 2. Analogously, two "c" values were calculated for each subject.

In order to enable meaningful comparisons between subjects, we normalized each "g" value ("c" value) by dividing it with the total number of coded references in that dialogue. We used the notation "gNorm" ("cNorm") to refer to the normalized "g" values ("c" values).

As described in Section 3.3, the 16 test subjects were divided into four groups, each group corresponding to a unique sequence of scenario-system pairs (AG-BS, BG-AS, AS-BG, and BS-AG). The first test performed was to investigate whether the scenario had any significance for the behavior of the subjects. We therefore compared the values for the "gNorm" and "cNorm" parameters for the AG-BS group with those of the BG-AS group, and similarly for the AS-BG and BS-AG groups. As we found no significant differences, we collapsed the AG-BS and BG-AS groups into one group called G-S (corresponding to the eight subjects who used system G first and system S second). The AS-BG and BS-AG groups were collapsed into another group called S-G (corresponding to the eight subjects who used system S first and system G second).

The next step was to compare the values of the parameters "gNorm" and "cNorm" between and within the G-S and S-G groups. More specifically, we were interested in the relations indicated by the arrows in Table 1 below. The horizontal arrows in the table correspond to possible changes in referential behavior within the same group, but between the subject's first and second dialogue. The vertical arrows correspond to possible differences between the two groups either in the subjects' first dialogue, or in their second dialogue. In Table 1-4 below, "D1" and "D2" denote the first and second dialogue, respectively.

Table 2 shows the relations that should hold for the data to support the weak convergence hypothesis of Section 3.1. The value of the "gNorm" parameter should be higher for the G-S group than for the S-G group in dialogue 1, since at that point in time the S-G group had not yet been subjected to "graphical" behavior from the system. Similarly, within the S-G group, the "gNorm" value should be higher in dialogue 2 (when the system starts to behave "graphically") than in dialogue 1. An analogous line of reasoning gives the required relations indicated in the "cNorm" part of Table 2.

Table 3 shows the additional relations, apart from those of the weak convergence hypothesis, that should hold for the data to support the strong convergence hypothesis. The value of the "gNorm" parameter should be higher for the S-G group than for the G-S group in dialogue 2, since in the second dialogue the system behaved "graphically" towards the S-G group but not towards the G-S group. Similarly, within the G-S group, the "gNorm" value should be higher in dialogue 1 (when the system behaves "graphically") than in dialogue 2. An analogous line of reasoning gives the required relations indicated in the "cNorm" part of Table 3 below.

Table 1 Relevant data relations

	gNorm			cNorm		
	D 1		D 2	D 1		D 2
G-S	?	↔	?	?	↔	?
	↕		↕	↕		↕
S-G	?	↔	?	?	↔	?

Table 2. Relations that would support the weak convergence hypothesis

	gNorm			cNorm		
	D 1		D 2	D 1		D 2
G-S	?			?	<	?
	>			<		
S-G	?	<	?	?		

Table 3. Additional relations that would support the strong convergence hypothesis

	gNorm			cNorm		
	D 1		D 2	D 1		D 2
G-S	?	>	?			?
			<			>
S-G			?	?	>	?

5. Results

The most important results of the experiment are summarized in Table 4 below.

Table 4. Mean values for the "gNorm" and "cNorm" parameters

	gNorm		cNorm	
	D 1	D 2	D 1	D 2
G-S	0.16	0.32	0.04	0.48
S-G	0.04	0.03	0.13	0.21

If we begin by examining the four relations relevant for testing the weak convergence hypothesis (cf. Table 2), we see our data supports the hypothesis in three cases. Only the decrease from 0.04 to 0.03 for the S-G group's "gNorm" parameter is inconsistent with the hypothesis (but on the other hand, the total number of graphical references is indeed very small in those dialogues). The other three relevant relations are consistent with the weak hypothesis. However, only the increase from 0.04 to 0.48 for the G-S group's "cNorm" parameter proved to be statistically significant using a correlated t-test (t(7)= -3.39, p<0.012), as well as a Wilcoxon signed rank test (W_{+}=1, p<0.028). We therefore conclude that we have found limited support for the weak convergence hypothesis.

In contrast, the strong convergence hypothesis is not supported at all by the data. Of the four relations indicated in Table 3, only the difference between the two groups for the "cNorm" parameter for the second dialogue (0.48 vs. 0.21) is consistent with the strong hypothesis, however not significantly so. There is even an almost significant difference (t(14)=2.12, p<0.053) between the two groups for the "gNorm" parameter for the second dialogue (0.32 vs. 0.03), something which speaks against the strong hypothesis.

The strong hypothesis is also contradicted by the tendencies within the groups between the first and second dialogues. The group which started out using System S increased their proportion of color references in their second dialogue (from 0.13 to 0.21), even though the system had changed its behavior. The same tendency could be shown for the group that started out using System G (0.16 to 0.32), i.e. the subjects amplified the behavior adopted in their first dialogue rather than allowing themselves to be "retrained".

The group who started using System G had a higher proportion of graphical references during both dialogues when compared to the other group (almost significantly so in the second dialogue, as discussed above). This might be seen as a "delayed" convergence effect from their first dialogue. However, a closer look at the data reveals that out of the 17 graphical references by this group in the second dialogue, ten are integrated with color ("cg"). Thus, the increased use of graphics did not occur at the expense of color references, but rather "hand in hand" with these.

Looking at the subjects' behavior across the two dialogues, what we have said above might be summarized as follows: Rather than the subjects replacing one type of behavior with another as an effect of modality convergence, their "converging" behavior in the first dialogues was amplified in the second dialogues. In addition, they showed clear potential for taking up and integrating a new form of converging behavior in one of the second dialogues, namely, with the system that used color references.

Another way of formulating this is that the added proportions of color and graphical references increased from dialogue 1 to 2 for both groups. In other words, there was a tendency for subjects to gradually converge to the two kinds of reference construction that the system used, ("c" and "g") at the expense of the other kinds of reference construction ("m" and "a", mentioned in Section 4.1 above).

An interesting observation is that none of the subjects had color ("c") as their a priori preference in their first dialogue; still, color ended up being the altogether most used reference construction.

6. Discussion

Our post-experimental interviews indicated that the function of mouse clicks was not entirely obvious to the subjects. One subject said: "I preferred to speak since what I could do with the mouse seemed so limited" and another reported: "He [the animated agent] understood what I said, but not what I meant by clicking". The post-experimental interviews also revealed that the graphical input mode was perceived by several subjects as being less efficient and concise: "The question one asks with a mouse click seems rather undefined", "I preferred to speak, it was easy", "It was faster (I think) to speak directly to the animated agent."

Intuitively, it seems that in order for the system to maximize its chances of successfully entraining the user, the manifestatons of the input and output reference constructions should be as similar or "symmetric" as possible. In our experiment, such a symmetry was trivially achieved for verbal references (through the spoken manifestations of the user and system), but less so for graphical references: User clicks on apartment icons were echoed by highlighting the selected icon, whereas the system's graphical references were indicated by shaking the icon. Because of this, the connection between the graphical output and input might not have been obvious to the subjects. One way of clarifying this connection might be for the system to produce a characteristic short sound as each icon is highlighted. The same sound could then be repeated as the user clicks on one of the icons on the screen.

Furthermore, it is worth noting that the dialogues, generally speaking, were quite short. Since the tasks given were deliberately vague, some of the subjects chose to speak about no more than a couple of different apartments. A tendency in our data was that those subjects who persisted in interacting with the system for a longer time were more likely to be affected by the system's behavior. Longer dialogues would most certainly have given us more datapoints, and possibly also more clear-cut entrainment effects.

Acknowledgements

The authors would like to thank the other members of the Adapt group at the Centre for Speech Technology. We also thank Nils Dahlbäck for helpful methodological advice in the process of writing this paper. We are grateful to Bo Delling and Mattias Heldner for their invaluable assistance with the statistical computations.

7. References

Brennan, S. (1996). Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*: 41-44.

Brennan, S. E. and H. H. Clark (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition* **22**(6): 1482-1493.

Clark, H. H. and D. Wilkes-Gibbs (1986). Referring as a collaborative process. *Cognition* **22**: 1-39.

Cohen, P. R., M. Johnston, et al. (1998). The efficiency of multimodal interaction: A case study. *Proceedings of the International Conference on Spoken Language.*

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

Oviatt, S., A. DeAngeli and K. Kuhn. (1997). Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction. *Proceedings of CHI '97.* Atlanta: 415-422.

Oviatt, S. and E. Olsen (1994). Integration Themes in Multimodal Human-Computer Interaction. *Proceedings of the International Conference on Spoken Language Processing.* Volume 2: 551-554.

Oviatt, S. and R. VanGent (1996). Error resolution during multimodal human-computer interaction. *Proceedings of the International Conference on Spoken Language Processing*: 204-207.

Oviatt, S. L. (1997). Multimodal interactive maps: Designing for Human Performance. *Human Computer Interaction* **12**: 93-129.

Communication and Cooperation among Agents

Guido Boella and Rossana Damiano

Dipartimento di Informatica and
Centro di Scienza Cognitiva - Università di Torino
C.so Svizzera 185 10149 Torino ITALY
email: {guido,rossana}@di.unito.it

Abstract

In this paper we consider the consequences for dialog modeling of a new proposal of cooperation based on the concepts of group utility, goal adoption and anticipatory coordination. On the one hand, the model accounts for the contextual occurrence of communicative acts during cooperative activity; on the other hand, it models grounding phenomena in dialog, seen as a cooperative activity, without explicitly prescribing them.

1. Introduction

One of the main aims of the definitions of cooperation is to explain when and why the members of a group have to communicate with each other. In particular, the attention has been focused on those parts of dialogs which have the function of coordinating helpful behavior (Grosz and Kraus, 1996) and the conclusion of the group's activity (Cohen and Levesque, 1991).

In (Cohen and Levesque, 1991), the definition of cooperation is built out of a number of examples which show how communication can be used to establish the right mutual beliefs among the agents. The role of communication in those examples is to ensure that the activity of the group is not jeopardised by a relevant divergence among the participants' mental states, especially for what the achievement of the joint goal is concerned. In order to ensure that the group ends its activity in a felicitous way, (Cohen and Levesque, 1991) add to the definition of joint intention a number of subsidiary goals: in particular, when the joint goal has been achieved, each agent has the goal that the other partners be aware of this situation; typically, but not necessarily, these goals are achieved via communication, thus explaining dialogic phenomena like notifications.

In this work we describe the consequences for dialog modeling of extending and generalizing the intuition of (Cohen and Levesque, 1991) that the lack of coordination leads to a worse performance of the group. We claim that coordination via communication can be explained by a sort of means-ends reasoning which aims at not decreasing the group performance.
In this way, it is possible to achieve the separation between the definition of cooperation and communication claimed by (Castelfranchi, 1998): communication and the related goals of making the knowledge mutual among the group members are just instrumental to the group joint action and not prescribed by it.

In order to do so we need two conceptual instruments for measuring the expected result of the group's action and for predicting what the partners are going to do. These two forms of reasoning are becoming increasingly exploited in the multiagent field; in particular, decision theory is used for making an agent choose the most promising course of action while having a look at the benefit of the whole group (Hogg and Jennings, 2000); on the other hand, game theoretic concepts are used for predicting the behavior of the other agents and choosing a suitable action accordingly (Ndiaye and Jameson, 1996), (Gmytrasiewicz and Durfee, 1995), (Boella et al., 2000).

The main ideas underlying the new definition of cooperation are the following ones: first, when a member of the group chooses an action, he has to take into account the benefit of the whole group and not only the utility achieved for himself. Second, the benefit for the group must be computed by taking into account how the agent's choices affect the choices of other agents; that is, an agent has to predict which actions the partners will choose after the execution of the selected actions (anticipatory coordination): he can do so, since in the contest of a group he knows (at least partially) the tasks assigned to the other partners of the group. Finally, an agent should consider whether to adopt the goals of his partners in order to provide them with help (not only material help, but also information); see (Castelfranchi, 1998) for the importance of adoption in modeling interaction among agents.

From this sketch of the model it is possible to understand why it is not necessary to include into the definition of cooperation goals concerning the achievement of a mutual belief, nor the goal to communicate with the partners. However, the model can explain the occurrence of communication during cooperative interaction, even if it is not explicitly prescribed. In particular, communicative acts aiming at enhancing the coordination among the members of a group are be performed depending on the utility they yield with respect to their cost. Notifications that a partner's goal has been achieved (or is unachievable) are motivated by the adoption of his goal to know whether the these conditions apply to the goal he is committed to: under some circumstances, this behavior can be useful from the group's perspective, because it prevents the partner from wasting resources. At the same time, agents can communicate to to signal the intention to adopt a goal of a partner, or to avoid conflicts in the use of shared resources: in the following we will detail the utility-driven reasoning underlying these phenomena.

On the other hand, the model can be applied as well to communication itself, seen as a cooperative activity: once two agents are involved in a communicative interaction, they choose subsequent dialogue moves in the light of the group utility they are expected to yield. Under this perspective, acknowledgements of understanding, repetitions, repairs and other dialog management actions are the result of a trade-off between the joint goal to ensure mutual understanding among the participants and the cost of these actions in terms of the resources they require.

2. The Definition of Cooperation

A group of agents GR composed of agents $G_1, \ldots, G_n$ cooperates to a shared plan[1] for α with an associated recipe R^x composed of steps $\beta_1^{x,i_1}, \ldots, \beta_m^{x,i_m}$ when:[2]

1. each step $\beta_r^{x,i}$ has been assigned to an agent G_i in GR for its execution;

2. each agent G_i of the group GR has the single agent intention to perform his part $\beta_r^{x,i}$ of the shared plan for α formed on the basis of the recipe R^x;

3. the agents of GR have the mutual belief that each one (G_i) has the intention to perform his part $\beta_r^{x,i}$ of the shared plan for α;

4. all agents mutually know that they share a utility function GF based on a weighted sum of the utility of the goal which the shared plan aims at and of the resource consumption of the single agents; each agent, when he plans his own part of the shared plan, has to consider also this global utility as part of the his individual utility function F_i;

5. when an agent G_i becomes aware that a partner G_j has a goal ϕ that stems from his intention to do his part $\beta_p^{x,j}$, G_i will consider whether to adopt it; if G_i believes that the adoption of ϕ produces an increase of the utility GF of the whole group, then he adopts that goal;

6. each agent remains in the group as long as the value of the utility function GF can be increased by executing his part of the shared plan for α or by adopting some of the goals of the partners.

Here, we don't address the problem of group formation: we assume that a group is already at work, with different tasks assigned to the various members of the group.
Moreover, the definition of cooperation presented above presupposes an appropriate model for planning, which includes the mechanism of anticipatory coordination.

3. The Anticipatory Coordination Planning

In order to plan and select actions, we exploited a decision theoretic planner, DRIPS (Haddawy and Hanks, 1998). DRIPS is a hierarchical planner which merges some ideas of decision theory with probabilistic planning techniques. Plans are organized along decomposition and abstraction hierarchies and have non-deterministic effects which produce a probability distributions over the action outcomes.
A utility function is used for evaluating how promising a given plan is. The utility function does not compute just the payoff of the possible outcomes of the plan: it is computed starting from simpler utility functions associated with the goals of the agent.[3]

However, the planner itself is not sufficient for our purposes. In fact, since the agent's world is populated by other agents, the consequences of an action may affect the subsequent behavior of other agents. So, in case of interaction, an agent has to consider the consequences of his behavior with respect to what the other agents will do afterwards: in order to evaluate the real expected utility of the plan, he must explore the outcomes that may result from the different reactions of other agents.

Moreover, as stated above, when an agents comes to know a goal of a partner, the agent considers whether to adopt it as part of his intentions: he does so, if the adoption of the partner's goal as a new intention amounts to an increase of the global utility notwithstanding the decrease of his partial utility due to the overhead of helping the partner.

The construction of a plan is carried out by an agent G_i in a stepwise fashion: if G_i is in charge of step $\beta_m^{x,i}$ of the recipe R^x shared with G_j[4] for achieving goal α, he first has to find the best recipe for $\beta_m^{x,i}$ (let's say R^y, with steps $\gamma_{m,1}^{y,i}$, $\gamma_{m,2}^{y,i}, \ldots, \gamma_{m,n_m}^{y,i}$), and then he can start refining $\gamma_{m,1}^{y,i}$. The approach of DRIPS to this process is to expand $\beta_m^{x,i}$ in all possible ways (i.e. applying to the current state S all existing recipes); then, it proceeds onward and expands the new partial plans. The search goes on in parallel, but the search tree is pruned using the utility function (applied to the state resulting from the potential execution of the recipe): so, the utility function acts as a heuristic able to exclude some possible ways (recipes) to execute an action.

In order to implement the ideas presented in the previous section, we had to make the evaluation of the heuristics somewhat more complex. Therefore, we have modified the DRIPS algorithm for allowing anticipatory coordination. In particular, since the planning agent has to predict the reactions of the partners, he must be endowed with the (assumed) knowledge concerning his partner's beliefs and utility functions. Moreover, we assume that the knowledge about actions is shared among all the group members.

The method we employed is the following:

- Using DRIPS (playing the role of G_i), we expand the current state S according to all alternative recipes for $\beta_m^{x,i}$, thus producing the states $S_1, S_2, \ldots, S_r$ (where r is the number of different recipes for $\beta_m^{x,i}$).

[1] Actually, we assume here that the initial shared plan has a particular structure; i.e., it is a one-level plan composed of a top-level action (α) decomposed into a sequence of steps. In a general plan, each step could in turn have been expanded into substeps, and so on recursively.

[2] The notation $\beta_l^{x,i}$ refers to the l-th step of the recipe R^x, a step which has to be executed by agent G_i.

[3] This aggregation of simpler utility functions in a global one is possible only if some independence assumptions hold (see (Haddawy and Hanks, 1998)).

[4] For simplicity we have assumed a single partner G_j.

- This set of states is transformed in the set of the same states as viewed by G_j, S'_1, S'_2, ..., S'_r.

- On each state S'_m (1 ≤ m ≤ r), we restart the planning process from the perspective of his partner G_j (i.e. trying to solve his current task $\beta_h^{x,j}$).

- This produces a set of sets of states $SS' = \{\{S'_{1,1}, \ldots, S'_{1,n_1}\}, \{S'_{2,1}, \ldots, S'_{2,n_2}\}, \ldots, \{S'_{n_r,1}, \ldots, S'_{r,n_r}\}\}$.

- The group utility function is applied to these states, and the best state of each subset is identified: SS'_{best}=$\{S'_{1,best(1)}, S'_{2,best(2)}, \ldots, S'_{n_r,best(n_r)}\}$. These states are the ones assumed to be reached by G_j's best action, for each of the possible G_i initial moves.

- The group utility function is applied to the states $S_{k,best(k)}$ $(1 \leq k \leq r)$ from G_i's point of view. This models the perspective of G_i on what could happen next.

- The best one of these states is selected $(S_{max,best(max)})$. This corresponds to the selection of the best recipe for $\beta_m^{x,i}$ of G_i (i.e. R^{max}).

Note that the algorithm above is just a modification of a two-level min-max algorithm: actually, it is a max-max, since at both levels the best option is selected, although at the second level it is evaluated from G_j's perspective. As in min-max, G_i, when predicting G_j behavior, assumes that his partner is a rational agent, i.e. that he will choose the plan that gets the highest utility for the group.

The problem of simulating another agent's planning is very difficult. For instance, in some situations, G_j could not be aware of R^x effects. In our implementation, we adopted the simplification that the initial state is shared by the agents, while, during the planning phase, G_j's knowledge of a state is updated in G_i's beliefs by an action of G_i only with the effects which are explicitly mentioned as believed by G_j (e.g., the result of a communicative action having G_j as receiver). However, the treatment of the changes of the beliefs of the partner would deserve a more accurate model, as the one proposed for multiagent systems by (Hideki and Hirofumi, 2000).

Some more words must be devoted to the probability that an effect holds after the execution of a recipe R^x. Note that if a recipe R^x of G_i makes a proposition *Prop* true only with probability *p(Prop)* the simulation of G_j's planning phase must be carried on starting from both "possible" worlds resulting from the execution of R^x (i.e. one where *Prop* is true and one where *Prop* is false).[5]

Therefore, we simulate separately (see figure 1) what G_j would plan if *Prop* were true and if *Prop* were false; since also G_j's recipes may involve uncertain effects, we adopted the solution of multiplying the probability of the different outcomes of G_j's actions with the probability of G_j's initial states in order obtain the set of worlds representing the possible outcomes of G_j's reactions to the plan R^x.

4. Consequences for Dialog

4.1. Communication as a Consequence of Cooperation

Consider the situation where an agent has just discovered that the goal of the group is impossible to achieve: he has two alternatives, banned the idea to go on with the now impossible shared plan; he can choose to give up the shared plan and do something else. Otherwise, he can consider adopting some of his partners' goals; banned again the idea of helping the partners in doing their part (remember that the shared plan is impossible to achieve, so no help can be fruitful), there remain, however, other goals available for adoption: if the partners still have the intention to achieve the joint goal, they consequently have the goals of knowing whether they succeeded or are prevented from achieving their aim (as it follows from the definition of (individual) intention in (Cohen and Levesque, 1991)). In order to choose among the alternatives of leaving the group and of communicating to the partners the impossibility of achieving their aims, the agent resorts to anticipatory coordination. This consists of simulating what the partners will (presumably) do in both situations: in the first case, being not aware of the new state of affairs, the partners will go on in doing their part; but if the agent chooses to inform them that the shared plan has become impossible, he predicts that they will give up the (now useless) joint action (provided that communication is successful).

The same holds for the information that the joint goal has been achieved. Consider, for example, the situation where two agents, G_i and $G_j$are looking for a particular piece of a jigsaw: in order to speed up the search, they divide the set of pieces into two parts, and each one searches one part. If the agent G_i finds the piece, he will inform the partner G_j, to prevent him from searching all his part unsuccessfully. Or, if he comes to know that the piece has gone lost, the model predicts as well that he will inform the partner (provided, in both cases, that the cost of communicating for G_i does not override the resources wasted by G_j in his useless search).

The same reasoning applies to the goals that the partners are in charge of, and to the goals that are instrumental to the achievement of these goals. Each time an agent knows that a partner has a goal, he can infer that the partner will also have the subsidiary goals of knowing if he succeeded. If he comes to know (without any further cost) that the partner's goal holds, in his next planning phase he has to consider whether to adopt the partner's goal of knowing if the goal holds; if he adopts this goal, he will communicate to the partner that the goal holds, an action that adds a little overhead to the group utility.

For example, if G_i and G_j have the shared goal of preparing tea, and G_j is in charge of boiling the water, G_i may come to know that the water is boiling before G_j does. At this point, G_i can inform the partner that his goal holds, by performing a communicative act ("The water is boiling"), or he can go on with his activity, without informing G_j.

In terms of utility, when the agent considers the alternative of going on with his own plan and not adopting the partners' goal, he may discover that the group utility would be lower,

[5] Using as G_j's initial world one where *Prop* has *p(Prop)* probability to be true, would correspond to the situation in which G_j is planning with uncertainty about *Prop*.

```
/* in input the one-level plan of agent $G_i$ (gi) for $\beta_k^{x,i}$ (plan-x-i-k), the identifier of
   agent $G_j$ (gj), the step in charge of $G_j$, i.e.  $\beta_m^j$ (action-j-m) and an initial world*/

plan-shared-actions(Gi, Plan-x-i-k, Gj, action-j-m, initial-world)
   begin
      /* refinement of plan-x-i-k by selecting an alternative or adding
      the decomposition of an action belonging to the plan */
      refined-plans := refine-plan (plan-x-i-k, gi, initial-world);
      final-worlds := nil;
      /* for each possible outcome of each possible alternative */
      for-each plan in refined-plans
      begin
         /* outcomes of a plan of $G_i$ from the initial worlds
           (their probability sums to one)*/
         for-each world in resulting-worlds(plan, initial-world)
         begin
            /* save the probability of the outcome of plan */
            prob := world.prob;
            /* simulate $G_j$ planning from an outcome of plan as it were the only
               possible one */
            world.prob := 1;
            primitive-plans-j := plan(action-j-m, gj, world);
            /* select best plan from $G_j$'s point of view:  $G_i$ considers only
               $G_j$'s best alternative */
            chosen-plan-j := best-plan-EU(primitive-plans-j, gj, world);
            resulting-worlds := resulting-worlds(chosen-plan-j, gj, world);
            /* restore the probability of the outcomes w that come after world */
            for-each w in resulting-worlds
            begin w.prob := w.prob * prob; end
            /* the probability of worlds in final worlds will sum to one */
            final-worlds := final-worlds + resulting-worlds;
         end
         /* assign to each $G_i$'s alternative the expected utility from $G_i$'s perspective */
         plan.EU := compute-EU(plan, final-worlds, gi);
      end
      /* eliminate plans that are not promising */
      return(eliminate-plans(refined-plans, gi));
   end
```

Figure 1: The function of the planner that given a plan, performs a step of refinement and discharges unpromising alternatives.

even if his own utility would be greater: the partners, being not aware of the relevant conditions, would waste their time going on in their activity or, at best, looking whether they succeeded or not.
In the previous example, if G_i does not warn G_j that the water is boiling, the consequence is that G_j will have to check again, while the water could evaporate and time could be wasted.

Communication is related to adoption also in a different sense; if an agent decides to adopt a goal of a partner, it can be useful for the group to communicate his intention to the partner. Although in many cases this intention can be easily inferred by the beneficiary by observing the adopting agent's actions, under some conditions it could be impossible for the adopting agent to display his adoption other than communicating it. If this is the case, the utility of communicating becomes higher than the utility of not communicating, since, even if communication adds a little overhead, the second alternative could result in a most disadvantageous situation where two agents (the adopting one and the beneficiary) waste resources - or even conflict - by trying to achieve the same task.
Consider the example where G_i and G_j cooperate to change a light bulb: one of them, G_i, will climb the ladder, while the other, G_j, is in charge of fetching a new light bulb. If G_i switches off the light (because he is nearer to the switch, for example) while G_j is away, and he knows that it was G_j's duty to do it, he should warn the partner ("I have already switched off the light for you!"). In fact, if G_j is not aware that G_i has already switched off the light, he may repeat the action, with the result of closing the circuit again, with potentially bad consequences for G_i: in this case the overhead added by the communicative action is clearly irrelevant if compared to the consequence it could provoke,

for both the beneficiary and the group.

Since agents share a group utility function, we can predict that they will (try to) avoid conflicts with other agents' intentions; in fact, performing an action that interferes with the plans of other team members decreases the utility of the whole team. But if they are aware of this possibility, they may resort to communication for coordinating the choice of single agents' plans ("Don't start painting the ceiling just now, I need the ladder to change the light bulb!"). The trigger of communication relies again on the utility function: if the partner is likely to choose an action that leads to the shared plan failure due to a conflict with the agent's intentions, the best strategy is to add a little overhead to communicate to the partner what he has planned to do.

In all the cases we have examined, the agent has to trade off the cost of communicating against the potential consequences of not communicating: i.e., communication is not compulsory, but that it is produced only if it is convenient to do so. In this way, the requirement of a more flexible communication posed by (Tambe, 1997) is solved by referring to the notion of utility for the whole group. If communication is expensive, it is not convenient for the group to waste resources in communicating. The same holds if communication is not reliable (the message gets lost) or slow: there is a probability that communication has not the desired effect (or it gets it too late). An important consequence is that, if an agent decides that is better (for the group) not to communicate, his choice does not disrupt the group: in fact, communication is not explicitly mentioned in our definition of cooperation.

Moreover, the utility-based model of cooperation allows to release the assumption of perfect communication: when evaluating the expected utility of a a communicative action, the agents takes into account the probabilities that it doesn't bring about the expected result. If the action fails, however, its repetition is not automatic: again, repetition is subject to utility evaluation and this evaluation cannot but take the previous failure into account.

4.2. Communication as Cooperation

The model of cooperation we propose can predict the occurrence of grounding phenomena when the group shared goal is constituted by the exchange of information itself. Once the shared goal of communicating has been established, the goal of reaching the mutual understanding is assumed as a natural consequence of it, stemming from the communicative competence of the group members: for the dialog to proceed coherently, in fact, the interactants need to share the same interpretation of the previous part of the interaction, or better, the mutual belief must hold that their respective interpretations are reasonably aligned.

The requests for acknowledgement of understanding, for example, are useful for the group only when the content to be conveyed is sufficiently relevant for the success of the exchange and there are not irrelevant probabilities that it has not been correctly conveyed. Otherwise, if the cost of the action of requesting for an acknowledgement - and the related cost of acknowledging by the interlocutor, in the speaker's anticipation of his reaction - is expected to be higher than the benefit it produces, the action is not executed, in favor of other, implicit grounding means (see the Principle of least effort in grounding presented in (Clark, 1996)).

The following exchange could seemingly take place in the light bulb example discussed above, due to the importance of the content conveyed by G_i in the first turn:

G_i:Be careful, I have already switched off the light

G_j:Pass me the light bulb, please

G_i:Did you understand what I said?.

Here, the explicit request for acknowledgement by G_i in the third turn is motivated by the relevance of the content of the turn to be conveyed, which has not been explicitly acknowledged by G_j in his reply (second turn).

Spontaneous acknowledgments of understanding by the interlocutor constitute the interlocutor's adoption of the speaker's goal of knowing if he succeeded in his communicative act (e.g. "I have already switched off the light" "I see"). Again, if the cost of the interlocutor adopting this goal is not balanced - according to the interlocutor's utility evaluation - by the utility of the outcomes it may produce, the acknowledgment is disregarded in favour of other, implicit grounding means and the conversation proceeds smoothly (compare the previous example with the following exchange, that we take to happen during an everyday conversation: "Do you know that John got married?" "What is the name of the wife?").

At the same time, the utility function, since it accounts for the reliability of communication, allows to model the frequency of acknowledgments and request for acknowldgements in certain interaction modalities, like telephone conversations.

The explanation of grounding that we have introduced so far is in line with the fact that interactants rarely acknowledge (or request to acknowledge) acknowledgements and notifications themselves: while in nested acknowledgments the relevance of the communicative acts decreases or at least remains the same, the probability that it doesn't succeed becomes lower, and the resource consumption increases.

Moreover, the utility function allows for accounting for the choice of actions that satisfy several goals at the same time, to different degrees: for this reason, the contextual preference for a smoother continuation of the dialog with respect to explicit grounding can be modeled.

In general, the lack of interpretation problems is, by itself, a symptom that there is a common interpretation, i.e., that participants' interpretations are reasonably aligned. This does not mean that the interpretation is really the same, but only that the potential differences fall within the standard individual differences. In absence of specific signs of misalignement, this normally makes the intrinsic utility of any active effort to check the interpretations' alignement very low, because this effort would not be compensated by the low risk of having to engage in a negotiation phase in the following, in order to restore the lost alignement.

On the contrary, the loss of dialog coherence normally means that a misunderstanding has occurred, i.e., that at least one of the two speakers has chosen a wrong interpretation of a turn (see (Ardissono et al., 1998) for a deeper analysis). If a participant realizes that a misunderstanding has occurred, he will compare the alternative of performing a repair action for addressing and solving the misunderstanding (Schegloff, 1992) with the alternative of going on with the next turn, without executing any extra action. If the comparison between the two alternatives ends in favour of the former one, the participant will act in order to find out who is the misinterpreting agent, and correct the wrong interpretation. Intuitively, the utility function here embodies the idea that a misunderstanding is addressed if it is deemed relevant, and is at risk of posing difficulties for the subsequent interaction.

1. if G_i realizes that his G_j is the misinterpreting agent, he will plan a request for repair ("No, I mean that..."); G_j, in turn, is expected to repair his interpretation as requested, and notify the execution of the repair to G_i ("Oh, ok ...").
2. if G_i realizes that he has misinterpreted G_j, he will plan a repair and the notification to B that now he holds the right interpretation ("Oh, you meant that ...").

While it is outside the scope of this work to explain how shared goals are established in a general way, the model of cooperation that we propose provides a framework where the establishment of the shared goal of communicating can be explained. When communicative acts are performed within the context of a group cooperating to a non linguistic shared goal, their occurrence is motivated by the utility they yield with respect to the shared goal. Under these circumstances, we postulate that the performance of a communicative act by one of the group participants normally sets up the shared goal of communicating.

Finally, we exploit a utility-based reasoning to explain the establishment of communicative cooperation outside the context of a group cooperating to a shared goal (Boella et al., 1999): when an agent is requested to cooperate, he is relatively free to refuse cooperation, but, even if he does so, he normally cooperates at the communicative level, by informing the partner about his refusal. In fact, an agent usually has the goal of not offending his partner and the refusal of communicative cooperation is interpreted as an offensive behavior: a notification is a low-cost action, and its omission could result in a repetition of the request or in a solicitation of feedback (rather expensive actions for the partner).

For example, if an agent G_i is asked by another agent G_j to tell him the time, and G_i's clock is broken, he will probably cooperate, at least conversationally, and reply with a communicative act of justification such as "I'm sorry, my watch is broken", instead of ignoring the request and letting G_j believe that his request has been ignored or refused.

The anticipation of the hearer's reaction, even if a logical and not decision-theoretic framework, has been proposed in (Ardissono et al., 1999) for dealing with the choice of polite forms of speech-acts.

5. Related Work and Conclusions

As we stated in the introduction, decision and game theoretic concepts are being increasely used in modeling multi-agent situation, even if they still present some limitations, like, for example, the difficulty of estimating the utility functions. In particular, the idea of taking into account the benefit of the group has been advanced in (Hogg and Jennings, 2000), but in that work the advantage of the group does not take into account anticipatory coordination. On the other hand, (Gmytrasiewicz and Durfee, 1995) exploit the prediction of what the other agents will do in a manner which is very similar to our approach, but they do not have a notion of group utility and goal adoption. In the subsequent paper (Gmytrasiewicz and Durfee, 1997) the consequences of their approach for dialog is considered.

Grosz and Kraus (Grosz and Kraus, 1996) propose a formal specification of the noti on of sharing a plan. They introduce the operator *Intend-that* in order to account for the commitment of each group member to the shared plan; from the intention-that the plan be performed, the agents derive that they have to avoid conflicts and to coordinate the group's behavior through communication. In addition, the definition of shared plans prescribes that agents intend that their partners are able to do their part in the plan. In our model, we have tried to obtain a similar effect by means of the interaction of the shared utility function with the mechanism of goal adoption. In particular, conflicts are avoided since a plan that interferes with the partners' actions normally makes the utility of the group decrease. The goal adoption mechanism makes an agent consider whether, by adopting the partners' goals, a gain for the group is achieved.

Finally, the ideas of goal adoption and anticipatory coordination have been put forth by (Castelfranchi, 1998) and exploited for the definition of cooperation in (Boella et al., 2000) and (Boella, 2000). A similar approach is used for modeling obligations in multi-agent systems in (Boella and Lesmo, 2000). The role of goal adoption and cooperation in communication has been analysed in (Ardissono et al., 2000); similar concepts have been used in (Boella et al., 1999) for dealing with obligations in dialog.

Acknowledgements

This work has been partially supported by the project "Conoscenza, intenzioni e comunicazione" (Knowledge, intentions, and communication) of the National Research Council and by the project "Approcci modulari all'analisi linguistica" (Modular approaches to linguistic analysis) of the Ministry for University and Scientific Research.

6. References

L. Ardissono, G. Boella, and R. Damiano. 1998. A plan-based model of misunderstandings in cooperative dialogue. *International Journal of Human-Computer Studies*, 48:649–679.

L. Ardissono, G. Boella, and L. Lesmo. 1999. The role of social goals in planning polite speech acts. In *Workshop on Attitude, Personality and Emotions in User-Adapted Interaction at UM'99 Conference*, pages 41–55, Banff.

L. Ardissono, G. Boella, and L. Lesmo. 2000. Plan based agent architecture for interpreting natural language dialogue. *International Journal of Human-Computer Studies*, (52):583–636.

G. Boella and L. Lesmo. 2000. Deliberate normative agents. In *Proc. of Autonomous Agents 2000 Workshop on Norms and Institutions.*, Barcelona.

G. Boella, R. Damiano, L. Lesmo, and L. Ardissono. 1999. Conversational cooperation: the leading role of intentions. In *Amstelogue'99 Workshop on Dialogue*, Amsterdam.

G. Boella, R. Damiano, and L. Lesmo. 2000. Cooperation and group utility. In N.R. Jennings and Y. Lespérance, editors, *Intelligent Agents VI — Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99, Orlando FL)*, pages 319–333. Springer-Verlag, Berlin.

G. Boella. 2000. *Cooperation among economically rational agents*. Ph.D. thesis, Università di Torino, Italy.

C. Castelfranchi. 1998. Modeling social action for AI agents. *Artificial Intelligence*, 103:157–182.

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

P. R. Cohen and H. J. Levesque. 1991. Teamwork. *Noûs*, 25:487–512.

P. J. Gmytrasiewicz and E. H. Durfee. 1995. Formalization of recursive modeling. In *Proc. of first ICMAS-95*.

P. J. Gmytrasiewicz and E. H. Durfee. 1997. Rational interaction in multiagent environments: Communication. In *Submitted for publication*, available at http://www-cse.uta.edu/~piotr/www/piotr.html.

B. Grosz and S. Kraus. 1996. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357.

P. Haddawy and S. Hanks. 1998. Utility models for goal-directed, decision-theoretic planners. *Computational Intelligence*, 14:392–429.

I. Hideki and K. Hirofumi. 2000. Observability-based nested belief computation for multiagent systems and its formalization. In N.R. Jennings and Y. Lespérance, editors, *Intelligent Agents VI — Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin.

L. Hogg and N. Jennings. 2000. Variable sociability in agent-based decision making. In N.R. Jennings and Y. Lespérance, editors, *Intelligent Agents VI — Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin.

A. Ndiaye and A. Jameson. 1996. Predictive role taking in dialog: global anticipation feedback based on transmutability. In *Proc. 5th Int. Conf. on User Modeling*, pages 137–144, Kailua-Kona, Hawaii.

E.A. Schegloff. 1992. Repair after the next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 7(5):1295–1345.

M. Tambe. 1997. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7:83–124.

First-Order Inference and the Interpretation of Questions and Answers

Johan Bos & Malte Gabsdil

Computerlinguistik
Universität des Saarlandes
Im Stadtwald, Postfach 151150
66041 Saarbrücken, Germany
{bos,gabsdil}@coli.uni-sb.de

Abstract

Building on work by Groenendijk and Stokhof, we develop a theory of question and answer interpretation for first-order formalisms. The proposed framework is less fine-grained than its higher-order ancestor, but instead offers attractive implementational properties as it deals with the combinatorial explosion problem underlying Groenendijk and Stokhof's original theory. To incorporate the treatment of questions and answers in a larger setting, we use an extension of Discourse Representation Theory to cover typical contextual phenomena such as anaphora and presupposition. The actual interpretation of the dialogue representation is done via a translation to first-order logic. A prototype implementation, using state-of-the-art theorem proving and model building facilities, supports the idea that this first-order approximation of the interpretation of questions and answers is indeed a useful one.

1. Introduction

This paper discusses the treatment of questions and answers in automatic dialogue understanding. Questions require answers, so an inference mechanism that determines whether an utterance is an appropriate answer to a question under discussion should be an elementary part of a dialogue system. Such a component obviously improves human-machine conversation.

We describe a theoretical account and its computational implementation of the interpretation of questions and answers in dialogue. Our first aim is to arrive at a formal definition of what counts as a proper answer to a posed question. Our second aim is to transfer the analysis of questions and answers into a framework that deals with other context-sensitive phenomena, such as pronouns and presuppositions. Our third aim, finally, is to implement these ideas in a prototype dialogue system.

More precisely, we show how first-order logic can be used to model questions and answers (building on work by Groenendijk and Stokhof), and present a computational framework, where state-of-the-art theorem provers and model builders perform the inferences required to determine the appropriateness of (possible) answers to questions. The entire framework is embedded in an extension of Discourse Representation Theory.

2. Modeling Questions

Questions are traditionally analyzed as sets of their possible answers (Hamblin, 1973; Karttunen, 1977). For instance, the question 'Who likes Paris', in a domain with two individuals Tim and Kim, denotes the set containing the answers:

(1) { 'Tim likes Paris', 'Tim does not like Paris',
'Kim likes Paris', 'Kim does not like Paris' }

These approaches are too weak to capture certain aspects of quantification, as they do not contain answers like 'Everybody likes Paris', or 'Only Kim likes Paris' (see Higginbotham (1996) for further discussion). Groenendijk and Stokhof (1984) argue that questions partition the logical space into mutually exclusive and jointly exhaustive sets of possible worlds which represent the different ways in which a question can be answered. Under this view, questions denote *sets of sets* of propositions. For 'Who likes Paris?', we have the answer set:

(2) { { 'Tim likes Paris', 'Kim likes Paris' },
{ 'Tim does not like Paris', 'Kim does not like Paris' },
{ 'Tim likes Paris', 'Kim does not like Paris' },
{ 'Tim does not like Paris', 'Kim likes Paris' } }

This approach adds more structure to the interpretation of questions and therefore offers a more sophisticated way for classifying answers. Following Groenendijk and Stokhof, answers remove those fields in the partition of a question with whom they are inconsistent. A *partial answer* is an answer inconsistent with at least one member of the question's partition, and consistent with all others. For instance, the answer 'Kim likes Paris' is only inconsistent with two members of the partition in (2). However, 'Tim has red hair' is not an answer as it is consistent with all fields in (2). A question is finally *resolved* (using Ginzburg's (1995) terminology) if only one consistent field is left. The answers 'Only Tim likes Paris' or 'Nobody likes Paris' are resolving answers, because they are consistent with only one field of partition (2).

We will use Groenendijk and Stokhof's approach for implementing questions and answers but modify it on two points. First, because we want to make use of first-order inference, we take propositions to denote truth-values instead of functions from states to truth-values. Second, computing the partitions in Groenendijk and Stokhof's theory is subject to a combinatorial explosion problem. To deal with this, we simplify the structure of partitions.

2.1. The Combinatorial Explosion Problem

From a computational perspective, the original approach of Groenendijk and Stokhof faces a serious problem. The size of partitions of single wh-questions grows exponentially in the size of the question's domain (by domain

	Consistency Checks				
Example	(3)	(4)	(5)	(6)	Result
'Where do you want to go? I go everywhere. '	yes	yes	no	no	proper answer
'Where do you want to go? I don't go to Paris.'	no	yes	yes	yes	proper answer
'Where do you want to go? I go to Paris'	yes	yes	yes	no	proper answer
'Where do you want to go? I go nowhere'	no	no	yes	yes	proper answer
'Where do you want to go? I go somewhere'	yes	yes	yes	no	proper answer
'Where do you want to go? I start in London'	yes	yes	yes	yes	improper answer
'Where do you want to go? Paris is beautiful'	yes	yes	yes	yes	improper answer
'Where do you want to go? Paris is a country'	no	no	no	no	improper answer

Figure 1: Illustration for determining proper answers to wh-questions.

we understand the individuals syntactically determined by the wh-clause, such as locations for 'where', persons for 'who', and so on). To check for consistency with every field of a wh-partition would mean making 2^n inferences (n being the size of the domain) which is computationally not feasible, except for toy domains.

To solve this problem, we assume that wh-questions are—much like quantifiers in natural language—segmented into a domain and a body. For wh-questions the domain is defined as above and is assumed to be nonempty, and the body is the property that should hold for the inquired members of the domain.

The general strategy pursued for interpreting answers is as follows: By using the domain and body of a question Q, it is possible to construct formulas of first order logic that coarsely describe Q's partition. Taking these formulas to represent the "semantics" of Q, we then determine whether a proposition A is a *proper answer* by checking for consistency of A and Q. The next sections describe this in more detail.

2.2. First-Order Semantics of Questions

Suppose that a wh-question Q is translated into a formula with domain D, body B, and principal variable referent x. Then an answer A is defined as proper for a question Q if at least one of the propositions (3)–(6) is consistent, and at least one of them is inconsistent.

(3) $\forall x[D(x) \rightarrow B(x)]$ & A

(4) $\exists x[D(x)$ & $B(x)]$ & A

(5) $\exists x[D(x)$ & $\neg B(x)]$ & A

(6) $\neg\exists x[D(x)$ & $B(x)]$ & A

Compared to the original ideas in Groenendijk and Stokhof (1984), these four formulas reduce arbitrarily large partitions for single wh-questions to partitions with only three fields. Formula (3) characterizes the answer that the body of the question holds for every individual in the question domain, (4) and (5) are compatible with possible answers that state that at least one individual either does or does not have this property, and (6) represents the answer that no individual in the question domain has the property expressed by the body (Figure 1 gives some examples). Thus, a proper answer resembles a partial answer (in the sense of Groenendijk and Stokhof) on such a reduced partition. Assume, for example, a model with three individuals a, b, and c and a unary property P. The partition for the question 'Who is P?' would have eight fields standing for the possibilities that either a, b, and c are P, only a, b, or c is P, nobody is P or that P holds of either a and b, a and c, or b and c. A graphical representation of such a partition is depicted in the left part of Figure 2. The four first-order formulas in (3)–(6), on the other hand, describe a partition of three fields for the same question and domain, as illustrated by the right part of Figure 2.

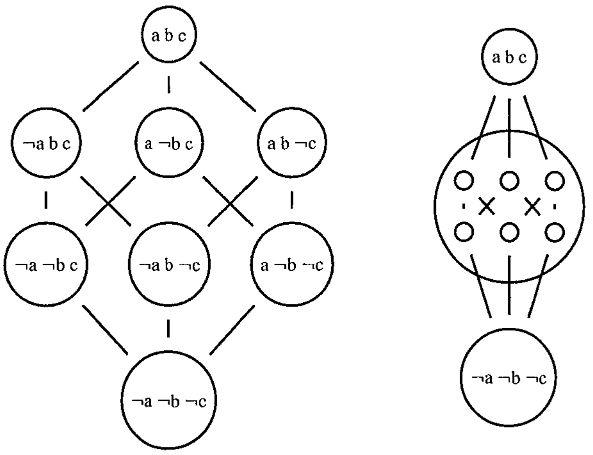


Figure 2: Two partitions for a wh-question with domain size 3, according to Groenendijk and Stokhof (left), and our simplified analysis (right). Note that since the two partitions have the same general structure, the right partition is included in the left partition by means of Groenendijk and Stokhof's partition-inclusion operator $\sqsubseteq$.

Of course, the modified analysis means a loss of finegrainedness with respect to Groenendijk and Stokhof's original work, because it is not able to determine strong exhaustiveness of answers for wh-questions. For instance, our analysis would classify both 'Only a is P' and 'a is P' as proper answers to the question 'Who is P?', but does not recognize that the former is strongly exhaustive, and the latter is not. Whether such a fine distinction is required is debatable. Ginzburg discusses several examples where strong exhaustiveness seems to be an inappropriate resolvedness criterion (Ginzburg, 1996).

The important feature of our new analysis is that it copes with the combinatorial explosion problem, and thereby opens the way to computational implementation. Moreover, we believe that the approach naturally extends to yes/no and choice-questions. For yes/no-questions, the four formulas in (3)–(6) collapse into two by equivalence. The remaining formulas represent the answers 'yes' and 'no', which in turn model the bipartitions for yes/no-questions assumed by Groenendijk and Stokhof. This means that in the case of yes/no-questions, a proper answer can be identified with a strongly exhaustive answer.

3. Questions and Answers in DRT

The previous section outlined our analysis of questions and the interpretation of answers. This section describes how we can embed it in a framework for dialogue analysis, namely Discourse Representation Theory (Kamp and Reyle, 1993). This shift does not mean that we say farewell to first-order logic. In fact, for the interpretation of Discourse Representation Structures (DRSs, the representations used in DRT), we use a translation to first-order formulas.

Originally, DRT focuses on discourse, and puts forward concrete proposals to deal with anaphora and presupposition. To deal with specific dialogue phenomena, we use some of the extensions to DRT as proposed by (Poesio and Traum, 1998), to wit the integration of dialogue acts in DRSs and operations on DRSs for modeling *grounding acts*. The implementation of grounding slightly deviates from Poesio & Traum (see below). The treatment of questions and answer is a novel extension to Poesio & Traum's model.

3.1. Defining Dialogue Representations

Using DRSs for the analysis of dialogue requires at least three simple extensions to the basic syntax of DRSs. First, we have DRS-merging for two DRSs K and K′ resulting in a new DRS (K;K′). Second, questions are represented by the DRS (K?K′), where K and K′ are DRSs, representing the domain and body of a question, respectively. Third, DRS-conditions can be formed by τ:K, where τ is a discourse referent and K a DRS. This latter extension associates discourse referents with DRSs, and hence allows us to connect discourse referents with questions and answers. A sortal ontology on discourse referents assures that discourse referents for questions or answers are disjoint from other entities in the domain of interpretation. This ontological information, and other supportive background knowledge, is assumed to be part of the main DRS in the following discussion.

3.2. Building Dialogue Representations

A new utterance is translated to a DRS K and appended to the dialogue representation by conditions 'x:K' and 'P(x)', where x is a fresh discourse referent associated with K, and P a relation symbol specifying sortal information (e.g., whether it is a question or a proposition). Further, there is a condition 'under-discussion(x)' that marks that x is currently under discussion. Additional conditions are introduced stating the dialogue act associated with the utterance. As basic dialogue acts we have 'ask', 'reask', 'check', and 'assert'.

Context-dependent phenomena are dealt with as described in (Blackburn et al., 1999), using Van der Sandt's resolution algorithm (Van der Sandt, 1992). Anaphoric and presuppositional elements are resolved with respect to the DRS of the dialogue so far, generating a set of potential readings. From this set those readings are chosen that obey the acceptability constraints: they should be consistent and informative.

Consider as example the dialogue in (7) with two participants A and B.

(7) A: 'Where do you want to start?'
B: 'I am leaving from Paris'.

The DRS for this mini-dialogue, from A's perspective, after hearing B's answer, is given in (8), where discourse referent y maps to participant A and x to B.

$$(8)\quad \begin{array}{|l|}\hline \text{q p x y u} \\ \hline \text{ask(y,x,q)} \\ \text{question(q)} \\ \text{obliged-to-address(x,q)} \\ \text{q:}\left(\begin{array}{|l|}\hline \text{z} \\ \hline \text{location(z)} \\ \hline \end{array} \; ? \; \begin{array}{|l|}\hline \text{e} \\ \hline \text{start(e,x)} \\ \text{in(e,z)} \\ \hline \end{array} \right) \\ \text{assert(x,p)} \\ \text{proposition(p)} \\ \text{paris(u)} \\ \text{p:} \begin{array}{|l|}\hline \text{f} \\ \hline \text{leave(f,x)} \\ \text{from(f,u)} \\ \hline \end{array} \\ \text{under-discussion(p)} \\ \hline \end{array}$$

The main DRS in (8) contains two conditions of the form τ:K, where the first occurrence represents the question, and the second an assertion.[1] These embedded DRSs are subordinated to the main DRS. This means that free variables appearing in them are actually bound by discourse referents occurring in the main DRS. To illustrate this idea, the proper name 'Paris' caused global accommodation of its discourse referent u (following Van der Sandt), which appears in the main DRS. This discourse referent binds the free occurrence of u in the DRS annotated by p in (8).

The main DRS represents the 'common ground', and is therefore subject to the process known as *grounding* (Traum, 1994). An optimistic instance of grounding is one where the hearer assumes that (s)he understood the utterance in the way it was intended, and as a result takes this information for granted (provided no contradictions arise). A pessimistic (or cautious) grounding scenario is one where the hearer is not sure what (s)he heard, does not accept the new information, and perhaps starts a clarification dialogue. This grounding behavior involves the speaker in a similar way.

[1] Note that temporal and modal information is left out from these examples. Events are represented by discourse referents.

Technically, the content of grounded utterances is accommodated to the main DRS, resembling an acceptance of the utterance. Coming back to our example (7), the DRS in (9) shows the situation after grounding the assertion 'I am leaving from Paris'.

(9)
$$\begin{array}{|l|}\hline \text{q p x y u f} \\ \hline \text{ask(y,x,q)} \\ \text{question(q)} \\ \text{q:}\left(\begin{array}{|l|}\hline \text{z} \\ \hline \text{location(z)} \\ \hline \end{array} \;?\; \begin{array}{|l|}\hline \text{e} \\ \hline \text{start(e,x)} \\ \text{in(e,z)} \\ \hline \end{array}\right) \\ \text{assert(x,p)} \\ \text{proposition(p)} \\ \text{p:}\begin{array}{|l|}\hline \text{f} \\ \hline \text{leave(f,x)} \\ \text{from(f,u)} \\ \hline \end{array} \\ \text{paris(u)} \\ \text{leave(f,x)} \\ \text{from(f,u)} \\ \hline \end{array}$$

The content of ungrounded utterances stays at the subordinate level until its status is clarified. Note that the actual content of questions is never 'grounded', only the content of propositions undergoes this kind of accommodation. This model of grounding is less elaborated than the one proposed by Poesio and Traum (1998) for DRT, but it suffices for our purposes.

Assuming that each new utterance is constructed by assigning fresh occurrences of discourse referents, it can be shown that clashes of duplicate discourse referents will never appear. As free variables in ungrounded DRSs are already bound by discourse referents declared in the universe of the main DRS, grounding will never introduce new free occurrences. Hence this grounding mechanism is safe from a computational semantic perspective.

3.3. Interpreting Dialogue Representations

The representations for dialogues that we use form an intermediate level of representation required for its interpretation, according to the principles of DRT. Interpretation of DRSs is required in our model to implement the consistency tests as formulated for the rules for questions and answers, but also for applying acceptability constraints put forward by the resolution algorithm for pronouns and presuppositions. To perform these inferences on DRSs, we appeal to the standard translation to expressions of first-order logic (Kamp and Reyle, 1993; Blackburn et al., 1999).[2] The translation is defined by the function $(.)^{fo}$ by the clauses presented in Figure 3.

This translation is meaning preserving (a DRS is *consistent* or *inconsistent* if and only if its first-order translation has the same property) and the computational overhead involved in translation is negligible (the translation is linear in the size of the input).

The standard translation is not defined for our three extensions of the DRS language, i.e. conditions of the form τ:K, and DRSs of the form (K;K′), or (K?K′). First, for DRSs of the form (K;K′) we use merge-reduction along the lines in (Muskens, 1996) to obtain standard DRSs. Merge-reduction is the process of combining the universes of K and K′ and their conditions respectively. This can be done safely as long as the intersection of the universes of K and K′ are disjoint and no free variables in K are bound by discourse referents in K′. Second, DRS-conditions of the form τ:K represent ungrounded utterances. We do not want to include ungrounded information in any inference tasks, and therefore do not need to extend $(.)^{fo}$ for this type of condition. Third, DRSs of the form (K?K′) represent questions, and are only interpreted with respect to possible answers. This is done by transferring the insights on question partitions for first-order logic to DRSs. Recall from the previous discussion that possible answers are grounded before they are subject to the process whether they constitute a proper answer. Hence, given an answer A and main DRS M, M contains the information of A after grounding. For a question (D?B), we then arrive at the following four consistency tests:

(10) $(\text{M};\begin{array}{|c|}\hline \\ \hline \text{D}\Rightarrow\text{B} \\ \hline \end{array})$

(11) (M;(D;B))

(12) $(\text{M};(\text{D};\begin{array}{|c|}\hline \\ \hline \neg\text{B} \\ \hline \end{array}))$

(13) $(\text{M};\begin{array}{|c|}\hline \\ \hline \neg(\text{D};\text{B}) \\ \hline \end{array})$

Applying these schemata to (9), where we want to check whether the proposition associated with discourse referent p is a proper answer to the question annotated by marker q, we get the following instantiations for (10)–(13):

(14)
$$\left(\begin{array}{|l|}\hline \text{q p x y u f} \\ \hline \text{ask(y,x,q)} \\ \text{question(q)} \\ \text{assert(x,p)} \\ \text{proposition(p)} \\ \text{paris(u)} \\ \text{leave(f,x)} \\ \text{from(f,u)} \\ \hline \end{array} ; \begin{array}{|c|}\hline \\ \hline \begin{array}{|l|}\hline \text{z} \\ \hline \text{location(z)} \\ \hline \end{array} \Rightarrow \begin{array}{|l|}\hline \text{e} \\ \hline \text{start(e,x)} \\ \text{in(e,z)} \\ \hline \end{array} \\ \hline \end{array}\right)$$

(15)
$$\left(\begin{array}{|l|}\hline \text{q p x y u f} \\ \hline \text{ask(y,x,q)} \\ \text{question(q)} \\ \text{assert(x,p)} \\ \text{proposition(p)} \\ \text{paris(u)} \\ \text{leave(f,x)} \\ \text{from(f,u)} \\ \hline \end{array} ;\left(\begin{array}{|l|}\hline \text{z} \\ \hline \text{location(z)} \\ \hline \end{array} ; \begin{array}{|l|}\hline \text{e} \\ \hline \text{start(e,x)} \\ \text{in(e,z)} \\ \hline \end{array}\right)\right)$$

[2] Some alternative translation are available (Van Eijck and De Vries, 1992; Muskens, 1996).

$$\left(\begin{array}{|l|}\hline x_1\cdots x_n\\\hline \gamma_1\\ .\\ .\\ \gamma_m\\\hline\end{array}\right)^{fo} = \exists x_1\cdots\exists x_n\,((\gamma_1)^{fo}\wedge\cdots\wedge(\gamma_m)^{fo})$$

$$\begin{array}{rcl}(R(x_1,\ldots,x_n))^{fo} & = & R(x_1,\ldots,x_n)\\ (x_1 = x_2)^{fo} & = & x_1 = x_2\\ (\neg K)^{fo} & = & \neg(K)^{fo}\\ (K_1 \vee K_2)^{fo} & = & (K_1)^{fo}\vee(K_2)^{fo}\end{array}$$

$$\left(\begin{array}{|l|}\hline x_1\cdots x_n\\\hline \gamma_1\\ .\\ .\\ \gamma_m\\\hline\end{array}\Rightarrow K\right)^{fo} = \forall x_1\cdots\forall x_n(((\gamma_1)^{fo}\wedge\cdots\wedge(\gamma_m)^{fo})\rightarrow(K)^{fo})$$

Figure 3: Translation of DRS to first-order expressions (Blackburn et al., 1999).

(16) $\left(\begin{array}{|l|}\hline \text{q p x y u f}\\\hline \text{ask(y,x,q)}\\ \text{question(q)}\\ \text{assert(x,p)}\\ \text{proposition(p)}\\ \text{paris(u)}\\ \text{leave(f,x)}\\ \text{from(f,u)}\\\hline\end{array};\left(\begin{array}{|l|}\hline \text{z}\\\hline \text{location(z)}\\\hline\end{array};\neg\begin{array}{|l|}\hline \\\hline \begin{array}{|l|}\hline \text{e}\\\hline \text{start(e,x)}\\ \text{in(e,z)}\\\hline\end{array}\\\hline\end{array}\right)\right)$

(17) $\left(\begin{array}{|l|}\hline \text{q p x y u f}\\\hline \text{ask(y,x,q)}\\ \text{question(q)}\\ \text{assert(x,p)}\\ \text{proposition(p)}\\ \text{paris(u)}\\ \text{leave(f,x)}\\ \text{from(f,u)}\\\hline\end{array};\begin{array}{|l|}\hline \\\hline \neg\left(\begin{array}{|l|}\hline \text{z}\\\hline \text{location(z)}\\\hline\end{array};\begin{array}{|l|}\hline \text{e}\\\hline \text{start(e,x)}\\ \text{in(e,z)}\\\hline\end{array}\right)\\\hline\end{array}\right)$

After applying merge-reduction to (14)–(17) these resulting DRSs can be fed to the standard translation arriving at ordinary first-order representations. For instance, the DRS (14) is reduced to (18) and translated to the first-order formula (19):

(18) $\begin{array}{|l|}\hline \text{q p x y u f}\\\hline \text{ask(y,x,q)}\\ \text{question(q)}\\ \text{assert(x,p)}\\ \text{proposition(p)}\\ \text{paris(u)}\\ \begin{array}{|l|}\hline \\\hline \begin{array}{|l|}\hline \text{z}\\\hline \text{location(z)}\\\hline\end{array}\Rightarrow\begin{array}{|l|}\hline \text{e}\\\hline \text{start(e,x)}\\ \text{in(e,z)}\\\hline\end{array}\\\hline\end{array}\\ \text{leave(f,x)}\\ \text{from(f,u)}\\\hline\end{array}$

(19) ∃q(∃p(∃x(∃y(∃u(∃f(ask(y,x,q) & (question(q) &
(assert(x,p) & (proposition(p) & (paris(u) &
(∀z(location(z)→ ∃e(start(e,x)&in(e,z))) &
(leave(f,x) & from(f,u)))))))))))))

If the resulting first-order formula is satisfiable, the DRS is consistent. If the negation of the resulting formula is a theorem, the DRS is marked as inconsistent. Of course, these inferences have to be supported by background knowledge. This background knowledge is a set of axioms derived from ontological information (an isa-hierarchy of concepts in the domain) and domain knowledge (such as leaving from a location implies starting in that location). Given the proper background knowledge, we find out that (14), (15), and (16) are consistent, and (17) is inconsistent. As at least one of the tests is consistent, and one of them is inconsistent, the question has been properly answered.

4. Implementation

The ideas presented above are implemented in one of the research prototypes developed in the Trindi Project (Traum et al., 1999a). MIDAS, as the system is called, covers the domain of route services, and aims to provide the user with a description of a route on the basis of a starting point, destination, and time. The implementation follows the model of dialogue moves and information state revision (Traum et al., 1999b). Utterances are treated as updates to the current state of the dialogue. This section describes the basic architecture of the system, how the information state (the DRS of the dialogue) is updated during dialogue processing, and how the inference tasks are implemented.

4.1. Basic Architecture

The system components of MIDAS are a parser, semantic construction, dialogue move engine, generator, and synthesizer. The start of the system initializes the information state, by loading a plan with actions it intends to perform. In the route service domain, these are questions that ask the user where (s)he wants to go, when (s)he wants to go, where (s)he wants to start, and so on. New utterances are analyzed by the parser, on the basis of which the semantic construction component builds a DRS. This DRS is integrated within the actual information state, by resolving pronouns, ellipsis, and presuppositions. Next, the dialogue move engine updates the current information state by applying a set of update rules to it. On the basis of this new information state, the system generates new utterances and feeds these to the synthesizer. Then the system waits for input of the user and the whole process is repeated until all intended actions are performed.

4.2. Information State Updates

The dialogue move engine from MIDAS changes the information state by either adding or removing information from it. These changes are triggered by *update rules*. Update rules consist of a name and three parts: a set of binders, a set of preconditions, and a set of effects. For any binding such that all the preconditions of an update rule holds, the dialogue move engine applies the effects to the information state. This iterative process continues until no further rules apply.

Figure 4 shows some of the update rules of MIDAS. Here we use the flat notation for DRSs, where $[U|C]$ stands for a DRS with discourse referents U and conditions C. 'K::D' binds the discourse referents in D with those in K. Further, 'K$\supseteq$ C' is short for 'DRS K contains conditions C' under the current binding, and 'K$\not\supseteq$ C' is short for 'DRS K does not contain conditions C'. The operations K+=C add conditions C to DRS K, and the operation K$-$=C remove conditions C from DRS K. Note that these operations are in the scope of the binders. The function 'consistent' returns true if its argument (a DRS) is consistent and false if it is inconsistent. $\oplus$ maps a list of boolean values to true if at least one member of the list is true, and one is false. The function 'consistent' calls the external inference component (see next section).

The rule for *optimistic grounding* has as preconditions that there is an asserted proposition under discussion, and that MIDAS is in optimistic mood. The effects add the content of the proposition to the main DRS, and cancel the status of being under discussion: after grounding the assertion, it is being dealt with. The rule for *pessimistic grounding* activates a check-question. The other rules deal with updating the intentions, and asking, addressing, or repeating questions.

The last update rule in Figure 4 deals with answer determination. One of the preconditions for this rule is the presence of a question that the user is obliged to answer. The other preconditions determine whether the main DRS contains a proper answer, by appealing to the consistency checks. If this is the case, the obligation expires. The next section describes how the consistency-checks are done in practice.

4.3. Inference

To implement inference, MIDAS makes use of current automated deduction techniques for first-order logic. As these, mostly, do not work on DRSs directly, we use the translation to first-order predicate logic with equality (Figure 3). The basic kind of inference we are interested in is checking for consistency. A formula is consistent if it is satisfiable, or if its negation is a theorem. Therefore, not only theorem provers are useful, but also model builders for detecting satisfiability.

MIDAS requires inference at two stages within processing the dialogue. First, the resolution component (dealing with anaphora and presupposition) generally produces several analyses, of which only the consistent ones are taken for further consideration (this is done by using the same techniques as in the DORIS system (Blackburn et al., 1999)). Second, the update rule for answer determination requires consistency checking of four formulas. So, generally, for both of these stages, we have many independent inference tasks for which we want an answer soon (to meet real-time constraints). Moreover, for each problem we need to find out whether it is a theorem or whether it is satisfiable, so it makes sense to call a theorem prover and a model builder in parallel.[3]

By making use of MathWeb (Franke and Kohlhase, 1999), inference problems can be solved in parallel in a competitive distributive framework, using local intra-nets or the Internet to spread the inference tasks on different machines. Currently, MathWeb runs inference services via the Internet at around 20 machines in Saarbrücken, Edinburgh, and Budapest. Of the inference arsenal offered by MathWeb, MIDAS uses the theorem provers Bliksem (De Nivelle, 1998), SPASS (Weidenbach et al., 1996), FDPLL (Baumgartner, 2000), and Otter (McCune and Padmanabhan, 1996), and the model builder MACE (McCune, 1998). It should be noted here that FDPLL and SPASS handle satisfiable problems, too.

As for answer determination, each question-answer pair results in four inference problems that are send to MathWeb. Computing times vary from 300–5000 msecs on each problem, where non-answers (300–1200 msecs) take less effort than proper answers (500–5000 msecs). For a set of four problems, MathWeb uses in average a total time of around 1.7 times the time of one problem (including Internet latency times, which are very low in general). These results clearly show the benefits of the MathWeb concept to distributed inference.

5. Conclusions and Future Work

We avoid the inherent combinatorial explosion in Groenendijk and Stokhof's theory of questions and answers by reformulating their approach in first-order logic and taking questions to denote partitions with only three different fields. We have illustrated this approach for wh-questions, and future work aims at extending the approach to deal with yes/no-questions and choice-questions.

The steps for deciding the properness of an answer are entirely based on first-order inference. This has computational advantages: many state-of-the art first-order inference services, such as theorem provers and model builders, offer high speed and coverage. The price for such a reduced analysis obviously is a loss of fine-grainedness in answer evaluation (i.e. if we wanted to we would have to find other means to detect (strongly) exhaustive answers), but so far we are not aware of important practical consequences.

[3] Incidentally, first-order logic is not decidable. In theory, this means that there is a chance that theorem provers for some input (when given enough resources) will never return with an answer. In practice, one uses time-constraints.

Name:	**Optimistic Grounding**
Binders:	M::[X,Y,P,K]
Preconditions:	M ⊇ [user(X), assert(X,Y,P), proposition(P), P:K, under-discussion(P), midas(Y), optimistic(Y)]
Effects:	M+= K M−=[][under-discussion(P)]

Name:	**Pessimistic Grounding**
Binders:	M::[X,Y,P,K]
Preconditions:	M ⊇ [user(X), assert(X,Y,P), proposition(P), P:K, under-discussion(P), midas(Y), pessimistic(Y)]
Effects:	M+= [Q][question(Q), obliged-to-address(X,Q), Q:K, check(Y,X,Q), under-discussion(Q)] M−= [][under-discussion(P)]

Name:	**Update Intentions**
Binders:	M::[X,Q]
Preconditions:	M ⊇ [question(Q), intend-to-ask(X,Q), answered(Q)]
Effects:	M−=[][intend-to-ask(X,Q)]

Name:	**Ask a Question**
Binders:	M::[X,Y,Q]
Preconditions:	M ⊇ [midas(X), user(Y), question(Q), intend-to-ask(X,Y,Q), unanswered(Q)] M ⊉ [obliged-to-address(Y,Q)]
Effects:	M+= [][ask(X,Y,Q), obliged-to-address(Y,Q), under-discussion(Q)]

Name:	**Repeat a Question**
Binders:	M::[X,Y,Q]
Preconditions:	M ⊇ [midas(X), user(Y), question(Q), obliged-to-address(Y,Q)] M ⊉ [under-discussion(Q)]
Effects:	M+= [][reask(X,Y,Q), under-discussion(Q)]

Name:	**Address a Question**
Binders:	M::[X,Y,Q]
Preconditions:	M ⊇ [midas(X), user(Y), ask(X,Y,Q), question(Q), under-discussion(Q)]
Effects:	M+= [][obliged-to-address(Y,Q)] M−= [][under-discussion(Q)]

Name:	**Determine Answer to Question**
Binders:	M::[Q,D,B,Y]
Preconditions:	M ⊇ [question(Q), Q:D?B, user(Y), obliged-to-address(Y,Q)] ⊕ (consistent(M;[|D⇒B]), consistent(M;(D;B)), consistent(M;(D;[|¬B])), consistent(M;[|¬(D;B)]))
Effects:	M−= [][obliged-to-address(Y,Q)] M+= [][answered(Q)]

Figure 4: Example Update Rules from MIDAS. Note that 'M' stands for the main DRS.

We implemented the ideas in a prototype system under the name of MIDAS. This dialogue system uses Discourse Representation Structures to serve as intermediate structures to model the ongoing dialogue. Questions are indirectly interpreted with their (possible) answers by translating them from the intermediate structure to first-order logic. The required inference tasks are carried out by the MathWeb society of inference agents.

Ginzburg (Ginzburg, 1995) has pointed out that to decide whether a question is finally resolved for a dialogue participant, one also has to take additional criteria into account. These constitute the goals associated with a question and the questioner's view of the world. In our approach, these factors can be integrated by performing additional inference tasks or by providing further axioms when checking for consistency. In particular, the mental state of a dialogue participant can be modeled by first order formulas that are send as additional axioms to the inference machinery. According to Ginzburg, a resolving answer has to entail the goals that a dialogue participant associates with a certain question. Consider:

(20) A: 'Where do you want to start?'
B: 'Germany.'

In our approach, B would properly answer A's question, because Germany is a location, and that's what A was asking for. But if A's goal was to find out in which *city* B intends to start, then B's answer does not provide the information A was looking for, although it is still a partial answer to A's question. In our view, an entailment relation between the questioner's goals and an answer should be modeled as an *additional* constraint on determining resolving answers. We argue that it is necessary to make a first coarse classification of whether a proposition addresses a question under discussion, before finally checking for resolvedness. Such a second step, taking into account the goals of the questioner for determining answers, would obviously improve our analysis and is currently under investigation.

6. Acknowledgments

This work has been carried out as part of the Trindi (Task Oriented Instructional Dialogue) EC Project LE4-8314. We thank Alexander Koller and Kristina Striegnitz, our colleagues from the Trindi project, as well as two anonymous reviewers, for valuable comments on earlier versions of this paper.

7. References

Peter Baumgartner. 2000. FDPLL – A First-Order Davis-Putnam-Logeman-Loveland Procedure. In David McAllester, editor, *CADE-17 – The 17th International Conference on Automated Deduction*, Lecture Notes in Artificial Intelligence. Springer. To appear.

Patrick Blackburn, Johan Bos, Michael Kohlhase, and Hans de Nivelle. 1999. Inference and Computational Semantics. In Bunt and Thijsse, editors, *IWCS-3*, Tilburg, NL.

Hans De Nivelle. 1998. A resolution decision procedure for the guarded fragment. In *CADE 15*. Springer Verlag.

Andreas Franke and Michael Kohlhase. 1999. System description: Mathweb, an agent-based communication layer for distributed automated theorem proving. In *16th International Conference on Automated Deduction CADE-16*.

Jonathan Ginzburg. 1995. Resolving Questions, I. *Linguistics and Philosophy*, 18(5):459–527.

Jonathan Ginzburg. 1996. Interrogatives: Questions, Facts and Dialogue. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 385–422, Oxford, UK/Cambridge, USA. Blackwell.

Jeroen Groenendijk and Martin Stokhof. 1984. On the Semantics of Questions and the Pragmatics of Answers. In Fred Landman and Frank Veltman, editors, *Varieties of Formal Semantics*, Groningen-Amsterdam Studies in Semantics (GRASS) 3, pages 143–170, Dordrecht/Cinnaminson, U.S.A. Foris Publications.

C.L. Hamblin. 1973. Questions in Montague Grammar. *Foundations of Language*, 10:41–53.

James Higginbotham. 1996. The Semantics of Questions. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 385–422, Oxford, UK/Cambridge, USA. Blackwell.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic.* Studies in Linguistics and Philosophy 42. Kluwer Academic Publishers, Dordrecht/Boston/London.

Lauri Karttunen. 1977. Syntax and Semantics of Questions. *Linguistics and Philosophy*, 1:3–44.

W. McCune and R. Padmanabhan. 1996. *Automated Deduction in Equational Logic and Cubic Curves.* Lecture Notes in Computer Science (AI subseries). Springer-Verlag.

W. McCune. 1998. Automatic Proofs and Counterexamples for Some Ortholattice Identities. *Information Processing Letters*, (65):285–291.

Reinhard Muskens. 1996. Combining montague semantics and discourse representation. *Linguistics and Philosophy*, 19:143–186.

Massimo Poesio and David Traum. 1998. Towards an Axiomatization of Dialogue Acts. In J. Hulstijn and A. Nijholt, editors, *Formal Semantics and Pragmatics of Dialogue, Proceedings of Twendial '98*, pages 207–221, Universiteit Twente, Enschede.

David Traum, Peter Bohlin, Johan Bos, Staffan Larsson, Ian Lewin, Colin Matheson, and David Milward. 1999a. Dialogue Dynamics and Levels of Interaction. Technical Report Deliverable D3.1p, Trindi.

David Traum, Johan Bos, Robin Cooper, Staffan Larsson, Ian Lewin, Colin Matheson, and Massimo Poesio. 1999b. A model of dialogue moves and information state revision. Technical Report Deliverable D2.1, Trindi.

David R. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation.* Ph.D. thesis, University of Rochester, Department of Computer Science, Rochester.

Rob A. Van der Sandt. 1992. Presupposition Projection as Anaphora Resolution. *Journal of Semantics*, 9(4):332–378.

Jan Van Eijck and Fer-Jan De Vries. 1992. Dynamic interpretation and hoare deduction. *Journal of Logic, Language and Information*, 1(1):1–44.

Christoph Weidenbach, Bernd Gaede, and Georg Rock. 1996. Spass & flotter, version 0.42. In *13th International Conference on Automated Deduction, CADE-13.* Springer. To appear.

8. URLs

MathWeb: `www.mathweb.org`
MIDAS: `www.coli.uni-sb.de/~bos/midas`
Trindi: `www.ling.gu.se/research/projects/trindi/`

Asynchronous Dialogue Management: Two Case-Studies

Johan Boye[1], Beth Ann Hockey[2], Manny Rayner[2,3]

[1] Telia Research
S-123 86 Farsta, Sweden
johan.boye@trab.se

[2]RIACS
Mail Stop 19–39, NASA Ames Research Center
Moffett Field, CA 94035-1000, USA
bahockey@riacs.edu

[3]netdecisions
Wellington House, East Road
Cambridge CB1 1BH, UK
manny.rayner@netdecisions.co.uk

Abstract

Much of the human-machine dialogue research in the literature tacitly assumes a "synchronous" dialogue model; user talks, system acts, system replies. In particular, the user is not supposed to interrupt the system, neither when it talks nor acts. In this paper, we argue that the synchronous model is not appropriate for most interesting real-life applications, but there is a need for asynchronous dialogue models where the user has the possibility to interrupt the system at any time. By referring to two implemented asynchronous dialogue systems, we try to pinpoint what implementation requirements such a dialogue model entails, and we also outline some of the theoretical implications of asynchronous dialogue management.

1. Introduction

The standard model for dialogue management in spoken language interfaces is based on the assumption of turn-taking: user utterances and system utterances will proceed in alternation. In the literature, this assumption is often made so automatically that one isn't even aware of it. Turn-taking typically extends not just to utterances but also to actions. If the system is capable of performing an action (for example, looking up information in a database) in response to a user command, then it is common to assume that the system's turn includes the relevant actions, and that the user will not speak until they are complete. That is, what is often assumed is something we will call the "synchronous dialogue model": user talks, system acts, system replies.

Although this assumption is popular, there are plenty of reasons for doubting that it is in general appropriate. Empirical investigations show that the strict turn-taking model often agrees badly with data from real conversations (Thompson, 1996). As far as practical system-building is concerned, it is now the case that many systems allow at least a limited ability to break the synchronous dialogue convention, and support so-called "barge-in" functionality: the user is allowed to speak before the system has finished talking, breaking it off in mid-stream.

In this paper, we will argue that "barge-in", far from being an isolated exception, is just the most common instance of a range of obviously reasonable dialogue strategies which break the synchronous convention. We will refer to such strategies as "asynchronous". We motivate our arguments with two case-studies drawn from implemented dialogue systems, each of which allows some degree of asynchrony. The basic questions we are asking are the following:

- What practical reasons are there for wanting to support asynchronous dialogue strategies in spoken language dialogue systems?
- What demands does asynchronous dialogue management place on system architecture?

In particular, we have found that the synchronous dialogue model does not account well for situations in which users are talking to a system that takes any appreciable time to do things. For instance, the system might be searching for information over the Internet, or the system might be a robot moving about in the physical world. In these cases, situations will arise where the human wants to interrupt the system not only when it is *talking*, but also when it is *acting*. There are at least two good reasons why this might happen[1]. The first is obvious: the human may not like what the system is doing, and want it to do something else instead. In particular, the human may want the the system to switch to a new behavior, or they may want it temporarily to suspend operations while it executes a new task.

The second reason is slightly more subtle, but becomes apparent as soon as one starts experimenting with a practical spoken language system. Humans get bored easily: even if the system is busy, they would still prefer to talk to it if they can find something useful to say. For example,

[1] A third, which we will not discuss further, is that the human may not even be aware that they are interrupting.

they may want to give the system a new task that can be executed in parallel with the current one, or after it.

The two points meet up when we consider how to organize confirmation strategies. There is generally a tension here between two conflicting goals; telling the user what you intend to do decreases the risk of a misunderstanding, but carries the penalty of slowing down the dialogue and consequently the execution of the task. An asynchronous dialogue architecture allows the possibility of a compromise, since the system can start executing the task and tell the user what it is going to do simultaneously. As long as choosing the wrong task is not actively dangerous, this tends to be a good way to operate: the user relies on being able to interrupt the system if necessary and correct it, but loses no time if the system understood correctly.

Conversely, the system may want to interrupt the human. Something may come up which it considers more important than its current task; alternately, it may be doing several things at once, and need to keep the user informed about their relative progress.

The rest of the paper is structured as follows. Section 2 describes our two sample applications, focusing on the question of how they realize some of the functionalities discussed above. Section 3 relates our findings to other established frameworks for dialogue management and concludes.

2. Two case studies

In this section, we present two case studies of systems that use asynchronous dialogue management strategies. SMARTSPEAK (Boye et al., 1999) is a travel planning system that fetches information from a web-server. Since this takes appreciable time (typically around 20 seconds to a minute), people want to be able to talk to the system while they are waiting for the web-server to return. This means that in general the system is in the middle of a new conversation by the time the web-server gets back.

The Personal Satellite Assistant PSA (PSA, 1999; Rayner et al., 2000) is a simulated version of a semi-autonomous speech-enabled robot intended for deployment on the International Space Station. The robot acts as a mobile sensor: it can go to different places and measure status variables such as temperature and pressure. Since the robot, once again, takes non-trivial time to carry out commands, the possibility arises that the user may want to interrupt them. We now describe the architectures of these two systems.

2.1. SMARTSPEAK

The architecture of the SMARTSPEAK system is based on having a set of independent modules (or agents) communicate asynchronously by message passing. In particular this entails that the agents have no a priori decided execution order; an agent starts executing when it receives a message, and as soon as it has finished executing it is ready to process the next message, regardless of the current state of the other agents. Hence in principle all the agents could run in parallel in different processes.[2]

The Dialogue Manager (DM) is the heart of the system. It can *receive* messages from the language analysis agents (parsed utterances), and from the database agent (database query results). The DM can *send* messages to the speech synthesizer (system utterances) and to the database agent (database queries). The DM maintains a *dialogue state*, which is transformed as a result to each incoming message. How the DM uses its dialogue state to interpret user utterances, resolve references, select system utterances, etc., is described in Boye et al. (1999).

The database agent (DA) receives messages (database queries) from the DM. The DA first checks its local state to see whether the query has been processed before, so that the results are cached. If so, the DA immediately sends a message back to the DM containing the results. If not, the DA sends a message to the DM indicating that the query will take some time to process, and then spawns a process that contacts the travel database web server via the Internet. When the search process returns its results to the DA, the DA will send the results in a message to the DM.

The asynchronous communication between the DM and the DA has several implications, most notably that it is possible that search results do not return to the DM in the same order the queries were sent (e.g. when the results of the second query were cached, but the results of the first query were not). This of course creates complications for the DM, but on the other hand the system can cope with dialogues like the following:

(a) U: I want to go from Stockholm to Gothenburg on Monday morning.

(b) S: I'm searching the database – it will take about 30 seconds. Do you want a single or a return trip?

(c) U: A return trip.

(d) S: When do you want to go from Gothenburg to Stockholm?

(e) U: On Tuesday afternoon.

(f) S: I have received information about trips from Stockholm to Gothenburg. There is a train at ...

(g) U: Fine, I want to book that please.

(h) S: I have booked a train on Monday at ... I have received information about trips from Gothenburg to Stockholm. There is a train at ...

As concerns the communication between the agents, the following points are worth noting.

1. The user utterance (a) makes the DM send off a search query to the DA.
2. The DA sends back a message that the search will take some time, which triggers the DM to generate the first sentence of utterance (b). The DM then initiates a conversation about a second topic (the return trip).

[2] In the current implementation there are four constantly running processes, which handle speech recognition, speech synthesis, database access, and everything else, respectively (where "everything else" essentially means language analysis and dialogue management). In addition the database agent will spawn short-lived processes; see the main text.

3. The user utterance (e) makes the DM send off a second query to the DA.

4. At some point between (b) and (f), the results concerning the outbound trip are sent from the DA to the DM. (The DM will not present the result as soon as they come in, but will wait until it thinks the moment is right.)

5. At some point between (e) and (h), the results concerning the return trip are sent from the DA to the DM.

Figure 1 shows the communication between the different agents in graphical form.

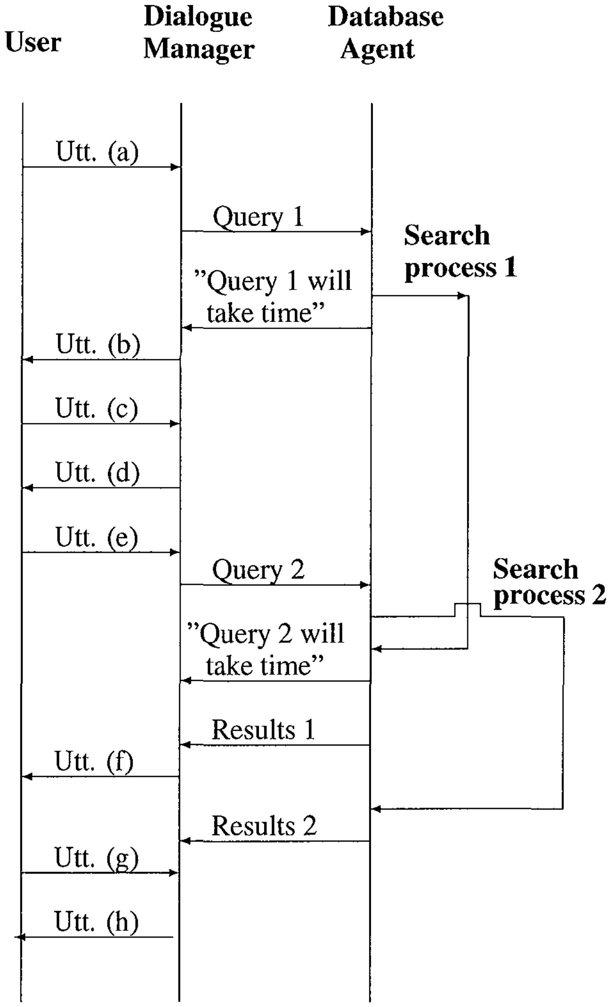


Figure 1: Agent communication in the Smartspeak example

2.2. PSA

The PSA system is configured as a set of independent agents connected using the SRI Open Agent Architecture OAA (Martin et al., 1998). In accordance with the usual OAA design philosophy, each agent is an independent process which maintains its own state. Agents communicate by means of calls in Interagent Communication Language (ICL), an extension of Prolog. Calls can be either synchronous (execution of the calling agent suspends until the call is complete, returning a value) or asynchronous (execution of the calling process continues, and no value is necessarily returned). The agents that are of interest to us here are the following:[3]

Speech recognition and parsing (SRP) The agent attempts to recognize input speech, and if successful produces an output semantic representation, which is sent as an asynchronous message to the dialogue manager agent.

Dialogue manager (DM) The dialogue manager is the central agent: it receives semantic representation messages from the SRP, and decides on the next dialogue action. There are a number of possibilities, of which the most important are the following. (1) Convert the semantic representation into an executable form (a "script"), and pass it to the Action Manager as an asynchronous message; (2) Send a request to the Action Manager specifying a modification to the executing script; (3) Send a request to the Generation agent to create a confirmation question (this can be combined with (1)), a clarification question, or some other verbal response.

The Dialogue Manager's behavior is explained in detail in Rayner et al. (2000).

Action manager (AM) The Action Manager consists of two interrelated pieces of functionality. The Executive subsystem receives scripts from the DM and executes them. Scripts are complex structures composed of atomic actions; execution is ultimately performed by sending synchronous atomic action messages to ROBOT.

The Action Manager also includes a second subsystem, the Interrupt Blackboard, which is used to hold state relevant to processing of interrupts and related requests. Other agents can write to and read from the Interrupt Blackboard at any time, irrespective of whether the Executive is busy.

Simulated PSA Agent (ROBOT) The back-end simulated PSA application, which executes the action messages produced by the AM.

Generation and speech synthesis (GSS) Turns abstract representations of utterances into speech.

The following simple dialogue illustrates how processing works.

(a) U: Go to flight deck.

(b) S: I am going to flight deck. *[Simultaneously starts moving towards flight deck]*

(c) U: Stop. *[Robot stops]*

(d) U: Measure temperature.

[3] In the interests of expositional clarity, we have omitted some agents and collapsed others into single agents. In particular, the Action Manager, which conceptually forms a single unit, is concretely realized as two OAA agents.

(e) **S:** The temperature at the current location is 19 degrees Celsius.

(f) **U:** Continue. *[Robot resumes moving to flight deck]*

In terms of message traffic between agents, the critical points are the following.

1. The user says (a). This causes SRP to send an asynchronous message to the DM, which in turn sends messages to AM (the script to execute) and GSS (the confirmation question to ask). Both messages are asynchronous, so SRP and DM are free to process new messages while AM and GSS are active.
2. The message sent to GSS causes it to say (b). AM sends a synchronous message to ROBOT to start the simulated move command.
3. When (c) is uttered, DM updates the state of AM's Interrupt Blackboard to ask for an interrupt. ROBOT, which is continually monitoring this blackboard, abandons processing of its current command and returns control to AM. Since the Blackboard has a stop message posted, AM pauses and waits for further instructions.
4. The user utters (d). DM now sends a new script to AM. Since AM is in a stopped state, it recursively invokes a new copy of the command interpretation loop from the interrupted point in the current execution. This executes the script (causing GSS to say (e)), and on completion returns control to the previous stopped state.
5. When the user utters (f), DM changes the state ofAM's Interrupt Blackboard from "stop" to "resume". This causes AM to continue from the stopped execution state.

Figure 2 presents the processing described above in graphical form; in order to highlight the correspondences with Figure 1, we have hidden SRP and GSS, which do no more than relay messages between the user and the other agents.

In the next section, we compare the architectures of the SMARTSPEAK systems and PSA systems with others described in the literature, and attempt to draw some general conclusions.

3. Conclusions: Architectures for Asynchronous Dialogue Management

To sum up, asynchronous dialogue management allows for the construction of spoken dialogue system that can *act* and *talk* at the same time. In particular, such systems are able to execute a command from the user, and at the same time receive the user's next command. We want to emphasize that this is not just a cute twist or a question of efficiency, but is rather a *fundamental requirement* when equipping real-time systems (like robots) with a spoken dialogue user interface. In many such applications, it will be unavoidable that the system is busy doing something when the user talks to it; hence for such applications the "synchronous" dialogue model tacitly assumed in much dialogue research is helplessly inappropriate.

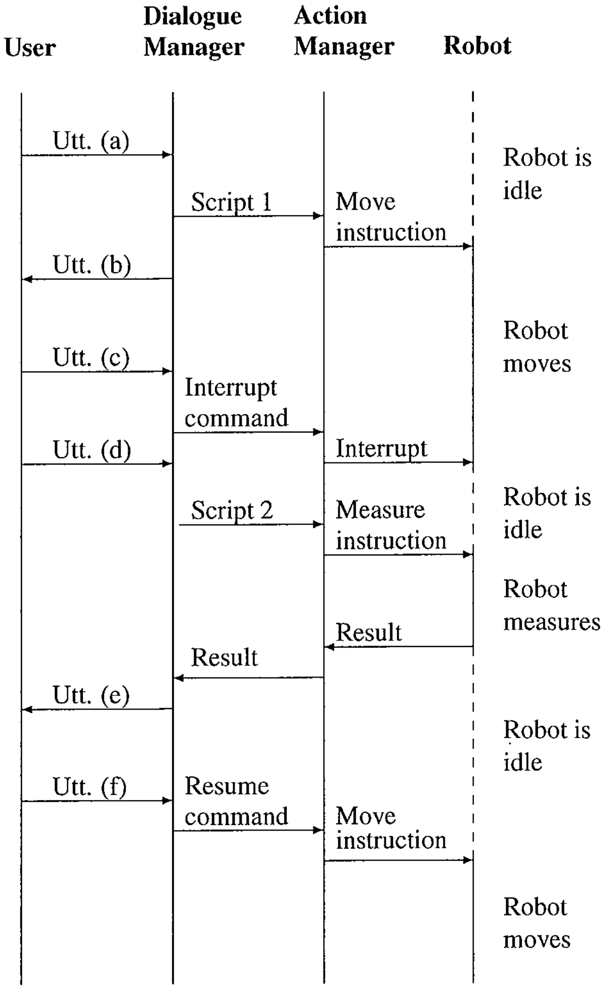


Figure 2: Agent communication in the PSA example

Furthermore, although we have emphasized the practical implementation issues in this paper, asynchronous dialogue management also poses some difficult theoretical problems which we have barely touched upon here (and to the best of our knowledge, have received little attention elsewhere). For instance:

- How should the system deal with a user command ordering the system to do action A, when it is currently busy performing action B? There are a number of possibilities: the system may execute A and B *in parallel*, or it may execute B after A is finished (*sequential* execution), or it may *abort* B, or *suspend* B (in order to resume B after A has finished executing). Or, finally, the system may choose to ignore the user's command altogether. The question is under which circumstances which alternative is preferable, and what criteria the system can use to decide the appropriate course of action.
- Assuming that there are several ongoing and/or suspended activities, how does the system determine the appropriate context in which to interpret the user's utterances, what are the preferred reference resolution strategies, etc.?

- How does the system schedule its utterances, so that the resulting dialogue is comprehensible for the user?

Although a few researchers pay lip-service to the concept of asynchronous dialogue management or handle some special cases, there seems to be surprisingly little acknowledgement that it is important. This is highlighted by the influential DARPA Communicator project DAR (1999), which is currently being used by a large number of research site in the US and Europe. The current Communicator architecture derives from GALAXY-II (Seneff et al., 1998), and organizes the system as a number of stateless "server" processes, controlled by a "script" run from a central "hub" process. This architecture is not aimed towards asynchronous communication between processes. For example, Aberdeen et al. (1999) contrasts Communicator with OAA as follows:

> Both schemes provide for flexible flow of control. However while flow of control is explicitly programmed in the Hub, in the OAA it is determined autonomously by interactions between the agents ... In sum, the Communicator allows for programmable but pre-determined flows of control while the OAA allows for dynamic but not pre-determined flows of control.

We are not claiming that the examples in sections 2.1. and 2.2. are startlingly novel or complex. The point we want to make is that being able to deal cleanly with this kind of thing makes certain demands on the architecture of a spoken language dialogue system. The processing is relatively simple because we have multiple asynchronously acting agents, each of which has independent state and is able to make requests of the other agents. In particular, we have separate agents which contain *dialogue state* and *action state* respectively. Each of these types of state constitutes a context which needs to be maintained, and which is essential to the interpretation of commands.

Although it would be possible to implement similar functionality using a centralized architecture like Communicator, this would be much less straight-forward: in particular, we would have to reify action state as an object which could be passed between the server taking the role of the Action Manager and the Hub. One could do so, but we feel that this is really somewhat beside the point. Our basic argument is that dialogue management is best conceptualized as a distributed and asynchronous process; if we are prepared to grant this, it certainly seems natural to conclude that it will be easiest to represent it in a distributed and asynchronous form.

4. References

J. Aberdeen, S. Bayer, S. Caskey, L. Damianos, A. Goldschen, L. Hirschman, D. Loehr, and H. Trappe. 1999. Implementing practical dialogue systems with the DARPA Communicator architecture. In *Proceedings of the IJCAI'99 Workshop on Knowledge And Reasoning In Practical Dialogue Systems*, pages 81–88.

J. Boye, M. Wirén, M. Rayner, I. Lewis, D. Carter, and R. Becket. 1999. Language-processing strategies for mixed-initiative dialogues. In *Proceedings of the IJCAI'99 Workshop on Knowledge And Reasoning In Practical Dialogue Systems*, pages 17–24.

1999. *DARPA Communicator Web Page.* http://fofoca.mitre.org. As of 14 February 1999.

D. Martin, A. Cheyer, and D. Moran. 1998. Building distributed software systems with the open agent architecture. In *Proceedings of the Third International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, Blackpool, Lancashire, UK.

1999. *Personal Satellite Assistant (PSA) Project.* http://ic.arc.nasa.gov/ic/psa/. As of 14 February 1999.

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

S. Seneff, E. Hurley, C. Pao, P. Schmid, and V. Zue. 1998. Galaxy-II: A reference architecture for conversational system development. In *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Australia.

H. S. Thompson. 1996. Why 'turn-taking' is the wrong way to analyse dialogue: Empirical and theoretical flaws. In *Proceeding of the 1996 International Symposium on Spoken Dialogue.*

Accommodating questions and the nature of QUD

Robin Cooper, Elisabet Engdahl, Staffan Larsson, Stina Ericsson

Department of Linguistics
Box 200
SE 405 30 Göteborg
Sweden
{cooper,engdahl,sl,stinae}@ling.gu.se

Abstract

We look at some real dialogue examples which can be analyzed as question accommodation. We argue that there is a need for a local QUD (representing questions that are currently being addressed in the dialogue) and a global QUD representing (possibly among other things) questions that have been addressed and that are available for reopening.

1. Introduction

In this paper we will look at some real dialogue examples which can be analyzed as question accommodation, that is the dialogue participant adds a question to QUD ("Questions Under Discussion") which has not be explicitly raised. We shall argue that there is a need for a local QUD (representing questions that are currently being addressed in the dialogue) and a global QUD representing (possibly among other things) questions that have been addressed and that are available for reopening.

We shall first present a sketch of our notion of information state. We will then discuss some examples which motivate the distinction between local and global QUD.

2. Information states

We present the basic ideas of question accommodation and some examples that we have presented in previous papers. Our approach is based on a view of dialogue analysis in terms of information states representing information about the dialogue that a dialogue participant has. Accommodation has to do with the nature of information state updates. The notion of information state we have proposed is basically a version of the dialogue game board which has been proposed by Ginzburg (Ginzburg, 1996a; Ginzburg, 1996b; Ginzburg, 1998). In TRINDI deliverables such as (Traum et al., 1999; Engdahl et al., 2000) we represent information states of a dialogue participant as a record of the type shown in figure 2..

The main division in the information state is between information which is private to the agent and that which is shared between the dialogue participants. The private part of the information state contains a PLAN field holding 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. The AGENDA field, 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. We have included a field TMP that mirrors the shared fields. This field keeps track of shared information that has not yet been grounded, i.e. confirmed as having been understood by the other dialogue participant. The SHARED field is divided into three. One subfield is a set of propositions which the agent assumes for the sake of the conversation. The other subfield is for a stack of questions under discussion (QUD). These are questions that have been raised and are currently under discussion in the dialogue. The LU field contains information about the latest utterance.

This is a simple notion of information state which makes no pretensions to completeness. We are using it as a baseline for further extensions in implementations we are developing using the TRINDIKIT dialogue move engine toolbox(Larsson et al., 1999)[1], a toolbox which allows experimentation with different definitions of information state and update rules.

The QUD here is a *local* QUD. That is, it represents questions that have been explicitly introduced, treated as a stack. In our current analyses we rarely, if ever, have more than two questions on this stack and the cases where there are two questions are normally such that an answer to the topmost question on the stack will also be an answer to the question beneath it, e.g. < *Where does A wish to fly to*, *Does A wish to fly?* >. In such a case a single answer will pop both questions off the QUD. In this paper we will argue that there should be a *global* QUD in addition to this and that the two QUDs give rise to different dialogue behaviour. Our argument will be based on the nature of accommodation of questions to QUD.

3. Question accommodation

Dialogue participants can address questions that have not been explicitly raised in the dialogue. In such cases, coherence is preserved if the agent is able to find a question which is relevant at that point in the dialogue which can then be accommodated onto the QUD. Here is an example[2]

[1] See the TRINDIKIT homepage: `www.ling.gu.se/research/projects/trindi/trindikit.html`

[2] from a dialogue collected by the Lund group in the SDS project. We are quoting the transcription made in Göteborg by Jens

$$\left[\begin{array}{lll}
\text{PRIVATE} & : & \left[\begin{array}{lll}
\text{PLAN} & : & \textsc{StackSet(Action)} \\
\text{AGENDA} & : & \textsc{Stack(Action)} \\
\text{BEL} & : & \textsc{Set(Prop)} \\
\text{TMP} & : & \left[\begin{array}{lll}
\text{BEL} & : & \textsc{Set(Prop)} \\
\text{QUD} & : & \textsc{Stack(Question)} \\
\text{LU} & : & \left[\begin{array}{lll}
\text{SPEAKER} & : & \textsc{Participant} \\
\text{MOVES} & : & \textsc{AssocSet(Move,Bool)}
\end{array}\right]
\end{array}\right]
\end{array}\right] \\
\text{SHARED} & : & \left[\begin{array}{lll}
\text{BEL} & : & \textsc{Set(Prop)} \\
\text{QUD} & : & \textsc{Stack(Question)} \\
\text{LU} & : & \left[\begin{array}{lll}
\text{SPEAKER} & : & \textsc{Participant} \\
\text{MOVES} & : & \textsc{AssocSet(Move,Bool)}
\end{array}\right]
\end{array}\right]
\end{array}\right]$$

(1) `$J (travel agent): // vilken månad ska du åka`
`& what month do you want to travel?`
`$P (customer): / ja: typ den: ä:h tredje fjärde april / nån gång där / så billigt som möjligt`
`& yeah, about the, eh, third or fourth of April, around there. As cheap as possible.`

The point here is that customer does more than answer the question that the travel agent has introduced onto QUD, namely *what month does P (the customer) want to travel?*. The additional information *as cheap as possible* is an ellipsis and addresses a different question, something like *How much does P want to pay?*. Note that this question, or something like it, has to be adduced in order to interpret the ellipsis. There is no mention of price in the previous part of the dialogue so this is not a matter of anaphoric interpretation. Rather the communication here is depending on the fact that both the customer and the agent know that price is a relevant question when booking a flight. The strategy we adopt for interpreting elliptical utterances is to think of them as short answers (in the sense of Ginzburg, (Ginzburg, 1996a; Ginzburg, 1996b; Ginzburg, 1998) to questions on QUD. The use of question accommodation enables us to generalize the treatment to cases where the elliptical utterance is not an answer to an overt question.

We have sometimes received the comment that it is unintuitive to consider the offering of additional information as answers to as yet unraised questions. Perhaps it is misleading to use the word "question" rather than "issue" here. The substitution of "issue" for "question" would be quite consistent with the way we are viewing QUD. Perhaps it sounds more intuitive to say that successful dialogues are those in which the dialogue participants can figure out what issues are being addressed rather than what questions are being answered and that a knowledge of what issues are being addressed can be essential for the interpretation of ellipses. Equally it seems more intuitive to say that the customer is raising the issue of price with her statement *as cheap as possible* rather than raising the question. We will nevertheless keep to the original terminology with QUD rather than changing it to the perhaps less happy abbreviation IUD (for "issues under discussion"). QUD is after all something you can chew on ... The substantive point here is that we do not at the moment see any reason to distinguish between issues and questions.

However, we do see reasons to keep track of questions that have been raised and that may be raised again later in the dialogue. Let us consider another similar dialogue.[3] The dialogue begins as in (2).

(2) `$B (travel agent): japp ///`
`& yes?`
`$A (customer): eh // jag undrar om eh // / / en resa till phuket // den tolfte trettonde december //`
`& er, I wonder if..., er, a trip to Pukhet, 12th or 13th December.`

At the start for the conversation the travel agent does not have a plan for the dialogue. She may have several plans in the background for the kinds of tasks that she is capable of such a booking a flight, changing the billing address on a booking, booking hotels, rental cars etc. It is not until the customer makes an initial contribution that the agent can form a plan. The initial contribution by the customer here seems to be fairly typical. It does not say exactly what the customer wants, although we as competent dialogue agents are so good at interpreting dialogues that it is not always obvious that this is the case. But we can test this by embedding the customer's contribution in another constructed dialogue where it gets a different dialogue interpretation and would suggest rather different dialogue plans. Consider (3).

Allwood's group where it is identified as dialogue A820101. We have added some annotations of our own, such as the English translation, and removed some annotations that are not relevant to the point at hand.

[3] From the same collection as the previous one, no. A821702 in the Göteborg transcription.

(3) Husband: What would you like for your birthday, darling?
Wife: er, I wonder if..., er, a trip to Pukhet, 12th or 13th December.

Or (4)

(4) Secretary: There's a line on the travel agent invoice that I can't match to any of your business trips
Executive: er, I wonder if..., er, a trip to Pukhet, 12th or 13th December.

Clearly, the activity (in something like Allwood's sense (Allwood, 1995)) can make a big difference to the way in which we interpret utterances. This is because different issues are being addressed, i.e. different questions are under discussion. The travel agent, therefore, has to accommodate questions onto QUD in order to be able to interpret the customer's contribution. She figures that this is an answer to questions like *Where do you want to go?*, *When do you want to go?* and that these are questions which are associated with the plan she has for booking trips and, on our analysis, she therefore loads this plan into her dialogue information state and thereby prepares to raise other relevant issues to the task. The dialogue continues:

(5) $B (travel agent): ja //
& yes
$A (customer): [1 fyra]1 personer /
& four people
$B (travel agent): [1 är det]1
& is it...
B accommodates: How many people will travel?

The travel agent has to accommodate the question *How many people will travel?* in order to be able to interpret *four people*. She interrupts her own speech in order to acknowledge the information as shown in (6)

(6) $B (travel agent): yes / är det charter du har sett där eller //
& Right... is it a charter you saw there or...
$A (customer): [2 n+ ja:]2
& well... (expressing doubt)
$B (travel agent): [2 eller vad har]2 ni tänkt för någonting /
& ...or what kind of thing have you been thinking of?
$A (customer): nej det det / jag tror att det blir för dyrt med hotell utan eh /
& no, it's...it's, I think it'll be too expensive with a hotel but, er...
B accommodates: Will it be too expensive for a hotel for this trip if it is booked together with the plane?

At this point it seems that the travel agent has to accommodate the question indicated in (6) onto her QUD.

(7) $B (travel agent): < / > / ja ni ska bara ha flyget /
& right, you just want a flight?

The question the customer raised is answered by the travel agent's *right* in (7) and is thus popped off QUD. The agent continues by raising the new closely related question *Does A just want a flight?*. The customer acknolwedges this question and answers it.

(8) $A (customer): ja /
& yes
$B (travel agent): okej / [3 m:]3
& okay, mm,..

At this point there is no question on QUD. But the customer is not entirely satisfied with the exchange and raises the second question again.

(9) $A (customer): [3 ja för]3 / vi / tror inte du också det att det blir väldigt eh höga summor om man ska [4 //]4
& yes, because... we... don't you also think that it would be very, er, high costs if you're...
B has to reaccommodate the question: How expensive will it be for a hotel for this trip if it is booked together with the plane?

There are two points we want to make about this utterance. Firstly, it is necessary to raise the question again. It cannot simply be addressed as a question already on QUD (that is, *local* QUD). Neither the customer nor the agent could answer this question at this point in the conversation by saying *Yes*. But, secondly, the raising of the question has the flavour of a second mention in that it is more reduced than would have been possible with the original raising. If we move the customer's utterance in (9) back in the conversation and substitute it for her last utterance in (6), where she first raises the question, it would not be successful. It would be incoherent, even if we remove the tell-tale *also*.

We shall call her utterance in (9) a *reraising* of the question. Note that such a reraising can have limits on how reduced it can be. It would not have worked here if the customer had only said: *Don't you also think (so)....* There is not enough information in this utterance to identify the intended question among the other questions that have been under discussion. We therefore conclude that, in addition to local QUD, a global QUD is needed which, possibly in addition to other things, keeps track of questions that have been under discussion. Reraisings have certain properties in common with definite descriptions in that they must provide at least enough information to uniquely identify the intended question among the restricted set of those on the global QUD.

4. Comparison with van Kuppevelt's approach

(van Kuppevelt, 1995) presents a theory which uses *topicality* as the main organising principle of discourse structure, and according to which topic is a context-dependent, question-based and dynamic notion. Topic and comment are characterised in the following way:

(10) *Topic:* A discourse unit U — a sentence or a larger part of a discourse — has the property of being, in some sense, directed at a selected set of discourse entities (a set of persons, objects, places, times, reasons, consequences, actions, events or some other set). This selected set of entities in focus of attention is what U is about and is called the topic of U

Comment: That which is newly asserted of the topic of U

van Kuppevelt assumes a close relationship between topic and comment, on the one hand, and (explicit and implicit) questions on the other. The topic part of a sentence is related to a question, whereas the comment part contains the answer. This is captured by van Kuppevelt's basic assumption as follows:

(11) *Basic assumption:* Every contextually induced explicit or implicit (sub)question Q_p that is answered in discourse constitutes a (sub)topic T_p. T_p is that which is being questioned; a set of discourse entities from which one is selected as an answer to Q_p. Comment C_p is provided by this answer and names or specifies the entity asked for

The inclusion of both explicit and implicit questions means that no actual question need be present in the dialogue or text, and that both situations — question present and question not present — can be treated on a par. An implicit question is assumed to be one "which the speaker anticipates will arise in the listener's mind on interpreting preceding utterance". The term *contextual induction* is used to indicate that a certain unit of discourse brings about a new (sub)question, and hence a new (sub)topic. A unit of discourse having this function is called a *feeder*, and is often topicless. In the following example, A's first utterance, which is the opening sentence of the dialogue, acts as a feeder since it causes, or contextually induces, B's (explicit) question:

(12) *Feeder:* A: Yesterday evening a bomb exploded near the Houses of Parliament

Question: B: Who claimed the attack?

Answer: A well-known foreign pressure group which changed its tactics claimed the attack

In a discourse, a topic-constituting question is then simply an explicit or implicit question raised as the direct result of a feeder.

A subtopic-constituting subquestion is hierarchically subordinate to some other question, and is essentially seen as the result of an unsatisfactory answer to that question or some other intermediate subquestion. van Kuppevelt formulates two principles which govern the behaviour of subquestions. The first is the *Principle of Recency*, which determines the order of subquestions in a given discourse. If a hierarchy of a discourse topics is seen as a tree with the initial feeder as the root node, the Principle of Recency seems to imply a depth-first strategy of topic development.

The second principle is the *Dynamic Principle of Topic Termination* which concerns the life-span of subtopics and topics. van Kuppevelt notes that even if a new (sub)topic is introduced, the old (sub)topic does not necessarily lose its actuality (the notion of topic discussed here is not limited to sentences but extends to discourse level). How to decide when a topic is terminated? The principle is formulated as follows:

(13) *Dynamic Principle of Topic Termination:* If an explicit or implicit (sub)question Q_p is answered satisfactorily, the questioning process associated with it comes to an end. As a consequence, topic T_p loses its actuality in discourse

This principle implies that as long as one of the conversation participants, or the writer/speaker in the case of a written or spoken monologue, is not satisfied with a certain answer (as indicated by more questions or answers to implicit questions), the topic is still valid.

The termination of a (sub)topic results either in the continuation of a, non-terminated, topic higher up in the topic hierarchy, or in a new feeder. According to van Kuppevelt's theory, then, it is not possible to return to some subtopic after it has been terminated and the participants have engaged in a new, non-subordinated, topic. If such a return to a terminated subtopic, say Q_k, nevertheless does occur, this can only be if the preceding answer, $A_k - 1$, functioned as a new feeder. In principle, then, every answer in a discourse can function as a feeder. Whether a given answer, not immediately followed by an explicit or implicit question, actually is to be seen as a feeder presumably cannot be determined until the end of the entire discourse, since feederhood is only determined at the moment an answer/feeder is followed by a question (otherwise the Principle of Recency would be violated). van Kuppevelt also formulates a Subordination Test for determining topic hierarchies. Put simply, Q_q is a subquestion of a preceding question Q_p if it is inappropriate to utter it if Q_p has been closed. Otherwise, if Q_q is appropriately uttered in the same circumstances, Q_q is a topic-constituting question. Question closure is tested by the insertion a sentence like *I now understand* Q_p.

As far as we can see, van Kuppevelt's approach puts a limit on which questions can be reraised in terms of topic hierarchies which we do not have in our theory. Instead we exploit the interaction between QUD, agenda and plan. What we have tried to emphasize in this paper is that there is a difference in nature between raisings and reraisings of questions. A potential advantage of our theory from the computational point of view is that it does not rely on the notion of feeder. It seems that feeders can best be recognized by having analyzed a contribution as an answer to a question and then reinterpreting the previous utterance as the feeder that introduced it. But once you have recognized something as an answer to a question you do not actually need to go back and find the feeder. Alternatively one could interpret everything as a potential feeder for a large number of questions. But this is computationally undesirable because it would mean that you would have to do a lot of computation of potential questions that actually never get taken up. This was a computational problem which we perceived with Ginzburg's original notion of QUD and which we have tried to avoid in our work, while maintaining the central ideas.

5. References

Jens Allwood. 1995. An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.

Elisabet Engdahl, Staffan Larsson, and Stina Ericsson. 2000. Focus-ground articulation and parallelism in a dynamic model of dialogue. Technical Report Deliverable D4.1, Trindi.

J. Ginzburg. 1996a. Dynamics and the semantics of dialogue. In Jerry Seligman and Dag Westerståhl, editors, *Logic, Language and Computation, Vol. 1*, volume 1. CSLI Publications.

J. Ginzburg. 1996b. Interrogatives: Questions, facts and dialogue. In *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.

J. Ginzburg. 1998. Clarifying utterances. In J. Hulstijn and A. Niholt, editors, *Proc. of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*, pages 11–30, Enschede. Universiteit Twente, Faculteit Informatica.

S. Larsson, P. Bohlin, J. Bos, and D. Traum. 1999. Trindikit 1.0 manual. deliverable D2.2, TRINDI.

David Traum, Johan Bos, Robin Cooper, Staffan Larsson, Ian Lewin, Colin Matheson, and Massimo Poesio. 1999. A model of dialogue moves and information state revision. Technical Report Deliverable D2.1, Trindi.

Jan van Kuppevelt. 1995. Discourse structure, topicality and questioning. *Journal of Linguistics*, 31:109–147.

When is a union really an intersection? Problems interpreting reference to locations in a dialogue system

Myroslava O. Dzikovska and Donna K. Byron

Department of Computer Science
University of Rochester
Rochester NY 14627, U.S.A.
myros/dbyron@cs.rochester.edu

Abstract

This paper describes issues that arose in our implementation of an interpreter for locative expressions in a spoken dialogue system. The expressions involve complex adverbial modification, include imprecise and innacurate content, and reflect conventional practices specific to the way English speakers refer to roads and highways. Our system utilizes semantic features in the parser to perform disambiguation, and domain-specific reasoning to resolve variables in the logical form to the correct referent. The functionality implemented to date was evaluated against a small corpus of naturally-produced expressions, of which 46% were correctly interpreted.

1. Introduction

This paper describes issues that arose in our implementation of an interpreter for locative expressions in the TRIPS-911 spoken dialogue system (Ferguson and Allen, 1998). In this system, the user is faced with a city map (Figure 1) and he must act as the dispatcher in an emergency-response center. Unlike our previous domain, in which the user could only refer to named map locations such as cities, the detailed nature of the new map necessitates that locations be described rather than named. Map objects described during a typical session include intersections (e.g. "where Main and Oak cross"), road segments (e.g. "all of Monroe between Main and Oak"), and regions of town (e.g. "in the north of the city"). Even objects with names, such as hospitals, may be described rather than named by users that are unfamiliar with the map.

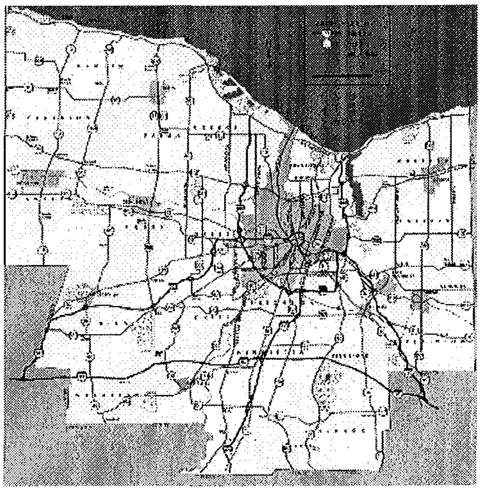


Figure 1: Map used in the TRIPS-911 system

In building an interpretation component for these expressions, our goal is to support natural language use in which many different expressions can describe the same object, and in which descriptive terms can be embedded to an arbitrary degree. To interpret these expressions requires first generating a logical form that accurately represents the semantics of the sentence, then using that logical form to constrain the search for the correct referent(s) of the expression.

Many reference resolution studies start from the logical forms, without considering the problem of relating them to natural language utterances. In building a practical dialogue system, these processes need to work together in order to quickly and accurately interpret the descriptions. This paper describes the entire process from natural language utterance to logical form to reference resolution. We first discuss some of the problems that exist in trying to interpret spoken natural language descriptions of locations; after discussing the relevant background, we then describe the interpretation process for locatives in our system, and evaluate the portion of the design that has been implemented to date.

2. Motivation

In preparing to develop a system to converse in this new domain, we collected a set of human-human problem solving dialogues involving emergency response tasks such as plowing roads, dispatching medical personnel, and repairing downed electrical lines (Stent, 2000). One of the most interesting aspects of these dialogues is the large variety of expressions that were used to refer to map objects. We were impressed

1) "the intersection of three eighty three and two fifty two A just below the airport"
2) "where Genesee, Brooks Road and three eighty three connect"
3) "at route thirty one at three ninety"
4) "the corner of Main street and East avenue downtown"
5) "at route two fifty two and the river"

Table 1: A variety of referring expression forms indicating junctions

1) "directly east of the inner loop a little bit"
2) "quite in the north"
3) "over near Gates"
4) "at three eighty three just past two fifty two A"
5) "the bridge near Gates where four ninety crosses the river there"

Table 2: Expressions involving imprecise descriptions and multiple constraints

by the richness of expressions we found in this corpus compared to the previous domain, which used an extremely simple (and fictional) map of a small island. A cursory examination of the corpus showed that complex locative expressions occur very frequently in this domain, and that special reasoning needed to be developed in the system to cover them. This section presents a few selected example phrases from our corpus to show the difficulty of developing automated methods to interpret these expressions.

References to streets, highways and junctions where streets cross can take many forms, some of which are shown in Table 1. These sentences demonstrate many of the problems we must tackle in converting all these different surface forms into the same logical form. From these examples, we can see that junctions can be indicated explicitly with nouns like "intersection" or propositions like "cross" and "connect", and implicitly, using a conjunction in the context requiring a location, such as the argument of the preposition "at". Junctions can involve just two or sometimes more than two streets, or even a street and another ribbon-shaped object such as a river.

Streets are often referred to by a proper name, such as "Main Street", so our intuition would be that streets can be processed simply as proper names, but this is too simplistic for a real application. Towns tend to include several streets of the same name, for example Genesee Avenue, Genesee Court, and Genesee Lane might all exist on the same map. When the speaker of sentence 2 says simply "Genesee", we have to determine which street he means, either by applying constraints from the remainder of the description (eg. which street named Genesee intersects with Brooks) or by applying a heuristic (prefer main roads over side streets). Also, in sentence 2, Brooks Avenue is incorrectly referred to as Brooks Road. We would like the system to be able to resolve this reference nonetheless.

Reference to highways in this domain have an additional set of problems because highways in the United States are typically given multi-digit numerical indicators. This introduces additional ambiguity, because depending on the context the numeral 104 (pronounced 'One oh four") can be interpreted as the highway 104, a time-point 1:04, maybe even an office number or some other kind of entity. Selecting the proper interpretation presents a challenge for a spoken dialog system.

Expressions that describe the location of an object on the map can be arbitrarily complex and often use imprecise operators. Converting these operators into well-defined functions in the interpretation process is not straightforward. Table 2 shows several such phrases from our corpus using the operators "east", "north", "near" and "past". These operators can be further modified by vague adverbials such as "just", "slightly" or "directly", which further limit the space from which the referents can be selected.

In the following sections we present some of the approaches we used to deal with the kind of descriptions presented above.

3. Background and Related Work

Much of the previous work in linguistics and psycholinguistics on human production and comprehension of location descriptions have focused on generating proper scene descriptions (cf. (Retz-Schmidt, 1988)). We found little of that work to be helpful in implementing an actual interpretation mechanism. Mainly this is because our system does not resolve spatially-located reference (objects located in a 3-D space and requiring the calculation of reference points, relative scale, etc.). Our requirements are simply to process locative expressions in a 2-D space from a fixed perspective. Also, many studies start from a disambiguated logical form rather than from natural language utterances, and do not relate their semantic theories with the needs of a parser.

Based on the study of spatial prepositions in different languages, Talmy (1983) identifies a set of possible spatial idealization schemas and properties that we associate with objects, such as idealization to a point, line or strip. He also enumerates the primary relations that can connect the objects in a scene and the possible restrictions on their arguments. Herskovits (1986) argues that simple relations are insufficient to explain the way locatives are used. She uses a similar set of schemas and properties and develops a formal geometric scene representation. In her theory, every preposition has an "ideal meaning" which is then transformed into actual meaning with the aid of pragmatic principles such as relevance, salience, topicality and tolerance. In a sense, our approach is similar, because the parser generates underspecified predicates in the surface form that can be seen as idealized meanings, and then the reference agent computes their actual meaning based on pragmatic considerations. However, her principles as described in the book are rather too vague to treat computationally.

Creary et al. (1989) present a logic of location predicates suitable to use in reference resolution. This work represents locative constraints as expressions over regions composed using intersection and inclusion operations. Their representation is computationally efficient and deals with scope ambiguity, permutability and ommisibility of locatives. We use a similar idea in our reference resolution algorithm. However, Creary assumes a fully disambiguated logical form, and therefore does not specify how to convert utterances into logical forms. In our case, the parser and reference resolution need to do extra work disambiguating the surface predicates.

4. The interpretation process in the TRIPS system

The TRIPS parser uses a chart-based best-first parser loosely based on the HPSG grammar formalism and described in (Allen, 1995). The parser receives a string of words from speech recognition or the keyboard and then obtains a syntactic analysis and a surface logical form. The final logical form includes tokens representing the binding for each referring expression. The reference agent then returns the object it believes the logical form refers to, along with the score indicating its confidence in the resolution result. In this section, we first describe the process by which the parser converts each referring expression into a logical form. Then we discuss how the reference resolution agent (RA) uses this logical form to resolve the referring expression to the correct entities.

4.1. Generating a Logical Form

Since the TRIPS parser is connected to an interactive system, speed and accuracy is of the essence. For the sake of efficiency, we would like to have as much disambiguation as possible to be done on early stages of interpretation, preferably during parsing, to eliminate (slow) reasoning about implausible analyses. On the other hand, we would like to keep the system portable to other domains, and therefore need to have a grammar and lexicon that are mostly domain-independent. As a way to balance these criteria, our system is heavily dependent on semantic selectional restrictions[1] that work to keep the parsing complexity down and to help disambiguate syntactic structure. An alternative to this solution would be implementing a statistical method to select among possible interpretations. However, statistical methods require large amounts of text for training, and this is not available in our domain. Moreover, the constructs we encountered in our corpus are not often found in more formal sources such as Wall Street Journal, precluding their use for purposes of training at this time.

Therefore, in designing the lexical semantic representation we have to worry not only about selecting a semantic representation that would accurately express the meaning of a word, but also about formulating the selectional restrictions that are useful for disambiguation. From the point of view of the system development, we found that separating those two issues to some extent helps to make lexicon maintenance and development easier. In order to do that, each word in the lexicon is characterized with a predicate, which corresponds to the (deep) meaning of the word and can be mapped to the corresponding entity in the domain knowledge representation, and a set of semantic features that express some basic lexical semantic properties of the word meaning that are used in formulating selectional restrictions and disambiguation.

Our initial feature set included most of the EuroWordNet top hierarchy features (Vossen, 1997), with the value sets somewhat modified to suit our needs. However, we discovered that these features were inadequate to provide reasonable selectional re-

[1] While selectional restrictions have a variety of problems that make their use in a general case impossible, we believe that they are a useful mechanism to control parser complexity in a particular domain.

```
at preposition
  LF AT-LOC ;; the meaning predicate to be used in the logical form
    SUBCATSEM (spatial-abstraction point) ;; the semantic restriction on the subcategorized NP
    ARGSEM    (spatial-abstraction point) ;; the semantic restriction on the object modified by the PP
  LF AT-TIME
    SUBCATSEM (function time-object)
    ARGSEM    (Aspect Bounded)

bridge noun
  LF BRIDGE
    SEM ;; the semantic features associated with the word
      (spatial-abstraction (OR point line)) (origin artifact)
      (form geographical-object) (function location)

Brooks name
  LF ROAD ;; for names LF carries the type of object to look for
  NAME BROOKS-AVENUE ;; the constant corresponding to the object
    SEM
      (spatial-abstraction (OR point line strip)) (origin artifact)
      (form geographical-object) (function location)
```

Figure 2: Parts of lexical semantic representation used to disambiguate "the bridge at Brooks".

strictions on locative predicates, so we augmented the set with *spatial-abstraction* and *form* features inspired by the work of Talmy and Herskovits. The selection of features is based on the idea that people abstract the actual shapes of objects to a (small) set of abstract geometric shapes, and these abstractions restrict what locatives can be used in connection with the scene.

Consider the sentence "Go to the bridge at Brooks". Among other things, in interpreting that sentence we need to decide that *at* in this context has locative meaning and that the prepositional phrase "at Brooks" modifies the noun phrase "the bridge." [2]

Figure 2 contains partial definitions for the words *bridge*, *at* and *Elmwood.*

The definition of the name Brooks states that it can be visualized as a point (in our system, all geographical objects can). This helps the parser to distinguish the AT-LOC sense of *at* from other possibilities, e.g. AT-TIME[3]. Moreover, the definition of AT specifies that it can only modify the objects that themselves can be visualized as points, and the definition of the *bridge* satisfies this condition. Thus, the parser sends the following objects to the RA to resolve the locative expression "the bridge at Brooks":

```
x: (AND(TYPE x ROAD)
        (NAMEOF x BROOKS-AVENUE))
y: (AND(TYPE y BRIDGE)(AT-LOC y x))
```

There is still a problem with defining selectional restrictions in terms of features in our system. Obviously, the granularity for spatial reference varies with the task. Depending on the scale, any physical object can be assigned almost any *spatial-abstraction* value. Consider words like *intersection* or *truck.* In a different domain, it is possible to imagine them visualized as areas in space, with something located ACROSS or OVER them. However, at the scale accepted in our domain, intersections and trucks are points, and such expressions do not often occur. If the selectional restrictions are loose enough to allow those other phrasings in our domain, the ambiguity of spatial expressions increases noticeably, requiring additional reasoning to disambiguate the obtained logical forms. Therefore, we selected a specific scale suitable for the domain, even though this excludes some of the interpretations allowable in other domains. Not surprisingly, this rigid restriction sometimes results in excluding expressions that should be acceptable in our task. The

[2] It is also possible for the "at" to modify the verb "go". In our domain this rarely happens and the lexical entry for "go" is set up so that this interpretation is excluded for the reasons similar to those discussed below

[3] our representation is set up so that the features corresponding to physical objects, such as *(spatial-abstraction point)* and features corresponding to abstract entities such as *(function time-object)* are mutually exclusive, which allows us to make this inference on the basis of information highlighted in the picture

solution could be to introduce softer selectional preferences, which is planned as a part of our future work in the system.

4.1.1. Pragmatic considerations

One can note that the feature-based representation is not suitable for fully representing spatial properties of an object (cf. (Jackendoff, 1983)). In fact, it is not the intended use of features in our system. A small set of features is obviously not sufficient to express all possible distinctions needed to obtain a fully disambiguated logical form. However, in the process of developing the TRIPS system we found that using a general knowledge representation that would allow us to obtain a completely disambiguated logical form before giving it to the reference resolution agent was costly computationally, and was making our lexicon too difficult to maintain. We use the semantic features as a representation that provides the information to cut down the number of possible sentence interpretations, and some basic properties often used by reasoning components, but that does not attempt to encode all the distinctions needed for full disambiguation.

For example, the preposition *from* is often ambiguous. For the phrase "an ambulance from Pittsford", the features of the word *Pittsford* are sufficient to eliminate the FROM-TIME sense of *from*. On the other hand, this phrase can mean an ambulance currently located in Pittsford, or an ambulance based in Pittsford, and the distinction is determined mostly by context and not by the form of the utterance. Therefore, it cannot be expressed with the kind of selectional restrictions used by the parser. For these cases, the parser outputs the predicate FROM, and the RA must decide on whether the correct interpretation is ORIGIN or AT-LOC, the domain predicates encoded in the general knowledge base.

Another case that requires special reasoning to be implemented in resolution is a practice of referring to highways by their numbers mentioned earlier. Selectional restrictions may be able to eliminate some of the ambiguity, for example, "the bridge on 104" definitely refers to a location. However, this is often not possible, for example, when "one oh four" is uttered as a short answer to the previous question. In this case, our parser will not select a type for numeric expressions but will leave that disambiguation up to the RA, which can use the context to determine the correct interpretation.

Yet another complication resulted from our assumption that street designations could be taken as proper names[4]. As described in the motivation section, users can often use abbreviated names that are ambiguous between a number of streets in the same town, or make mistakes in street names. Therefore, we treat street designations more like definite descriptions than proper names. Lexical entries for streets contain the full name, for example BROOKS-AVENUE. If the speaker uses the entire correct name, the parser interprets it as a name. If the user says "Brooks Road" but only "Brooks Avenue" is in the lexicon, the parser will generate a representation (AND (TYPE x ROAD)(ASSOC-WITH x BROOKS-AVENUE)) (where the predicate ASSOC-WITH denotes some association between objects) and let the RA decide how the objects are associated. The RA includes the heuristic that in our domain if two locations are associated, then they share at least some space, and this allows it to make the inference that the road in question is indeed Brooks Avenue. A more difficult case is when objects of different types share a name, (e.g. a street and a body of water named "Ontario") selectional restrictions may filter out some possibilities. But if more than one interpretation is left and the tie cannot be resolved by the weights associated with grammar rules, the system may simply arrive at an incorrect interpretation. Currently the problem is partially solved by trying the two most likely parses and using interpretation confidence scores to determine which object is a better referent. We plan to modify the system in the future so that the parser outputs a more general type for these cases, such as (TYPE x LOCATION), allowing the RA to reason out the correct referent.

4.2. Interpreting the logical form

Once a logical form representing the referring expression is selected, it is passed to the RA to be resolved. The logical form includes the semantic type of the variable and the description constraints needed to select the referent from among the objects of that type. The RA has access to a geographical database containing all objects on the map, and each referring expression is resolved to one or more map objects. Simple bindings can be resolved by table lookups, e.g. (AND (TYPE w CITY) (NAME-OF w GATES)) is resolved by finding an object of type CITY and named GATES in the map database.

The RA contains domain-independent reasoning to correctly interpret different referring expression forms (eg. definite versus indefinite phrases), and has access to a domain-specific reasoner. This domain-specific reasoner is employed to determine which map objects satisfy the predicates given in the logical form, a pro-

[4] Creary *et al* claim that the analysis of streets as proper names is uncontroversial.

cess that varies based on the task to which the system is currently being applied. For the "ambulance from Pittsford" example presented above, it will try to locate both the ambulances currently in Pittsford and those originating from Pittsford, and if more than one entity is returned, will sort the results according to a heuristic taking into account salience and contextual factors.

Additionally, the domain-specific reasoner is responsible for translating the more domain-independent representation produced by the parser into the representation that matches the way objects are stored in a geographical database. For example, in this domain the predicate MIDDLE (e.g."in the middle of Elmira") actually means (NEAR (CENTER ELMIRA)), not the epicenter of Elmira, and the precision with which objects can be considered in the middle of another object varies with the domain. The domain-specific reasoner available to the RA translates MIDDLE into its domain-specific use. Other operators may need to be coerced based on the internal representation used in the map. The expression "the end of highway five ninety" produces a logical form (END x highway-590). In our system, the database contains information about the segments that together constitute a road, but there are no facts in the database in the form (END x y). Therefore the domain reasoner must know that to find the END of a road, you must find a point on the highway that is the end of one road segment and that is not the beginning of another segment.

Many locative predicates are interpreted as regions. There are standard techniques to judge whether a point belongs to a region (cf. Gapp (1994)). See Figures 3 and 4 for a simple example of region definitions that represent several variants of the predicate NORTH-OF defined in the geographical database, depending on the shape of the reference object. As the first drawing shows, the predicate NORTH-OF is not interpreted strictly as describing a line running due north from the reference object. The region is relaxed to within plus or minus a few degrees, and objects in this region are considered to satisfy NORTH-OF. The predicate (NORTH-OF x POLICE-STATION1) will be interpreted as the set of objects that fall within this region using POLICE-STATION1 as the reference object. Some modifiers impact the shape of this region, for example (DIRECTLY NORTH) narrows the width of the region, and WAY requires matching objects to be in the upper portion of the region. When the reference object is not a point, the NORTH-OF defines a region along the northern boundary of the object.

Vague modifiers such as "just", "slightly" or "quite" are applied to objects in the list when assigning the RA's confidence in the object as the correct referent. The confidence scores of objects that satisfy the description ((SLIGHTLY NORTH-OF) x GATES) would be assigned so that items close to Gates have a higher confidence, while the objects satisfying ((WAY NORTH-OF) x GATES) would be sorted so that objects further away have the highest confidence.

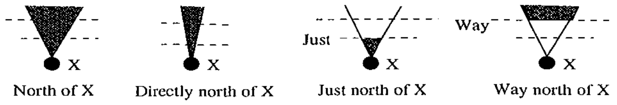


Figure 3: Definition of NORTH if X is a point

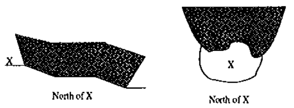


Figure 4: Definition of NORTH if X is a line or a region

4.2.1. Combining the constraints

Earlier we presented some examples where locative descriptions can contain an arbitrary number of constraints. The RA processes these constraints in bottom-up order based on the embedding, which typically corresponds to right-to-left order in the surface expression. Table 3 contains a sample resolution for "The bridge near Gates at four ninety and the river."

A wide variety of expressions can be used to talk about intersections (see Table 1 above). Many of these constructions can be fully interpreted by the parser, in which case it sends the RA a fully specified intersection description of the form (AND (TYPE z JUNCTION) (JUNCTION z y)), where y corresponds to a set of objects forming the junction. In other cases, additional reasoning is required. For example, in our current implementation, it is difficult for the parser to decide whether the word "and" creates the set with a union of two streets, like "Go down Monroe and Elmwood", or refers to their common point, as in "it is located at Monroe and Elmwood". Therefore, the fragment "STREET1 and STREET2" is initially interpreted as a type UNION, corresponding to a set, but the RA knows that in this domain, a union of two ribbonal objects corresponds sometimes to the point at their intersection. As long as the parser can identify the entity as some type of LOCATION rather than a ROUTE or true set (as in the last line of Table 3), a junction object will result. In cases where the junction is described by a proposition (e.g. "where Genesee

Request	Action
Resolve x: (TYPE x RIVER)	Returns objects of type RIVER
Resolve y: (AND (TYPE y ROAD) (NAME-OF y 490))	Returns one item, the road named 490
Resolve z: (UNION z (x y))	The RA makes a set containing street1 and street2
Resolve w: (AND (TYPE w CITY) (NAME-OF w GATES))	Returns one object, the city Gates
Resolve v: (AND (TYPE v BRIDGE) (NEAR v w) (AT-LOC v z))	Returns all bridges matching this form. Now that *z* is described as the location of *v*, we know it is a JUNCTION. At this outermost level is an error signaled if the result is still a set.

Table 3: Sample Resolution Process for the phrase:"The bridge near Gates at four ninety and the river".

and Elmwood connect") the agent would need to have a set of rules that locate the intersection based on the set of objects mentioned in the description.

5. Evaluation

For our evaluation, we examined 3 dialogues from the Monroe corpus(Stent, 2000), and isolated out all phrases that contained location descriptions. The dialogues included 559 utterances and the resulting list of location descriptions contained 196 phrases (including some duplication). We tested the portion of our interpretation process implemented to date on these expressions. In this evaluation, the expressions were treated (and evaluated) as stand-alone expressions rather than as connected text. Expressions that were anaphoric in the original dialog were judged as correct if the RA bound the variables to map objects matching the descriptive content of the expression. The current implementation includes parser coverage for numeric highway names and underspecified street names (such as "Genessee" for Genesee Avenue). The domain-specific reasoner for reference resolution at this time resolves logical forms for intersections, the end of roads and highways, map directions NORTH/ABOVE, SOUTH/BELOW/UNDER, EAST, and WEST, and FROM.

Table 4 contains the results of this evaluation. The parser found a syntactic analysis and produced a logical form suitable for further interpretation in about 2/3 of the cases. Out of these, we found the correct referents for 92 expressions. Expressions that could not be resolved typically involved unimplemented predicates such as NEAR and various uses of the ASSOC-WITH predicate. Not surprisingly, we are doing better on short phrases that contain fewer modifiers. At the same time, a reference resolution algorithm in the old TRIPS system that could only look up the named locations that are defined in the lexicon could resolve only 47 of those expressions, so adding the special processing for locatives is helpful in our system.

	Number of phrases	Words per phrase
Total	196	5.90
Correctly parsed	131	5.07
Acceptable partial parse	6	10.17
Incorrectly disambiguated	13	8.31
No parse found	46	7.02
Found correct referents	92	4.59

Table 4: Evaluation Results

6. Future Directions

Currently, we have only implemented a small number of predicates. Adding more functionality to the RA, including implementing confidence sorting based on the linguistic hedges is will be accomplished in the near future.

An important problem to be solved is that many locative descriptions produced by people are redundant, and some constraints can be dropped without impeding understanding. Such locative expressions are intended to give a human addressee a visual clue to locate the referent on the map, but are not strictly necessary for the system to understand the reference. These expressions should be applied as soft constraints on the resolution process; handling them properly requires discourse-level processing. For example, there is only one Beahan Road on our map, but the descriptions like "Beahan Road just below the airport" show up quite often in the corpus. Since the name "Beahan Road" can be resolved uniquely, the system could just stop there without checking the rest of the description. A better strategy would be to try to make sure that the object that was found satisfies the rest of the constraints. If this is not the case, this may indicate that the user is confused or simply that the system's idea of 'just below the airport' differs from the user's. In that case, a clarification a move could be generated, something like "Is this the road you mean?" (the road blinks). We could not implement this strategy due to deficiencies in the system dialogue manager, but we are currently in the process of implementing a new system architecture (Allen et al., 2000) that would make such processing possible.

Another problem results from the fact that objects on the map being used in the current task cause nearby objects to become salient, and our current model of context management does not account for this. For example, in some of the tasks presented to the users in the Monroe dialogues, they had to evacuate the people from the North-West corner of the city. As soon as this was established, often just by locating an object in that area, people would start using expressions like "go to the end of 390". This expression is ambiguous, because the road has two ends, and neither has been previously mentioned. However, in those cases it appears that the competitors outside of the region at which the attention is centered are not salient and therefore these references are understood as unambiguous. Therefore, a reference resolution algorithm in this domain should not only take into account recently mentioned objects, but also incorporate information from other sources, for example, the regions where the user tends to look or regions adjacent to the objects on the map that have been recently brought to the user's attention in the process of creating a plan. Extending the definition of salience in this manner would help us to resolve the anaphoric expressions similar to those mentioned above, and is planned as an important part of our future work.

7. References

J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. 2000. An architecture for a generic dialogue shell. *To appear in the Natural Language Engineering Journal special issue on Best Practices in Spoken Language Dialogue Systems Engineering.*

J. Allen. 1995. *Natural Language Understanding, 2nd edition.* Benjamin/Cummings, Menlo Park.

L. Creary, J. Gawron, and J. Nerbonne. 1989. Refence to locations. In *Proceedings of ACL '89*, pages 42–50.

G. Ferguson and J. Allen. 1998. Trips: An intelligent integrated problem-solving assistant. In *Proceedings of AAAI '98.*

K. P. Gapp. 1994. Basic meanings of spatial relations: Computation and evaluation in 3d space. In *Proceedings of AAAI '94*, pages 1393–1398.

A. Herskovits. 1986. *Language and Spatial Cognition: an Interdisciplinary Study of the Prepositions in English.*

Ray Jackendoff. 1983. *Semantics and Cognition.* Number 8 in Current Studies in Linguistics. MIT Press, Cambridge, MA.

Gudula Retz-Schmidt. 1988. Various views on spatial prepositions. *AI Magazine, Summer 1988*, pages 95–105.

Amanda Stent. 2000. The monroe corpus. Technical Report 728, University of Rochester Computer Science Department.

L. Talmy. 1983. How language structures space. In *Spatial Orientation: Theory, Research, and Application.*

P. Vossen. 1997. Eurowordnet: a multilingual database for information retrieval. In *Proceedings of the Delos workshop on Cross-language Informatio n Retrieval*, March.

Overlaps and interruptions: Towards a hearer's model of turn taking

Mika Enomoto*, **Syun Tutiya**†

*Graduate School of Science and Technology,Chiba University
1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 163-8522, Japan
enomoto@icsd4.tj.chiba-u.ac.jp
† Faculty of Letters, Chiba University
1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 163-8522, Japan
tutiya@chiba-u.ac.jp

Abstract

The model we proposed based on the analyses of the functionalities of the overlapping utterances has focused on the next speaker, or the hearer who starts the overlapping utterance . There remained room for further investigation as to how the overlapped utterances interact with the overlapping utter ances. In the research reported here, Three types of overla pping phenomena, viz., invited interruption, strategic interruption, and collaborative interruption are related to the functionalities of the overlapped utterances. 36 students have been asked to classify the functionalities of the overlapped utterances in 4 dialoges in Japanese Map Task Corpus, the coincidence rate of 82.6%. The classification revealed that the funcitionalities of overlapping utterances are constrained by those of the overlapped utterances. We have proposed a revised model of the interruption which reflects the correlation between the overlapped and overlapping utterances.

1. Introduction

The goal of this study is to propose a "hearer's model" in conversation, in such a way that it takes into consideration the notion of "hearer" in order to account more naturally for the conversational phenomenon called "overlap," namely that of one speaker starting his/her utterance before the other speaker ends his/her utterance. Overlaps do not necessarily imply interruptions. One speaker can start talking while the other is speaking, without interrupting the interlocuter, who may continue on speaking or stop talking yielding the turn to the one who interrupts. Or one may acknowledge or backchannel while the other is talking, in which case the resultant overlaps are clearly meant not to interrupt the interlocuter's utterance. All that suggests a need for the analysis and explanation of the overlapping phenomena in terms of the cognitive model of the participants of dialog[1]. We start from the model which has been widely accepted: i.e., the model consisting of turn allocation rules as stipulated in Sacks, Shegloff & Jefferson(1974)[2]. The turn allocation rules are to the effect that the current speaker(C) selects who to be the next speaker(N)[3]. We call this the CsN model. The intuition behind this model was based on the analysis of the dialogs where the speakers orderly take turns, and it was empirically supported by observations like Ervin-Tripp's(Ervin-Tripp, 1979), viz., that less than 5% of the speech is "overlapping/overlapped."

The CsN model entails an apparently feasible principle: namely only one speaker speaks at a time in conversation. This "one-at-a-time" principle is a result of accepting the assumption that, in "normal" conversations, turns are taken in an orderly manner. Given this

[1] The same and similar kinds of phenomena have been dubbed not only overlaps, but "simultaneous speech," "cospeech," etc. Authors have distinguished diffenent kinds of exemptions from different points of view and , though in this paper we begin with the brute fact of two participants of a dialog speak at the same time

[2] The rules remain virtually the same since the time of publication. The nature of rules is an arguable issue. One way of interpreting the rules, which most of the serious researchers subscribe to is that they are a set of normative rules in the sense that the participants of conversation are not cognitively but only by perceived abberations, made aware of their presence.

[3] The rules are
(1)For any turn, at the initial transition-relevance place of an initial turn-constructional unit:
(a)If the turn-so-far is so constructed as to involve the use of a 'current speaker selects next' technique, then the party so selected has the right and is obliged to take next turn to speak; no others have such rights or obligations, and transfer occurs at that place.
(b)If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then self-selection for next speakrship may, but need not, be instituted; first starter acquires rights to a turn, and transfer occurs at that place.
(c)If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then current speaker may, but need not continue, unless another self-selects.
(2)If, at the initial transition-relevance place of an initial turn-constructional unit, neither 1a nor 1b has operated, and, following the provision of 1c, current speaker has continued, tehn teh rule-set a-c re-applies at the next transition-relevance place, and recursively at each next transition-relevance place, until transfer is effected. (Sacks et al., 1974),

principle, overlaps are generally considered to be deviations from the norm the occurences of which require recovery to the orderly course of turn-taking. Although the overlaps take place often enough to be noticed to be significant phenomena that deserve scrutiny[4], they have hardly been considered as among the normally permissible forms of conversational interchange.

Now in the Japanese Map Task dialog corpus, which has 128 map task dialogs recorded for about 22 hours, we observe that nearly 45 % of the 53,111 utterances are either overlapping or overlapped, or both. Furthermore, Clark(Clark, 1994) points out that the turn allocation rules failed to account for a number of strategies that are common in conversations. It naturally follows that the rules proposed by Sacks et al.(Sacks et al., 1974) are inadequate as a general model as understood as a cognitive model, and thus, it is necessary to revise or argment it with further rules that would account for the overlaps. The cognitive model of the "hearer" that we propose is viewed as an addition to extend the turn allocation rules. We have investigated 56 dialogs in the Japanese Map Task Corpus, which is basically a replication in Japanese of the Edinburgh HCRC Map Task Corpus, differing from the HCRC, among other things, in that, in the transcriptions, overlaps are marked with accurate time stamps and that utterances of different speakers can be played separately. We then classify 6173 overlaps into the categories we borrow from Herb Clark's proposal with necessary modifications. Based the analysis of the results of classifying the overlaps, we propose a cognitive model of adialog in which the hearer plays substantial part in the turn-taking system.

In this paper, we present three increasingly detailed analyses of the overlapping phenomena from different points of view. In Analysis 1, we classify the overlaps into 5 categories, and argue these 5 categories can, in turn, be viewed as constitutive of three types of interruptions. Analysis 1 makes it clear that overlaps are not anomalous, and that the patterns of occurrences of overlaps are neatly categorizable. In Analysis 2, we focus on the next speaker's "reasons" for beginning to talk while the other still talking, and conclude that, in a significant portion of overlaps, the next speaker is "invited" to interrupt by the current speaker. In Analysis 3, wetag each overlapped utterance with a speech act type and suggest that the other portion of overlaps which have turned out not to be clearly invited in Analysis 2 are still significantly associated with a well-defined set of speech act types of the overlapped utterances.

2. Phenomenal categories of overlaps

2.1. Clark's classification

Clark(Clark, 1994) introduced six phenomenal categories to classify overlaps. His system of classifying overlaps certainly offers an advanced standpoit in understanding the overlapping phenomena. Roger and others(D. B. Roger and Smith, 1988) developed a comprehensive system for classifying interruptions, assuming that Clark's classificatioin scheme consists of 6 phenomenal categories:namely (1)Acknowledgments, (2)Collaborative completions, (3)Recycled turn beginnings, (4)Invited interruption, (5)Strategic interruptions, (6)Nonlinguistic actions. Acknowledgements are short interjections like "yes," which have been also called backchannels. Collaborative completions are utterances deliberately begun in the middle of a turn-constructional unit, contrary to Sacks and others' rules 1a and 1b. Recycled turn beginnings are the starts of utterances made in order to signal they want the next turn. Invited interruptions are such interruptions made by the addressees in response to the invitios on the part of the current speakers. Strategic interruptions are interruptions by the next speakers which interrupt current speakers mid-turn for other reasons they consider legitimate. Nonlinguistic actions are a variety of nonlinguistic signals used in conversation, but they are typically accompany verbal behavior simultaneously, thereby defying analysis into turns.

2.2. Revised classification

In applying the categories to the actual data from the Japanese Map Task Dialog Corpus, we wanted the categories to be reliable with high inter-rater agreement so revised Clark's overlap classification into the one with 5 categories which are shown below. We suspect the reason that we did not achieve a high enough inter-rater agreement was that the classification scheme is basically a mixture of speech act types and conversational strategies, thereby confusing the raters in the preliminary rating tasks. The principle of the revision is to take a two-step decision procedure, in which speech act types, or the types of communicative functions, are judged for each relevant utterance first and in which then strategies are taken into consideration. The reliability of the new classification scheme was confirmed by checking the agreement rates for a part of the overlaps in question. The new classification consists of the following five categories, with a decision tree to help raise the objectivity and consistency. (1) Early Response (2) Request for Quick Response and Utterance providing New Information (4)"acknowledgement marker"(5) Inarticulate utterance.

2.2.1. Early Response

This is a type of utterance with which the person-to-be-the-next-speaker responds to the current speaker while the current speaker is still trying, sometimes inexplicitly, to select the next speaker.
[5]

[4] There have been sizable amount of literature which discusses overlaps or simultaneous speech in the tradition where people think in terms of the "one-at-a-time" principle(Jefferson and Schegloff, 1975),(Jefferson,).

[5] Clark suggested that traditionally many acknowledgments, such as *yes*, are not considered turns because a overlapping speaker doesn't intend to take a turn, and criticizes those who do not take acknowledgements as full-fledged ut-

We group acceptances and answers into one category when they are uttered while the current speaker is still speaking and call the categroy *Early Responses.* The following two excerpts server to illustrate the category[6].

```
(1) j1n1:0205/0206
0205:AA:eto migisita ni naruwake desu * ne hai
  (Well)(it is on the light side)(isn't it?)
0206:AB:                            *soudesu ne hai
                                (Yes, it is.)
```

```
(2) j1n4:0090/0092
0090:AC:ki ni sotte teppen made   itte kuda*sai
  (along the tree)(to the top)(go)(please)
0092:AD:                                *hai
                                        (Yes)
```

To be sure, in (1), AB's utterance is an acknowledgement because B starts his utterance before A finishs his confirmation, while, in (2), D's utterance is a a answer because C complete his/her order. Just because the current speaker's speech act doesn't finish before next utterance, it does not follow that one is a second pair part and the other is a no second pair part. Viewed in this light, these two speech act types can be regarded as the same type of speech act.

2.2.2. Requests for quick response

When the interlocutor interrupts the current speaker to take turn by uttering the first pair part of an adjacency pair in Schegloff's sense, i.e., an interjection, a question, a confirmation, a suggestion, a request or an order, this overlapping utterance is considered to be a request for quick response. An example is

```
(3)j1n2:0354-0366
0354:AE:tuirakugenba<128>teari*masen ka<128>
      (acrash place)    (is there?)
0355:AF:*ariasu arima*su
      (I see it)(I see it)
0356:AE:*arimasu yo ne<720>
       (sure, you see it)
```

Here with 0356, AE triggers off AF's utterance(0355) in order to confirm the landmark in his map. This is different from Early Response in that the overlapping utterance does not only respond to the current speaker but initiate a new move in conversation.

2.2.3. Uttrances providing new information

With an utterance providing new information, the interlocutor overlaps the current speaker to take a turn by a statement, a discourse marker or a self-interrogation. In the excerpt (4) below, when AH answers AG's question(0097), with 0100 AG interrupts his utterance to convey new information. In the excerpt (5), too, AJ provides new information. In our classification, (4) and (5) are basically of the same type, though later it will be shown that (5) has to be given special attention[7]. It is to be noted that, in (5), the first word of AG's utterance(0336), "te", completes AI's "too," though AI utters the "te" himself.

```
(4)j1n1:0097-0100
0097:AG:eto   haioku no          migigawa?
        (well)(a deserted house's)(right?)
0099:AH:sou desu ne   *hai
        (that's right)(yes)
0100:AG:              *ni it<256>te <224>
        (go there)
        sorede haioku ni tuite
       (and)  (the deserted house)(you reach)
```

```
(5)j1n3:0333/0336
0333:AI: -wo too*te--
        (through-)
0336:AJ:        *te<240>de sorede youturuni
        (through-) (and) (in short)
        hanntokeimawari ni iku to
       (counterclockwise) (if you go)
```

2.2.4. "Acknowledgement markers"

"Acknowledgement markers," which are brief utterances that backchannel, complete or prompt for continuation, are uttered not by being invited by the current speaker or by intending to take the next turn. The term "Backchannel" can be defined, in this case, as the utterance which appears during the current speaker's utterance in order that the hearer expresshis/her hearing or understanding with the word 'hai','uhm',etc.

```
(6) j1n1:0082-0083
0082:AK:de sitani tuki masi ta*ra<416>
  (if you reach the bottom)
0083:AL:                     *hai
                             (uhm)
0082:AK:sositara kondo<128>--
       (and)    (then--)
```

In (6), AL merely responds to AK's utterance by way of an acknowledgement marker, "hai" in this excerpt, assuring him that he is listening to or following him. Acknowledgement in this sense is different from that in the first category in that utteraces in this category do not merely accept or reply but prompt the current speaker to keep speaking. Some "completions" do prompt the current speaker to keep speaking. Next speaker obviously shows that he understands and prompts the current speaker to continue.

terances. We take Clark's criticism seriously and take such acknowledgement into consideration. It should be noted that *yes* are generally of two types, which we introduce as Early Response and "Acknowledgement marker."

[6] In what follows, the transcription texts are given in latinized Japanese together with rough English glosses underneath. Aligned asterisks indicate the approximate time points where the overlapping utterances begin. The colon separated indication right of the excerpts number read the dialog ID and the utterance IDs with in the dialog.

[7] It seems that, in Clark's clsssification, overlapping utterances of this kind are not treated as constitutive of a consistent category

Table 1: Incidents and Residuals of That Types of Speech Act of Overlapping Utterances Every Conditions of Preceding Utterances

Conditions of preceding utterances	Types of speech act of overlapping utterances			total
	Invited Interruption	Strategic Interruption	Collaborative Interruption	
Selectedness	821	284	440	1545
Nonselectedness	456	721	1516	2693
total	1277	1005	1956	4238

```
(7) jln1:0008/0010
0008:AM:--no sugusita ni oo*tokyanpuzyou
    ( at bottom      autocamp place)
0010:AN:                  *tokyanpuzyou
                    (to autocamping park)
```

Here in (7), AM and AN utter the same expression at the same time. AN does not interrupt the current speaker but is considered to show that he is following AM by means of completing AM's utterance.

2.2.5. Inarticulate utterance

Under this rubric, we group echoes, fillers and laughs, because it is impossible to ask the raters to judge the next speaker's intention of the utterance. Echoes are repetitions of what has been just spoken by the interlocuter. Fillers are the expressions like "ee" and "aa" in Japanese[8].

3. Analysis 1

Analysis 1 attempts that classification of overlaps and the relative occurrence of each category. In Analysis 1, 56 dialogues were tagged with respect to 6173 overlaps therein. The duration of each dialogue ranges from 4' 10" to 23' 44", with the mean around 11'. Twenty-one students on campus served to judge the types of overlaps in the dialogues, each working is 8 dialogues. Their task was to listen to the digitized sounds while reading the transcription around overlaps and to classify each of them into the 5 types discussed above, using a decision tree with each node presented at a time on the computer screen for the yes/no answer. Each dialogue was worked on by three students, and the agreement rate was 83.2%. We use, for this study, the 5136 overlaps whose classification were agreed on by at least two students. The percentages of (1)Early Response, (2)Request for Quick Response, (3)Utterance providing New Information, (4)Acknowledgement markers, and (5)Inarticulate utterances are 25%, 11%, 8.6%, 25%, and 13% respectively. The last category has been excluded from analysis without the loss of generality.

These five phenomenal categories can be viewed as divided into the following 3 types from a strategic standpoint, where the existence of the current speaker's invitation and the existence of a real turn taken by the interlocuter. The three types are (A) Invited Interruption, (B) Strategic Interruption, and (C) Collaborative Interruption. (1)Early Response is of the type (A) in that the utterer of utterances of this category is invited by the current speaker and does not intend to stop the current speaker. (2) Request for Quick Response and (3) Utterance providing New Information are of the type (B) because the speaker who begins to speak will eventually take the next turn, whether the interruption is invited by the current speaker or not. (4) Acknowledgement markers are of the type (C), where utterances are not invited by the current speaker or intend to deprive the current speaker of the next turn. It follows that the three types of interruption account for the major cases of overlaps and that it is well motivated to give separate explanations for them.

It should be concluded from what has been said above that overlaps have a number of functionality in the dialogues, and we may go on from this to infer that speech act types affect the variety of turn-taking strategies. Thus we see that the turn-taking model suggested by Sacks et.al.(ibid.), which necessarily takes overlaps as deviations from the application of their rules, fails to account for what role the hearer plays in taking turns in conversation. We need a model that takes into serious consideration the role played by the hearer, or the next speaker.

This conclusion leads us to the next question why the next speaker start to his/her speaking before the current speaker stops his/her talk, and how the next speaker decides on the time point at which he or she begins to utter. Given that the current speaker tries to select the next speaker, and then the next speaker begins to speak, based on the projectability of next turn, before the current speaker stops his/her talk, there should be a marker for a next-speaker selection around overlapped utterance. We are concerned with this point in Analysis 2.

4. Analysis 2

Analysis 2 is about the same set of dialogues and overlaps. In this analysis, the question is whether the current speaker has selected the next speaker by inviting responses from him/her at the point when the other speaker starts. Students were asked to tell whether

[8] We can not dicide whether H repeats G's words to show a confirmation or only to show an understanding.The reason classified this category is that this repeating utterance has two aspect of speaker's speech act.One aspect is a confirmation or for the before utterance, the other is an understanding. The observer felt these difference is not clear by utterance. One explanation for this difference may be that a strength and a risen of a intonation . There is room for further investigation.

Table 2: Incidents of That Detailed Types of Speech Act of Overlapping Utterances Every Conditions of Preceding Utterances

	Answer	Acknowledgment	Call	Question	Suggestion	Order	Completion	Backchannel
Selectedness	265	50	3	90	3	18	14	84
Non-selectedness	97	108	1	150	5	15	36	363
	362	158	4	240	8	33	50	447

	Statement	Discourse Marker	Self-integration	echo	Filler	Admiration	Unclear
Selectedness	46	1	3	14	11	37	14
Non-Selectedness	108	11	3	66	9	100	30
	154	12	6	80	20	137	44

each preceding utterance invites the next speaker to start, base on the following criteria: it invites the interlocutor to interrupt either if it is a 'first pair-parts' in the sense of conversation analysis, if it has a question intonation, or if it ends with a tag question. It is clear that, if the utterance preceding the overlapping utterance passes the criteria above, overlaps are expected to occur according to the projectability or predictability for next turn. Such techniques as eye-contact and the use of certain particles do not seem to influence the selection of the next speaker in the map task dialogues, so, in the criteria above, they are not taken into account. So we have a list of overlaps with selectedness, where the current speaker has selected the next speaker by using a certain set of linguistic means, and those with non-selectedness, where he/she has not done so.

Now χ^2 test was applied to see the statistical association between these two characteristics of the overlapped utterances and the 3 types of overlapping utterances - (A) Invited Interruption, (B) Strategic Interruption, and (C) Collaborative Interruption, as we have seen in Analysis 1. The result confirmed a statistical significance of the incidence between the conditions(χ^2(3)662.08,p<.001), justifying the use of residual analysis. Table 1 is a result of a statistical analysis applied to the incident of each utterances type. As can be easily seen, (A) Invited Interruption significantly increases in the selectedness condition, and (B) Strategic Interruption and (C) Collaborative Interruption markedly grow in the non-selectedness condition.

It follows from what has been demonstrated that only (A), Invited Interruption, especially Answer(Table 2),as like example (2) different from the other types of overlaps statistically, is likely to take place while a current speaker tries to select next speaker using the techniques with which to select next speaker such as a 'first pair-part,' a question intonation or a tag question around the nearby TRPs. (B)Strategic Interruption, such as Statement, Discourse Marker, and (C) Collaborative Interruption, such as Backchannel, account for about 60% of all are likely to occur in non-selectedness condition. If that is the case, the next natural question would be: Can these two types of overlaps be totally independent of what the current speaker is talking or doing?

5. Analysis 3

Analysis 3 aims to explain the interruptions which are not invited in the sense of Analysis 2 in terms of the conversational strategies taken by the speaker of the preceding, overlapped utterance. The strategies in question are (a) Responding Strategy, (b) Opening Strategy, and (c) Cooperating Strategy. It should be obvious that the speech act types of utterances in general are determined by the strategies taken by the speaker. We naturally assume that the types of the utterances that reply or respond to preceding utterances are motivated by the Responding Strategy, and that Opening Strategy elicits the utterances of the speech act type of request and providing new information, and that acknowledgement markers, namely backchannels, completions and prompts, are due to Cooperating Strategy of the speaker.

Using the same classification scheme, 36 students have been asked to classify the speech act types of the overlapped utterances in 4 dialogs that contained 547 overlaps. The duration of each dialog ranges from 7' 03" to 10' 52", with a mean around 9'. The students were each given the task of working on one-third of each dialog, with the agreement rate of 82.6%.

As shown in Table 3, the number of occurrences of acknowledgement markers is small, whereas the occurrences of acknowledgement marker account for approximately one-third of overlapping utterances. This remarkable frequency makes it clear that the next speaker can't trigger off the utterances of this speech act type.

We compared, with the χ^2 test, the proportion of the three types of speech acts of overlapping utterances and the two types of speech acts of overlapped utterances except for this category. The classification revealed that the speech act types of overlapping utterances are determined by those of the overlapped utterances(χ^2(6)15.89,p<.001). Table 3 also gives the incident of each utterances type. (B) Strategic Interruption increases in the (b) Responding Strategy, while (C) Collaborative Interruption rises in the (B) Opening Strategy.

Several things the tables do not reveal but which we believe we have to take into consideration are (1) that most of the respondinig utterances are those utterances of the natural categories expected from the types of

Table 3: Incidents of That Types of Speech Act of Overlapping Utterances Every Types of Speech Act of Preceding Utterances

That Types of Speech Act of Overlapping Utterances	Types of Speech Act of Preceding Utterances				Total
	Responding Strategy	Opening Strategy	Collaborative Strategy	etc.	
Invited Interruption	17	95	0	14	126
Strategic Interruption	21	58	0	12	91
Collaborative Interruptions	14	114	0	6	134
etc.	15	51	2	7	73
Total	67	318	2	39	424

the previous utterances. and (2) that the types of the hearer's utterances respond to or correspond to those of the utteraces made by the person who is apparently addressed to[9].

This result makes it clear that utterances of the two speech act types in the non-selectedness condition are, to a certain extent, controled by the types of the current speaker's utterances, and hence the conversational strategies behind. That means that the other speaker can not interrupt when current speaker's utterance is of the type associated with (c)Cooperating Strategy, though overlaps of the (B) Strategic Interruption take place when the current speaker speaks who takes (a)Responding Strategy,

6. The hearer's model of turn taking

rtbreawWe can recognize from these three analyses the extent to which the interaction between the types of the communicative functions of overlapped and overlapping utterances results in turn taking with overlaps. We are now in a position to propose a hearer's model in which the hearer starts to speak while the current speaker is still talking. The model can be stated by stipulating the three cognitive rules the hearer apparently follows. The rules are expressed in such a way that the next speaker(N) may, or is allowed to, start talking even while the current speaker(C) is talking.

Rule(A) Next speaker(N) may start talking in the middle of the current speaker(C)'s utterance, either if N is selected or if N's interruption is invited.

Rule(B1) Whether C invites N in the middle or not, N may interrupt C's utterance if C takes Responding Strategy. If C stops speaking, N takes the turn.

Rule(B2) Whether C invites N in the middle or not, N may utter an acknowledgement marker, if C takes Opening Strategy. When this rule is applied, the turn is not taken.

These rules conjointly explains most of the examples discussed above. But it is to be noted the explanations here do not provide sufficient conditions for the occurrences of overlaps but necessary conditions, due to the nature of the rules we have introduced. We will see how the 3 rules above explain how the hearer is allowed to start talking even while the current speaker is talking, thus overlaps taking place.

The type of overlaps illustrated by the excerpt (1) will take place when Rule(A) is applied. AB(Next speaker) starts talking in the middle of the AA(Current speaker)'s utterance, when AB's interruption is invited. The type of overlaps illustrated by the excerpt (2) can be explained by using the Rule(A), too. AD(N) starts talking in the middle of the AC(C)'s utterance as he is being invited.

Rule(B1) explains the type of overlaps illustrated by the cecerpt (3). AE(0356)(N) interrupts AF(0355)(C) in the middle of the his utterance, because AF understands in the middle of AE's utterance that he is requested to answer AE(0354) by the utterance(0356), thereby taking the Responding Strategy. The type of overlaps illustrated by the cecerpt (4) will also take place because of the existence of the Rule(B1). AG(0100)(N) interrupts AH(0099)(C) in the middle of the his utterance, when AH(0099) answers AG(0097).

The type of overlaps illustrated by the cecerpt (6) will take place thanks to Rule(B2). AL(0083)(N) utters an acknowledgement marker, when AK(0082)(C) takes the Opening Strategy and starts to provide new information. Rule(B2) accounts for The type of overlaps illustrated by the excerpt (7). AN(0010)(N) utters an acknowledgement marker, when AM(0008) states new information.

The type of overlaps illustrated by the excerpt (5) is not necessarily explicable by way of the three rules above alone. When AJ(0336)(N) interrupts AI(0333) in the middle of the utterance, the utterance 0333is not an acknowledgement marker in the sense defined in this paper[10].

7. Discussion

The model for orderly turn taking derivable from Sacks et al(Sacks et al., 1974) is one in which the cur-

[9] Acknowledements account for more than a half of the utterances that are obviously made for the purpose of self-defence.

[10] This is actually a "real" case of interruption in the "realest" sense of the word. AJ's interruption is irrespective of AI's attitudes. The set of rules we have proposed can be said to put more stress of hte collaborative nature of conversation and dialog.

rent speaker **selects** the next speaker. The model proposed in this paper, on the other hand, takes hearer's role seriously. What has been made our kind of hearer's model possible is obviously the distinctions we have made between the speech act typs and the communicative functions of utterances involved. The rules mention the strategies the speakers might take, and the strategies taken are determined by the kind of communicative functions, or the speech act types those utterances have.

The CsN model only mentions the selection by the current speaker, regardless of what kind of speech acts the speakers are performing. Jefferson(Jefferson, 1984a) apparently notices the nature of the CsN model but does not seem to have taken a step in our direction. In the "one-at-a-time" priciple entailed by the CsN model, each participant of a conversation is supposed to independently determine whether he or she starts to talk or not, and also who speaks next or nobody. In the model proposed here, the paricipants are somehow committed to a certain kind of collaborative action, and the map task in particular. The nature of task and the course of dialog jointly determine who starts speaking when. The model explains, in short, the fact that in such task oriented dialogs, there are a large number of overlaps observed.

This model assumes the the set of turn allocation rules we have identified as the CsN model is functional in constructing the turns we obseverd taken in conversation, and add a explanatory power for a certain kind of conversations or dialogs in which overlaps abound. In the course of the analyses and arguments it has been suggested that such an abundance of overlaps has something to do with the collaborative nature of the map task, during the performance of which our target dialogs have been recorded. The sparcity of forceful interruptions on the part of the hearer in the map task dialog corpora comes from the collaborative nature of the task, but the fact that there still are some "real" interruptions could be a problem for our model, as we discuss below.

8. Conclusion and remaining problems

The model proposed in this paper explains the patterns of overlaps tagged in the Japanese Map Task Dialog Corpus and those reported by Clark. Thus the model accounts for the cases which can not be explained by the CsN model, providing a basis for the theory that deals with the interaction between the participants in a dialog in general. More particularly, it suggests that conversation must be organized by not only the current speaker but also the hearer and sometimes by both speakers collaboratively. Further investigation is necessary as to the conversational system into which turn allocation rules for the current speaker and this hearer's model are to be integrated, and which can consistently treat all other related phenomena.

The most nagging problem for the model proposed is the timing and the motivation of the "real" interruptions, which is a subset of what we classified as the Strategic Interruption. The fact that most of the interruptions are explained by the model here means that what has been vaguely classified as "interruptions" are not very much of the interruption but results of hearer's collaborative efforts which, most of the time, end up with the turns taken. Those "real" interruptions are tough to handle for researchers as well as the participants of the conversation on the spot.

9. References

H. H. Clark. 1994. Discourse in production. In Gernsbacher. A. M., editor, *Handbook of psycholinguistics*, pages 985–1018. Academic Press, San Diego.

P. E. Bull D. B. Roger and S. Smith. 1988. The development of a comprehensive system for classifying interruptions. *Hournal of Language and Social Psychology*, 7:27–34.

S. Ervin-Tripp. 1979. Children's verbal turn-taking. In E. Ochd and Schieffelin. B.B., editors, *Developmental Pragmatics*, pages 391–414. Academic Press, New York.

G. Jefferson. Notes on.

G. Jefferson and E. A. Schegloff. 1975. Sketch: Some orderly aspects of overlap onset in natural conversation. *Paper presented at the 74th Annual Meeting of the American Anthropological Association.*

G. Jefferson, 1984a. *Notes on some orderlinesses of overlap onset*, pages 11–38. CLEUP.

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

The Pragmatics of English Dialogues in the Chinese Context

Zongxin FENG

Post-doctoral fellow, Beijing Foreign Studies University
Foreign Languages Research Centre, Beijing Foreign Studies University, Beijing, 100089, P R China
zxfeng@bj.col.com.cn

Abstract

This paper is part of a cross-cultural study on the differences in the principles governing English dialogues between native speakers and Chinese speakers in a special context where both sides share some linguistic and cultural backgrounds. From the findings, it distinguished the abstract cooperative code and the concrete maxims of the Cooperative Principle, holding that the former is prescriptive and the latter descriptive. In view of form and function of speech codes, the paper points to the universality of CP by showing that politeness can not override the action of the maxims in certain cross-cultural contexts.

Introduction

While discussions on the relationship between language and culture started in China in the 1950s (Luo, 1950), pragmatic studies of cross-cultural communication started only in the 1980s (Huang, 1984; Yan & He, 1985). Such pragmatic studies since then have largely concentrated their attention on listing Chinese learners' common 'errors' of interaction in English or instances of misunderstanding and miscommunication with native speakers in the name of 'pragmatic failure' following Jenny Thomas (1983). Some discuss the problems in terms of cross-cultural conflicts of pragmatic principles and maxims. Others attribute the problems exclusively to Chinese speakers' communicative incompetence in a second language. Thus, remedies were suggested such as more emphasis on communicative approaches in foreign language teaching, or an inclusion of culture-related courses in the curriculum, with few studies probing into the workings of the pragmatic principles in real-situation cross-cultural dialogues.

The present study aimed to find out how the general Cooperative Principle (Grice, 1975) is actualized in the choice of speech strategies and how the specific maxims work in accounting for what is actually said in such a particular context of verbal interaction. An investigation was conducted of everyday dialogues between individuals belonging to two different groups of speakers: Chinese English-speakers and native English-speakers, at a national university in Beijing. The subjects stay in the same complex and their daily dialogues are in both English and Chinese. They have approximately the same degree of second language proficiency, and presumably similar degrees of awareness to linguistic and cultural relativity. Instances of what would have been labelled as pragmatic failure or miscommunication due to inappropriate verbal behaviour or sociopragmatic match of speech acts, as is defined by Thomas (1983), have been observed in English data (from Chinese English-speakers) in similar patterns in comparison with Chinese data (from native English-speakers). Unlike common assumptions and expectations alerted by other existing studies in China, such risky or problematic speech acts (i.e. either Chinese-culture oriented ones in English data or the other way round) successfully accomplished practical communicative tasks in particular situations.

The Investigation

This investigation was a continuation of the research that was first started in early 1995. The university where the research was conducted, like many other national universities in Beijing, has an English Department and a Chinese Language Training Centre, the former preparing Chinese students of English and the latter training foreign students of Chinese, mostly from America and Great Britain. There is much out-class communication between students from the two departments.

In everyday communication, the two types of language students use either Chinese or English. Since both types of students prefer to use their second languages more for practice, the ratio of conversations in Chinese and English is approximately half and half. In both cases, similar patterns of verbal behaviour have been observed. On the one hand, speech acts in second languages seem to be strongly influenced by those in the speakers' first languages. On the other hand, with limited knowledge of their cross-cultural interlocutors' ways of saying and doing things, both types of speakers tend to over-generalize. For instance, in the initial stage, Chinese students of English tend to speak English in a typical Chinese way and native English speakers tend to speak Chinese in their own way. Some native English speakers take 'Have you had your meal?' as a common greeting among the Chinese people and greet their Chinese schoolmates in this way regardless of time and place. When speakers have got more familiar with their second languages, they can learn more 'proper' ways of saying and doing things, but they still tend to go a bit too far because of stereotypical assumptions and over-generalizations. While Chinese speakers of English generally assume that Americans are informal and British people are more formal, native English speakers assume that the Chinese are too indirect and sometimes unfathomable. In spite of the cultural differences, the two types of speakers have obviously been making efforts in seeking common grounds. Therefore, instances of inappropriate verbal behaviour as are found in cross-cultural pragmatics and ELT studies (Huang, 1984; Hu, 1988, 1992, 1993) did not necessarily lead to pragmatic failure or communication breakdowns.

After these two groups of speakers had known each other for some more time, the frequency of miscommunicataion was greatly reduced. However, the so-called 'inappropriate' verbal behaviour on either side did not significantly diminish. At the same time, native English speakers arriving each year are increasingly more proficient in the Chinese language and Chinese students of English are better equipped with English before they enter college. Then, the communication was based on more common factors, and new features started appearing in the process. While typical 'cultural mistakes' and 'pragmatic failure' were no longer evident as the speakers became more communicatively competent in their second languages, certain risky and problematic speech acts had not been wiped out but seemed to be on the rise due to cross-linguistic and cross-cultural pragmatic factors. Therefore, a further attempt was made to narrow down the data on English dialogues between these two groups of speakers.

A long list was prepared consisting of the reported instances of pragmatic failure due to 'inappropriate verbal behaviour', including Chinese speakers' ways of greeting, initiating conversations, asking questions, giving and accepting advice, and responding to thanks and compliments, etc. Meanwhile, recordings were made of real conversations through some Chinese volunteers interacting with native English speakers. A comparison showed that the listed instances of 'inappropriate verbal behaviour' have positive pragmatic values in the conversations under scrutiny. Tracking-down interviews in the native English-speaker group showed that the transcribed items in question are acceptable and communicable, with the comments that they are 'intuitively nothing abnormal'.

The transcribed data were then put in questionnaires and distributed among a separate body of native speakers of English, who have an average of two years' experience of studying Chinese as a second language. 95% valid responses are retrieved (38 out of 40) through one of their Chinese language instructors and the results are as follows.

Initiating Talks

Ways of initiating talks consists of greetings and asking questions, ranging from the most impersonal chat to the more personal 'How are you?' and 'Are you all right today?', from 'I haven't seen you for ages' to 'What have you been doing these days?' and 'Where have you been recently?'. The last two are often regarded as being inquisitive and nosy by Chinese scholars on cross-cultural communication. However, not even a single native English speaker marked asterisks on any that he/she would in all instances regard as communicatively unacceptable. To my question whether they would take the last two as instances of invasion into their privacy, except for three subjects giving comments 'It depends', all the rest responded negatively. They commented that 'One is usually able to tell the difference between nosiness and curiosity and chit-chat', 'A colleague you know well can certainly ask these questions. But if someone you barely know is in the intention of demanding details, it is considered as rude'. The consensus is that these questions are not likely related to nosiness or invasion of privacy. The only minor exception is, 'If the questions are repeated, I would think he is being inquisitive'.

Receiving Compliments

Typical Chinese responses to compliments on one's excellence in certain skills, one's remarkable accomplishment, or the like are "There is nothing worthy of note", "You are over-praising me", "You may be joking", or "It is really difficult for me to do it well", "I have been really working hard on it", etc. These responses are typical Chinese sayings when receiving compliments, and students are therefore told, stereotypically, that they are associated with impoliteness in respect to the hearer. While the first three may carry strong implications of questioning the judgement or evaluation of the one who compliments, the last two may indicate a strong sense of immodesty of the one who responds to the compliment. But results of the questionnaires show that these responses are not inherently impolite or arrogant in the specified contexts of situation.

Two subjects reported that the Chinese speakers are being humble and reserved. But all the rest made the similar comments that the responses are 'normal', 'acceptable', and 'no problem' because the speaker is, or is being, modest. One subject commented, 'One not familiar with Chinese customs would generally find these responses [to compliments] annoying or impolite at best. Generally "Thank you" will suffice, or sometimes "Thank you, but it's nothing really"'. Another subject's comment was, 'I understand the Chinese customs of modesty, so I don't get offended. I continue to praise my students or friends if they are worth because in America it's seen as being supportive or helpful to praise.'

Receiving Thanks

In many cases, when a native speaker thanks the Chinese English-speaker for his/her kindness or a favour that has been done, the Chinese, out of his own cultural norms, may respond with 'This is what I should do', 'Don't mention it' and "It's really nothing" (which are literal translations of Chinese utterances in receiving thanks), although he is taught to respond with 'With pleasure' or 'I'm glad to be of help', etc.

The majority of the subjects took the 'This is what I should do' type responses as being 'the same as "You're welcome"', 'communicatively OK' and "very polite ways to acknowledge an act of kindness". One subject wrote, "We do this in America. We say "Don't think about it", "Not a problem", so I am QUITE USED TO IT'. All native speakers commented that the first response is 'great' and 'normal', and that the speaker was modest and polite. However, three subjects thought a little differently on the first response. Two responded that it 'sounds a bit awkward' and 'a bit too strong', and one commented that 'it is unacceptable because it is too humbling, so "Don't mention it" is the best'. In tracking-down interviews, all the three subjects admitted that they would appreciate the favour done to them and knew what politeness was meant in the responses. The one who regarded it as unacceptable commented that when she heard this, she was aware of the polite message there. Only when a non-native English learner asked her whether it was good English, she would reject it as unacceptable.

Accepting Offers

During visits, when a native English speaker as host shows hospitality and offers a Chinese visitor a cup of coffee or a tin of sprite, the visitor often replies 'Please do not take the trouble' or 'No, thanks' when he actually means to accept the kind offer. These are typical Chinese responses to offers: The guest pretends to save the host trouble making such offers, and the host will appear even more hospitable if he insists. While politely rejecting, the guest understands well that the host will show further politeness and continue the offer. In fact, he is still expecting the coffee or sprite to be brought unless he repeats his utterance or explains why he does not want to have any. As a norm, he will make further polite protestations when coffee or sprite is brought and then take it after being repeatedly urged. However, among close friends, a simple 'Thank you' in Chinese would mean a happy acceptance.

Chinese speakers of English are taught not to say anything in a typical Chinese way in English. While most students avoid using 'Please do not take the trouble' as a proper response to polite offers, a significant number of students do use it. While they are taught that a 'No, thanks' would definitely mean a rejection, these students do use it when they really wish to accept the offer. However, all the native speakers of English in my research commented that responses of this type in given situations are 'polite' and 'normal', although a few of them added that 'Most Americans prefer a more straightforward approach'. Typical comments are 'It is expected. If I do not hear it, I began to think they do not appreciate my trouble', 'He is being polite and only means it if he insists', 'One should take the trouble, unless it makes the guest uncomfortable. Or if they just don't want it, they'll probably say it out directly', 'My friends in America are the same. They say it and I still give them a cup of coffee', 'I understand culturally why they do this', and 'I would just bring them a cup of coffee for fear that they are just being polite', etc.

This result is surprising because I myself used to strongly believe that these responses are literal translations of a Chinese version *Qing Bie Keqi*, and so I had expected more opposing comments on these utterances of accepting offers from native English speakers.

Giving Advice or Showing Concern

A Chinese may say to a native English speaker 'You should really go and visit the beautiful city of Harbin' (before Christmas), 'You must get ready by . . .' (before a winter holiday), and 'Be careful!' or 'Look out!' (when climbing a mountain).

These typical Chinese ways of giving advice or showing concern led to no controversy on the issue whether the speakers are being imposing. Responses range from a simple 'No' and 'Gosh, no' to the most emphatic 'No, not at all!!!'. One subject added 'I'm used to them, and I realize most Chinese never mean to impose', and another added 'I'd appreciate such information if it was with good reason'. Still others commented 'I'd think they are being helpful', 'That's normal everyday talk', 'Only pushiness and repetitiveness in this case are considered unreasonable'.

In this particular context of exchange involving cross-cultural speakers, possible problematic implications on the side of the Chinese speakers of English are comprehensibly and empathetically ruled out by the native English speakers, especially those of nosiness (in asking questions for phatic purposes) and pushiness (in giving earnest advice). Significantly, when the attention of these informants was deliberately drawn to reconsidering the issues, the subjects made notations that the risk of pragmatic failure or misunderstanding is only a matter of degree. For example, they reported that Chinese students initiate talks with native English speakers by compliments or questions. They use compliments on various subjects just for phatic communion, as the native English speakers often do. And the native speakers of English understand that their Chinese interlocutors are choosing the safest topics possible or available, so they do not take their remarks at the face value. Incidentally, the native English speakers admitted themselves sometimes deliberately communicating in a typical Chinese way in either English or Chinese. Even if the topic for compliment or initiation of a talk is not the best chosen by the Chinese, no native English speaker reported having been offended.

However, on the relation between conversation implicature and politeness in certain cases, there is a drastic difference between their interpretations. For instance, He (1988: 84) rightly interprets the following conversation

Student: Beirut is Peru, isn't it?
Teacher: Rome is in Romania, I suppose.

as the teacher's violation of Grice's maxim of quality. However, a few pages later his interpretation goes that the teacher is being polite, in order not to hurt the student's feelings, for the teacher could have made a direct comment 'It's absolutely ridiculous!' (p. 101).

More than 90% of Chinese college freshmen under an earlier survey took the latter interpretation for granted, except for a few who thought that the teacher is not necessarily polite. In contrast, the pattern is just reversed among native speakers of English. Except for only two, all the rest of the 38 subjects gave the comments on the questionnaires that the teacher is not being polite in whatever circumstances, but being 'sarcastic' and 'rude'. Interviews with the two subjects responding 'Yes' showed that they would think so only if the teacher is speaking in a most humorous tone and a friendly atmosphere with obvious intentions to cause a hearty laugh. However, a survey from Chinese sophomores produced a pattern similar to that of the native English speakers.

What is significant is that, in this special context, typical Chinese ways of talking in English are mostly accepted rather than rejected as incommunicable or labelled as instances of Chinese speakers' communicative incompetence in their second language. Certain unpleasant and impolite conversational implicature, which may otherwise have arisen, has been reasonably cancelled because the speech acts concerned are not inherently polite or impolite without considering specific contexts of situation.

Discussion

As communication is a two-way process, what lies behind may be the functioning of a general code, or what Leech calls 'Interpersonal Rhetoric' (Leech, 1983), which consists of Grice's Cooperative Principle and his own Politeness Principle with their respective maxims. The general principle of co-operation serves as a binding force between the interlocutors who, whether conscious or unconscious, successfully bridge the gap between what is intended and what is actually said in each communicative event, in terms of the speaker's deliberate choice of speech strategies and the hearer's cooperative attempt to infer the communicative intentions therein. The distinction of pragmatic competence and performance brings us to the discussion of a distinction between the abstract 'code of cooperative behaviour' which already presumes politeness, and the concrete maxims in the Cooperative Principle put forward by Grice.

Few works discussing Grice's theory distinguish the two related aspects of the Cooperative Principle: a general principle and four maxims. For example, Levinson (1983: 101-2) treats the general principle and the maxims as the same in nature. Likewise, Leech (1983) makes no point of treating the 'principle' and the 'maxims' differently, saying that the maxims are 'special manifestations' of the principle. But Grice originally implies a different status of the principle and the 'attendant' maxims (Grice, 1975), or as others may call them as 'concomitant' maxims (Mao, 1994).

This distinction is closely related to the nature of the Cooperative Principle with its maxims. By the 'code of cooperative behaviour' which 'organizes the way interlocutors interpret each other's speech', Grice only emphasizes the interpretative efficacy of his maxims for conversational implicature. This may partly account for a negligence of the other related aspect of Grice's principle. For example, Leech, in a later work, characterizes the maxims of the Cooperative Principle as 'purely descriptive', 'postulated for the purpose of explaining observed behaviour' (Leech, 1992: 261). However, it is the abstract code that has normative, if not moralistic, values on the speaker's having to be cooperative and assuming his interlocutor's cooperative act in order to accomplish a sensible communication event. Against this backdrop, the so-called 'conflict' of the individual maxims, or groups of maxims in CP and PP in Leech's discussion, can be better explained as a matter of explicitness in actualizing a cooperative (and polite) act in achieving certain communicative goals.

Conclusion

In cross-cultural circumstances, where there are always special problems, the normative and prescriptive values of the abstract code of cooperation and politeness may go beyond cultures. The assumption of mutual cooperation and politeness between conversation participants is the key point in securing contextually appropriate encoding and decoding of communicative intentions. Thus, it is the social message being conveyed that actually counts, and it is the good or bad intentions of the speaker that determines whether a speech act is polite. In spite of cultural differences, there seems to be implicit norms for the cooperative principle to be observed and the sub-maxims to be specifically actualized (with or without violations). Thus, pragmatic functions that contextual utterances serve can outweigh isolated speech forms, whether superficially polite or impolite.

References

Grice, H. P. (1975). 'Logic and Conversation', in Cole and Morgan (eds) *Syntax and Semantics III: Speech Acts* (pp. 41-58). New York: Academic Press.

Gu, Yueguo. (1992). 'Politeness, Pragmatics and Culture'. In *Waiyu jiaoxue yu yanjiu* (Foreign Language Teaching and Research). No. 4.

He, Ziran. (1988). *A Survey of Pragmatics.* Changsha: Hunan Education Press.

Hu, Wenzhong. (1992). 'Cultural Factors in ELT'. In *Waiyu jiaoxue yu yanjiu* (Foreign Language Teaching and Research). No. 3.

Hu, Wenzhong. (1993). 'On Cross-Cultural Communication Studies in FLT'. In *Waiyu jiaoxue yu yanjiu* (Foreign Language Teaching and Research). No. 1.

Hu, Wenzhong. (ed) (1988). *Cross-cultural Communication and English Teaching.* Shanghai Translation Publishing House.

Huang, Cidong. (1984). 'Pragmatics and Pragmatic Failure'. In *Waiguoyu* (Foreign Languages). No. 2.

Leech, G. (1983). *Principles of Pragmatics.* London: Longman.

Leech, G. (1992). Pragmatic Principles in Shaw's *You Never Can Tell.* In M. Toolan (ed) *Language, Text and Context* (pp. 259-78). Routledge: London and New York.

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

Luo, Changpei. (1989 [1950]). *Language and Culture.* Beijing: Language Press.

Mao, R. L. (1994). 'Beyond Politeness Theory: "Face" Revisited and Renewed'. In *Journal of Pragmatics*, Vol. 21, No. 4.

Thomas, J. (1983). 'Cross-Cultural Pragmatic Failure'. In *Applied Linguistics.* Vol. 4, No. 2. Oxford University Press.

Thomas, J. (1995). *Meaning in Interaction: an introduction to pragmatics.* London: Longman.

Yan, Zhuang & He, Ziran. (1985). 'Pragmatic Failure of the Chinese Learners in Communication with English Native Speakers'. In *ELT in China* (pp. 185-98). Beijing: Foreign Language Teaching and Research Press (1990).

What Is a Situation?

Kerstin Fischer

University of Hamburg
Fachbereich Informatik
Arbeitsbereich Natürlichsprachliche Systeme
Vogt-Koelln-Str.30
D-22527 Hamburg
fischer@nats.informatik.uni-hamburg.de

Abstract

Our everyday use of the term *situation* suggests that it is unproblematic to decide what constitutes a situation in which language happens. At the same time it seems to be important to account for what a situation is like since situational factors have been found to influence systematically the properties of the language which occurs in that situation. We are investigating a corpus of (simulated) human-computer interaction in which the simulated system's utterances are produced independent of what the speaker says according to a fixed schema with recurring phases. This way, what one communication partner utters is both intra- and interpersonally comparable. The corpus thus provides a unique opportunity to study linguistic behaviour in a completely fixed situation in which even the contributions of one of the speakers can be controlled. Although situational variables are kept constant in this corpus, the speakers' linguistic behaviour varies both intra- and interpersonally. Given the hypothesized strong connection between situation and linguistic behaviour, the variability of the linguistic properties observable constitutes a problem. With respect to intrapersonal variation, the variability can be attributed to a change in speaker attitude towards the system. In order to explain interpersonal variation, the corpus is checked for indicators of how speakers conceptualize the situation. What speakers believe about their communication partners can be shown to correlate with aspects of their linguistic behaviour. Instead of defining a situation by means of extra-linguistic variables, it is suggested to treat a situation as a speaker category, that is, as a complex concept to which the speakers themselves can be shown to attend.

1. Introduction[1]

Everyone can answer questions like 'what situation are you in,' 'in which situation do you record your corpora,' or 'what would be a situation in which this or that sequence would make sense.' The term *situation* therefore seems to be quite unproblematic. At the same time, discourse and text typologies, the number of different 'kinds' of dialogues, for instance, map task -, appointment scheduling -, or instruction - dialogues, as well as the obligatory section on the corpus data used in empirical studies of discourse suggest that the situation in which language happens has an impact on the language used in this situation. Furthermore, concepts like register and sublanguage assume an essential relationship between situational factors and the linguistic properties observable.

While situational factors seem to be of influence on the language observable, what can be found regarding dialogues which are recorded in absolutely identical circumstances with 36 speakers (19 women, 17 men) is that there is an enormous intra- and interpersonal variation as to their linguistic behaviour. The, old yet unanswered, question to be asked here is thus, if the situation influences the use of language made by the speakers, why does their linguistic behaviour differ so much, during time and between speakers?

The focus in this study will be on what the speakers' linguistic behaviour can tell us about what they think the situation is about themselves. The corpus investigated provides us with the unique opportunity to study the role of individual speakers' conceptualizations of what the situation is like because one of the speakers' linguistic behaviour is absolutly identical – both through time and between speakers. Thus in the current corpus not only the external variables, participants, task, domain, activity type, etc., are kept constant, but also what one of the participants contributes. In particular, an automatic speech processing system is simulated by a human 'wizard' (Fraser and Gilbert, 1991), and its linguistic output is created according to a fixed schema with fixed recursively recurring sequences (see section 3.). Thus, while dialogues in general are interactively achieved (Clark, 1996), the current corpus allows us to analyse controlledly the contribution of a single speaker's conceptualization of what the situation is like.

2. The Impact of Situational Variables on Linguistic Behaviour

On the basis of Firth's contextual theory of meaning (Firth, 1957), Halliday et al. (1964) develop the concept of register, the "systematic variation by use in relation to social context" (Lyons, 1977, p.584) to account for the fact that particularly for language teaching it is essential to consider that not all linguistic forms are equally well suited to be used in all situations: "It is only by reference to the various situations, and situation types, in which language is used that we can understand its functioning and its effectiveness" (Halliday et al., 1964, p.89). The relationship between situation and linguistic properties appropriate in it are taken to be conventional: "Linguistic features of registers can sometimes be seen to have language-external

[1] The research for this paper was supported by the German Federal Ministry of Education, Science, Research and Technology (BMBF) in the framework of the Verbmobil project under grant number 01 IV 701 F7. The responsibility for the contents lies with the author.

causes, (...) but otherwise they must be accepted as being in the same arbitrary type of relation to the situational features they correlate with as, in general, linguistic items are to the situational items they 'mean'." (Ellis and Ure, 1969, p.251-259). Language variation 'according to use' can therefore be accounted for by creating a situation typology and by associating conventionally particular linguistic features to the different situation types.

With respect to the communication partner automatic speech processing system, Krause and Hitzenberger (1992) have proposed a register 'computer talk'. They describe 'computer talk' quantitatively as a variety in contrast to 'normal' language by means of the increase and decrease of particular linguistic features. That is, they propose that the situation is determined by the particular communication partner automatic speech processing system, and that knowing about this suffices to predict the linguistic properties that can be observed in the speakers' linguistic behaviour in the communication with such systems, at least in comparison with 'normal' speech.

In addition to the role of the communication partner, a number of criteria have been identified that determine the linguistic properties of speakers' utterances, which are of relevance to automatic speech processing. For this reason, to increase the reusability of corpora as resources for the development of such systems, dialogues were proposed to be classified according to these criteria. The dialogue typology suggested by the EAGLES group on *Integrated Spoken and Written Language Resources*, for instance, suggests the following criteria by means of which corpora can be classified (EAGLES, 1998):

- Number of Participants
- Task Orientation
- Applications Orientation
- Domain Restriction (also: which domain)
- Activity Types (cooperative negotiation, instruction, etc.)
- Human-Machine Participation
- Scenario
  - Speaker Characteristics (gender, age, geographical provenance, smoking habits, etc.)
  - Channel Characteristics (spoken vs. written)
  - Other Environment Conditions (special recording conditions, such as Wizard-of-Oz)

Besides these practical considerations, the influence of situational variables on language use are investigated, for instance, in sociolinguistics (Fasold, 1990) and in approaches to linguistic variation (Chambers and Trudgill, 1998), as well as in the ethnography of communication (Hymes, 1972). It can be concluded that there is a strong connection between situational variables and the properties of the language occurring in them. Dialogues that do not differ with respect to the situational factors can conversely be expected to be homogeneous regarding their linguistic properties.

3. Corpus

The data are 36 dialogues of 18 to 33 minutes length that were transcribed and prosodically, conversationally and lexically annotated (Fischer, 1999a). Each dialogue consists of 248 turns on the average,[2] 124 of which are uttered by the human speaker.[3] Participants (19 women, 17 men) are between 17 and 61 years old and all native speakers of German.[4]

As a methodology for controlling inter- and intrapersonal variation, a fixed dialogue schema has been created which determines the utterances made by the system. Thus certain sequences of system output have been defined which are combined in a fixed order, all sequences occurring at least twice. These recursively recurring dialog phases make it possible to analyse the reactions to the same sequences of utterances at different stages of the dialogue. The system output is thereby completely independent of the users' utterances. For instance, the system may ask the user to make a proposal for a day when to meet. Irrespective of the user's reaction, the system will then utter that the first of January is a holiday, simulating a speech recognition error. After the next speaker utterance, the system will assert that it is impossible to meet at four o'clock in the morning. This sequence may occur four times in each dialogue. The impression the speakers have during the dialogue is that they are talking to an automatic speech processing system that repeatedly misinterprets their utterances, and that sometimes fails to understand completely. Furthermore, the system produces long pauses (30 secs.) and 'wrongly synthesized,' not understandable, utterances.

Speakers are instructed to schedule ten appointments with the system. Before speakers are confronted with the (simulated) malfunctioning system, they are involved in a 'test phase' (ca. 20 turns) of which they are told that it is necessary so that the system can adjust to the quality of their voices. In this phase the wizard is, contrary to the 'real' dialogues, cooperative. After this phase, the speakers are confronted with the fixed dialogue schema. Each of the recordings is ended by a sequence of system output 'I did not understand' and is then interrupted by the experimenter with the comment that the machine is obviously 'hung up'. The speakers are then asked to answer some questions about their satisfaction with the system, whether they believe to have been emotionally engaged and whether they have believed to be talking to a computer. None of the 36 speakers has reported that s/he has realized that the system output was created by a human 'wizard'. Afterwards

[2] The variation, in spite of the fixed dialogue schema, is due to the fact that speakers sometimes wait until "the system has recovered", and thus the 'wizard' has to initiate a nonscheduled request to propose a date in order to continue the dialogue.

[3] The 124 turns were divided into six phases of approximately 20 turns each for the quantitative analyses presented in the following sections. Phase 0 refers to the cooperative phase.

[4] The transcription conventions are as follows: <B> = breathing, <P> = pause, +/ter=/+ = repaired and aborted utterance, <uh> = hesitation. Speakers are named e0xx. Each turn is furthermore labelled according to the subdialogue and phase in which it occurs. For instance, e0585302 is the second turn uttered by speaker e058 in subdialogue 5, phase 3.

they are informed about the real purpose of the recording (that is to elicit speech in the interaction with a malfunctioning system). To sum up, the 'situation' in which the corpus is recorded is the following:

- Number of Participants: 2
- Task Orientation: appointment scheduling
- Applications Orientation: non-applications-oriented
- Domain Restriction: appointment scheduling
- Activity Types: cooperative negotiation
- Human-Machine Participation: human-to-computer
- Scenario
    - Speaker Characteristics: 36 speakers; 19 women, 17 men; native speakers of German; mostly from Northern Germany.
    - Channel Characteristics: spoken
    - Other Environment Conditions: Wizard-of-Oz scenario; speakers are 'alone with the system' during the recording.

An exerpt from a dialogue that illustrates the lacking cooperativity of the simulated system, producing its utterances according to a fixed schema, and the speaker's attempts to schedule an appointment in spite of the communication partner's uncooperative behaviour can be found below:

(1) e0582305: wie wär's denn mit dem elften ersten, von acht bis vierzehn Uhr? [*how about the 11th of January, 8am to 2 pm?*]

s0582306: Freitag, der achte erste paßt sehr gut. [*Friday the 8th of January suits me well.*]

e0582306: elfter erster, du Rübe. [*11th of January, you turnip*]

s0582307: die erste Märzwoche ist schon belegt. [*the first week of March is already occupied*]

e0582307: <Noise> <P> du brauchst mal einen Systemadministrator. elfter Januar. [*you need a system administrator. 11th of January*]

s0582308: ich habe den Termin für Sie notiert. [*I have noted down the appointment for you*]

e0582308: welchen? <P> +/ob/+ welcher Termin ist +/ter=/+ <uh> notiert für den elften Januar? [*which one? <P> whether which time was uh noted down for the 11th of January?*]

The same sequence of system utterances causes the speaker to behave completely differently at a later stage in the dialogue:

(2) e0585305: <B> fünfter Januar, zehn bis zwölf. [*5th of January, 10 to 12.*]

s0585306: Freitag, der achte erste paßt sehr gut. [*Friday the 8th of January suits me well.*]

e0585306: fünfter Januar, zehn bis zwölf. [*5th of January, 10 to 12.*]

s0585307: die erste Märzwoche ist schon belegt. [*the first week of March is already occupied*]

e0585307: fünfter Januar, zehn bis zwölf. [*5th of January, 10 to 12.*]

s0585308: ich habe den Termin für Sie notiert. [*I have noted down the appointment for you*]

e0585308: siehst du, man muß nur oft genug sagen. [*you see, one only has to repeat often enough*]

4. Variation in Speakers' Linguistic Behaviour

It can be assumed that, given the laboratory conditions of the recording and the predetermined behaviour of one of the communication partners, the situation is stable. The question to be answered now is whether speakers also behave similarly regarding the linguistic properties of their speech. If not, we have to determine what the variation is determined by. If language use is indeed determined by situational factors such as the ones listed in the discourse typology described in section 2., the linguistic properties observable may vary only with respect to the variables speakers' age and gender.

The 36 lexically, conversationally, and prosodically annotated dialogues are now analysed for the speakers' intra- and interpersonal variation.

4.1. Intrapersonal Variation

For all of the properties annotated, systematic intrapersonal variation can be found. Regarding the lexical material used, there are items, such as *wunderbar ('wonderful')*, which mostly occur in earlier dialogue phases. Other items, such as *interessant ('interesting')*, mainly occur in the middle of the dialogues. Items involved in cursing, such as *Gott ('Lord')* or *Scheiß ('shit')*, can only be found in later phases of the dialogues.

Regarding conversational strategies, especially in later phases of the dialogues, speakers may repeat their utterances irrespective of the speech act uttered by the system. Examples (1) and (2) illustrate how the speakers conversational strategies change from metalanguaging, reformulation and clarification questions to simple repetition, irrespective of what the system utters. A statistical analysis of all of the dialogues shows that while in the cooperative phase there are no repetitions at all, there are 137 occurrences of repetitions in phase two, that is, on the average, 3.81 of 20 turns in this phase are repetitions. Furthermore, when we consider the reactions to a particular utterance, for example, the system's statement that holidays will be in June and July (when the task is to find a date for an appointment in January), the likeliness that a speaker reacts by means of a repetition increases from 14% when this utterance occurs for the first time to 43% when it is uttered a third time towards the end of the dialogue. Likewise, if the system produces a sequence of incomprehensible utterances, the probability that the speakers will only repeat their utterances is five times higher when it occurs for the fifth time than when speakers are confronted with it for the first or even the second time (Fischer, 1999b). While in early phases of the dialogues speakers react directly to the system's output, that is, acknowledging what has been said and reacting relevantly, for instance, by means of reformulations and metacommunicational statements, they cease to try out different conversational strategies when they are more frustrated. Figure 1 displays the distribution of repetitions in the different phases for female and male speakers.

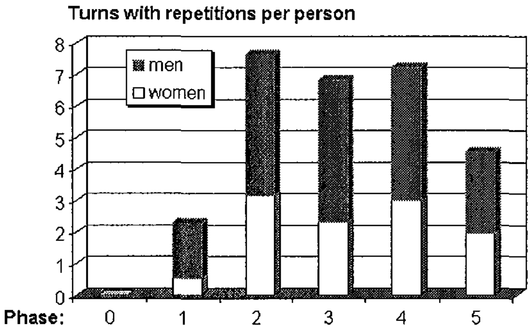


Figure 1: The Distribution of Repetitions throughout the Dialogues

Similarly, besides strictly repeating their utterances after having given up on more cooperative strategies, speakers often do not react to the system's output any more at all, producing conditionally irrelevant utterances themselves which do not relate to their partner's contributions. The distribution of these utterances is shown in Figure 2.

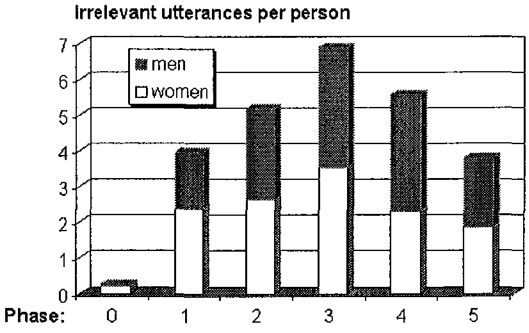


Figure 2: The Distribution of Irrelevant Utterances throughout the Dialogues

As a phonological property, hyperarticulation has been found to occur mainly in the later phases of the dialogues, such that 70% of all instances of hyperarticulation occur in the second half of the dialogues (Fischer, 1999b). Figure 3 shows the distribution of the prosodic peculiarities hyperarticulation and pausing inside words within the different phases of the dialogues.

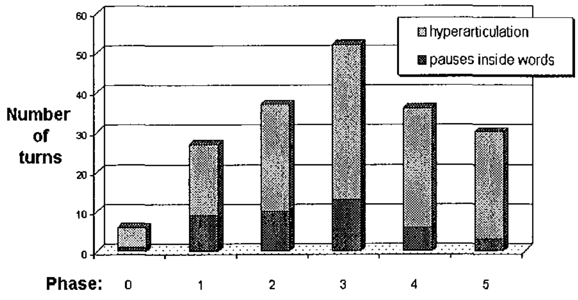


Figure 3: The Occurrence of Prosodic Peculiarities in Different Phases of the Dialogues

4.2. Interpersonal Variation

In addition to the result that speakers change their linguistic behaviour systematically during time, there are furthermore interpersonal differences between speakers. For instance, speakers can be distinguished according to which communicative strategy they prefer: There are some who prefer to reformulate and to use metalanguage even in later phases of the dialogues, and there are others who begin to repeat their utterances much more readily. There is a negative correlation of -0.6 between the use of reformulations and metalanguage on the one hand and repetitions on the other.

Regarding lexical material, only 36% of the speakers have been found to use swear words for the system, some of them however produced as many as 14 instances of such words.

Similarly, with respect to phonological and prosodic properties, while the speakers on the average have been found to produce 5.14 turns per dialogue which contain instances of hyperarticulation, there is one speaker for whom as many as 74 of such turns (of 124) can be found. This relationship is illustrated in Figure 4.

Regarding age, no significant influence could be determined. With respect to the variable speakers' gender, there are indeed systematic differences in their linguistic behaviour. The swear words discussed above, for instance, were produced by only three women but by ten men. Correspondingly, women use much fewer communicative acts in which they directly or indirectly evaluate the system's communicative behaviour. Thus, while in the speech of the 19 women investigated only eight instances of such evaluations could be found, 42 were identified for the 17 men. Prosodically, no significant gender-related differences could be found. To sum up, however, the interpersonal variation observed cannot be solely attributed to differences in gender or age.

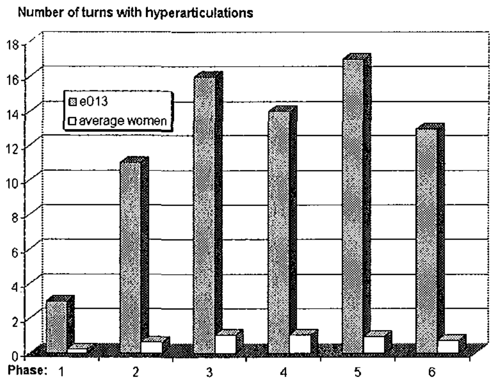


Figure 4: The Interpersonal Variation regarding Hyperarticulation for Female Speakers in Different Phases of the Dialogues

4.3. Consequences of Intra- and Interpersonal Variation for the Concept of *Situation*

Systematic intrapersonal variation could be found such that cooperative behaviour is slowly replaced by uncooperative linguistic behaviour. This is obvious for the conversational strategies employed which aim at increasing understandability for the computer in earlier phases, such as reformulation. However, also hyperarticulation, strong emphasis, syllable lengthening, inclusion of pauses etc. can be interpreted as partner-oriented, cooperative, strategies which aim at increasing the understandibilty of utterances (Oviatt et al., 1998). This is supported by the fact that speakers often cease to use these strategies after some time when they have found out that they are not helpful. Besides the varying attempts to increase understandability by means of particular strategies, the systematic intrapersonal variation can be attributed to a change in speaker attitude (Fischer, 1999c). This interpretation is supported by the fact that all subjects have reported afterwards that they were emotionally engaged.

While intrapersonal variation can be attributed to changing speaker attitude, regarding interpersonal variation an explanation for the differing linguistic behaviour is still missing. The claim made in this paper is that instead of explaining the speakers' use of linguistic properties on the basis of extralinguistic aspects of the situation, that what really determines their linguistic behaviour is how they UNDERSTAND the situation. The variable linguistic behaviour observable is thus proposed to result from different ways of conceptualizing the situation. This view is in accordance with Gumperz (1982) who writes that instead of relying "on a priori identification of social categories" (Gumperz, 1982, p.130), "linguistic diversity serves as a communicative resource in everyday life in that conversationalists rely on their knowledge and their stereotypes about variant ways of speaking to categorise events, infer intent and derive expectations about what to ensue. All this information is crucial to the maintainance of conversational involvement and to the success of persuasive strategies" (Gumperz, 1982, p.130).

5. What Speakers Understand as the Situation

The methods used to determine how speakers understand the situation include first of all the analysis of the assertions speakers make in the questionnaire they fill out after the recording. Thus, some write in the questionnaire that they have found the interaction amusing, explaining that 'he couldn't annoy me, since it is a machine without a soul,' while most speakers find it enervating and annoying. This information can indeed be found to have consequences for the linguistic properties of the dialogues: While speakers who report to have been annoyed produce 47.68 turns on the average which include instances of hyperarticulation, syllable lengthening and pausing between syllables, those who report to have been amused produce no more than 13.25 turns on the average with such linguistic properties. This difference is significant ($p < .05$, two-sided). Figure 5 illustrates this:

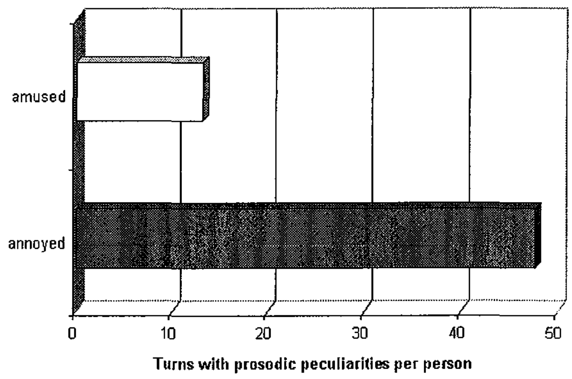


Figure 5: Number of Turns Containing Prosodic Peculiarities for Amused and Annoyed Speakers

Methodologically more importantly, however, speakers display their understanding of what the situation is like in their conversational behaviour (Sacks et al., 1974), by designing their utterances for the communication partner. For instance, the speaker in the following example displays in her repetition that she believes increased loudness, a change in her pattern of emphasis (indicated by capital letters), syllable lengthening (<L>) and pauses between the words (<P>) to increase the understandability of her speech:

(3) e0118204: am MONtag, dem VIERten ersten, von ZWÖLF Uhr bis vierzehn Uhr. (*on Monday, the 4th of January, from 12am to 2pm.*)

s0118205: Mittwoch, der sechste erste, von acht bis zehn Uhr ist schon belegt. (*Wednesday, the 6th of January, from 8 to 10 am is already occupied.*)

e0118205: <B> <:<very loud> AM MONtag:>, dem VIERten ERSten, von ZWÖLF bis VIERzehn UHR. (*on Monday, the 4th of January, from 12am to 2pm.*)

s0118206: Donnerstag, von acht bis zehn Uhr ist schon belegt. (*Thursday, from 8 to 10am is already occupied.*)

e0118206: am Mo<L>nta<L>g, <P> dem VIERten ERSten, <P> von ZWÖLF bis VIERzehn UHR. (*on Monday, <P> the 4th of January, <P> from 12am to 2pm.*)

While this speaker even increases her strategy of syllable lengthening during the dialogue, the instance of increased loudness in this example remains the only one in the dialogue. Furthermore, the speaker ceases to hyperarticulate towards the end of the interaction. Thus, the linguistic properties observable can be seen as strategies speakers try out in the interaction with their communication partner, and they cease to employ them when they realize that they do not help. That is, by using particular strategies, speakers display how they design utterances for their recipients and thus they indicate their expectations about their communication partners to them (Hausendorf, 1993), as well as to the analyst (Sacks et al., 1974).

Besides this implicit information on what speakers' may believe about the capabilities of their communication partners, speakers may also comment on the system more directly. In example (2), for instance, the speaker displays his theory that repeating his utterance often enough will lead to successful understanding. Likewise, in the following example, the speaker explicitly announces his strategy to speak very slowly:

(4) e0375303: ja, aber ich rede jetzt von abends. so. noch mal, ganz langsam. Montag, der elfte erste, von <laugh> zwanzig bis zweiundzwanzig Uhr. [*yes, but I'm talking about the evening. so. once again, very slowly. Monday, the 11th of January, 8 to 10 pm*]

In example (5) the speaker shows that she expects the system to be consistent in its utterances:

(5) e0067303: ja, eben sagten Sie aber, da wäre noch was frei. [*yes, but a minute ago you said that there is still something free*]

Besides expecting consistency, speakers assume the system also to remember what has been said before:

(6) s0025304: Freitag, der fünfzehnte erste, von sechzehn bis achtzehn Uhr ist noch frei. [*Friday the 8th of January from 4 to 6 pm is still free*]

e0025304: das sagt' ich bereits, +/da/+ zu diesem Termin kann ich nicht ... [*I've already told you that I have no time then*]

Some time later in the dialogue, the speaker, in reaction to the same proposal by the system, repeats that he has been rejecting the proposal throughout the dialogue and thus displays that he expects his communication partner to recall that he had already rejected that date:

(7) s0025304: Freitag, der fünfzehnte erste, von sechzehn bis achtzehn Uhr ist noch frei. [*Friday the 8th of January from 4 to 6 pm is still free*]

e0025304: das geht schon die ganze Zeit bei mir nicht ... [*this has not been working for me the whole time*]

In contrast, in the following example, the speaker displays her understanding of the system's behaviour as helpful:

(8) e0323204: Dienstag, vierzehnter erster neunzehnhundertneunundneunzig, von achtzehn bis zweiundzwanzig Uhr. (*Tuesday the 14th 1999 from eighteen to twenty-two hundred hours.*)

s0323205: Donnerstag von acht bis zehn Uhr ist schon belegt. (*Thursday from eight to ten is already occupied.*)

e0323205: (...) am Donnerstag, den vierzehnten ersten neunzehnhundertneunundneunzig, achtzehn bis zweiundzwanzig Uhr. (*on Thursday the 14th 1999, eighteen to twenty-two hundred hours.*)

s0323206: ich habe den Termin für Sie notiert. (*I have noted down the appointment for you.*)

e0323206: jetzt versteh' ich. ich muß deutlich sagen den Unterschied zwischen acht und achtzehn. danke für den Hinweis. (*now I understand. I have to say clearly the distinction between eight and eighteen. thank you for the tip.*)

In another instance, she suggests that the problems they may have with each other may result from the fact that they do not work with the same calendar. At the same time, she treats the system as if it was another, embodied, human being, having to have the calendar 'in front of him:'

(9) e0322203: <B> Montag, der achtzehnte erste neunzehnhundertneunundneunzig, von acht bis dreizehn Uhr. haben Sie da Zeit? (*<B> Monday the 18th of January 1999 from 8am to 1pm. do you have time then?*)

s0322301: der siebte Februar ist ein Sonntag. (*the 7th of February is a Sunday*)

e0322301: jetzt haben wir uh große Mißverständnisse. <B> uh sind Sie nicht auch haben Sie nicht den Plan von Januar vor sich liegen? <P> wir müssen Termine im Januar vergeben. (*now we have uh great misunderstandings. <B> uh aren't you don't you also have the calendar for February in front of you? <P> we have to schedule appointments in January.*)

Her understanding of her communication partner as a human being is mirrored in her linguistic behaviour which mainly consists in reformulating and explaining her intentions. Her understanding of her communication partner also becomes clear through the following utterance:

(10) e0324304: so, so. Sie verwechseln jetzt den Wochentag mit der Uhrzeit. da haben wir ja beide Probleme. vielleicht sind Sie Ausländer. [*oh well. you are mixing up the day with the time. seems we both have problems with that. maybe you are a foreigner*]

As a last example, speaker e004 switches from the formal form of address in German to the informal one during the interaction:

(11) e0043205: <B> ich würde dann gerne mit **Ihnen** diesen vier–stündigen Termin am Montag, den achtzehnten Januar, um acht Uhr morgens vereinbaren. [*I would like to make this four-hour meeting with you on Monday the 18th of January at 8am*]

s0044101: ich habe den Termin für Sie notiert. <P> bitte machen Sie einen Vorschlag. [*I have noted down the appointment for you <P> please make a suggestion*]

e0044101: könnten wir uns nicht am Sonntag, den zehnten Januar, +/a/+ gegen Abend, achtzehn Uhr treffen? [*couldn't we meet on Sunday the 10th of January in the evening, 6pm?*]

s0044102: dieser Termin ist schon belegt. [*this date is already occupied*]

e0044102: und eine Woche später, Sonntag, der siebzehnte Januar. [*and a week later, Sunday, the 17th of January*]

s0044103: ich habe Sie nicht verstanden. [*I did not understand*]

e0044103: am Sonntag, den siebzehnten Januar, hast **Du** denn da noch Zeit, um **Dich** mit mir zu treffen, sagen wir vierzehn Uhr.[*on Sunday, the 17th of January, would you have still time to meet me, say at 2pm*]

The speaker's linguistic behaviour is characterized by reformulations and, correspondingly, the lack of repetitions, assessments, and criticism. Thus, her conversational behaviour is cooperative till the end. Prosodically, there are no more than three instances of particularly strong emphasis on certain syllables, only two instances of syllable lengthening, and no instances of hyperarticulation or pauses inside words. Thus, her informal address of her communication partner (note also the *wir* 'we' in e0044101) corresponds to her treatment of the system as a familiar person. This is furthermore mirrored in her following suggestion:

(12) e0045206: <B> können Sie denn Ihre Mittagspause auch erst um vierzehn Uhr machen, so daß wir uns dann treffen können? [*could you perhaps take your lunch as late as 2 pm so that we could meet then?*]

While she has switched to the formal form of address again at this later stage of the dialogue, she pretends her communication partner to eat lunch, that it has a particular time when to eat lunch, and that it would mean a concession for it to postpone it. Thus, she treats her communication partner as if it was an embodied person.[5] This conceptualization of her partner determines not only what the situation means to her, namely a cooperative negotiation of dates when to meet, it also influences the lexical, conversational, and prosodic properties of the language she uses.

6. Conclusions

In this paper it was shown that what a particular situation consists in is not entirely defined by external variables. Instead it depends also on how speakers conceptualize the situation. The corpus investigated thereby provides a unique opportunity to study the contribution of a single speaker's conceptualization of the situation. As a means of identifying what speakers believe about the situation, data from questionnaires were used as well as implicit and explicit information from the interactions with their communication partner. The different conceptualizations of the situation have been found to have correlates in the linguistic behaviour observable. However, the conclusion cannot be that if the conceptualization of the situation is taken into account, the speakers' linguistic behaviour could be completely predicted. Instead, the speakers have been found to constantly define and, if necessary, redefine what they understand the situation to consist in, depending on their current hypotheses about their communication partners and on their own emotional state.

Methodologically, the results from this study point to ways of how to determine what a situation consists in, at least for the speakers themselves: Since speakers constantly display their understanding of the situation to their communication partners, as well as to the analysts (Sacks et al., 1974), we just need to look at the speakers linguistic behaviour in order to determine what the situation is all about. This would help us to construe the category *situation* not as an *a priori* descriptive category based on our intuitions, but as a speaker category, that is, as a complex concept to which the speakers themselves can be shown to attend.

7. References

J.K. Chambers and P. Trudgill. 1998. *Dialectology*. Cambridge University Press, 2nd edition.

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

H.H. Clark. 1999. How do real people communicate with virtual partners? In *Proceedings of AAAI-99 Fall Symposium, Psychological Models of Communication in Col-*

[5] Clark (1999) argues that human-computer interaction is a matter of 'joint pretense.' Speakers' e004 and e032 treatment of the supposed system, and these speakers, as much as all the others, answered in the questionnaire that they did not doubt the existence of such a system, as another human being may support Clark's hypothesis that human-computer interaction is actually carried out on two different layers: one in which the speakers create a layer of joint pretense, and a second one in which the speakers communicate.

laborative Systems, November 5-7th, 1999, North Falmouth, MA. Menlo Park, Calif.: AAAI Press; Cambridge, Mass.: MIT Press.

EAGLES. 1998. Representation and annotation of dialogue (recommendations by the eagles workpackage 4 on integrated spoken and written language resources). Technical report, to appear as supplement to the Handbook of Standards and Resources for Spoken Language Systems, Berlin, New York: Mouton de Gruyter, EAGLES homepage.

J. Ellis and J.N. Ure. 1969. Language varieties: Register. In A.R. Meetham and R.A. Hudson, editors, *Encyclopedia of Linguistics. Information and Control*, pages 251–259. Oxford: Pergammon.

R. Fasold. 1990. *Introduction to Sociolinguistics*. Oxford: Blackwell.

J.R. Firth. 1957. Ethnographic analysis and language with reference to malinowski's views. In F.R. Palmer, editor, *Selected Papers of J.R.Firth, 1952-1959, republished 1968*, pages 93–118. London: Longman and Bloomington: Indiana University Press.

K. Fischer. 1999a. Annotating emotional language data. Technical Report 236, Verbmobil.

K. Fischer. 1999b. Discourse effects on the prosodic properties of repetitions in human-computer interaction. In *Proceedings of the ESCA-Workshop on Dialogue and Prosody, September 1rst - 3rd, 1999, De Koningshof, Veldhoven, The Netherlands*, pages 123–128.

K. Fischer. 1999c. Repeats, reformulations, and emotional speech: Evidence for the design of human-computer speech interfaces. In Hans-Jörg Bullinger and Jürgen Ziegler, editors, *Human-Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany.*, pages 560–565. Lawrence Erlbaum Ass., London.

N. Fraser and G.N. Gilbert. 1991. Simulating speech systems. *Computer Speech and Language*, 5:81–99.

J. Gumperz. 1982. *Discourse Strategies*. Number 1 in Studies in Interactional Sociolinguistics. Cambridge University Press.

M.A.K. Halliday, A. McIntosh, and P. Strevens. 1964. *The Linguistic Sciences and Language Teaching*. London: Longman.

H. Hausendorf. 1993. Die Wiedervereinigung als Kommunikationsproblem. Technical Report 6/93, 'Nationale Selbst- und Fremdbilder in osteuropäischen Staaten - Manifestationen im Diskurs.' Zentrum für interdisziplinäre Forschung der Universität Bielefeld.

D. Hymes. 1972. Models of the interaction of language and social life. In John J. Gumperz and Dell Hymes, editors, *Directions in Socolinguistics. The Ethnography of Communication*. Holt, Rinehart and Winston, New York et al.

J. Krause and L. Hitzenberger, editors. 1992. *Computer Talk*. Hildesheim: Olms Verlag.

J. Lyons. 1977. *Semantics*. Cambridge University Press.

S. Oviatt, J. Bernard, and G.-A. Levow. 1998. Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech*, 41(3-4):419–442.

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

Towards an Analysis of Dialogue Acts and Indirect Speech Acts in a BDI Framework

Andreas Herzig, Dominique Longin, Jacques Virbel

Institut de Recherche en Informatique de Toulouse (IRIT),
Université Paul Sabatier,
118 Route de Narbonne, F-31062 Toulouse Cedex 4
mailto: {herzig,longin,virbel}@irit.fr

Abstract

We study the dynamics of belief in cooperative task-oriented man-machine dialogues. We introduce a modal logic of action, belief and intention, where intention has a non-normal modal logic. We focus on two aspects of speech acts: we define a semantics to take into account a feedback from the adressee of a speech act; we characterize indirect speech acts.

1. Introduction

Task-oriented man-machine dialogue is one of the most important challenges for AI. Participants in such dialogues have one major common goal, viz. to achieve the task under concern. Each of the participants has some information contributing to the achievement of the goal, but none of them can achieve it alone.

Cooperativity is a fundamental and useful hypothesis. Informally, an agent is cooperative w.r.t. another one if the former helps the latter to achieve his goals (cf. Grice's cooperation principles, as well as his conversation maxims (Grice, 1989)).

Cooperativity does not always entail *sincerity* because *unsincerity* may serve cooperativity (Longin, 1999). We suppose here that each participant is *sincere.* This means that his utterances faithfully mirror his beliefs: if a participant says "the sky is blue" then he indeed believes that the sky is blue. Such a hypothesis entails that contradictions between the presuppositions of an utterance and the hearer's beliefs about the speaker cannot be explained in terms of lies.

We finally suppose that utterances are *public*: all the agents perceive the performance of an utterance (but might misinterpret them).

The background of our work is an effective generic realtime cooperative dialogue system which has been specified and developped by the France Telecom R&D Center, as an instantiation of a rational agent technology called ARTIMIS (Sadek et al., 1997, 1996). This approach consists in first describing the agent's behaviour within a logical *theory of rational interaction* (Sadek, 1991a, 1991b, 1992, 1994), and second implementing this theory by an *inference engine* (Sadek et al., 1997; Bretier, 1995). The latter is the kernel of ARTIMIS. For a fixed set of domains, this system is able to accept nearly unconstrained spontaneous language as input, and react in a cooperative way. The activities of the dialogue system are twofold: to take into account the speaker's utterances, and to generate appropriate reactions. The reactive part is completely defined in the current state of both the theory and the implementation. On the other hand, the consommation of an utterance is handled only partially, in particular its belief change part.

In the next sections we introduce the ingredients of our BDI framework, summarizing (Herzig and Longin, 1999a). First we define our multimodal language (Sect. 2.). Then we give a simple theory of topics: we associate a set of topics to every formula (its subject), every agent (his competence), and action (its scope) (Sect. 3.). Then we define a topic-based possible worlds semantics of dialogue acts, with an appropriate semantics of intention and of the update of the agents' mental states after an action. It also integrates (possibly non-linguistic) actions of feedback (Sect. 4.). Finally we sketch how indirect speech acts can be inferred in that framework (Sect. 5.).

2. The multimodal language

As most of the authors, we work in a multi-modal framework, with modal operators for belief, mutual belief, intention and action. Our language is that of first-order multimodal logic without equality and without function symbols.

Let AGT be the set of agents. For every $i, j \in AGT$, there are *modal operators* Bel_i, $Intend_i$ and $Bel_{i,j}$. $Bel_i A$ is read "agent i believes that A", and $Intend_i A$ is read "agent i intends that A". $BelIf_i A$ abbreviates $Bel_i A \vee Bel_i \neg A$. $Bel_{i,j} A$ is read "i and j mutual believe that A". For example in a ticket selling scenario, $Bel_u Dest(\text{Göteborg})$ expresses that the agent u believes that the destination is Göteborg.[1]

Speech acts (Austin, 1962; Searle, 1969) are represented by tuples of the form $\langle \mathsf{FORCE}_{i,j}\ A \rangle$ where

- $\mathsf{FORCE} \in \{\mathsf{Assert}, \mathsf{Inform}, \mathsf{Request}, \mathsf{QueryYN}, \mathsf{QueryWh}\}$ is the illocutionary force of the act,
- $i, j \in AGT$, and
- A is a formula (the propositional content of the act).

[1] In our ticket selling scenario, u is the user of a man-machine dialogue system s.

For example, $\langle \mathsf{Inform}_{u,s}\ Dest(\text{Göteborg})\rangle$ represents a declarative utterance of the user u informing the system s that his destination is Göteborg; $\langle \mathsf{QueryYN}_{u,s}\ Bel_s Price(150\,€)\rangle$ means « the user asks the system if he believes the price is 150 € ».

Actions are either speech acts, or physical actions such as $\langle \mathsf{Give}_{i,j}\ \text{salt}\rangle$ (the agent i gives the salt to the agent j). Let ACT be the set of all actions, containing in particular all speech acts. With every $\alpha \in ACT$ we associate *modal operators* $Done_\alpha$ and $Feasible_\alpha$. $Done_\alpha A$ is read "α has just been performed, before which A was true". $Done_\alpha \top$ is read "α has just been performed". $Feasible_\alpha A$ is nothing but the $\langle\alpha\rangle A$ of dynamic logic (Harel, 1984), while $Done_\alpha A$ is $\langle\alpha^{-1}\rangle A$.

$Bel_s Dest(\text{Göteborg})$ is an example of a formula. Another one is $Bel_s(Dest(\text{Göteborg}) \wedge Class(\text{1st}) \rightarrow Price(150\,€))$. The latter is also a nonlogical axiom (alias domain law), allowing the system to deduce the price from information about destination and class. Another meaningful example is the formula $Done_{\langle \mathsf{Inform}_{u,s}\ p\rangle} Bel_u p$ expressing the sincerity of u (which is also a domain law) : the agent u has just informed the agent s that p, before that u believed that p.

Atomic formulas are noted $p, q, \ldots$ or $P(t_1, \ldots, t_n)$. Atm is the set of all atomic formulas. Formulas are noted $A, B, \ldots$.

3. Adding topics

The competence of agents and the influence of speech acts. Which mental attitudes of an agent can 'survive' the performance of a speech act α? In our approach, we proceed in two steps: the hearer always accepts the indirect and intentional effects, but not all of their consequences. We consider that if there exists a relation of influence of α towards an attitude, then the latter cannot be preserved in the new mental state.

A topic-based approach. All this presupposes that we are able to determine the competence of an agent and the influence of a speech act. In our approach, we base both notions on the concept *topics*. This is a natural and intuitively appealing concept, and it allows us to fine-tune the consommation of speech acts.

Topics are well studied in linguistics and philosophy (Ginzburg, 1995; van Kuppevelt, 1995; Lewis, 1972). Epstein (Epstein, 1990) associates to a formula its *subject matter*, and defines two formulas as being related if they have some subject matter in common. Generalizing his idea, we associate a set of topics to every agent i, speech act α, and formula A. Then we consider that i is competent at a topic if and only if that topic is associated to i. And a speech act α influences a topic if that topic is associated to α.

We have developped a metalinguistic theory of topics in (Herzig and Longin, 1999a). We give a brief overview in the rest of the section.

Themes and topics. Topics are themes in context. We suppose that there is a nonempty set of *themes*, such as destinations, classes, and prices in a ticket selling scenario. Our notion of theme is very closed to the Epstein's *subject-matter*. But we need here a more subtil notion and we introduce now the notion of context.

For $i \in AGT$, ma_i is called an *atomic context*. ma_i stands for "the mental attitude of agent i". A *context* is a possibly empty sequence of atomic contexts, noted $ma_{i_1}{:}ma_{i_2}{:}\ldots ma_{i_n}$. A theme t together with a context c makes up a *topic of information*, noted $c{:}t$. For example, $ma_u{:}price$ is a topic consisting in the user's mental attitude at prices, and $ma_s{:}ma_u{:}price$ is a topic consisting in the system's mental attitude at the user's mental attitude at prices.

The empty context is noted λ. By convention

$$\lambda{:}c = c{:}\lambda = c, \tag{1}$$

$$\lambda{:}t = t. \tag{2}$$

Moreover, we require $ma_i{:}ma_i = ma_i$. This identity is justified by principles of introspection that are valid in standard modal logics of belief.

Given a set of themes and a set of agents we note $\mathbb{T}$ the associated set of topics.

The subject of a formula. The *subject of a formula* is the set of topics the formula is about: $\mathfrak{subject}(A) \subseteq \mathbb{T}$. For example, $\mathfrak{subject}(Bel_s\, Class(\text{1st})) = \{ma_s{:}class\}$, and expresses that $Bel_s\, Class(\text{1st})$ is about the system's attitude at classes.

By (1) and (2), every theme is a topic and then, e.g. $\mathfrak{subject}(Dest(\text{Göteborg})) = \{\lambda{:}dest\} = \{dest\}$.

The competence of an agent. We suppose that we can associate to each agent a set of topics, the topics he is *competent* at: $\mathfrak{competence}(i) \subseteq \mathbb{T}$. For example, in our ticket selling scenario, the user is competent at destinations and classes, but not at prices.

The scope of an act. The *scope* of an act α tells us which mental attitudes of an agent are influenced by this act: $\mathfrak{scope}(\alpha) \subseteq \mathbb{T}$. An act always influences the hearer's beliefs about the speaker's attitude towards the propositional content. Consider e.g. the speech act where the user informs the system about the ticket price. This speech act influences the system's beliefs about the user's attitude towards prices. Hence $ma_s{:}ma_u{:}price \in \mathfrak{scope}(\langle \mathsf{Inform}_{u,s}\ Price(150\,€)\rangle)$. Formally, if $t \in \mathfrak{subject}(A)$ then

$$ma_j{:}ma_i{:}t \in \mathfrak{scope}(\langle \mathsf{FORCE}_{i,j}\ A\rangle) \tag{3}$$

for every illocutionary force FORCE. In the case of request, these mental attitudes are typically the only ones that are influenced.

Interactions. Is there a relationship between these functions? We propose the following axiom for acts of the informative type.

$$\text{If } \alpha = \langle \mathsf{Inform}_{i,j}\ A\rangle \text{ and } t \in \mathfrak{themes}(A) \cap \mathfrak{competence}(i)$$
$$\text{then } t \in \mathfrak{scope}(\alpha) \text{ and } ma_j{:}t \in \mathfrak{scope}(\alpha). \tag{4}$$

Suppose e.g. the user informs the system about his destination. As the user is competent at destinations, this influences the system's factual beliefs about the destination. And also about prices: a new destination means a new price. Hence $\mathfrak{scope}(\langle \mathsf{Inform}_{u,s}\ \mathit{Dest}(\text{Göteborg})\rangle)$ contains *dest*, *price*, ma_s*:dest* and ma_s*:price*. Postulates for other illucutionary forces are in (Longin, 1999).

4. Towards a semantics of dialogue acts

We aim at a semantics having both a complete axiomatization and an associated automated deduction procedure. This has motivated several choices, in particular a Sahlqvist-type possible worlds semantics (Sahlqvist, 1975), for which general completeness results exist, and a notion of intention that is primitive (contrarily to the complex constructions in the literature). Intentions have a non-normal modal logic, reflecting that they are not closed under conjunction and implication. They can nevertheless be reduced to the Sahlqvist framework (Fariñas del Cerro and Herzig, 1995).

Semantics is in terms of *possible worlds models* $\langle W, \mathcal{S}, D, V\rangle$, where

- W is a set of worlds;
- D is the domain of objects;
- V is a valuation mapping variable and constant symbols to elements of D, and associating to each possible world $w \in W$ an interpretation V_w of predicate symbols;
- $\mathcal{S}$ is a collection of structures on W, consisting of
  - partial functions

$$\mathcal{D}_\alpha : W \longrightarrow W \text{ for every } \alpha \in ACT, \tag{5}$$

  - mappings

$$\mathcal{B}_i : W \longrightarrow 2^W \text{ for every } i \in AGT \tag{6}$$

  - and mappings

$$\mathcal{I}_i : W \longrightarrow 2^{2^W} \text{ for every } i \in AGT. \tag{7}$$

The $\mathcal{B}_i$ are accessibility relations as usual. The set of possible worlds $\mathcal{B}_i(w)$ is called the *belief state* of i. The partial functions $\mathcal{D}_\alpha$ correspond to deterministic accessibility relations. $\mathcal{D}_\alpha(w)$ represents the possible result of α. The $\mathcal{I}_i$ are *neighbourhood functions* (Chellas, 1980, Chap. 7). Every set of worlds $U \in \mathcal{I}_i(w)$ stands for an intention of i.

The satisfaction relation $\Vdash$ is defined as usual. A useful abbreviation is $[\![A]\!] = \{w \in W : w \Vdash A\}$, called the *extension* of the formula A. Then

$$\begin{gathered} w \Vdash P(t_1, \ldots, t_n) \\ \text{if } \langle V_w(t_1), \ldots, V_w(t_n)\rangle \in V_w(P); \end{gathered} \tag{8}$$

$$\begin{gathered} \text{the standard truth conditions for the connectives} \\ \text{of classical logic are still true;} \end{gathered} \tag{9}$$

$$\begin{gathered} w \Vdash \mathit{Feasible}_\alpha A \\ \text{if } \mathcal{D}_\alpha(w) \text{ is defined and } \mathcal{D}_\alpha(w) \in [\![A]\!]; \end{gathered} \tag{10}$$

$$\begin{gathered} w \Vdash \mathit{Done}_\alpha A \\ \text{if } \mathcal{D}_\alpha^{-1}(w) \text{ is defined and } \mathcal{D}_\alpha^{-1}(w) \in [\![A]\!]; \end{gathered} \tag{11}$$

$$w \Vdash \mathit{Bel}_i A \text{ if } \mathcal{B}_i(w) \subseteq [\![A]\!]; \tag{12}$$

$$w \Vdash \mathit{Intend}_i A \text{ if } [\![A]\!] \in \mathcal{I}_i(w). \tag{13}$$

Contrarily to previous work of ours in (Herzig and Longin, 1999a), our notion of intention is not necessarily closed under logical truth, logical consequence, conjunction, and material implication.[2] This is in accordance with Bratman's (1987), Cohen&Levesque's (1990a; 1990b) and Sadek's (1992) analyses of intention. Contrarily to these approaches, intention is primitive here, as in (Rao and Georgeff, 1991) and (Konolige and Pollack, 1993). We thus generalize the semantics in (Konolige and Pollack, 1993), where only closure under logical consequence had been given up. The only principle that is valid is

$$\frac{A \equiv B}{\mathit{Intend}_i A \equiv \mathit{Intend}_i B} \tag{14}$$

We have chosen this solution for three reasons. First, building intention from other primitive notions such as goals or desires leads to various sophisticated notions of intention, with subtle differences between them. We have kept here only those properties common to all of them, viz. extensionality. Second, the definitions in the literature are rather complex, and it is difficult to find complete automated theorem proving methods for them, while our analysis enables more or less standard completeness techniques and proof methods. Third and most importantly, we think that our simplified notion of intention is sufficient in most applications.

Indeed, we think that rather than the interaction of intentions with goals or desires, it is their interaction with belief which is crucial. For example, an agent i should abandon his intention that A when he starts to believe that A. This is guaranteed by the semantical constraint

$$U \cap \mathcal{B}_i(w) = \emptyset \text{ for every } U \in \mathcal{I}_i(w). \tag{15}$$

Syntactically, this corresponds to validity of the axiom schema

$$\mathit{Intend}_i A \rightarrow \mathit{Bel}_i \neg A \tag{16}$$

[2] Hence our semantics does **not** validate

$$\frac{A}{\mathit{Intend}_i A}$$

$$\frac{A \rightarrow B}{\mathit{Intend}_i A \rightarrow \mathit{Intend}_i B}$$

$$\mathit{Intend}_i A \wedge \mathit{Intend}_i B \rightarrow \mathit{Intend}_i(A \wedge B)$$

$$\mathit{Intend}_i A \wedge \mathit{Intend}_i(A \rightarrow B) \rightarrow \mathit{Intend}_i B$$

which are all valid in any normal modal logic.

This axiom together with the standard axiom for belief

$$Bel_i \neg A \rightarrow \neg Bel_i A \tag{17}$$

entails

$$Bel_i A \rightarrow \neg Intend_i A \tag{18}$$

as expected.

Models must satisfy several other *constraints*. In particular, if an atom is independent of α then its truth value should be preserved. Hence we would like to validate

$$A \rightarrow \neg Done_\alpha \neg A \text{ if } \mathfrak{scope}(\alpha) \cap \mathfrak{subject}(A) = \emptyset \tag{19}$$

This is guaranteed by the semantical constraint

$$\begin{array}{c}\text{If } \mathcal{D}_\alpha(w) \text{ is defined and } \mathfrak{subject}(p) \cap \mathfrak{scope}(\alpha) = \emptyset \\ \text{then } \{p \in Atm \mid w \Vdash p\} = \{p \in Atm \mid \mathcal{D}_\alpha(w) \Vdash p\}\end{array} \tag{20}$$

We have defined other constraints warranting preservation of beliefs and intentions that are independent of a given act. We have also defined a topic-based belief adoption constraint stipulating that belief should amount to knowledge in the case of competence; formally we thus validate the axiom schema

$$Bel_i A \rightarrow A \text{ if } \mathfrak{subject}(A) \subseteq \mathfrak{competence}(i) \tag{21}$$

The semantical constraint associated to the above axiom is as following :

$$\begin{array}{c}\text{For every } w \in W \text{ and every agent } i \\ \text{there is some } v \in \mathcal{B}_i(w) \text{ such that} \\ Atm_W(w, \mathfrak{competence}(i)) = \\ \qquad Atm_W(v, \mathfrak{competence}(i)).\end{array} \tag{22}$$

This means that in the belief state of i there is a "witness world" mirroring the part of the actual world i is competent at.

Following Sadek (1991b), we associate with each speech act its preconditions and effects. Consider e.g. the informative act $\langle \mathsf{Inform}_{i,j}\ A \rangle$. It has the sincerity precondition $Bel_i A$ and the precondition of relevance to the context $\neg Bel_i BelIf_j A$. $\langle \mathsf{Inform}_{i,j}\ A \rangle$ has an intentional effect (in the Gricean sense, viz. $Intend_i Bel_j Intend_i Bel_j A$), an indirect effect (viz. the persistence of preconditions after the performance of the speech act), and a perlocutionary effect (expected effect).

Our agents being autonomous, the expected effect of an act does not obtain systematically. This means that the perlocutionary effect is not always consumed, in the sense that the propositional content is not necessarily added to the hearer's belief state. In the case where the new mental state (obtained by the admission of a speech act and the consommation of his indirect and intentional effects) entails the perlocutionary effect, then we say that the latter has been consumed.

The indirect effect of precondition preservation is problematic in the case of the relevance precondition: it means that the speaker believes that the expected effect of his speech act has not been consumed by the hearer. Our idea is to integrate a *feedback* from the hearer into the speech act. Thus, the latter is considered to be completely performed when the feedback has been consumed by the speaker.

In this perspective, we propose that the relevance precondition should not be preserved, but should be (transiently) replaced by a particular effect. This effect must express that transitionally, the speaker doesn't know anything about the hearer's attitude towards the propositional content of his act. For example, after the informative speech act $\langle \mathsf{Inform}_{i,j} A \rangle$, the speaker i ignores whether the adressee j believes A, i.e. neither $Bel_i \neg BelIf_j A$ nor $BelIf_i Bel_j A$ should hold. This means that $\langle \mathsf{Inform}_{i,j} A \rangle$ should influence the speaker's mental attitudes towards those of the hearer about the subject of A. Formally, if $t \in \mathfrak{themes}(A)$ then

$$ma_i{:}ma_j{:}t \in \mathfrak{scope}(\langle \mathsf{Inform}_{i,j}\ A \rangle) \tag{23}$$

Feedback actions are not necessarily speech acts. Being about some proposition A, they take the form

$$\langle \mathsf{Feedback}_{h,s}\ A \rangle.$$

Semantically, such actions are very similar to informative speech acts. In particular they have the same scope:

$$\mathfrak{scope}(\langle \mathsf{Feedback}_{i,j}\ A \rangle) = \mathfrak{scope}(\langle \mathsf{Inform}_{i,j}\ A \rangle). \tag{24}$$

They are related to dynamic logic test actions. Indeed, we may consider that $\langle \mathsf{Feedback}_{h,s} A \rangle$ amounts to testing truth of the formula $Bel_h A$.

5. Towards a formalisation of indirections

In this section we propose an analysis of indirect speech acts in our framework. We illustrate our purpose by the case of directive speech acts. Consider the following utterances:

1. *Give me the salt!*
2. *I ask you to give me the salt!*
3. *You give me the salt.*
4. *You can give me the salt.*
5. *Can you give me the salt?*
6. *I want you to give me the salt.*
7. *You must give me the salt.*

The *main speech act* is the act intended by its author. It can be performed directly or indirectly. In the above utterances, the main speech act is an order from the speaker s to the hearer h to give him the salt. This main speech act is direct in the cases of (1.) and (2.), and indirect in the cases of utterances from (3.) to (7.). With respect to Searle's point of view of indirection notion, if the mean of the utterance and the mean of the speaker match, the speech act performed is purely direct, and indirect(ly performed) else (Searle, 1979).

Let $\alpha_1 = \langle \mathsf{Request}_{s,h}\ p \rangle$ be the speech act corresponding to the utterance (1.). Suppose $\beta = \langle \mathsf{Give}_{h,s}\ \text{salt} \rangle$ is the action requested in α_1. (Hence the propositional content p of α_1 is $Done_\beta \top$.) From Searle's point of view, Σ_{α_1}, the *preparatory precondition* of α_1, is $Feasible_\beta\top$ (viz. the hearer can perform β). And its *sincerity precondition* Ψ_{α_1} is $Intend_s Done_\beta \top$ (s wants h to perform β).

Roughly speaking, α_2 (representing (2.)) can be identified with α_1. More precisely, the latter is an *implicit direct speech act*, while the former is an *explicit direct speech act*. The main speech act of (1.) and (2.) being direct, their main speech act is identified with α_1 (and α_2).

Suppose now the main speech act is obtained indirectly. How can we infer it from the associated litteral act? According to Virbel (Virbel, 1999), a small set of rules is enough to characterize a large set of indirect speech acts[3]. For example, to obtain the main speech act of (1.) (viz. α_1), we can:

- assert its propositional content p (as in (3.))
- assert its preparatory condition (as in (4.))
- query about its preparatory condition (as in (5.))
- assert its sincerity condition (as in (6.))
- assert on the obligation to perform β (as in (7.))

We thus have a set of simple and elegant properties of indirect speech acts. Utterances from (3.) to (7.) can be respectively represented by the following speech acts:

- $\alpha_3 = \langle \mathsf{Assert}_{s,h}\ p \rangle$,
- $\alpha_4 = \langle \mathsf{Assert}_{s,h}\ \Sigma_{\alpha_1} \rangle$,
- $\alpha_5 = \langle \mathsf{QueryYN}_{s,h}\ \Sigma_{\alpha_1} \rangle$,
- $\alpha_6 = \langle \mathsf{Assert}_{s,h}\ \Psi_{\alpha_1} \rangle$,
- $\alpha_7 = \langle \mathsf{Assert}_{s,h}\ MustPerform(h, \beta) \rangle$.

When can we infer an indirect act α' from a direct α? For example, consider $\alpha_5 = \langle \mathsf{QueryYN}_{s,h}\ \Sigma_{\alpha_1} \rangle$, querying the preparatory precondition $\Sigma_{\alpha_1} = Feasible_\beta\top$. The precondition of context relevance of α_5 is $\neg BelIf_s Feasible_\beta \top$. If h presupposes that s has satisfied the preconditions of α_5, then $Bel_h Done_{\alpha_5} \neg BelIf_s Feasible_\beta \top$ holds. Suppose now that we are in a situation where $Bel_h Done_{\alpha_5} Bel_h Bel_s Feasible_\beta \top$ holds in the memory of the hearer. Then this type of indirection is characterized by the inconsistency of the preservation of (a part of) the memory with the preservation of the presuppositions. This makes it possible to infer (at the meta-level) $Done_{\alpha_1}\top$ from $Done_{\alpha_5}\top$.

We can generalize this approach to the other indirections, and we hope also to other classes of speech acts (such as indirect assertive speech acts).

Our analysis is compatible with our approach to belief dynamics: all we have to do in the belief preservation and adoption process is to take into account the set of preconditions and effects. Thus, as the feedback is viewed as a particular action, the hearer can admit it in the same way as the other actions.

6. Conclusions

We have defined a multimodal logic of actions, beliefs, and intentions, integrating both speech acts and physical actions. We have provided a simple non-normal semantics for intentions, and thus removed some problems concerning closure of intentions under logical consequence and conjunction.

We have also refined the analysis of speech acts towards a two-step process, taking into account feedback actions.

Finally, we have sketched how to characterize indirect speech acts within that framework.

7. Thanks

This work has been supported by France Telecom - CNET, scientific area "Intelligentes Interactions and Dialogue", in the framework of contract 97 1B 046. Thanks to David Sadek, Philippe Bretier, Rémi Faure and anonymous referees for useful discussions and/or comments.

8. References

John L. Austin. 1962. *How To Do Things With Words*. Oxford University Press.

Michael E. Bratman. 1987. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA.

Philippe Bretier. 1995. *La communication orale coopérative : contribution à la modélisation logique et à la mise en œuvre d'un agent rationnel dialoguant*. Thèse de doctorat en informatique, Université Paris Nord, Paris, France.

B. F. Chellas. 1980. *Modal Logic: an introduction*. Cambridge University Press.

Philip R. Cohen and Hector J. Levesque. 1990a. Intention is choice with commitment. *Artificial Intelligence Journal*, 42(2–3):213–261.

Philip R. Cohen and Hector J. Levesque. 1990b. Persistence, intentions, and commitment. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, chapter 3, pages 33–69. MIT Press, Cambridge, MA.

R. L. Epstein. 1990. *The Semantic Foundations of Logic Volume 1: Propositional Logic*. Kluwer Academic Publishers.

Luis Fariñas del Cerro and Andreas Herzig. 1995. Modal deduction with applications in epistemic and temporal logics. In Dov M. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 4: Epistemic and Temporal Reasoning, pages 499–594. Oxford University Press.

J. Ginzburg. 1995. Resolving questions I,II. *Linguistics and Philosophy*, 18:567–609.

[3] A large set of (meta-)rules describing indirections can be found in (Bretier, 1995). Although the Bretier's approach and the Virbel's approach have been both developped separately the one from the other, there are several similarities between them. We adopt here the Virbel's approach because it seams to us more general

H. Paul Grice. 1989. *Studies in the way of words.* Harvard University Press, USA, 3rd edition.

David Harel. 1984. Dynamic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume II. D. Reidel Publishing Company.

Andreas Herzig and Dominique Longin. 1999a. Belief dynamics in cooperative dialogues. In Jan van Kuppevelt, Noor van Leusen, Robert van Rooy, and Henk Zeevat, editors, *Proceedings Amsterdam Workshop on the Semantics and Pragmatics of Dialogue (Amstelogue'99).* Available from web site. 13 pages (extension de (Herzig and Longin, 1999b)) – To appear.

Andreas Herzig and Dominique Longin. 1999b. Belief dynamics in cooperative dialogues. In Noor van Leusen, Robert van Rooy, and Henk Zeevat, editors, *Preproceedings Amsterdam Workshop on the Semantics and Pragmatics of Dialogue (Amstelogue'99)*, Amsterdam, May. 5 pages.

Kurt Konolige and Martha E. Pollack. 1993. A representationalist theory of intention. In *Proceedings of the 13rd International Joint Conference on Artificial Intelligence (IJCAI'93)*, pages 390–395, Chambery, France. Morgan Kaufmann Publishers.

D.K. Lewis. 1972. General semantics. In D. Davidson and G. Harman, editors, *Semantics of natural language.* D. Reidel Publishing Company.

Dominique Longin. 1999. *Interaction rationnelle et évolution des croyances dans le dialogue : une logique basée sur la notion de topique.* PhD thesis, Université Paul Sabatier, Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France, November.

Anand S. Rao and Michael P. Georgeff. 1991. Modeling rational agents within a bdi-architecture. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484, San Mateo, CA. Morgan Kaufmann Publishers.

M. D. Sadek, A. Ferrieux, A. Cozannet, P. Bretier, F. Panaget, and J. Simonin. 1996. Effective human-computer cooperative spoken dialogue: The AGS demonstrator. In *Proceedings of ICSLP'96 International Conference on Spoken Language Processing*, Philadelphia, USA, October.

David Sadek, Philippe Bretier, and Franck Panaget. 1997. ARTIMIS: Natural dialogue meets rational agency. In Martha E. Pollack, editor, *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI'97)*, volume 2, pages 1030–1035. Morgan Kaufmann Publishers, August.

M. D. Sadek. 1991a. *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication.* PhD thesis in Computer Sciences, Université de Rennes I, Rennes, France, June.

M. D. Sadek. 1991b. Dialogue acts are rational plans. In *Proceedings of ESCA/ETRW, Workshop on The Structure of Multimodal Dialogue (Venaco II)*, pages 1–29, Maratea, Italie, September.

M. D. Sadek. 1992. A study in the logic of intention. In Bernhard Nebel, Charles Rich, and William Swartout, editors, *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 462–473, Cambridge, Massachusetts, October. Morgan Kaufmann Publishers.

M. D. Sadek. 1994. Towards a theory of belief reconstruction: Application to communication. *Speech Communication Journal'94, special issue on Spoken Dialogue*, 15(3–4):251–263.

H. Sahlqvist. 1975. Completeness and correspondence in the first and second order semantics for modal logics. In S. Kanger, editor, *Proceedings of 3rd Scandinavian Logic Symposium 1973*, volume 82 of *Studies in Logic*, North-Holland.

John R. Searle. 1969. *Speech acts: An essay in the philosophy of language.* Cambridge University Press, New York.

J. R. Searle. 1979. *Expression and Meaning.* Cambridge University Press.

J. van Kuppevelt. 1995. Discourse structure, topicality and questioning. *Linguistics*, 31:109–147.

Jacques Virbel. 1999. Contributions de la théorie des actes de langage à une taxinomie des consignes. In J. Virbel, J-M. Cellier, and J-L. Nespoulous, editors, *Cognition, Discours procédural, Action*, volume II, pages 1–44. PRESCOT, May.

Dialogue Games are Recipes for Joint Action

Joris Hulstijn

Department of Artificial Intelligence, Faculty of Sciences
Vrije Universiteit Amsterdam
De Boelelaan 1081a, 1081 HV Amsterdam
joris@cs.vu.nl

Abstract

The alleged opposition between dialogue games and plans and goals as approaches to the study of natural language dialogue is false; the two approaches are complementary. On the one hand, plans and goals may function as a semantics to interaction patterns studied as dialogue games. On the other hand, dialogue games are 'compiled-out' recipes for joint action. This claim is illustrated by a dialogue model for inquiry and transaction that uses aspects of both approaches.

1. Introduction

Dialogue models of a more or less formal nature are needed for the design and evaluation of natural language dialogue systems. With respect to such models 'dialogue engineers' can design a system for a particular application and later assess its performance with respect to the specifications (Hulstijn, 2000; Bernsen et al., 1998). Three aspects need to be modeled: the *information structure* of the objects and attributes that play a role in the application domain, the *task structure* of the available actions and plans that play a role in the application, and finally the *interaction* or *dialogue structure* that determines the linguistic form, and order among utterances and the way they are related. In this paper we look at the relationship between these aspects. Different types of dialogue models take different aspects to be the dominant one.

First we make some modelling assumptions (but see section 5. below). A dialogue is considered to be a coherent sequence of utterances. The information conveyed or requested by an utterance is called the *semantic content.* An utterance consists of one or more phrases with a single purpose or function: the *communicative function* of the utterance. The function is usually related to the task, but often some of the function is to maintain the interaction process. Each utterance is analyzed as a *dialogue act.* A dialogue act is fully characterized by a semantic content and a communicative function. Dialogue acts with an interaction-related communicative function, are called *dialogue control acts* (Allwood et al., 1992). Many dialogue acts do double duty. Groups of utterances that naturally 'belong together' are called dialogue segments. The purpose of a dialogue model is to structure a dialogue into meaningful segments, and to find the relationships between utterances that explain their coherence.

The first approach we consider is the *plan-recognition* approach (Carberry, 1990; Litman and Allen, 1987). Here the task structure is taken to be the dominant modeling aspect. Each utterance is understood as an attempt to convey an intention of the speaker, which plays a part in a larger plan to achieve some underlying goal that apparently motivated the dialogue. Given a sufficiently detailed model of plans and goals for the task at hand, this may then be used to determine an appropriate response.

The second approach we consider is based on the observation that certain patterns of utterance types are frequent in naturally occurring dialogue. For example, questions are usually followed by an answer; greetings are returned; proposals accepted or rejected. Such re-occurring interaction patterns can be studied either in terms of *discourse* or *dialogue grammars* (Polanyi and Scha, 1984; Prüst et al., 1994; Jönsson, 1993) or in terms of *conversational games* or *dialogue games* (Levin and Moore, 1978; Mann, 1988; Carletta et al., 1997). In both the games and grammar approaches, rules describe admissible sequences of utterance types and the way they are related.

In the dialogue system community these approaches are presented as rivals, e.g. (Jönsson, 1993, ch 4). However, recent developments in the theory of planning and action explicitly consider *joint actions*, i.e. actions carried out by different agents in cooperation. The need to cooperate creates a need to coordinate. Typical examples of joint actions are carrying a piano or playing a duet. The success of a joint action crucially depends on the combined success of the actions of the participating agents, called participatory actions. It is because of this mutual dependency that synchronization and coordination is necessary. Coordination can be achieved by conventions, by explicit agreement or else by communication of some sort. Moreover, in various publications Clark and others have argued that communication processes themselves can be seen as joint actions between speaker and hearer (Clark and Schaefer, 1989; Clark, 1996). So on the one hand coordination requires communication; one the other hand communication processes can be seen as coordinated actions. Can we build a theory of dialogue along these lines?

Current formal theories of joint planning and action stress the importance of existing schematic joint plans called 'recipes' and of conventional protocols to regulate cooperation (Grosz and Kraus, 1996; Wooldridge and Jennings, 1999). Now the idea is that the dialogue game rules used to describe interaction patterns are precisely the kinds of recipes for joint action that the theorists are looking for, at least where it concerns human-human and human-computer interaction. Moreover, a theory of joint actions may serve as a motivation or 'semantics' to the interaction patterns described as dialogue games. This is similar to hybrid architectures suggested in agent theory.

The paper is organized as follows. In section 2 we look at constraints for joint planning and commitment. Section 3 describes the key ideas of dialogue games. In section 4 we describe ideas of Clark to view communication as a combination of joint actions, scheduled at different levels. Under this view, a combination of the two approaches is becomes very natural. Section 5 shows how a combination of the two perspectives can be used to model dialogues for inquiry and transaction (Hulstijn, 2000). The paper ends with conclusions and related research.

2. Plans, Goals and Commitment

How are dialogue acts to be combined? Like any action, a dialogue act can be characterized by its preconditions and intended effect, sometimes extended with applicability conditions, which determine if the action is applicable at all, and failure conditions, which determine what must remain the case, even if the action fails. Take a question for example: the intended effect is to get to know the answer. A precondition is that the responder knows the answer and is willing to tell it. An applicability condition would be that the responder is paying attention. A failure condition is that even if the responder does not know the answer, at least the contact between asker and responder is maintained. Why does someone ask a question? The answer to the question may be needed for the task. So like other actions, dialogue acts are applied as part of a plan to accomplish some goal. Therefore, the task-related communicative function of an utterance, analyzed as a dialogue act, corresponds to the role of the act in some plan.

From basic actions, more complex plans can be constructed by production rules and by connectives like ';' for sequential composition, '||' for parallel composition, '|' for indeterministic choice and by renaming, substitution and abstraction. How do you construct a plan to achieve a given goal? Reasoning backwards, you can determine the actions that would accomplish a goal, turn their preconditions into sub-goals, and repeat the procedure until you have reached the current state of affairs. This is a sketch of a naive algorithm for means-end reasoning. Obviously, if the plan involves actions of other participants the scheduling of subtasks becomes much more complicated. There must be no conflicts about shared resources.

Luckily, plans do not have to be re-calculated every time. Agents may select a suitable plan from a plan-library. The exact way a plan will be carried out can not always be known in advance. In particular in dialogue situations, you do not know how the other participants will respond. Dialogue is opportunistic, so it pays to have only a partial and indeterministic specification of the actions to undertake. Details can be filled in as the interaction progresses. Also the order in which the different actions are to be carried out may be left unspecified, unless there are some logical constraints. Such previously calculated partial specifications of combinations of actions to reach some goal, are called *recipes* (Grosz and Kraus, 1996). Recipes for joint action typically contain *roles* to be assigned to the participating agents, and for each role a particular sub-task.

Once a plan is adopted, an agent will typically persist in trying to achieve the goal of the plan until it is reached, becomes unreachable or becomes undesirable. This aspect of goals or intentions is called *commitment* (Cohen and Levesque, 1990). It stabilizes behavior and makes agents more reliable. This aspect is crucial for cooperating agents. Commitments made to other agents are the 'glue' that keeps joint actions together. There is a social obligation on agents to keep their commitments. But also aspects like trust, common decency and the desire to be liked play a role.

A joint action is an action consisting of sub-actions, called participatory actions, carried out by different agents in cooperation. What are the essential characteristics of a joint action? Bratman (1992) lists three constraints for what he calls *cooperative activity*: (a) mutual responsiveness, which can be compared to a particular aspect of cooperativity, (b) commitment to the joint action, and (c) commitment to mutual support, which forces agents to assist other agents in their participatory actions. These constraints define what cooperative action is, but it remains unclear how, for example, help is elicited or how mutual responsiveness is established and maintained.

Grosz and Kraus (1996) describe a different set of constraints on plans for joint action, which they call *shared plans*. In particular, (i) agents must do means-end reasoning involving joint actions, (ii) there must be decision procedures for selecting a recipe and for assigning agents to roles and thus to tasks, (iii) there must be communication procedures for reaching agreement, (iv) to assess their abilities, agents must compute the context in which they plan to carry out an action, and (v) the intended effects of sub-actions must be consistent. Note that constraint (ii) and (iii) refer to conventions for interaction, and that (i) and (iv) involve reasoning about other agents' likely behavior.

In the account of commitments of Wooldridge and Jennings (1999) a joint commitment of a group of agents is characterized by an *immediate goal*, a long-term *motivation goal*, a set of *preconditions* and set of *conventions* ρ that defines an interaction protocol. Each of these notions can be given a formal semantics in terms of a modal logic. This analysis stresses the fact that commitments are goals that are meant to be kept, but only relative to a motivation and as long as it is practical. The conventions specified by ρ indicate, among other things, the circumstances under which a commitment can be abandoned. A convention is modeled as a list of production rules. Each time a trigger condition is true, the agent will adopt the corresponding goal. In effect, these production rules constitute what we have called a recipe above, although Wooldridge and Jennings do not account for partiality. The parameter ρ can be filled in different ways. For example, it may contain sanctions upon not keeping a commitment. Or to take a different example, under a Cohen and Levesque-like account of commitments, the immediate goal may only be dropped when the agent believes that the motivation behind it has been achieved or has become unachievable or undesirable, or else when the agent believes that a better plan has become available. In each of these cases, the agent is held to communicate these changes of insight to the other participants.

So these accounts of joint action and commitment involve communication. If we view communication processes themselves as joint actions, what recipes are used?

Game	Illocutionary Point	Goals-of-R	Conventional Conditions
information seeking	I knows $?\varphi$	I knows $?\varphi$	R knows $?\varphi$
information offering	R knows φ	R knows φ	I knows φ ; R's information and φ are consistent
information probing	I knows whether R knows $?\varphi$	R informs I of R's knowledge about $?\varphi$	I knows $?\varphi$
helping	I is able to perform α	I is able to perform α	R is able to cause I to be able to perform α; I has the right to perform α
dispute	R believes φ	R justifies that I might not believe φ	I believes φ; R does not believe φ
permission seeking	I knows that R grants the right that I performs α	R grants the right that I performs α or not, and I knows this	I wants to perform α; I does not have the right to perform α; R can grant the right to perform α
action seeking	R causes α to be performed	R causes α to be performed	R would not cause α to be performed in the normal cause of events

Figure 1: Examples of dialogue games (Mann, 1988; p515)

3. Dialogue Games and Grammars

Utterances do not need to be analyzed in terms of the underlying intentions. Looking at a corpus, an exchange of greetings simply marks the beginning or end of an interaction; acknowledgments just follow assertions. Such stereotypical sequences of dialogue acts are called *interaction patterns*. It is possible to analyze a pattern as evidence of a plan, and work out a cooperative reply in this way, e.g (Litman and Allen, 1987). But many types of utterance seem not to be consciously deliberated, but conventionally triggered by the circumstances.

3.1. Dialogue Games

A useful metaphor for studying interaction patterns is that of a *dialogue game*. Each participant plays a role and expects the others to play their respective roles. Each participant makes one of the moves that are appropriate for its role at that stage in the game, which is determined by the rules of the game. The stage in a game, the scoreboard if you like, determines which moves are allowed. Based on corpus annotation, Carletta et al. (1997) have found a number of hierarchically ordered interaction patterns, analyzed as *games*. Games are sequences of *moves*. Each move corresponds to a type of utterance. So what we called the interaction-related communicative function of a dialogue act, is defined as the move in a game. Moves can be either *initiatives* or *responses*. Typically, each initiative must be followed by an appropriate response, although there may be other exchanges first. For example, a clarification sequence may precede the answer to a question. Initiative-response units have been successfully used in the design of spoken dialogue systems, e.g. (Bilange, 1991).

The dialogue game metaphor has been applied in different ways. Interestingly, early work by Levin and Moore (1978) and particularly Mann (1988) lists the following parameters for the definition of a game. (i) There are roles, in this case initiator and responder. (ii) The game has an illocutionary point: the goal of the initiator in starting the game. (iii) The responder has goals upon accepting the game, and (vi) there are a number of constraints: the initiator and responder must pursue their respective goals, goals must be believed to be feasible, the illocutionary point must not already be achieved, the initiator must have the right to initiate the game, and the responder must be willing to pursue its goals upon accepting the game. This looks very similar to the constraints for successful joint action discussed above. The 'goals upon accepting a game' are nothing but the commitments of the initiator and responder.

In figure 1 we reproduced the list of game types suggested by Mann. The list is by no means complete, and does not constitute an analysis of a particular type of dialogue. We replaced Mann's notation for required information specifications Q (the content of a question) with formulas $?\varphi$, propositions P with φ, and action-type formulas A with α. This notation will be explained below. An information seeking game corresponds to a question-answer sequence. However, a formulation in these terms stresses that information seeking can be expressed, for example, by a declarative question like "I would like to know if ...". An information offering game corresponds to an inform act followed by a acknowledgment. A prolonged sequence of information seeking and information offering games about a single topic constitutes a game of inquiry (Hulstijn, 2000, ch3). The information probing game corresponds to a confirmation sequence or check. In different settings it could correspond to an exam question. So the point of the initiator is not to find out some information, but to check if the other participant agrees. Carletta et al.(1997) call this an *align* move. A helping game corresponds to a clarification question and subsequent explanation. The fact that clarifications are needed at all, shows that exchanges are not guaranteed to succeed. In some cases misunderstandings have to be cleared up. The dispute game illustrates that dialogue participants do not have to be cooperative nor in agreement. Obviously, the dispute game is important for activity types like debates or arguments. It may be used in dialogues for inquiry and transaction too, to resolve misunderstandings. Finally, an action seeking game corresponds to what we call a proposal, suggestion or request followed by an counter proposal, acceptance or rejection. Again, there may be various ways to express such acts. For example, there is a close relationship between statements of preference and requests.

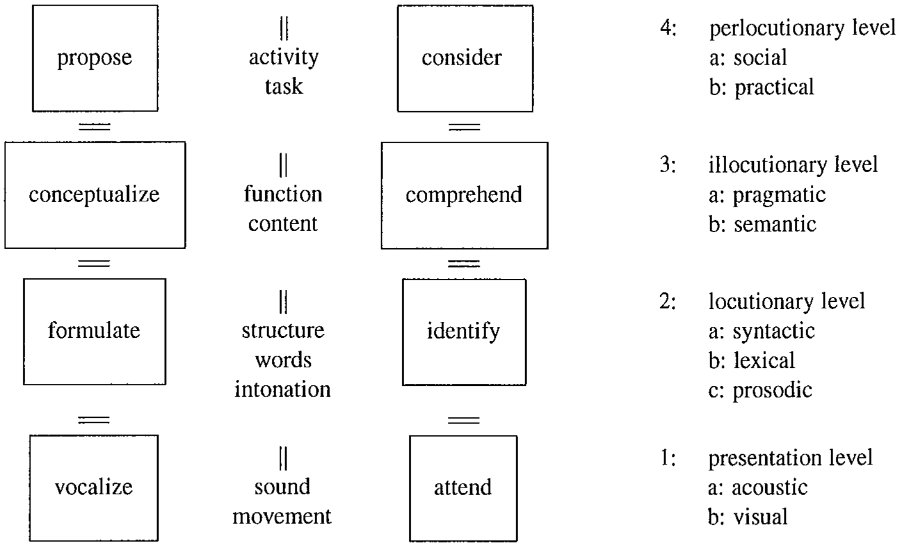


Figure 2: Coordination *at* and *between* and different levels of communication

3.2. Grammars and Coherence Principles

Another important approach to interaction patterns makes use of *discourse grammars* (Polanyi and Scha, 1984; Prüst et al., 1994). These have been particularly successful in studying ellipsis and parallel elements in adjacent utterances. Discourse grammar rules define well-formed sequences of utterances and the relationships between them. The resulting parse-tree may function as a discourse representation structure. At the leaves of the tree we find the content of the utterances, represented by some appropriate representation scheme. The branches are labeled by *coherence relations* that express the functional relationship between segments of the discourse. For discourse in general, typical coherence relations are *contrast, causation* or *continuation*. A similar approach has also been applied to dialogue. Asher and Lascarides (1998) propose coherence relations and coherence principles for dialogue which are based on typical interaction patterns. The role of the discourse grammar rules to define well-formed sequences is taken over by the coherence principles: abductive rules for meta-reasoning about the way a new dialogue segment can be coherently attached to some part of the dialogue representation structure. For example, the occurrence of a cue word like 'but' is evidence for the application of 'contrast' as the most appropriate coherence relation for attachment. In a similar way, the occurrence of an utterance that answers some question that was previously asked or implied, suggests the use of the *indirect question answering pair* (IQAP) as a coherence relation to attach the answer, to its question. So coherence principles define relationships between dialogue segments, that are analogous to agreement constraints in a sentence grammar. Coherence relations may be defined at different linguistic levels, involving form-related aspects such as intonation or parallel syntax and semantics, but also involving content- or task-related aspects. An IQAP-like coherence relation requires for instance that the answer is relevant to the question, non-trivial, consistent and 'to the point'. Such coherence constraints can be expressed in a logic (Groenendijk, 1999).

4. Communication Levels

Communication processes can be seen as a combination of joint actions consisting of processes that are scheduled in parallel *at* and *between* different communication levels (Clark, 1996). This proposal resembles the layered communication protocols that are applied in distributed and concurrent programming. A particular interpretation of this idea is depicted in figure 2. The '||' sign illustrates parallel composition. The tags in the middle indicate the aspects used to coordinate on. Processes within one agent are coordinated between different levels, indicated by the '==' sign. What communicative levels and functions can be distinguished?

At the *presentation level* speaker and hearer coordinate on mutual attention and on the delivery of the utterance. This constitutes a joint action of vocalizing-attending an utterance. A similar remark can be made about gestures, which are deliberate movements.

At the *locutionary level* participants coordinate on the wording, structure and prosody of an utterance or on the shape and size of a gesture. Such form-related aspects convey information about the semantic content and communicative function of the utterance. For example, form-related aspects like parallelism and intonation indicate the scope of a focus sensitive operator, such as 'only' or 'not'.

At the *illocutionary level* participants establish the semantic content and communicative function of an utterance, in relation to the context of surrounding utterances. At this level an utterance can be described as a *dialogue act*. A dialogue act is fully characterized by a semantic content and a communicative function which has a task-related aspect and an interaction-related aspect (see below).

At the *perlocutionary level* participants coordinate on their task: to carry out participatory actions to accomplish a joint goal. The goal is partly determined by the social activity they are engaged in. For different types of activities different conventions apply.

Opening | Inquiry | Negotiation | Confirmation | Closure

Figure 3: Phases of a transaction task

5. Towards a combination

What would a dialogue theory that uses dialogue games as the recipes for joint action look like? We follow figure 2.

First, we need some general account of the social activity that underlies the type of dialogue. Activity types are determined by the roles of the participants, their private and public goals, the physical and social settings and a number of phases that determine the progress of the activity. In our example of ticket reservation in a theater, we have a ticket agent who wants to sell and a client who wants to buy a number of tickets for a certain performance, at a certain date and time. The social institution is that of ownership. Important artifacts are money tokens and the tickets. By convention each ticket gives a right to a seat. The typical setting is that of box office with a counter, which can be replaced by a telephone or Internet setting in case we are dealing with reservations. A reservation is a kind of transaction in which the delivery is postponed. The phases in a transaction task (figure 3) are an inquiry phase, of information exchange about the product, a negotiation phase in which there is bargaining on the relationship between price and attributes of the product or service, followed by a confirmation phase in which the deal is closed. Like any interaction, a transaction has an opening phase in which mutual contact is established, and a closing phase in which the contact is closed successfully. Note that each phase is delimited by some mutual understanding. The opening phase results in a mutual commitment to maintain contact. The inquiry phase results in mutual knowledge about the available possibilities. The negotiation phase results in several mutual agreements on the proposals discussed and the confirmation phase reaffirms these agreements and makes some implicit consequences explicit. These signs of mutual understanding must be grounded by an exchange of utterances. Obviously, this transaction model represents an idealized case. Actual dialogues may skip some phases by convention or return to a previous phase. For example, a user may have wrongly thought that there is a discount on Mondays. During negotiation this misunderstanding may force the participants to jump back to an 'inquiry' exchange. Pricing conventions that were earlier skipped and expected to be known implicitly, now have to be dealt with explicitly.

Second, we need a repertoire of dialogue acts to model the different types of utterances. Figures 4, 5, 6 and 7 show examples of dialogue acts that are typical for the opening, the inquiry, the negotiation and the confirmation phases. Acts of one phase may be used as part of another. So an inform act may be used as part of the opening phase, to establish the goal of the interaction. Other acts can be defined that are appropriate at any stage, like asking for clarification and a subsequent explanation, time management acts like "wait a minute" or expressions of emotional attitudes. In each case, a dialogue act is defined as a semantic content combined with a communicative function.

Action	Description	Applicability
greet(a, b)	establish contact	requires or in response to greeting
meet(a, b)	identify and establish social relationship	a and b have contact, requires or in response to meeting

Figure 4: Opening acts

Action	Description	Applicability
inform(a, b, φ)	state fact φ	requires ack, φ uncontroversial
ack(a, b)	indicate receipt and understanding	in response to inform, suggest
assert(a, b, φ)	state opinion φ	requires assent
assent(a, b)	indicate receipt, understanding and agreement	in response to assert, check
deny(a, b)	indicate receipt, understanding and disagreement	in response to assert, check
correct(a, b, φ)	deny information in focus, and replace with φ	in response to assert, inform; requires assent
ask$(a, b, ?\varphi)$	request information pertinent to $?\varphi$	requires answer
answer(a, b, φ)	provide information pertinent to $?\varphi$	in response to ask
check$(a, b, ?\varphi)$	request agreement	requires assent

Figure 5: Inquiry acts

The *semantic content* needs to be defined in terms of some semantic theory. Asher and lascarides (1998) choose DRT. We haven chosen to use a version of *update semantics* (Veltman, 1996) combined with a semantics of questions and answers (Groenendijk and Stokhof, 1996). In an update semantics, the meaning of an utterance is equated with the difference it makes to the information in the dialogue context. Different types of dialogue acts affect different aspects of the information in a dialogue context. We distinguish *assertives* φ, that add factual information, *interrogatives* $?\varphi$ that raise issues and thereby structure the information, and *directives* $!\varphi$, that affect the commitments being made by agents and thereby partly determine future actions α (Hulstijn, 2000). In our account, information is modeled as a set of possible worlds , that are compatible with the information. Adding information means the elimination of possible worlds. The set of worlds is structured by a partition, or equivalently, by an equivalence relation. The partition models the distinctions made by the current 'questions under discussion' or *issues* (Ginzburg, 1995; Hulstijn, 1997). An issue is an abstract entity that represents the content of a question, just like a proposition represents the content of an assertion. Each of the equivalence classes or blocks in the partition correspond to an alternative answer or solution. We say that an issue is resolved, when only one alternative remains. An issue is partially resolved, when at least one

Action	Description	Applicability
request$(a, b, !\varphi)$	get b to do $!\varphi$	requires acceptance
propose$(a, b, !\varphi)$	offer to do $!\varphi$	requires acceptance
suggest$(a, b, !\varphi)$	bring possibility of $!\varphi$ to attention	requires acknowledgment
counter propose$(a, b, !\varphi)$	reject current offer and replace with $!\varphi$	in response to request, suggest or proposal; requires acceptance
accept(a, b)	indicate receipt, understanding and agreement with current offer	in response to request or propose
reject(a, b)	indicate receipt, understanding and disagreement with current offer	in response to request or propose

Figure 6: Negotiation acts

Action	Description	Applicability
ask-confirm$(a, b, !\varphi)$	ask commitment on complete offer	a, b have contact; requires confirm
confirm(a, b)	indicate commitment on complete offer	a, b have contact; there is a complete offer; in response to ask-confirm
disconfirm(a, b)	indicate absence of commitment on complete offer	a and b have contact; there is a complete offer; in response to ask_confirm

Figure 7: Confirmation acts

Action	Description	Applicability
thank(a, b)	strengthen relationship	b did a a service; optionally followed by ack_thanks
ack_thanks(a, b)	indicate receipt and understanding of thanks	in response to thanks
bye(a, b)	close contact successfully	a and b have contact, requires or in response to bye
break(a, b)	close contact unsuccessfully	

Figure 8: Closing acts

alternative has been eliminated. In this way, we can make a distinction between information that is relevant or irrelevant with respect to a particular issue. Factive information is relevant to an issue, when it eliminates at least one of the alternatives. This relevance constraint can be combined with constraints like consistency, that the resulting set of possible worlds is non-empty and informativeness, that some possible worlds have been eliminated. Groenendijk (1999) argues for a fourth constraint, that utterances must be *licensed*, i.e. not over-informative. It means that worlds must be eliminated 'block wise'; differences between worlds that are irrelevant to the current issues are neglected. Although we can define relevance of information with respect to some previously raised issues, the relevance of bringing up an issue itself, can only be assessed relative to the underlying task or topic of the dialogue.

The *communicative function* of a dialogue act is often to exchange information about the underlying task. So a question is used to raise an issue, and thus to invite the hearer to specify information that is needed to achieve some goal defined at the task level. However, the function of a dialogue act is not necessarily related to the task; there are dialogue control acts such as greeting and acknowledgment whose function is merely related to the interaction process (Allwood et al., 1992). A greeting invites a greeting in response. Such a mutual greeting establishes mutual contact between the participants. An acknowledgment signals receipt and understanding of an utterance; it helps to establish a basis, in this case the dialogue history, upon which a common ground can be build (Lewis, 1969). Obviously, many acts have both a task and an interaction-related communicative function. For example, an answer to a question also functions as an acknowledgment of the question. In general, we can say that the interaction-related function of an utterance corresponds to the role it can play in the available exchanges or games. The task-related function corresponds to the role it can play in a plan that is suitable for the task.

Third, we need appropriate ways to combine dialogue acts into coherent and useful exchanges. Now crucially, the intended effect of a single dialogue act is not very interesting, from the point of view of a plan. What counts is the effect of the exchange the act is a part of. The intended effect of a question is to get to know the answer. But the question is only the initiative of a question-answer exchange. Only after the response one could tell if the information was or was not provided. In other words, the success conditions of a question involve the answer. And this seems to be true of most single-agent actions that are used as part of joint actions. When planning, dialogue acts only make sense as part of an exchange, or initiative-response unit. So actually exchanges are the building blocks of joint actions. They are the smallest possible actions that can still be called joint.

We found that initiatives and responses are often related by a triangle structure such as the one depicted in figure 9. A complete initiative-response-unit is composed of an initiative, followed by either a positive or a negative response, or else a retry. For example, a proposal is an initiative, an acceptance is the corresponding positive response, a rejection is the negative response and a counter proposal is an example of a retry. In case of a negative response some precondition for a successful completion of this exchange may have been violated. In that case a repair may be needed of parameters that were fixed in a previous phase. A combined initiative-response unit results in some information being added to the common ground. A response of the right type has a grounding effect (Clark and Schaefer, 1989).

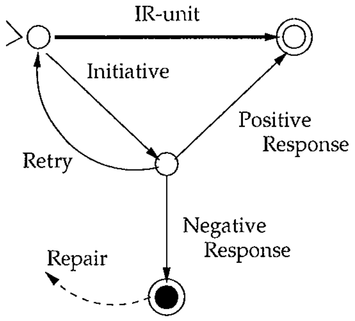


Figure 9: Initiative-response units

Interestingly, the content of most initiatives corresponds to an issue: it opens alternatives. The content of a response usually narrows down the number of alternatives. The figure may suggest that a response must occur after the initiative. This is not so. Many types of non-verbal feedback, such as nodding, may occur simultaneous to the utterance they give feedback to. Initiative shifts between exchanges are allowed, for example when an 'information seeking game' is continued with an 'information offering game', in terms of Mann (1988). Not all exchanges need to have a binary triangle shape like this one. Another common type of exchange starts with a so called *pre-sequence* (Schegloff et al., 1977). It invites the other participant to join in actual exchange. Consider "Can I ask you something? – Yeah – Could we come over this weekend?". This pre-question probes one of the applicability conditions, namely that the responder is willing to respond or willing to comply. Or consider: "So, good-bye then – Yeah, bye! – Bye!" where a pre-sequence announces the closing sequence. In a similar way there are also post-sequences. More importantly, a response may function itself as an initiative to respond to. Hence the possibility of counter proposals and corrections

In the description of a response, the so called *scope* of the response should be clear from the context. For example, a question is an act of type $\text{ask}(a, b, ?\varphi)$, where agent a asks b a question with the issue $?\varphi$ as its content. The act of answering this question is indicated by $\text{answer}(b, a, \psi)$, where we require that ψ is pertinent to $?\varphi$. In that case, the issue $?\varphi$ is called the scope of the answer. In a similar way, acceptance, rejection and counter propose expect some recent offer as their scope, and assent, deny and correct expect information that is 'in focus'. The responses confirm and disconfirm expect a complete offer. Apart from these differences in scope, accept and reject, assent and deny, and confirm and disconfirm behave in a similar way. They all establish some kind of agreement. These responses are stronger than an acknowledgment, which merely indicates receipt and understanding.

Both initiatives and responses can be made implicitly. Initiatives may for instance be triggered by the situation. At a box office, an utterance like "3 for tonight, please" is appropriate, because it resolves the issue that is raised by a box office situation. Moreover, a single utterance can perform several communicative functions at once. This mechanism is used by indirect speech acts. A question like "Do you know the time?" asks for a precondition for successful completion of the question-answer exchange that was intended. This only makes sense in case the speaker actually wants to start the question-answer exchange. Also acknowledgments can be implicit. In general, a pertinent continuation of the dialogue counts as an acknowledgment of the previous utterance.

Finally, we need a theory that translates the linguistic features of utterances, including parallel elements and intonation, into a proper representation format. Here we may benefit from the insights of discourse grammars. In dialogue in particular we may expect ellipsis, continuation and other constructions that depend on the immediate context. Recognizing semantic content and various communicative functions of an utterance is not easy. However, we can combine aspects from different communication levels into one interpretation. Some of this ambiguity may be deliberate. The speaker may leave it to the responder how an utterance should be 'taken up'. So part of the meaning of the original initiative is only properly established, when the whole exchange is completed. This type of underspecification would be impossible in traditional one-step accounts.

6. Conclusion and Related Research

That concludes the comparison of accounts of joint action with interaction patterns. What it shows is that the perceived opposition between a plan-based and a pattern-based approach to natural language dialogue is false. The two approaches are complementary. Once the constraints of joint planning and action are taken seriously, a need for conventions to regulate the interaction arises. These conventions are represented by recipes. The smallest recipes for joint action are precisely the exchanges described by dialogue game rules. On the other hand, plans and goals may function as a semantics for dialogue game rules. They motivate the illocutionary point of initiating a game and explain various aspects of cooperativity in dialogue.

Several people have reached similar conclusions. We mention Nicolas Maudet (p.c.), Traum (1997) and the developers of the Trindi dialogue system architecture (Traum et al., 2000). The insight that pattern-directed approaches need to be combined with higher-level notions like plans and goals, is compatible with a general trend towards hybrid architectures for agents. Agent architectures increasingly combine high-level deliberative features with low-level reactive features. Examples can be found in recent editions of ATAL conferences or in the literature on the RoboCup, where low-level skills like aiming and shooting must be combined with higher level reasoning about team-work and tactics (Stone and Veloso, to appear). In some applications, a team makes use of so called 'locker room agreements' that fix the strategy of the team. This reduces the need for communication while playing. The general principle that underlies the trend seems to be that frequently occurring activities that can be 'automated' are often dealt with by fixed pattern-directed protocols, recipes or rules. Infrequent activities or failure and misunderstanding require higher-level deliberation. It seems that dialogue is no exception to this principle.

7. References

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, 9:1–26.

Nicholas Asher and Alex Lascarides. 1998. Questions in dialogue. *Linguistics and Philosophy*, 21:237–309.

Niels Ole Bernsen, Hans Dybkjaer, and Laila Dybkjaer. 1998. *Designing Interactive Speech Systems. From First Ideas to User Testing.* Springer-Verlag, Berlin.

Eric Bilange. 1991. A task independent oral dialogue model. In EACL-91, pages 83–88. Berlin.

Michael E. Bratman. 1992. Shared cooperative activity. *Philosophical Review*, 101:327–341.

Sandra Carberry. 1990. *Plan recognition in natural language dialogue.* MIT Press, Cambridge, Mass.

Jean Carletta, Amy Isard, Stephen Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational linguistics*, 23(1):13–32.

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

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

Philip R. Cohen and Hector J. Levesque. 1990. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261.

Jonathan Ginzburg. 1995. Resolving questions, I. *Linguistics and Philosophy*, 18:459–527.

Jeroen Groenendijk and Martin Stokhof. 1996. Questions. In *Handbook of Logic and Language.* North-Holland, Elsevier.

Jeroen Groenendijk. 1999. The logic of interrogation: classical version. In SALT-9, Santa Cruz. also ILLC research report PP-1999-19, University of Amsterdam.

Barabara J. Grosz and Sarit Kraus. 1996. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357.

Joris Hulstijn. 1997. Structured information states: Raising and resolving issues. In Mundial'97, pages 99–117. University of Munich. also CTIT Report 97-18, University of Twente.

Joris Hulstijn. 2000. *Dialogue Models for Inquiry and Transaction.* Ph.D. thesis, University of Twente, Enschede.

Arne Jönsson. 1993. *Dialogue Management for Natural Language Interfaces.* Ph.D. thesis, Linköping University.

James A. Levin and James A. Moore. 1978. Dialogue-games: Metacommunication structures for natural language interaction. *Cognitive Science*, 1(4):395–420.

David Lewis. 1969. *Convention: A Philosophical Study.* Harvard University Press, Cambridge.

Diane L. Litman and James F. Allen. 1987. A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11:163–200.

William C. Mann. 1988. Dialogue games: Conventions of human interaction. *Argumentation*, 2:511–532.

Livia Polanyi and Remko Scha. 1984. A syntactic approach to discourse semantics. In COLING10, pages 413–419. Stanford, CA.

Hub Prüst, Remco Scha, and Martin H. van den Berg. 1994. Discourse grammar and verb phrase anaphora. *Linguistics and Philosophy*, 17:261–327.

Emanuel A. Schegloff, G. Jefferson, and H. Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2).

Peter Stone and Manuela Veloso. to appear. Task decomposition, dynamic role assignment and low-bandwidth communication for real-time strategic teamwork. *Artificial Intelligence.* to appear.

David Traum, Peter Bohlin, Johan Bos, Stina Ericsson, Staffan Larsson, Ian Lewin, Colin Matheson, and David Milward. 2000. Dialogue dynamics and levels of interaction. Technical report, TRINDI, LE4-8314.

David Traum. 1997. A reactive-deliberative model of dialogue agency. In *ECAI'96 Workshop on Agent Theories, Architectures and Languages (ATAL)*, LNCS 1193, pages 157–172. Springer-Verlag, Berlin.

Frank Veltman. 1996. Defaults in update semantics. *Journal of philosophical logic*, 25(3):221–262.

Michael. J. Wooldridge and Nicholas. R. Jennings. 1999. The cooperative problem-solving process. *Journal of Logic and Computation*, 9(4):563–592.

Acknowledgments

This research was conducted at the Parlevink Group for language technology of the University of Twente in Enschede, The Netherlands.

Obligations, Intentions, and the Notion of Conversational Games

Jörn Kreutel*, Colin Matheson †

* SAIL LABS S.L.
Barcelona
joern.kreutel@sail-labs.es

† Language Technology Group
University of Edinburgh
colin.matheson@ed.ac.uk

Abstract

We argue that in order to capture uncooperative behaviour in dialogue it is necessary to model the way the intentions of the participants are related to each other. We show how this intentional structure can be determined in an approach which uses DISCOURSE OBLIGATIONS as basic structural means. Given the assumption of correspondence between dialogue structure and intentional structure, we can demonstrate that from our point of view CONVERSATIONAL GAMES can be seen as macro-structures which are decomposable into smaller functional units where the coherence between the latter is explained in terms of obligations. We further suggest that an obligation-driven approach to dialogue provides a more satisfying account for conversational cooperation than current intention-based models

1. Introduction

One argument motivating the use of DISCOURSE OBLIGATIONS in dialogue modelling is that they allow an account of the behaviour of agents who are unable or unwilling to act in the expected manner, but nevertheless *act* (Traum and Allen, 1994):

(1) A[1]: Did Pete drive here?
B[2]: I don't know / I don't want to talk about that

Here, intention-based dialogue models fail to explain why B actually responds to A's question with either reply. An obligation-driven approach, on the other hand, predicts B's behaviour appropriately by assuming that participants in a dialogue (DPs) are socially *obliged* to respond, no matter what their intentions are.

Whereas most approaches that use discourse obligations preserve a representation of the DPs' intentions, Kreutel and Matheson (1999) propose a dialogue model that captures a range of subdialogues initiated by questions or assertions by exclusively referring to the obligations imposed on the DPs, leaving aside their intentions. Conceiving thus of 'cooperativity' as the DPs' willingness to act according to the obligations imposed on them, obligations are established as the basic explanatory means which reconstruct the essentially rule-governed behaviour exhibited by cooperative DPs.

One can then ask, however, what kind of representation of the DPs' intentions an obligation-based dialogue model has to incorporate in order to cover uncooperative actions. Note the following dialogue, in which an uncooperative action occurs in the middle of a discussion:

(2) A[1]: Jack and Helen will split very soon.
B[2]: How do you know that?
A[3]: She doesn't love him anymore.
B[4]: That's not true.
A[5]: I don't want to discuss that.
B[6]: Then please don't try to make me think
that they will split.

In this case, A's unwillingness to discuss the question whether Helen still loves Jack or not results both in A's not being able to convince B of the claim in [3] and also in the failure of the initial attempt to establish the assumption that Jack and Helen are about to split as shared belief. Such examples show that the representation of intentional structure has to allow for the DPs' reasoning about the way their respective intentions are related. For example, considering A's options after [4] above, it is perfectly possible that the unwillingness to discuss the issue raised in [3] could be abandoned in favour of convincing B of [1], even though in a neutral context A might never want to discuss the issue.

Our main aim here is to outline how the model in Kreutel and Matheson (1999) can be extended to account for the DPs' reasoning over their intentions. After defining our use of INFORMATION STATE (IS) and discourse obligations, we propose a set of inference rules which allow us to construct an intentional structure reflecting the relationships between the DPs' intentions at a given state of a dialogue. The structures generated by these rules are then discussed in terms of CONVERSATIONAL GAMES (see for example Carletta et al. (1996)). Given the fundamental importance of discourse obligations we will show how the structuring aspect of games can be reconstructed in terms of obligations. We will then discuss the advantages of the obligation-based approach to cooperativity with respect to recent developments in intention-based models which overcome some of the problems pointed out in Traum and Allen (1994).

The immediate theoretical background to this paper is the work of Poesio and Traum (1998) on their theory of discourse obligations and grounding. We ignore grounding here to concentrate on the relationship between obligations and intentions. The broader context includes a wide range of research, including work on the semantics and pragmatics of questions (Ginzburg, 1995a), (Ginzburg, 1995b), cooperativity (Di Eugenio et al., 1997), and, above all, the issue of how the linguistic structure of discourse and its intentional structure are related to each other (Grosz and Sidner, 1986), (Moser and Moore, 1996), (Lascarides and Asher, 1999). Our central claim is that some of the problems that are well understood in this broad context can be handled in a way which minimally extends the machinery assumed by Poesio and Traum, particularly the use of discourse obligations, which we assume to be independently motivated. The fact that these problems appear tractable in a framework which uses obligations as part of the basic expressive means testifies to the descriptive power and flexibility of this notion.

2. Information States and Intentional Structure

2.1. Information States

Our use of IS follows quite closely the approach proposed in the TRINDI project.[1] See Cooper (1998), Bohlin et al. (1999) and Cooper and Larsson (1999) for general discussions on the application of the notion of IS in dialogue research.

The basic assumption is that, from a very general position, it is useful to view dialogues in terms of the relevant information that the dialogue participants have at each stage in the discourse. The main effect of an utterance is thus to change this information, and an immediate question is what kind of information is appropriate in characterising this process. The TRINDI project has therefore experimented with various versions of IS, looking at various issues including the representation of common mental attitudes such as beliefs and intentions, and at more dialogue-specific notions like turn, question under discussion, and dialogue history. A software tool for constructing and manipulating ISs was developed during the project in order to provide a common platform for comparing dialogue management techniques (the TrindiKit; Larsson (1999)). Note that many aspects of the theoretical concerns which we discuss here have been implemented, both directly in Prolog (Kreutel, 1998), and using the TrindiKit (Bos et al., 1999), (Matheson et al., 2000).

Among other things, viewing utterances in terms of the way they update ISs allows the notion of dialogue move to be 'decomposed'; it is possible to characterise the information change arising from the utterance in terms of the particular effects on the IS without attempting to classify the utterance in terms of a particular move, or moves. In general, however, it is often still useful to talk of the moves associated with an utterance, and we shall occasionally do so here.

The representation of ISs in the TRINDI project is generally done using feature structures (attribute-value matrices) to represent the components that are assumed to constitute the DPs' knowledge of 'the state the dialogue is in'. In an obligation-driven model, typical components of an information state are, for instance, representations of the common ground between the DPs and of the obligations that are imposed on them. (See, for example, Poesio and Traum (1998), Kreutel and Matheson (1999)) The common ground will normally hold a representation of the history of the dialogue, which in the case of Poesio and Traum means a structure containing the DIALOGUE ACTS which represent the main contents of utterances.

Given this concept of information states, updates can be formalised as operations which may be as simple as pushing values into the components of an IS. Popping the contents of a component would be another basic update which may serve, for example, to formalise the notion of an obligation being discarded. However, in more elaborated models of information states the updates can be much more complicated, involving pushing whole new structures into the IS and operations such as the merging of parts of the IS. The latter merge operation is used by Poesio and Traum in cases where information which is not considered part of the common ground gets acknowledged. In these circumstances Poesio and Traum assume that the grounding process merges the ungrounded material with the information in the common ground.

2.2. Information States in Question-Answer Contexts

Kreutel and Matheson (1999) assume a very basic notion of IS which includes just the DIALOGUE HISTORY (DH) as a sequence of moves, and a representation of the DPs' obligations (OBL). The dialogue acts used are a subset of those proposed in Poesio and Traum (1998), where the acts are sub-classified into CORE SPEECH ACTS (`assert`, `ask`, `accept`, etc.) and ARGUMENTATION ACTS (`answer`, `info_request`, etc.). This classification captures the idea that any speech act may appear in a wider discourse context, where argumentation acts characterise the context-dependent actions of core speech acts. Obligations are represented as stacks of `address` and `answer` elements.

The updating of information states is done based on the notion of INFORMATION STATE UPDATE SCENARIOS. Update scenarios are meant to specify certain 'constellations' of IS, corresponding to situations like the turnholder's replying to an assertion, to a question, or to an assertion that is meant as an answer to a question. Each scenario can be defined based on the information in IS, i.e. on the overall structure of the DPs' obligations and the history of the dialogue.

In particular, Kreutel and Matheson (1999) provide a detailed analysis of questions in the framework of the obligation-based approach. Based on examples such as (3) below, the authors argue for the retention of the obligation to `answer`, which is introduced by a question, until the question has been resolved, and also propose that the evaluation of an answer should be seen as a twofold process that takes into account first the 'assertive' and then the 'answerhood' properties of the answer:

(3) 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]: Ok. But she could have come by bicycle.
A[5]: I stayed there until 4 o'clock in the morning and she didn't show up.
B[6]: Ok.

[1] Telematics Applications Programme, Language Engineering Project LE4-8314.

Here, B's turn in [4] expresses acceptance of the assertive content of [3], but at the same time rejects [3] as an assertion which resolves the request for evidence in [2]. As the obligation for A to `answer` [2] is still present after [4], A comes up with an alternative answer which now is accepted by B's *ok* in [6], both as an assertion *and* as an answer to [2]. Kreutel and Matheson therefore propose to analyse [6] as a move that performs the two core speech acts `accept` and `accept_answer`.

As for the details of updating IS, the following example of a simple question-answer sequence will illustrate how DH and OBL are managed. Note that for each move, for brevity, we only represent the elements that are added to DH by the update:

```
(4)  A[1]:  Did you see Peter at the party?
            DH:    <ask(A,B,q)>
            OBL:   <answer(B,[1])>
     B[2]:  Yes.
            DH:    <assert(B,p),
                    answer(B,[2],[1]) | ...>
            OBL:   <address(A,[2]), answer(B,[1])>
     A[3]:  Ok.
            DH:    <accept(A,[2]) | ...>
            OBL:   <answer(B,[1])>
            DH:    <accept_answer(A,[2],[1]) | ...>
            OBL:   <>
```

The analysis shows that A's *ok* in [3] first of all serves as an acceptance of the assertive content of [2] and discards the obligation imposed on A to `address` [2]. Secondly, it accepts [2] as an answer to [1], and this results in the obligation for B to `answer` [1] finally being removed from OBL. Assuming this basic mechanism for managing OBL, Kreutel and Matheson demonstrate how more complex sequences of questions and assertions such as those in (3) can be dealt with without the need to refer to the intentions of the DPs, assuming they act cooperatively.

2.3. Assigning the Intentional Structure

2.3.1. Basic Assumptions

The intentional structure INT is built up by means of inference rules given the information in DH. This assumes that from the occurrence of certain dialogue acts the DPs can infer each other's intentions and the way they are managed (i.e. whether they are satisfied, dropped, and so on). For each move, update of INT takes place after DH has been updated, and it is the 'whole' move, i.e. that object that comprises all the actions that are performed, which is the basis of the update. We assume that argumentation acts are processed before core speech acts.

Whereas we see intentions as objects that are individualised with respect to the DPs,[2] we consider the content of intentions as interactive goals to be achieved in the course of a discourse. The contents of the intentions associated with assertions and questions are assumed to be the goals `shared_belief(p)` and `resolved(q)`, where `p` and `q` are the contents of the assertion and the question, respectively.

We use the quadruple $< I, \gg, sat, drop >$ to represent INT formally, where I is the set of the DPs' intentions, $\gg$ is the two-place relation of *immediate dominance* in I, and *sat* and *drop* are 1-place relations in I that correspond to the sets of satisfied and dropped intentions, respectively. The relation $\gg$ is defined as follows: $i_j \gg i_k$ iff i_j and i_k are members of I, $i_j > i_k$ and there is no i_m in I such that $i_j > i_m$ and $i_m > i_k$, where $>$, the relation of *dominance*, is defined as proposed in Grosz and Sidner (1986), i.e. an intention i_m *dominates* an intention i_n iff an action that satisfies i_n also contributes to the satisfaction of i_m.

2.3.2. Determining INT

As mentioned above, the performance of dialogue acts which update the DH allows the DPs to make inferences about each other's intentions. We propose the following rules associating intentions with dialogue acts:

- `assert`: update I with i : *shared_belief*(p), where p is the content of the assertion.
- `ask`: update I with i : *resolved*(q), where q is the content of the question.
- `answer`: update I with i : *resolved*(q), where q is the content of the question the answer replies to.
- `info_request`: update I with i : *resolved*(q), where q is the content of the question the argumentation act replies to.

The simplest method of updating is to see this as an operation that adds elements to a set. However, as intentions are personalised with respect to the DPs, but not individualised with respect to the time axis, some updates will then map I onto itself; namely whenever the intention that is added already is a member of I. This case occurs, for instance, when a DP needs two attempts to answer a question. Even though on the level of dialogue acts two different `answer` argumentation acts are performed, the second answer will not introduce a new intention in I, but only 'reaffirm' the DP's intention to resolve the question (which itself can be inferred from the first answer).

In order to determine the set of satisfied intentions (*sat*) we assume the following inference rules which add information to the DH. We define an intention as satisfied iff its content occurs in the DH:

- `accept`: infer that `shared_belief(p)`, where `p` is the content of the assertion accepted.
- `accept_answer`: infer that `resolved(q)`, where `q` is the content of the question that is resolved.

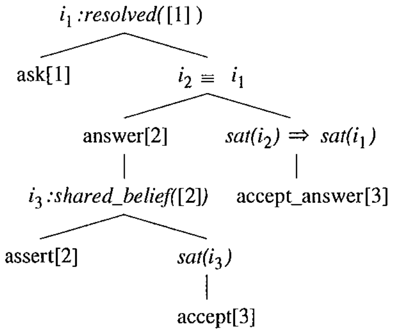
Finally, for updating $\gg$, we attach a new intention as subordinated to the last intention which has neither been satisfied nor dropped:

- When I $\neq \{\}$ and an update of I results in an intention i_j actually being *added* to I then $\gg$ is updated in the following way: $i_k \gg i_j$ where i_k is that element of I such that $\neg sat(i_k)$ and $\neg drop(i_k)$ and for all i_m for which $i_k > i_m$: $sat(i_m)$ or $drop(i_m)$.

[2] Different DPs can intend to do the same thing, but these are of course different intentions

Given these rules, we assume the structure below as the representation of the ordinary question-answer sequence below:

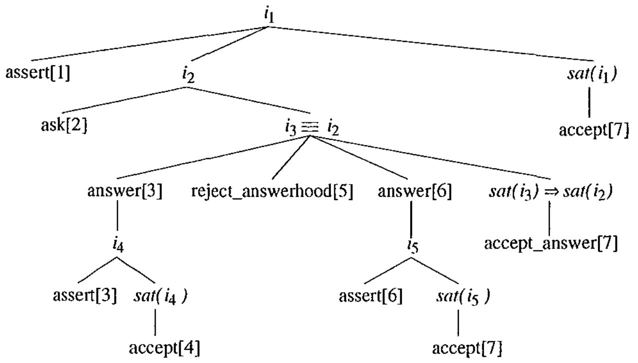
(5) A[1]: Did Helen come to the Party?
B[2]: Yes.
A[3]: Ok. Thanks.



As the tree shows, we assume that the intentions associated with dialogue acts reformulate the idea that an ASKEE adopts an ASKER's desire for information when answering a question (see Cohen and Levesque (1990)): i_1 and i_2 are identical with respect to their content, but individualised over the two DPs. The tree also captures the notion that the intention i_3 associated with the assertion in [2], which is offered as an answer to the question, is *dominated* by the askee's intention that the question should be resolved, and thus significantly differs from a discourse-initial assertion due to its context. As i_1 and i_2 have the same content, satisfaction of the latter by A's acceptance of B's answer automatically means satisfaction of i_1.

Now consider a more complex example in which a question appears in response to an assertion, requesting evidence for the latter's propositional content. The askee, however, cannot immediately come up with an answer which is acceptable to the asker (for reasons of simplicity we do not further analyse the assertion in [5] 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]: Ok.
B[5]: But she could have come by bicycle.
A[6]: I stayed there until 4 o'clock in the
morning and she didn't show up.
B[7]: Ok.



Here, A's intention i_1 to establish the assumption that Helen did not attend the party as part of the common ground constitutes the background which motivates the DPs' actions. It dominates B's intention i_2, which attempts to resolve B's uncertainty about the truth of [1] and, via i_2, also prompts A's reply to B's request. Notice that the intention associated with [6] – namely A's desire to resolve [2], which underlies the alternative answer – is identical to i_3 and does not constitute a distinct member of I, as we have pointed out above. Finally, B's *ok* in [7] not only expresses acceptance of [5] as an answer to the question in [2]; as [2] is a request for evidence, [7] is also interpreted as acceptance of the assertion in [1].

As the tree for (6) shows, the equivalence between the resolution of a request for evidence and the satisfaction of the intention associated with the assertion which the request replies to is reconstructed at the level of the mapping from dialogue acts to intentions: as pointed out in Kreutel and Matheson (1999), we assume that requests for evidence introduce a conditional in the DH that expresses the idea that the asker will accept a preceding assertion by the askee if the latter is able to come up with an answer that is accepted by the former. We then can infer $sat(i_1)$ via the resolution of the conditional introduced by [2], for which B's `accept_answer` act in [7] provides the antecedent, and via our rules for INT update, which claim that `accept` acts allow the inference that the content of the accepted assertion constitutes part of the DPs' shared knowledge.

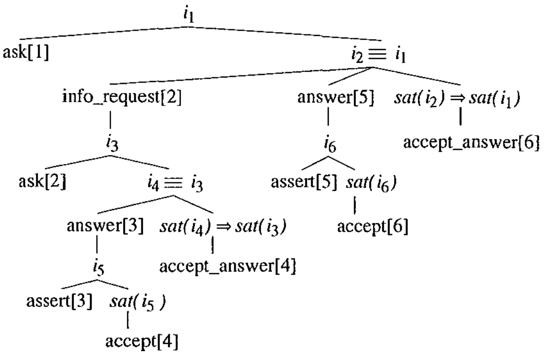
Having thus outlined how the model proposed in Kreutel and Matheson (1999) can be extended to allow for an account of the intentional structure of dialogue, the following sections discuss two issues concerning the advantages of the obligation-driven model as opposed to the notion of conversational games, on the one hand, and recent developments in intention-based approaches to dialogue, on the other.

3. Reconstructing Conversational Games

The notion of conversational games has proved to be very useful in dialogue modelling from the descriptive (Eklundh, 1983), (Carletta et al., 1996) as well as from the computational (Lewin, 1998), (Pulman, 1998) point of view. Of particular interest here is the way games can model the structure of dialogue as a reflection of intentional structure, a basic paradigm for discourse analysis postulated by Grosz and Sidner (1986). However, in terms of explanatory power the games approach runs into similar difficulties as intention-based approaches to dialogue in that capturing uncooperative actions is at best very awkward *within* the concept of a game. We argue that our own approach reconstructs the structural aspects of the games model and extends the coverage to include some data for which an approach using games is arguably too rigid.

In a games model, one could associate the games `question-answer`, `answer` and `assert` with i_1, i_2 and i_3 in (5), respectively (see Eklundh (1983) for a similar proposal), thus labelling the discourse segments associated with the intentions with types of games. So far, then, on a descriptive level our treatment of INT coincides with the basic structural mechanisms of the games model. However the latter has problems in providing a satisfactory explanation of sequences of moves that deviate significantly from the 'canonical' structure of a game, but which nevertheless can be seen as very common in human interaction. For instance:

(7) A[1]: Did Helen come to the Party?
B[2]: Did Jack come?
A[3]: Yes.
B[4]: Ok.
B[5]: Then Helen didn't come.
A[6]: Ok.



In terms of dialogue processing, either by humans or machines, the idea of a conversational game allows a distinction of the moves that follow the initial one in terms of 'preferred' or 'dispreferred' moves (see Levinson (1983), Eklundh (1983), Lewin (1998)) and assumes an increased processing effort for all the cases which are considered to be dispreferred. According to this point of view, for `question-answer` games an assertion which provides an answer will always be considered the preferred follow-up move to a question.

The possibility of the askee replying to a question with a question, as in (7), which can be considered a reasonably common case in information-oriented interaction (see INSERTION SEQUENCES in Levinson (1983)), will thus have to be treated as an exception along with requests for clarification or utterances that express the askee's inability or unwillingness to answer. Apart from the fact that the latter falls completely outside the scope of the games model (if one does not assume the contradictory notion of an 'uncooperativity game'), this way of prioritising the range of possible follow-up moves in a dialogue must be seen as too strict. The obligation-driven approach, on the other hand, simply assumes that a question obliges the askee to provide an answer which resolves the question, and thus a situation where an askee replies with a question in order ultimately to provide an answer can be seen as just an alternative to the 'canonical' case.

However, in spite of their explanatory weakness conversational games have proved to be of great use in the area of language technology and can still be seen as one of the leading notions in dialogue related research. In particular, the possibility of modelling games as recursive transition networks (see for instance Lewin (1998)), and thus of determining significant 'states' of a game, has made the games model an attractive candidate for dialogue systems which make use of probabilistic heuristics to influence the behaviour of, for example, modules for speech and speech act recognition (see Wright, Poesio and Isard (1999), Poesio and Mikheev (1998)). However, as we have pointed out above, our model defines a set of update scenarios as the framework for determining or interpreting the DPs' actual actions. In terms of scenarios, the ordinary question-answer sequence in (5) can be analysed as follows:

(8) A[1]: Did Helen come to the Party?
REPLY_QUESTION
B[2]: Yes.
REPLY_ANSWER ⊃ REPLY_ASSERT
A[3]: Ok. Thanks.

While B's move in [2] takes place in the context of a scenario in which B has to answer a question, [2] itself results in an information state in which A is obliged to `address` [2] and B is obliged to `answer` [1]. This constellation of the DPs' obligations characterises a scenario in which A's actions can be interpreted as expressing his evaluation of the answer provided by B and is identical with the situation after A's move [6] in the more complex example (6) above.

Our model thus assumes the classification of information states in terms of update scenarios as part of the expressive means, and we propose an alternative to the games approach also in terms of the model's capacity to feed back heuristic analyses to other modules in a dialogue system in the same way that a games-based account allows follow-up moves to be ranked. So, for each scenario, probabilities can be assigned to each admissible subsequent action thus allowing for the interpretation of ambiguous or 'noisy' responses. This way we are able to introduce a notion of prioritisation – which we have argued is problematic for the games approach in explanatory terms – at a higher level in our model. It should be noted that we have not undertaken such an analysis; however, it seems clear that the scenario-based account can be used to provide the same formal properties as conversational games, and hence similar predictive power can be assumed.

We have suggested that the obligation-based approach is able to reconstruct games structures at the descriptive level and also to flexibly integrate probabilities as developed in the games framework. In addition to this, however, taking into account the equivalence between 'states' in a games model and 'scenarios' on the one hand, and the definition of scenarios in terms of the DPs' obligations on the other, we claim that conversational games should be seen as structures that emerge from the DPs' acting according to the obligations imposed on them rather than as primitives in the theory of dialogue modelling.

4. Conversational Cooperation Revisited

One of the main challenges for obligation-driven dialogue models concerns the account of cooperative or uncooperative actions, as noted in recent work by Boella et al. (see Boella et al. (1999) for the central arguments). Based on a notion of POLITENESS, to which DPs are said to commit themselves for the sake of the success of the interaction, Boella et al. suggest an explanation in the framework of an intention-based dialogue model of the classic example which prompted Traum and Allen to introduce obligations:

(9) A[1]: Did Pete drive here?
B[2]: I don't want to talk about that

Distinguishing a DP's 'conversational' and 'domain goals',[3] Boella et al. assume that DPs will always try not to offend their conversational partners in order for the dialogue to proceed smoothly. This maxim is then used as an explanation for why B replies to A and thus *acts* instead of not doing anything: even though in (9) a discussion about whether Pete drove contradicts B's 'domain goals', B would have to consider A's being offended as a possible reaction if the question is simply ignored. If B's intention is the cooperative continuation of the dialogue, the unwillingness to talk about the issue raised by A must be expressed.

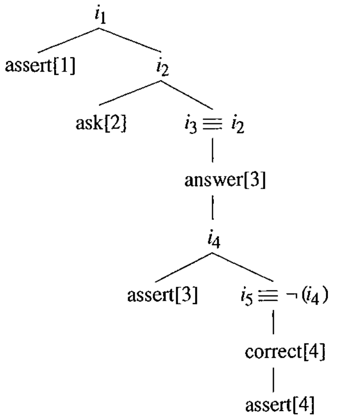
Even though the proposals in Boella et al. (1999) can be considered an elegant solution at the descriptive level to the classic 'no response' problem inherent in intention-based approaches, the idea of modelling cooperation in terms of politeness – or in terms of the DPs' intending not to offend each other – lacks explanatory power. Boella et al. (1999) themselves note that the maxim of politeness should be seen as a 'social goal', thus introducing another type of goal in their model. However, the way in which the social goal of not offending a dialogue partner and the linguistic goal of achieving a smooth interaction are related is left unspecified, even though the agents' reasoning about the linguistic goals takes into account the partner's reactions to impolite actions and therefore operates itself with a notion of politeness. Further, there is no explanation of the existence of a 'politeness goal' itself, and this would be desirable given the use of such a fairly sophisticated notion like politeness as basic explanatory means in the model.

The strength of the idea of discourse obligations as opposed to intentions in this context lies in the ability to overcome the shortcomings of intention-based accounts in explaining the existence of intentions to respond. Being 'demands for the DPs to act', obligations provide a more basic account of the reasons why DPs finally decide to act, and can be seen as the triggers of reasoning processes such as the ones outlined in Boella et al. (1999). Thus it is due to the existence of an obligation to act that DPs take account of the implications of acting in a more or less polite way.

As far as the descriptive power of obligation-driven dialogue models is concerned, Kreutel (1998) and Kreutel and Matheson (1999), as well as the examples provided above, show that obligations allow an account of considerably complex interactions without the need to refer to the DPs' intentions, given the presupposition of cooperativity. In explanatory terms, however, we can provide further arguments for conceiving of cooperative behaviour as the DPs' acting according to the obligations imposed on them. We thus provide an account of cooperativity which is more profound and at the same time less idealised then the notion of conversational cooperation employed in Boella et al. (1999), where cooperativity is defined as the DPs' trying not to offend each other.

Furthermore, supplying an account of the intentional structure of dialogue via inference rules on information states as proposed above allows the obligation-driven approach to model genuine uncooperative scenarios such as the one below, by providing the necessary structure for reconstructing the DPs' reasoning over their intentions:

(10) A[1]: Jack and Helen will split very soon.
B[2]: How do you know that?
A[3]: She doesn't love him anymore.
B[4]: That's not true.
A[5]: I don't want to discuss that.
B[6]: Then please don't try to make me
think that they will split.



After B starts a discussion about the content of [3], A knows that participation in the discussion is necessary if B is to be convinced of the claim in [1]. However, if A acts as above, expressing an unwillingness to discuss the issue raised by [3] and not coming up with an alternative answer to B's request for evidence, then A has to take into account the failure of the attempt to make B think that [1] is true.[4]

It is clear that modelling how A finally decides to act given the scenario after [4] has to involve a mechanism for evaluating the DPs' goals. However, as we have shown in this paper, the obligation-based approach is able to generate the necessary structures which feed the DPs' decision-taking processes by reconstructing the knowledge of the way their intentions – as they have become evident in the course of the interaction – are related to each other. Cooperative as well as uncooperative actions thus can be dealt with in a coherent way using the notion of discourse obligations as the basic expressive means for accounting for the DP's actions.

[3] Boella et al. (1999) also specify 'linguistic goals' which refer to the correct understanding and comprehension of utterances, and which can be used in an intention-based approach to grounding actions.

[4] Notice that in the tree for (10) i_4 and i_5 cannot be thought of as related by the *dominance* relation as they are mutually exclusive. A complete formal treatment of this kind of discussion scenario is still pending at the current stage of our model, but this does not affect the possibility of A reasoning about whether to pursue i_4 in the context of i_1 and i_3 as referred to above, given that A at least knows that in order to satisfy i_4 the problematic issue has to be discussed.

5. Conclusion

Beginning with a dialogue model which uses discourse obligations as basic means to predict the actions of participants in a dialogue, we have proposed a set of inference rules which operate on the information in the dialogue history to determine a representation of the DPs' intentions, enabling our model to cover uncooperative actions. We have shown that the structures generated by our rules reconstruct some intuitions about how to conceive of the intentional structure of dialogue. In particular, we demonstrated that the structures we assign can be reinterpreted in terms of the notion of conversational games thus showing that games can be thought of as macro structures that emerge from the DPs' acting according to the obligations imposed on them. We further suggest that the obligation-based approach to dialogue also offers a more satisfying account of the phenomenon of cooperativity in general than current intention-based models.

We neither deny the importance of prioritising the possible actions of DPs at a given state of a dialogue in an actual implementation of a dialogue system nor the necessity of complex reasoning processes about the DPs' goals in a theoretical dialogue model that is more complex than the one presented here. However, with respect to the symbolic foundations of a computational model of dialogue, we advocate the expressive power and the flexibility of the obligation-driven approach, taking into account its ability to handle a wide range of interactions exhibited in human communication.

6. Acknowledgments

The authors would like to thank the two anonymous Gotalog reviewers of this paper for their constructive criticism. We would also like to note the support of Sail Labs and the TRINDI project partners.

7. References

Guido Boella, Rossana Damiano, Leonardo Lesmo, and Liliana Ardissono. 1999. Conversational cooperation: the leading role of intentions. In *Proceedings of Amstelogue 99, the 3rd Workshop on the Semantics and Pragmatics of Dialogue*. University of Amsterdam, May 1999.

Peter Bohlin, Robin Cooper, Elisabet Engdahl, and Staffan Larsson. 1999. Information states and dialogue move engines. In *IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

Johan Bos, Peter Bohlin, Staffan Larsson, Ian Lewin, and Colin Matheson. 1999. Dialogue dynamics in restricted dialogue systems. Technical report, TRINDI Deliverable D3.2, University of Gothenburg, 1999.

Jean Carletta, Amy Isard, Stephen Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon, and Anne Anderson. 1996. HCRC dialogue structure coding manual. Research Paper 82, Human Communication Research Centre, University of Edinburgh, June 1996.

Philip R. Cohen and Hector J. Levesque. 1990. Performatives in a rationally-based speech act theory. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 79–88, Pittsburgh, Pa., June 1990. University of Pittsburgh.

Robin Cooper. 1998. Information states, attitudes, and dialogue. In *Proceedings of ITALLC-98*.

Robin Cooper and Staffan Larsson. 1999. Dialogue moves and information states. In *Proceedings of the Third IWCS*.

Barbara Di Eugenio, Pamela Jordan, Rich Thomason, and Johanna Moore. 1997. Reconstructed intentions in collaborative problem-solving dialogues. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*.

Kerstin Severinson Eklundh. 1983. The notion of language game: A natural unit of dialogue and discourse. University of Linköping, Department of Communication Studies.

Jonathan Ginzburg. 1995a. Resolving questions, I. *Linguistics and Philosophy* 18(5).

Jonathan Ginzburg. 1995b. Resolving questions, II. *Linguistics and Philosophy* 18(6).

Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, July-Sep 1986.

Jörn Kreutel. 1998. An obligation-driven computational model for questions and assertions in dialogue. Master's thesis, Department of Linguistics, University of Edinburgh, Edinburgh, 1998.

Jörn Kreutel and Colin Matheson. 1999. Modelling questions and assertions in dialogue using obligations. In *Proceedings of Amstelogue 99, the 3rd Workshop on the Semantics and Pragmatics of Dialogue*. University of Amsterdam, May 1999.

Staffan Larsson, Peter Bohlin, Johan Bos, and David Traum. 1999. Trindikit 1.0 manual. Technical report, TRINDI Deliverable D2.2, University of Gothenburg, 1999.

Alex Lascarides and Nicholas Asher. 1999. Cognitive States, Discourse Structure and the Content of Dialogue. In *Proceedings of Amstelogue 99, the 3rd Workshop on the Semantics and Pragmatics of Dialogue*. University of Amsterdam, May 1999.

Stephen C Levinson. 1983. *Pragmatics*. Cambridge University Press, Cambridge.

Ian Lewin. 1998. The Autoroute Dialogue Demonstrator: Reconfigurable architectures for spoken dialogue understanding. Technical report, prepared by SRI International Cambridge Computer Science Research Centre for the UK Defence Evaluation and Research Agency, Malvern.

Colin Matheson, David Traum, and Massimo Poesio. 2000. Modelling grounding and discourse obligations using update rules. To appear in *Proceedings of NAACL 2000*, Seattle, April 2000.

Megan Moser and Johanna Moore. 1996. Toward a Synthesis of Two Accounts of Discourse Structure. Computational Linguistics 22(3), 1996.

Stephen Pulman. 1998. The TRINDI project: Some preliminary themes. In *Proceedings of the Twente Workshop on Language Technology*.

Massimo Poesio and David Traum. 1998. Towards an axiomatisation of dialogue acts. In *Proceedings of the Twente Workshop on Language Technology*, pages 207–222.

Massimo Poesio and Andrei Mikheev. 1998. The Predictive Power of Game Structure in Dialogue Act Recognition: Experimental Results Using Maximum Entropy Estimation. In *Proceedings of ICSLP-98*, November 1998.

David Traum and James Allen. 1994. Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual meeting of the Association for Computational Linguistics*, pages 1–8, June 1994.

Helen Wright, Massimo Poesio, and Stephen Isard. 1999. Using high level dialogue information for dialogue act recognition using prosodic features. In *Proceedings of the ESCA Workshop on Prosody and Dialogue*, Eindhoven, September 1999.

A formal model of Conversational Game Theory

Ian Lewin

SRI International
Cambridge Computer Science Research Centre
23 Millers Yard
Mill Lane
Cambridge CB2 1RQ
United Kingdom
email: ian@cam.sri.com

Abstract

Conversational Game Theory represents a thread of extant research work which embodies an intuitive picture of dialogue including ideas from both plan-based rational agency and dialogue grammars. Defining the combination of these two ideas is potentially one of its most attractive features. The theory has remained a somewhat abstract architecture however with various degrees of freedom. This paper presents a detailed formalization of the theory and discusses various pressures that arise during formalization and their resolution. The formal picture is fully implemented as a dialogue manager component within trindikit - a Dialogue Move Engine framework developed for the Trindi project.

1. Introduction

Conversational Game Theory (CGT) represents a line of research (Power, 1979; Houghton, 1986; Houghton and Isard, 1987; Kowtko et al., 1992; Carletta et al., 1996) in which the following points are highlighted. First, dialogues consist of exchanges called games whose internal structure is shared knowledge between conversants. Paradigm structures are question-answer and initiate-response-feedback. Such structures form a linguistic resource that is known and exploited by conversants, for example in predicting what should happen next or interpreting what has just been uttered. Secondly, game have goals and are planned by rational agents along with other non-linguistic actions. Discourse planning occurs at the game level, not the utterance level. This planning of games may induce higher levels of structure over dialogue and this permits accommodation of familiar claims concerning determination of dialogue structure by task structure(Grosz and Sidner, 1986). Such higher level structures are not a central feature of CGT. Additional structure within games is permitted since games may nest.

The general conception is appealing but how should one fill it in? Various different forms of theoretical support are appealed to in CGT literature. (Houghton and Isard, 1987) is clearest in suggesting that moves should be speech-acts in the tradition of (Austin, 1962; Searle, 1969) and that games are conventional sequences of speech acts. (Carletta et al., 1997) does not actually mention speech-acts but does claim that games are 'often also called dialogue games (Carlson, 1983; Power, 1979), interactions (Houghton, 1986), or exchange structure (Sinclair and Coulthard, 1975).' The most that can be fairly claimed is a loose family resemblance. Sinclair and Coulthard are very keen to distinguish their approach from speech-act theory, for example. Their classification is generally less fine-grained than that of speech-act theory and is designed to exhibit interactive functions such as *eliciting X with Y*, *acknowledging X with Y*, *evaluating X with Y*. They are also very keen not to permit embedding structures (but reluctantly admit the possibility of some in (Coulthard, 1977)).

Rather than relying on pointers to elucidation, one can examine ways in which CGT *has* been filled in. The most widely known use of CGT is the markup of the Map Task Corpus (Carletta et al., 1996). A corpus of 128 task-oriented dialogues is marked up with a basic set of 12 conversational move categories: instruct, explain, align, check, query-yn, query-w, acknowledge, clarify, reply-y, reply-n, reply-w and ready. Of these, the first 6 are potential game-initiating moves. The corpus is also marked up with actual game-initiators, game-closers (points at which a game finishes or is abandoned) and an embedding flag indicating whether or not a game is top-level or embedded. Unfortunately, this particular way of filling in the general conception appears to be just that: one particular way. What, for example, motivates the particular move categories chosen? Why are reply-y, reply-n and reply-w distinguished as different moves? Why should checks be distinguished from other sorts of queries? Why are embedding levels, rather than just an embedding flag, not annotated? The reliability of the coding scheme (how well coders can reproduce it) has also been tested (Carletta et al., 1997). Move coding proved more reliable than game coding. The latter was "promising but not entirely reassuring". Agreement on embedding was poor. These results are suggestive: the notion of move (at least, those moves prescribed for the Map Task corpus) seems learnable but those of game and nested game seem much less stable. Yet it is the notion of game that is central to CGT.

The other uses of CGT are in working implementations of dialogue systems. CGT has been realized in a series of computer programs (Power, 1979; Houghton, 1986; Carletta, 1992) designed to model agents who share and use game definitions. Without a clear conception of the underlying theory however, it can be unclear whether parts of an implementation represent real features of the theory or just simplifications or essential modifications of it. It may be unclear what remains unimplemented and whether this is important. Processing issues and descriptive issues may be problematically entangled.

In the following sections, we describe one possible formalization of Conversational Game Theory and give an algorithm for playing games. A discursive section then motivates our particular design choices and discusses a number of issues and pressures that arise and how these issues are resolved or at least clarified by the formalization.

2. Formalization

Conversational moves are objects with a category (e.g. *reply*) and a content (*that london is the destination*). Conversational games are rule-governed sequences of moves also with a category and a content (e.g. finding-out, giving-information). So, let G and M be disjoint finite sets of game types and move types. We suppose contents are identified by sets of propositions. P is the set of all propositions. Then a *move* is a pair $\langle\beta, p\rangle$ and a *game* is a pair $\langle\gamma, p\rangle$ where $\beta \in M$, $\gamma \in G$ and $p \subseteq P$. For example, {query,inform,greet,pardon,interrupt} $\subset G$, {qw,rw,ack,cnf,ryes,inf} $\subset M$.

With one eye on implementation, we suppose the game rules can be coded up in simple recursive transition networks. Each game type $\gamma \in G$ has a grammar RTN_γ, whose arcs are labeled with conversational move and game types. $RTN_\gamma = \langle s, t, r, s_0, f\rangle$, where s is a set of states, $t \subseteq (M \cup G)$ is a set of arc labels, $r \subseteq s \times t \times s$ is the transition relation, s_0 is the initial game state and $f \subseteq s$ is the set of final game states. Some example games are illustrated in figure 1.

An *instance* of a game γ_k is a paired game and function $S : s \rightarrow Pd$. That is, for each state of the game, we identify a set of propositions under discussion at that point in the game. Pd is defined below.

Moves are realized by utterances, and, in general, we suppose the categorization of an utterance may depend on its form, content and relation to the current propositions under discussion. Let realize(u,$\langle\beta, p\rangle$,q) mean utterance u realizes a move of type β and content p in context q ($q \in Pd$). Then, a sequence of utterances $u_0, u_1, \ldots, u_n$ realizes an instance of a game $\langle\gamma_k, p_0\rangle$ just in case there is a sequence of game states $s_0, s_1, \ldots, s_n$, move types $\beta_0, \beta_1, \ldots, \beta_n$ and sets of propositions $p_0, p_1, \ldots, p_n$ such that, for all i, realize(u_i,$\langle\beta_i, p_i\rangle$,$S(s_i)$), and the sequence of move types is accepted by RTN_γ. The content of the first move of a game is also the content of the game.

Semantically, we suppose that games are commitment slate updaters and that moves are updaters of propositions under discussion. A commitment slate is just a set of propositions ($c \subseteq P$) and a set of propositions under discussion is another such set with possibly one distinguished member: $\langle p, d(p)\rangle$, where $p \subseteq P$ and $d(p) \in p$, if there is a distinguished member else $d(p) = 0$. The distinguished member is one highlighted by confirmation moves. If Cm is the set of all commitment slates and Pd the set of all sets of propositions under discussion then a dialogue model is $\langle I_1, I_2, Pd, Cm\rangle$ where I_1 interprets move and game types M $\cup$ G and I_2 interprets game types G.

We suppose move type tells one *how* to update the propositions under discussion ($x \in Pd$) and the move content ($y \subseteq P$) tells one *what* to update it with. The result is another set of propositions under discussion ($y \in Pd$). Consequently, $I_1(\beta) : Pd \times P \rightarrow Pd \quad (\beta \in M \cup G)$. The value of a move $[\![\ ]\!]_1$ is a function $Pd \rightarrow Pd$ and is defined by $[\![\langle\beta, p\rangle]\!]_1 = \lambda x.[I_1(\beta)(x, p)]$. That is, take the *type* of operation on Pd denoted by the move type β and apply it to the current set of propositions under discussion and the content of the move p, thereby generating a new set of propositions under discussion. A sequence of moves determines, by functional composition, a function $Pd \rightarrow Pd$. Applying this function to the propositions under discussion at the start of a game {} results in the propositions under discussion after that sequence. Let $Ag(\gamma_k)$be the propositions left when the game completes. $Ag(\gamma_k)$ will be $S(s)$ for the terminating state $s \in f$ reached in playing the game.

We adopt similar definitions for game semantics. $I_2(\gamma) = Cm \times Pd \rightarrow Cm$ (game type tells you *how* to update the commitments). $[\![\gamma_k]\!]_2 = \lambda x.[I_2(\gamma)(x, Ag(\gamma_k))]$ (Apply the operation denoted by the game type to the current set of commitments and the propositions left at the end of the game). Finally, a dialogue consisting of a sequence of games determines by functional composition a function from $Cm \rightarrow Cm$ and applying this function to {} (the commitments before conversation begins) gives the value of the dialogue so far.

The model permits fully incremental processing of dialogue meaning in that there is a function which determines the meaning of a discourse (its commitments and propositions under discussion) after any initial substring of utterances within it.

3. A simple instantiation

Figure 1 shows the move and game definitions implemented in the Trindi Autoroute Demonstrator. [1] The demonstrator is a simple dialogue system in which the system queries the user for various route parameters (origin, destination, time and so forth) before making a web query to one of a number of route planning services.

A query game opens with a query move (qw) or a "restricted query move" (qw-r) which should be followed by a reply (rw). The querier may then either terminate the game with an acknowledgment (ack) or make a confirmation move (cnf). Possible responses to a confirmation include a "reply-yes" move (ryes), a "reply-no" move (rno) and a "reply-modifier" move (rmod). The latter move encodes a correction such as "No, to Leicester" following a confirmation of "You want to go to Bicester. Is that correct?".

An information-giving game consists simply of an information move (inf) followed by an acknowledgment (ack).

A pardon game consists of one player making a move that is (deemed to be) indecipherable (unrec) and the other player saying "pardon" (pdn).

An interruption game consists of one player saying something that is (deemed to be) unimportant (unimp) and the other player playing an information game (inf).

Although not shown in the diagram, pardon and interruption games can link any node in any game back to itself.

[1] TRINDI (Task Oriented Instructional Dialogue), EC Project LE4-8314, http://www.ling.gu.se/research/projects/trindi

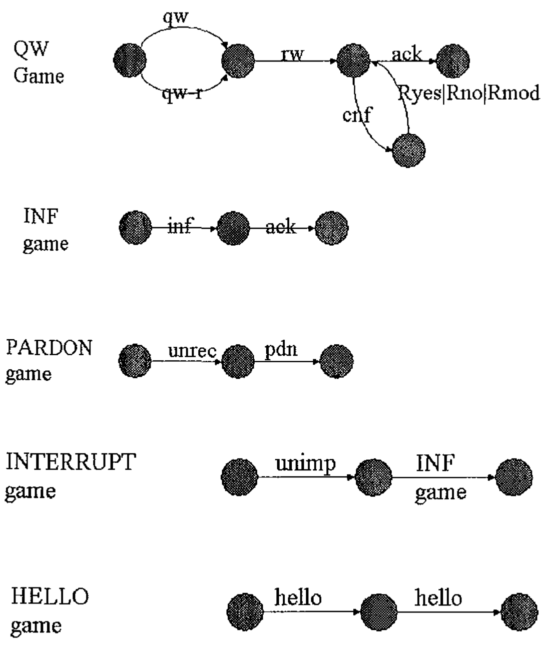


Figure 1: Game Definitions

For the game semantics, qw and inf games simply add their content to the commitment slates. That is, if the game began with a commitment slate of Cm and it terminates with propositions under discussion of $\langle p, d(p)\rangle$, then the new commitments are just $Cm \cup p$. All other games represent the identity function on commitments. There are no games for commitment revision in the demonstrator.

For the move semantics, all initiating moves initialize the current propositions under discussion with the content of the move without distinguishing any member. That is, if the move content is p, then the propositions under discussion become $\langle p, 0\rangle$. The rw and inf moves behave similarly. That is, if Pd is the set of propositions under discussion, and p is the content of the move, then the propositions under discussion become $\langle p, 0\rangle$. Ack, ryes and pardon moves represent the identity function on propositions under discussion. Cnf distinguishes a proposition. So, if $\langle p, d(p)\rangle$ is the set of propositions under discussion and c is the content of the cnf move ($c \in p$), then $\langle p, c\rangle$ is the new propositions under discussion. Rno simply removes a proposition p from the set (and, if p were distinguished then $d(p)$ becomes 0). Rmod replaces the distinguished proposition under discussion with the content of the move. That is, $\langle p, d(p)\rangle$ updated by move content c and becomes $\langle p - d(p) + c, c\rangle$.

4. Game Playing

Games are represented by simple RTNs so game-playing is viewed as 'parsing' such a network. Of course, it cannot be just parsing since the system must generate as well as interpret utterances. The overall structure we employ is that of a dialogue monitor (something which, utterance by utterance, parses the dialogue) alternating with a dialogue contribution generator. The monitor maintains a picture of the current game state. The contribution generator either looks for user input, if the game state indicates it is the user's turn or else it generates an output itself appropriate to the game state. Either way, a new token is generated and the dialogue monitor updates its picture to incorporate it.

The monitor functions as a parallel, no-look-ahead, incremental parser. After each token appears, the set of next possible states is calculated, each of which is at the end of a possible path in the network covering all the utterances in the current game. All the possible paths are then sorted according to a utility based preference mechanism. The topmost item on the agenda represents what the system actually believes the current state to be and the path to it. The other paths are maintained however and each is extended to incorporate new tokens when they appear. The appearance of a new token may result in some paths being discarded and new ones being generated. In particular, the topmost path may not be extendible to incorporate a new token and, even if it is, the result may not be most preferred by utility measurement.

5. What do Conversational Games and Moves do?

Games and moves, like speech acts, are most naturally thought of as objects that do things and formalized as context update functions. What is the nature of the contexts to be updated?

In speech act theory, two sorts of contexts have been used. In a tradition beginning with (Hamblin, 1971), speech acts are commitment slate updaters. A commitment slate is not to be thought of as a set of mental attitudes but as public objects that one can be held to. One might say the stress is on what you say committing you, rather than your committing to what you say. Researchers interested in formalizing legal argumentation (notably (Mackenzie, 1979), (Lodder, 1998) is a recent review) have further developed the idea. The picture contrasts strongly with the "speech acts as plan operators" approach (beginning with (Cohen and Perrault, 1979; Perrault and Allen, 1980)) which defines utterances as actions with pre-conditions and effects all describing the mental states of the dialogue participants. In that account, the primary and intended effects of our dialogue utterances are to update each other's mental states. Understanding a dialogue becomes immediately a matter of discerning other people's mental states.

In our approach to CGT, we apply the commitment idea at the level of games. The primary effect of games is to commit dialogue participants publicly to the truth of certain propositions, to undertake certain actions and so forth. Furthermore, games commit all the participants. The point of moves is to negotiate the objects to commit to. In many ways, this is a natural development of Hamblin's original picture (see especially his 'System 7'). Hamblin's contexts are sets of commitment slates, one per discourse participant and each locution act is an updater of everyone's commitment slate. If I assert p, everyone becomes committed to p. Hamblin assigns some locution acts, including inquiries and retraction demands, the identity function on commitment slates. However, he also adopts a rule (the rules define the set of legal dialogues) that after a question there may follow only an assertion which answers it or an 'I don't know' locution. Consequently, the significance of a question is really given derivatively as the set of assertions that may legitimately follow it rather than in its own context change potential. It is very natural, indeed more perspicuous, to recast this into a games framework in which questions and answers are equally moves that negotiate a set of objects that form the final commitments to update with. The picture of moves as updaters of 'propositions under discussion' has evident affinities with earlier suggestions including especially contract games (Dahl, 1981) and also questions under discussion (Ginzburg, 1995). Suggestions along similar lines can also be found in (Stalnaker, 1978).

6. Identifying moves

The picture of negotiation moves in pursuit of publically agreed commitments promises a much more satisfactory account of the role of certain paradigmatic dialogue moves such as acknowledgments, replies and answers than one based on speech-act theory (cf. (Houghton and Isard, 1987)). An initiating move may propose a public commitment or ask for one. An acknowledgment accepts one. A rebuttal rejects one. An answer supplies one. In contrast, it seems very hard to give a speech-act characterization of moves such as replies and answers. As (Levinson, 1983) remarks, such categories depend on possibly complex meaning and sequential relations between utterances and not on their content plus an illocutionary force. There is no illocutionary force of answering. The possibility of such complex relations is built into our formalization through the realization relation realize(u,$\langle\beta, p\rangle$,q). Some moves are discernible purely on the basis of form - e.g. 'pardon' realizes a pardon move simply in virtue of its form. However, 'I want to go to London' realizes a *reply* move only in virtue of its meaning and the fact that the propositions concerning the desired destination have been put under discussion by a previous question. In different circumstances, 'I want to go to London' might best be interpreted as an *inform* move.

We should note too that the general picture also offers an answer to some of our original questions, for example, why are reply-y and reply-n distinguished? The answer is simply that in one case an offered commitment is accepted and in the other it is not.

7. Identifying Games

When do games begin and end and when do they *nest*? In general, CGT literature simply notes that the guiding idea for detecting a game's beginning and end is that a game has a purpose and lasts until it is achieved or abandoned. Perhaps unsurprisingly, and as we noted earlier, achieving reliable agreement on game-nesting has proved difficult. The reason it is unsurprising is that purposes can be discerned at various levels of granularity and one may be unclear what level of granularity to use. Does one game (the outer bracketing) cover the whole of Dialogue 1 (Figure 2)? Certainly, the purpose of obtaining a route is not achieved until utterance 10.

Alternatively, is the structure covering 1 to 10 not a game structure but a higher level structure? What really distinguishes these two ideas?

1. U: I'd like a route please.
2. S: Okay.
3. S: Where do you want to travel?
4. U: from Bideford to Exeter
5. S: When would you like to travel.
6. U: I want to arrive at 4 p m.

. . .

10.S: The quickest route from Bideford to Exeter is . . .

Figure 2: Dialogue 1

It is noteworthy that CGT implementations tend to employ very little nesting themselves. (Carletta, 1992) discusses repair strategies in the context of CGT. In her analysis, if a problem arises in understanding an initiating move (A says, 'the first section of the route goes between the palm beach and the swamp' but B has no knowledge of any swamp), then B simply closes the game with 'I don't understand'. Then either A or B, depending on whether B cedes the turn, may initiate a new game. B may reflect on *partial information* he generated during the game in order to reconstruct the speaker's goals in starting it, viz. A wants to give a route, and that requires giving the first section and that requires identifying the palm beach and something else ('the swamp'). B may then adopt these goals himself, spot the problem and start a repair game with 'Where is the swamp?'. However, although the discourse or intentional structure can be highly complex (involving structured goals and sub-goals) this is not reflected at all in game structure. In the example, no game is opened for 'giving a route' or 'giving the first section'. The repair game 'Where is the swamp?' is not a nested game either. Indeed, there does not appear to be any game nesting. Every game is essentially initiate-response or initiate-response-feedback. The rest of the structure comes via planning.

(Houghton and Isard, 1987) also only permits one sort of embedding. In Houghton's system, after a game has been initiated, either the reply is of the expected type in which case the current game continues or an interruption is assumed.

Clearly, the general formalization of section 2. permits arbitrary embedding. The RTN framework permits any game to occur at any point within any other game. Whilst this may seem satisfactorily general, reflection on the meaning of embedding structures suggests otherwise.

If nesting is permissable at all, then, from our formal perspective, games must have a dual role: 1) in updating propositions under discussion 2) updating commitment slates. However, these two roles are not orthogonal. It makes no sense to suppose that a nested game could result in α being added to a commitment slate and yet remain under discussion in the embedding game. Similarly, if α is under discussion in a game, what could a nested game do? Put it under discussion again? Perhaps the paradigm example of a nested game is the confirmation sub-dialogue

1. U: I'd like a route to Cambridge please
2. S: You want a route to Cambridge?
3. U: Yes
4. S: Okay

Figure 3: Dialogue 2

It seems the most natural thing in the world to identify the sequence of 2 and 3 as a sub-dialogue. But what really is meant by so identifying it? It is worth stressing first of all that the mere fact that 2 sets up an expectation for 3 is quite insufficient to deem the sequence a game. After all, a reply sets up an expectation for an acknowledgment but they do not form a game. A perfectly flat sequence of moves encodes the dependency as well as a nesting structure.

It is also insufficient that the sequence of 2 and 3 might occur as a game *in another context*. For example, one can imagine S encountering a tourist in St. Ives looking hopelessly lost but clutching a map and a guidebook to Cambridge. The sequence of 2 and 3 then follow. Again, the fact that 'I want to route to London' could be an *inform* move in another context does not entail that it so functions in the current context.

The issues are delicate. But without some sort of understanding of the meaning of structures, what are we discerning them for?

There is of course another pressure in CGT towards minimal embedding. This is the element of the picture that games are intended to encode shared linguistic knowledge of possible structures of moves. The knowledge is intended to be distinct from general considerations of 'rational agency' that agents may employ at the game level. There is something *missing* in a dialogue with an initiation but no response. It is perhaps worth stressing that, just as with sentential grammar, communication can still proceed perfectly adequately. Half utterances and ungrammatical utterances are often good enough for beings with enough intelligence and pragmatics. Consider again the issue of the possible embedding surrounding the whole of Dialogue 1 (figure 2). The questions for us are: does some proposition remain under discussion until 10 or, rather, is some commitment established early; is the role of utterance 10 predictable on the grounds of shared knowledge of linguistic structure – or just on the basis that a rational and cooperative agent will address an agreed commitment at some point in the future? (Recall that S himself might not like to think of himself as committing so early. Perhaps he withholds the *Okay* in utterance 2, adopts a worried look and asks his follow-up questions in a highly quizzical tone. But S's mental state does not settle the matter. Is it reasonable to *hold* S as having committed early?)

Of course, 10 might not occur at all. It seems that the content of the conversation after 6 might simply remove any need to further address utterance 1. Consider the continuation of Dialogue 1 shown in figure 4

7a. S: What sort of car will you be driving?
8a. U: I don't want to go by car
9a. U: I want to go by bus and train
10a S: I cannot give bus and train information
11a.U: Okay
12a.U: Bye

Figure 4: Dialogue 3

The general idea is that, rather than positing a piece of incomplete structure (to be dealt with in discourse pragmatics), the role of the initial utterance is best construed as establishing some content (namely, a commitment of a certain sort) which content is later addressed or becomes obsolete in virtue of the way the rest of the conversation proceeds. Again, the issue is really quite fine but only by clarifying and refining some reasonable notion of content can it become in any way tractable.

Finally, we should note that *arbitrary* embeddings seems far too general. Nesting is permissible; but there are different sorts of embedding. Can a game begin with another game, for example? A focus on purely on purpose and intention, as in the discourse segment theory of (Grosz and Sidner, 1986), might encourage this outlook. Focussing on the structural expectations generated by an utterance however reduces this pressure. (Carletta et al., 1996) states that a game may nest if it serves the goal of a game which has already been initiated. (Houghton and Isard, 1987) similarly refers only to the *immediate and intentionally apparent goal in initiating the interaction.* It is also interesting to note that LINLIN dialogue structures (also fairly small and simple Initiate-Response structures) permitted recursion but only certain examples such as (I (I R) R), (I (I R) (I R) R), (I (I (I R) R) R) are ever reported (Jönsson, 1992).

7.1. Planning for Commitments versus Beliefs

Plan based approaches to dialogue management can involve highly complex reasoning about beliefs about others' beliefs and intentions. If A asserts x and B acknowledges it then B will believe A believes x (in virtue of the assertion) and conversely A will believe B believes x (by the acknowledgment). But if B rejects x, then all one can say is that A will believe B believes *not* x. A's beliefs about x itself will depend on how A responds to the rejection and B's beliefs about A's beliefs about x will depend on how he models those thought processes of A.

The "public joint commitment" picture avoids the need, at the level of game playing, for such representation and reasoning. A game serves to establish propositions to commit to publically and the moves serve that purpose. In the case where B rejects A's assertion, no joint public commitments result from the game. The picture offers nothing about what mental states people may have after a game concludes. Indeed, even if B had acknowledged x, he might have done so even though he himself did not believe x. A might have asserted x, whilst disbelieving it. Nevertheless A and B have committed to x. Whatever their mental states are, they have laid themselves open in a certain way - parties can justifiably complain if later actions or utterances are inconsistent with declared commitments. The useful simplification is to separate at the discourse level what an agent does from *why* he does it.

Of course, in dialogue systems generally, it is held that modeling others' beliefs and intentions is potentially valuable in building *helpful* systems: e.g. systems which can respond to what are discerned to be the user's underlying though not expressed goals ; or systems which can attempt to understand how dialogue problems have arisen and thereby repair them. It is therefore a very interesting question how far one can get without such modeling. For instance, in our implementation, if a game fails, the commitments are unaltered and agents must plan their next move on just this basis. This is clearly overly simple for a complete analysis of human-human dialogues but it might be very appropriate for practical computer based systems. If one does not worry about why the last game went wrong then one evidently risks encountering the same problem again, but this should be compared with the risk of incorrectly diagnosing the problem in the first place.

(Traum and Andersen, 1999) discusses a number of examples in which A makes a request, B carries out an action and then A corrects B. For instance,

(A train called Metroliner has just moved from Boston to Albany)
A: send the Boston train to New York
B: *(sends it)*
A: No, send Metroliner

Did A mis-speak "Boston" for "Albany", or B mishear? Did A *mean* Metroliner by "the Boston train"? Has A just changed his mind about what he wants? Traum notes that, whatever the truth, the best continuation strategy is in fact just to carry out A's subsequent request. Traum also discusses more complex examples in which, he suggests, that just carrying out the subsequent request is not optimal.

A: send the Boston train to New York
B: *(sends it)*
A: No, send the Boston train

Traum thinks B should calculate that "the Boston train" cannot mean what he thinks its means and choose another candidate (presumably Metroliner). This is not an easy calculation - maybe A didn't realize B *had* sent the Boston train so sending it again is the right thing to do. Clearly, the value added to systems by incorporating such reasoning needs careful evaluation.

8. Execution strategies and Re-analysis

A formalization of *what* conversational games are enables a clean separation from processing issues.

Most work on conversational games has indeed been procedurally oriented and followed an execution model exemplified by (Houghton and Isard, 1987). Houghton defines games by a goal, a precondition for attempting the game, an executable procedure (e.g. search one's memory, alter the world in some way) and a reply type. If no game is current and an actor receives input, it must decode the goal, load the appropriate game definition into memory, execute the procedure and issue a reply of the right type. The initiating actor must then compare the reply with the expected type. In Houghton's system, either the reply is of the right type and the game closes or an interruption is assumed. In such a system, what recovery strategies are there should one discover a misunderstanding? If an initial goal is decoded incorrectly then the reply may not be of the expected type and so an interruption will be assumed. It is still assumed however that the initial goal *was* correctly decoded and so a 'correct' response to it will be expected.

Our work has been undertaken within the Trindi project which has been promoting an Information State Update view of dialogue. Utterances are to be analyzed as information state transformers. Primary interest attaches to identifying different notions of information state and their updates. The picture is undertaken for intrinsic interest, to attempt to compare alternative dialogue theories re-cast into it and also to build a generic 'dialogue move engine' for experimentation. How much can you do with small information states? How much does more state buy you? The general architecture is shown in figure 5.

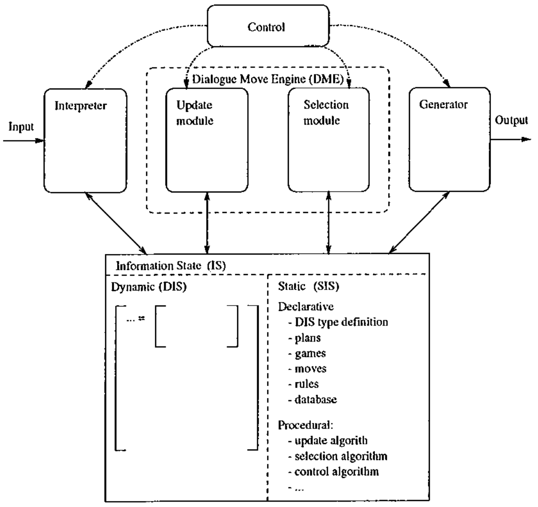


Figure 5: The TRINDIKIT architecture

The Information State functions like a blackboard and contains both static parts and dynamic parts. The dynamic parts are those that can change during the course of a dialogue. Apart from the overall control module, the functions of the modules are meant to correspond to the following highly general account of what a dialogue manager must do:

- interpret inputs, classifying them as instances of known moves
- update the information state based on the classified move
- choose a move (or action) to carry out next
- update the information state based on that move

In perhaps the most natural instantiation of this model, the choice of which move was made by an utterance is taken prior to operation of the update rules in the update module. The update rules encode how to update the information state given a particular choice of move. That is, one first recognizes an utterance as constituting, for example, a question and then applies the update rule for questions (or, one of the rules for questions, if the rules are also context sensitive) thereby updating the information state.

Once again, such a model implies a certain degree of early commitment. For how can the system undo or redo an update, if a problem arises later which is caused by an earlier incorrect choice? In order to allow a revision or 'undo' of an early choice, the information state must carry around within it not just the new information generated by the early choice but also information that might be required if the choice is to be un-made or re-made.

In our implementation, the Information State is actually the current agenda of a parallel parse which contains alternative interpretations of what was said. (The parallel feature is clearly not psychologically realistic). Subsequent information can cause the system to change its mind about the best parse of earlier inputs. For example, when the system asks a question 'When do you want to travel?', it analyses its own output as possibly being an `unrecognized` move (that is, a publically indecipherable output which put no propositions under discussion at all), as well as being a question move. Subsequently, should the system hear 'pardon', it will determine that the current interpretation of its own original utterance is no longer viable since 'pardon' is not a valid move in the qw game. The interpretation in which its own utterance was not recognized becomes preferred. No question was ever asked. No such move was actually made.

The point is of course a rather obvious one once games have been declaratively defined. There is a more substantive point available though. Having raised the issue of re-analysis, how much re-analysis might be necessary or possible? How open to re-interpretation are all our earlier utterances? By adopting the commitment picture of conversational games, we effectively limit the amount of re-interpretation permitted. The *point* of a game is to agree something so, once the game is completed, the path by which it was agreed is no longer needed. If I assert something and you agree, then we are entitled to behave on the basis of what (we think) we have agreed. Of course, problems may still arise later but, if they do, then that is a basis for *complaint* not re-interpretation: But you asserted x!; But you agreed to y!. One will not attempt to re-interpret any exchanges in the dialogue which followed the identified earlier problem. Of course, as an academic exercise, sometimes one can be interested in understanding how such a mis-understanding remained hidden for so long, but reconstructing the mis-construal is not a strategy for repairing the dialogue.

9. Conclusion

We have developed a formal account of CGT in an attempt to more precisely delineate its main features and possible contributions to dialogue analysis. In order to sharply delineate the roles of games and moves, we have taken games to be update functions on "publically agreed commitments" whereas moves update "propositions under discussion" in order to establish those commitments. We have explored some of the reasons for using these notions and their possible impact on several salient issues for CGT: the relative roles of rational agency and shared knowledge of game-structure; the extent and amount of embedding structures; the amount of re-analysis that might be required in understanding dialogue.

10. References

J.L. Austin. 1962. *How to do things with words*. Oxford: Clarendon Press.

J. Carletta, A. Isard, S. Isard, J. Kowtko, and G. Doherty-Sneddon. 1996. Hcrc dialogue structure coding manual. Technical report, Human Communication Research Centre, University of Edinburgh. HCRC research paper TR-82.

J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, forthcoming.

J. Carletta. 1992. *Tisk Taking and Recovery in Task Oriented Dialogues*. Ph.D. thesis, University of Edinburgh, Scotland, UK. unpublished PhD Thesis.

L. Carlson. 1983. *Dialogue Games: An approach to Discourse Analysis*. D.Reidel.

P Cohen and C.R. Perrault. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212.

M. Coulthard. 1977. *An Introduction to Discourse Analysis*. Applied Linguistics and Language Study. Longman.

Ö. Dahl. 1981. The contract game. In J. Groenendijk, T. Janssen, and M. Stokhof, editors, *Formal Methods in the Study of Language*, pages 79–86. Mathematical Center, Amsterdam.

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

B.J. Grosz and C.L. Sidner. 1986. Attention, intentions and the structure of discourse. *Computational Linguistics*, 12:175–204.

C.L. Hamblin. 1971. Mathematical models of discourse. *Theoria*, 37:130–155.

G. Houghton and S.D. Isard. 1987. Why to speak, what to say and how to say it: Modelling language production in discourse. In P. Morris, editor, *Modelling Cognition*. John Wiley and Sons.

G. Houghton. 1986. *The Production of Language in Dialogue*. Ph.D. thesis, University of Sussex.

A. Jönsson. 1992. *Dialogue Management for Natural Language Interfaces*. Ph.D. thesis, Linköping University. Linköping Studies in Science and Technology, No. 312.

J.C. Kowtko, S.D. Isard, and G.M. Doherty. 1992. Conversational games within dialogue. HCRC research paper RP-31.

S.C. Levinson. 1983. *Pragmatics*. Cambridge University Press.

A.R. Lodder. 1998. On structure and naturalness in dialogical models of argumentation. In J.C. Hage, Bench-Capon T.J.M., Koers A.W., de Vey Mestdagh C.N.J., and Grtters C.A.F.M., editors, *Legal Knowledge-based Systems. JURIX: The Eleventh Conference, GNI, Nijmegen*, pages 45–58.

J.D. Mackenzie. 1979. Question begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133.

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

R. Power. 1979. The organization of purposeful dialogues. *Linguistics*, 17:107–152.

J.R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge Univeristy Press.

J. M. Sinclair and R.M. Coulthard. 1975. Towards an analysis of discourse: The english used by teachers and pupils.

R.C. Stalnaker. 1978. Assertion. In P. Cole, editor, *Pragmatics*, volume 9 of *Syntax and Semantics*, pages 315–332. New York: Academic.

D.R. Traum and C.F. Andersen. 1999. Representations of dialogue state for domain and task independent meta-dialogue. In *Proceedings of the IJCAI'99 Workshop on Knowledge And Reasoning In Practical Dialogue Systems, Stockholm*, pages 113–120.

Empirical study of the anaphoric accessibility space in Spanish dialogues

Patricio Martínez-Barco and Manuel Palomar

Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
Carretera de San Vicente del Raspeig - Alicante - Spain
Tel. +34965903653 Fax. +34965909326
{patricio, mpalomar}@dlsi.ua.es

Abstract

This paper shows an empirical study about the anaphoric accessibility space in Spanish dialogues. According to this study, antecedents of pronominal and adjectival anaphors can almost always (95.9%) be found in the noun phrases set taken from spaces defined using a structure based on adjacency pairs. Furthermore, a proposal of a reliable annotation scheme for Spanish dialogues is presented in order to define this anaphoric accessibility space. Using this annotation scheme, anaphora resolution algorithms can locate the adequate set of anaphor antecedent candidates.

1. Introduction

Anaphora resolution is one of the most active areas of research in Natural Language Processing (NLP). The comprehension of anaphora is an important process in any NLP system, and it is among the toughest problems to solve in Computational Linguistics and NLP.

According to Hirst (1981): *"Anaphora, in discourse, is a device for making an abbreviated reference (containing fewer bits of disambiguating information, rather than being lexically or phonetically shorter) to some entity (or entities) in the expectation that the receiver of the discourse will be able to disabbreviate the reference and, thereby, determine the identity of the entity."*

The reference to an entity is generally called an anaphor (e.g. a pronoun), and the entity to which the anaphor refers is its referent or antecedent. Moreover, it is well-known that anaphora is a mechanism used by speakers in conversation to achieve the common ground. Thus, NLP systems need to both resolve and generate anaphora and they generally resolve it by constructing a set of possible antecedents and then choosing the best one. For this, it is necessary to decide the adequate anaphoric accessibility space, i.e. the space where any anaphora has its candidate set of possible antecedents.

According to Dahlbäck (1991), the efforts made so far towards resolving anaphora can be divided into two basic approaches : Traditional and Discourse-oriented. The traditional approach generally depends on linguistic knowledge. In the discourse-oriented approach, however, the researcher tries to model the complex structure of discourse. Anaphora, accepted as discourse phenomena, is resolved with the help of that complex structure. These works are mostly focused on defining anaphora resolution algorithms, both the traditional approaches (Hobbs, 1986), (Baldwin, 1997), (Mitkov, 1998) and the discourse-oriented ones (Grosz et al., 1995), (Strube and Hahn, 1999).

However, the former do not perform a defined proposal about anaphoric accessibility space, and the latter constraint the space for possible antecedents to the previous utterance. Although, this strategy is adequate for English processing, its application to other languages such as Spanish is not such suitable. For instance, Spanish personal pronouns contain more morphological information than English ones. This makes Spanish speakers to expect larger anaphoric accessibility spaces.

This paper shows that in Spanish dialogues, antecedents of pronominal and adjectival anaphors can almost always be found in the set of noun phrases taken from the anaphoric accessibility space. This space is defined according to an structure based on adjacency pairs (or synchronizing units according to Eckert and Strube (1999a)). Furthermore, a proposal of an annotation scheme for Spanish dialogues is presented, in order to define this anaphoric accessibility space. Moreover, a detailed study of this space and the antecedents we have found in it has been carried out.

Our proposal has been evaluated on the *Corpus InfoTren: Person*, a corpus of Spanish dialogues provided by the BASURDE (1998 2001) Project. These dialogues are conversations between the telephone operator of a railway company and a user of the company.

2. A proposal for an annotation scheme for dialogue structure

For the successful processing and resolution of anaphora in dialogues, we believe that the proper annotation of the dialogue structure is necessary. With such a view, we propose an annotation scheme, for Spanish dialogues, that is based on the work carried out by Gallardo (1996), who applies, to Spanish dialogues, the theories put forward by Sacks et al. (1974) about the taking of speaking turns (conversational). According to these theories, the basic unit of knowledge is the *move* that can inform the listener about an action, request, question, etc. These moves are carried out by means of *utterances*[1]. Therefore, utterances are joined together to become *turns*.

Since our work was done on spoken dialogues that have been written (transcribed), the turn appears annotated in the

[1] An *utterance* in dialogues would be equivalent to a sentence in non-dialogues, although, due to the lack of punctuation marks, utterances are recognized by means of speaker's pauses.

texts and the utterances are delimited by the use of punctuation marks. The reading of a punctuation mark (., ?, !, ...) allows us to recognize the end of an utterance.

As a conclusion, therefore, we propose the following annotation scheme for dialogue structure based on Gallardo (1996):

Turn (T) is identified by a change of speaker in the dialogue; each change of speaker supposes a new speaking turn. On this point, Gallardo makes a distinction between two different kinds of turns:

- An **Intervention Turn (IT)** is one that adds information to the dialogue. Such turns constitute what is called *the primary system of conversation*. Speakers use their interventions to provide information that facilitates the progress of the topic of conversation. Interventions may be **initiatives ($\mathbf{IT}_I$)** when they formulate invitations, requirements, offers, reports, etc., or **reactions ($\mathbf{IT}_R$)** when they answer or evaluate the previous speaker´s intervention. Finally, they can also be **mixed interventions ($\mathbf{IT}_{R/I}$)**, meaning a reaction that begins as a response to the previous speaker's intervention, and ends as an introduction of new information.
- A **Continuing Turn (CT)** represents an empty turn, which is quite typical of a listener whose aim is the formal reinforcement and ratification of the cast of conversational roles. Such interventions lack information.

Adjacency Pair or **Exchange (AP)** is a sequence of turns T headed by an initiation intervention turn (IT_I) and ended by a reaction intervention turn (IT_R). One form of anaphora which appears to be very common in dialogues is the reference within an adjacency pair (Fox, 1987).

Topic (TOPIC) is the main entity in the dialogue. According to Rocha (1998) four features are taken into account in the selection of the best candidate for discourse topic: frequency, even distribution, position of first token, and semantic adequacy. The topic must be a lexical item which is frequently referred to.

According to the above-mentioned structure, the following set of tags is considered necessary for dialogue structure annotation: IT_I, IT_R, CT, AP and TOPIC. AP and TOPIC tags will be used to define the anaphoric accessibility space and the remaining will be used to obtain the adjacency pairs. The $\mathrm{IT}_{R/I}$ tag standing for mixed interventions is not considered because mixed interventions can be split into two different interventions: IT_R and IT_I. This task will be done in the annotation phase.

For this experiment, the corpus has been manually annotated. However, nowadays there are some works performing an automatic adjacency pair tagging, such as the BASURDE (1998 2001) Project. On the other hand, there are other works performing automatic topic tagging (e.g. Reynar (1999)) or automatic topic extraction (e.g. the method for anaphora resolution shown in Martínez-Barco et al. (1999)).

An example of an annotated dialogue with such tags is presented in figure 1. It should be pointed out that the tag (OP) indicates the turn of the operator of a railway company, and the tag (US) indicates the user's turn. The transcribed dialogue provides these tags.

The annotation of conversational dialogues is carried out, as shown above, and the evaluation of the proposed anaphoric accessibility space accomplished. An important aspect of this type of annotation is the training phase, which assures the reliability of the annotation.

The annotation phase is accomplished in the following way: a) two annotators are selected, b) an agreement[2] is reached between the two annotators with regard to the annotation scheme using 5 dialogues (training corpus), c) the annotation is then carried out by both annotators in parallel over the remaining 35 dialogues (test corpus) and, d) finally, a reliability test is done on the annotation (see Carletta et al. (1997)). The reliability test uses the *kappa* statistic that measures the affinity between the annotations of the two annotators by making judgements about categories. See Siegel and Castellan (1988) for *kappa* statistic (k) computing.

Because of turns are marked during the transcription phase, all the annotator must do in relation to the adjacency pair is to classify turns according to the above classification, and then to relate each initiative intervention turn IT_I to its reaction intervention turn IT_R. As a result, the adjacency pair is defined. Thus, this task was limited just to a classification task that is easily measured using the *kappa* statistic.

Another task is the topic definition. According to the corpus structure, this task is trivial because the corpus is organized into short dialogues, and each dialogue has only one main topic or theme. This topic is introduced clearly by means of some user's intervention at the beginning of the dialogue. Consequently, we have not detected discrepancies between both annotators with regard to the topic definition, and because of this, this task was not measured using the *kappa* statistic.

According to Carletta, a k measurement such as $0.68 < k < 0.8$ allows us to make encouraging conclusions, and $k > 0.8$ means total reliability between the results of both annotators.

Once both annotators have carried out the annotation, the reliability test of the annotation has been run, with a *kappa* measurement of $k = 0.91$. We therefore consider the annotation obtained for the evaluation to be totally reliable.

In those cases where some discrepancy between the annotators was found, the following criteria was applied: each dialogue has a main annotator whose criteria with regard to the annotation is considered definitive although there were discrepancies between both accounts. In order to guarantee the results, each annotator was the main annotator in only 50% of the dialogues.

As this annotation would be processed by some

[2] This agreement is about what every tag means to every annotator when it is applied to the corpus

TOPIC			tren *(train)*
AP1		IT_I (OP)	información de Renfe, buenos días *(Renfe information, good morning)*
		IT_R (US)	hola, buenos días *(hello, good morning)*
		CT (OP)	hola *(hello)*
AP2		IT_I (US)	me podéis decir algún tren que salga mañana por la tarde para ir a Monzón *(could you tell me about any train that leaves tomorrow evening for Monzon)*
		IT_R(OP)	si, vamos, mira hay un talgo a las tres y media de la tarde *(let me see, there is a talgo at half past three)*
AP3		IT_I (US)	sí tiene que ser más tarde *(it has to be later)*
		IT_R (OP)	más tarde. Hay un intercity a las cinco y media, un expreso a las seis y media *(later. There is an intercity at half past five, an expreso at half past six)*
AP4		IT_I (US)	el de las seis y media ¿llega a Monzón? *(the half past six one, does it go to Monzon?)*
	AP5[a]	IT_I (OP)	a ver. El de las seis y media me ha preguntado ¿verdad? *(let me see. You have asked about the half past six one, haven't you?)*
		IT_R (US)	si *(yes)*
		IT_R (OP)	a las nueve y veinticinco *(at twenty-five past nine)*
AP6		IT_I (US)	a las nueve y veinticinco está en Monzón *(at twenty-five past nine it is in Monzon)*
		IT_R (OP)	si *(yes)*
		CT (US)	vale, pues ya está. Esto ya es suficiente. *(ok, that's all. That's enough.)*
		CT (OP)	hum, hum (simultáneo)
AP7		IT_I (US)	gracias, ¿eh? *(thank you, ok?)*
		IT_R (OP)	muy bien a usted. Hasta luego *(thanks. Bye)*

[a]This adjacency pair is included in AP4

Figure 1: An example of an annotated dialogue from *Corpus InfoTren: Person*

anaphora resolution system, we propose an SGML tagging format such as the one that can be seen in figure 2.

The SGML markup will have the following form:

```
<ELEMENT-NAME ATTR-NAME="VALUE" ...>
text-string
</ELEMENT-NAME>
```

Thus, the following notation is provided in each case:

- Topic:

```
<TOPIC>
topic-entity
</TOPIC>
```

- Adjacency pairs:

```
<AP ID="number">
Adjacency-pair
</AP>
```

ID contains an identification number for arranging the adjacency pairs in sequential order.

- Intervention turns:

```
<IT TYPE="R|I" SPEAKER="speaker">
Intervention-turn
</IT>
```

```
<TOPIC>                           tren
                                  (train)
</TOPIC>
                                  ...
<AP ID="4">
<IT TYPE="I" SPEAKER="US">        el de las seis y media ¿llega a Monzón?
                                  (the half past six one, does it go to Monzon?)
</IT>
<AP ID="5">
<IT TYPE="I" SPEAKER="OP">        a ver. El de las seis y media me ha preguntado ¿verdad?
                                  (let me see. You have asked about the half past six one, haven't you? )
</IT>
<IT TYPE="R" SPEAKER="US">        si
                                  (yes)
</IT>
</AP>
<IT TYPE="R" SPEAKER="OP">        a las nueve y veinticinco
                                  (at twenty-five past nine)
</IT>
</AP>
                                  ...
```

Figure 2: SGML annotation example

TYPE may be "R" or "I" (Reaction or Initiative), and SPEAKER is the mark for the participant that is speaking this turn.

- Continuing turns:

```
<CT SPEAKER="speaker">
Continuing-turn
</CT>
```

3. Accessibility space proposal

Based on the above-mentioned annotation, an anaphoric accessibility space is proposed in order to solve anaphors generated by Spanish personal pronouns, demonstrative pronouns and adjectival anaphors [3].

According to Fox (1987) the first mention of a referent in a sequence is done with a full noun phrase. After that, by using an anaphor the speaker displays an understanding that sequence has not been closed down. Then, we consider that two different sequences generate mostly of the anaphors to be found in dialogues: the adjacency pair and the topic scope. The former generates references to any local noun phrase, and the later generates references to the main topic of the dialogue.

Based on this, we propose the anaphoric accessibility space as the set of noun phrases taken from:

- the same adjacency pair as the anaphor, plus
- the previous adjacency pair to the anaphor, plus
- another adjacency pair including the anaphor adjacency pair, plus
- the noun phrase representing the main topic of the dialogue.

4. Empirical study

In order to carry out the evaluation of the anaphoric accessibility space, the global process shown in figure 3 was performed.

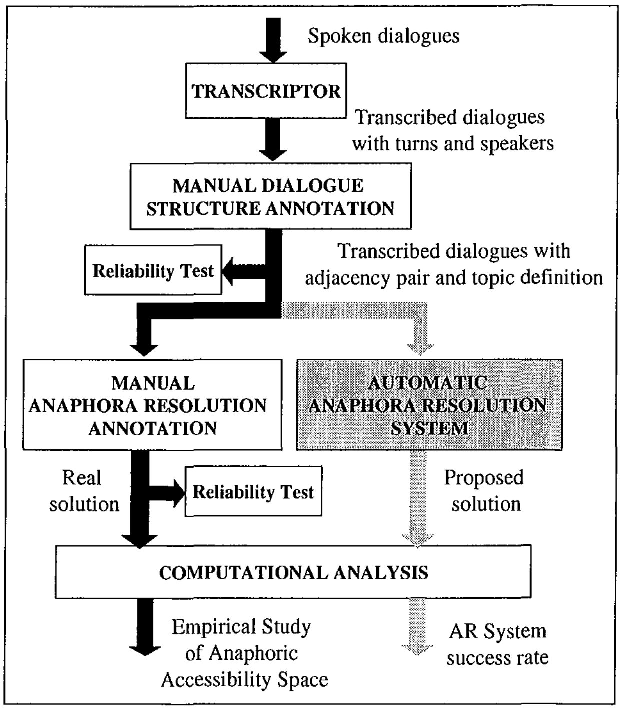

Spoken dialogues
TRANSCRIPTOR
Transcribed dialogues with turns and speakers
MANUAL DIALOGUE STRUCTURE ANNOTATION
Reliability Test
Transcribed dialogues with adjacency pair and topic definition
MANUAL ANAPHORA RESOLUTION ANNOTATION
AUTOMATIC ANAPHORA RESOLUTION SYSTEM
Real solution
Reliability Test
Proposed solution
COMPUTATIONAL ANALYSIS
Empirical Study of Anaphoric Accessibility Space
AR System success rate

Figure 3: Global process

In this experiment, 40 transcribed spoken dialogues were selected from the 200 afforded us by the *Basurde* project. The transcriptor used in the *Basurde* project provides written dialogues with turn and speaker marks.

[3] the Spanish adjectival anaphor is a kind of English one-anaphora where the word *one* is omitted. For instance, *el de las seis y media (the half past six one)*.

	Same AP[a]	Previous AP[b]	Included AP[c]	TOPIC[d]	Others[e]
Pronominal	60.6%	24.6%	8.2%	4.9%	1.7%
Adjectival	44.7%	28.9%	5.2%	13.4%	7.8%
Total Results	Anaphoric accessibility space proposal: 95.9%				4.1%

[a] The antecedent is found in the same Adjacency Pair as the anaphor one

[b] The antecedent is found in the previous Adjacency Pair to the anaphor one

[c] The antecedent is found in another adjacency pair including the anaphor adjacency pair

[d] The antecedent is found in the main Topic of dialogue

[e] The antecedent is found in other sources

Table 1: Empirical results

Afterwards, these selected dialogues were manually annotated according to the proposed annotation scheme. From the 40 dialogues, 5 were randomly selected for the annotators' training and the remaining 35 were reserved in order to carry out the final evaluation. Then, the reliability test of this annotation was performed in order to guarantee the final results.

Following this, a manual annotation of the anaphor solutions was performed over pronominal and adjectival anaphors in the corpus. This annotation relates each anaphor to the correct antecedent. Again, in order to guarantee the results, this annotation was performed by two annotators in parallel, and a reliability test of the annotation was carried out. In this way, the annotation was considered a classification task consisting in defining the adequate solution from the candidate list (we estimated an average of 6.5 possible candidates per anaphora after applying restrictions). Once the reliability test over the manual anaphora resolution annotation was run, a *kappa* measurement of $k = 0.87$ was achieved.

After that, a study of each pronominal and adjectival anaphora was developed to obtain the antecedent location, as shown in table 1. This study was made applying a computational analyzer that obtains information about an automatic anaphora resolution system[4]. As a result, the analyzer compares the output of this AR system with the real solution in the manual annotation and provides several statistics about it. One of these statistics is the study presented in this paper[5].

According to this study, 95.9% of the antecedents were located in the proposed anaphoric accessibility space. Remaining antecedents (4.1%) were estimated to be located in subtopics of the dialogues. In order to incorporate these antecedents to the anaphoric accessibility space, a basic strategy based on the use of the full space (i.e. all the noun phrases from the beginning of the dialogue to the anaphor) could be proposed. As shown in table 2, our proposal of anaphoric accessibility space works with an average of 10.5 antecedents per anaphor (before applying restrictions) instead of 35 antecedents per anaphor that could be obtained if we consider the full space. That means a decreasing of 70%. Evaluating the advantages and the disadvantages, considering the full space implies a) great computational efforts and b) 70% more possibilities to obtain an incorrect response in the anaphora resolution algorithm that uses this anaphoric accessibility space. Notice that our experiments had been performed over a collection of short dialogues (around 332 words per dialogue). This difference will increase in longer dialogues.

	Full space	AAS proposal
Total antecedents	3245	1025
Antecedents per anaphor	35	10.5
Reduction	70%	

Table 2: Anaphoric accessibility space vs full text

5. Conclusions

This paper shows that in a corpus of Spanish dialogues, the antecedent of pronominal and adjectival anaphors can almost always be found in the set of noun phrases taken from the same adjacency pair as the anaphor, the previous adjacency pair, any containing adjacency pair, plus a noun phrase representing the main topic of the dialogue when the anaphor occurs.

Furthermore, an annotation scheme of dialogue structure for Spanish has been presented, allowing us to define the adequate anaphoric accessibility space. Starting with the study performed over a dialogue corpus, it has been shown that this proposed space allows us to locate 95.9% of anaphoric antecedents. We consider that anaphora resolution in Spanish dialogues needs to have a dialogue structure and define the adequate space that improves this resolution.

In this work, we only deal with individual anaphora, i.e. anaphors whose antecedents are noun phrases. There are several studies about deictic anaphora, that is, anaphors having abstract antecedents, showing the importance of this kind of anaphora in dialogues (see Eckert and Strube (1999b)). Thus, a full study of spaces for deictic anaphora and other kinds of anaphora (surface-count anaphora, definite descriptions, one-anaphora, etc.) must be performed.

6. Acknowledgments

The authors wish to thank N. Prieto, F. Pla and A. Molina (Universitat Politecnica de Valencia) for having

[4] This anaphora resolution system uses an algorithm based on the proposed anaphoric accessibility space (see Martínez-Barco and Palomar (2000)).

[5] Notice that the study about anaphoric accessibility space was not developed using the AR system proposal, but the manual annotation of anaphors, (i.e. real solutions).

contributed their tagger and E. Segarra (Universitat Politecnica de Valencia) for affording us the *Corpus InfoTren: Person* from the BASURDE Project. We are also grateful to several anonymous reviewers of *Gotalog'2000* for their helpful comments of earlier drafts of the paper. And finally thanks also to Rafael Muñoz and Maximiliano Saiz Noeda who discussed several issues about the final paper.

This paper has been supported by the Comisión Interministerial de Ciencia y Tecnología (CICYT) with project number TIC97-0671-C02-01/02.

7. References

B. Baldwin. 1997. CogNIAC: high precision coreference with limited knowledge and linguistic resources. In *Proceedings of ACL/EACL workshop on Operational factors in practical, robust anaphora resolution*, Madrid (Spain), July.

Proyecto BASURDE. 1998–2001. *Spontaneus-Speech Dialogue System in Limited Domains.* CICYT (TIC98-423-C06). http://gps-tsc.upc.es/veu/basurde/Home.htm.

J. Carletta, A. Isard, S. Isard, J.C. Kowtko, G. Doherty-Sneddon, and A.H. Anderson. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 23(1):13–32.

N. Dahlbäck. 1991. *Representations of Discourse-Cognitive and Computational Aspects.* Ph.D. thesis, Department of Computer and Information Science, Linköping University, Linköping, Sweden.

M. Eckert and M. Strube. 1999a. Dialogue Acts, Synchronising Units and Anaphora Resolution. In *Proceedings of Amsterdam Workshop on the Semantics and Pragmatics of Dialogue (AMSTELOGUE'99)*, University of Amsterdam, Holland, May.

M. Eckert and M. Strube. 1999b. Resolving Discourse Deitic Anaphora in Dialogues. In *Proceedings of 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL'99)*, Bergen, Norway.

B. Fox. 1987. *Discourse Structure and Anaphora.* Written and conversational English. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.

B. Gallardo. 1996. *Análisis conversacional y pragmática del receptor.* Colección Sinapsis. Ediciones Episteme, S.L., Valencia.

B. Grosz, A. Joshi, and S. Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

G. Hirst. 1981. *Anaphora in Natural Language Understanding.* Springer-Verlag, Berlin.

J. Hobbs. 1986. Resolving pronoun references. In B. Grosz B.L. Webber and K. Jones, editors, *Readings in Natural Language Processing.* Morgan Kaufmann, Palo Alto, CA.

P. Martínez-Barco and M. Palomar. 2000. Dialogue structure influence over anaphora resolution. In O. Cairo, L.E. Sucar, and F.J. Cantu, editors, *MICAI 2000: Advances in Artificial Intelligence*, volume 1793 of *Lecture Notes in Artificial Intelligence*, Acapulco, México, April. Springer-Verlag.

P. Martínez-Barco, R. Muñoz, S. Azzam, M. Palomar, and A. Ferrández. 1999. Evaluation of pronoun resolution algorithm for Spanish dialogues. In *Proceedings of the Venezia per il Trattamento Automatico delle Lingue (VEXTAL'99)*, pages 325–332, Venice (Italy), November.

R. Mitkov. 1998. Robust pronoun resolution with limited knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*, Montreal (Canada), August.

Jeffrey C. Reynar. 1999. Statistical Models for Topic Segmentation. In *Proceedings of 37th Annual Meeting of the Association for Computational Linguistics*, pages 357–364, Maryland, USA, June.

M. Rocha. 1998. *A corpus-based study of anaphora in dialogues in English and Portuguese.* Ph.D. thesis, University of Sussex, Sussex. UK.

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

S. Siegel and J. Castellan. 1988. *Nonparametric Statistics for the Behavioral Sciences.* McGraw-Hill, 2nd edition.

M. Strube and U. Hahn. 1999. Functional Centering - Grounding Referential Coherence in Information Structure. *Computational Linguistics*, 25(5):309–344.

Processes of Collaboration and Communication in Desktop Videoconferencing: Do They Differ From Face-to-Face Interactions?

Alison Newlands*, Anne H Anderson†, Jim Mullin†, Anne-Marie Fleming†.

*University of Strathclyde, Glasgow, G1 1QE
alison.newlands@strath.ac.uk
†University of Glasgow, Hillhead St, Glasgow
Anne@psy.gla.ac.uk, jim@mcg.gla.ac.uk, annemarie@mcg.gla.ac.uk

Abstract

The impact of desktop videoconferencing (DVC) upon interpersonal communication are explored to determine if this communicative context affects the processes required for effective communication. Twenty undergraduate participants acted as Clients during a stimulated service encounter (The Travel Game) in one of two contexts, DVC or face-to-face. The content and structure of the interactions are examined using Conversational Games Analysis. The results of the analysis show that participants in the DVC and face-to-face contexts interacted and collaborated in different ways. In the DVC dialogues participants elicited a greater amount of listener feedback (Align Games), offered more information about the task and their activities (Explain Games), and sought more information by asking a greater number of yes-no questions (Query-yn Games). These results are discussed within the framework of a collaborative model of communication.

1. Introduction

Establishing mutual understanding, or 'common ground', is required for effective communication. This is referred to as the 'process of grounding' (Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989). Grounding is a collaborative, interactive process, which ensures that participants have understood a previous utterance, to a level sufficient for their current purposes. The process of grounding can be affected by several factors. Clark and Schaefer (1989) suggest that different conversation purposes impact on grounding, so task related conversations might require stronger evidence of understanding than social dialogues. It has also been proposed that the process of grounding changes with communicative context (Clark and Brennan, 1991). This is 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. Some methods of grounding appear to require very little effort in communicatively rich contexts, but using the same grounding constraints in another context may take considerably more effort. For example, while it is easy to use non-verbal behaviour to show agreement and understanding in face-to-face communication, this is not so easily achieved during a videoconference, where the visual channel is often impoverished. The effort required to maintain the process of grounding would therefore vary dramatically with communicative context (Clark and Brennan, 1991). In video-mediated communication (VMC), attenuation of visual signals can make it difficult to time the effective use of non-verbal signals to show understanding (Heath and Luff, 1991). Users of VMC systems should use the grounding constraints that require the least collaborative effort. The question being addressed in this paper is whether they do so or not.

Although there have been a number of studies of the impact of VMC on users (see for example, Sellen, 1995; Anderson et al. 1996; O'Conaill et al., 1997), very little research has investigated whether collaboration over the content of communication and establishment of common ground is affected by communication medium. This paper reports the results of detailed discourse analysis of dialogues that took place in two of the communicative contexts explored by Anderson et al. 1996, face-to-face interactions and desktop video-conferencing. The aim is to see if the content and structure of dialogues from these contexts differed in terms of observable patterns of pragmatic function. The research builds upon earlier research, which has examined the effect of a range of communicative contexts upon task performance and the structure of dialogues in collaborative task-oriented interactions (for example, Anderson et al, 1997; Doherty-Sneddon et al 1997).

Conversational Games Analysis (Kowtko, Isard and Doherty-Sneddon, 1992), which is used in this paper, provides a framework for looking at the communicative functions (conversational goals and sub-goals) that speakers attempt to convey in their contributions. Conversation 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. It 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 (Newlands et al., 1996; Doherty-Sneddon et al., 1997). CGA can be used to elucidate the process of grounding in dialogues. The distribution of the Conversational Games and Moves can highlight the ways in which grounding may differ between face-to-face interactions and desktop videoconferencing.

2. Goal of the Paper

The paper attempts to add to the literature on video-mediated communication, by exploring the effects of a DVC system which provides low quality visual images but high quality (full duplex) audio signals. The effects of this VMC context are compared with face-to-face communication, to see what impact the impoverished video channel had upon communication and collaboration. The structure and content of the VMC dialogues were explored using Conversational Games Analysis to see if the quality of the visual signals available in the DVC context affected the processes of communication and collaboration.

3.1 Design and Procedure

Using an experimental paradigm, twenty pairs of participants took part in a collaborative problem-solving task in a simulated travel agency (The Travel Game, Anderson et al., 1996; Newlands et al., 1996). Participants were asked to plan an itinerary around the United States; their goal is to visit as many destinations as possible given the restrictions imposed by available connecting flights. Ten pairs of participants undertook the task in a face-to-face context, using paper maps of the USA which showed the position of available airports. They were assisted by a Travel Agent who had details of flight times and possible connections. Participants in the VMC context were presented with a multi-media version of the Travel Game. The map was displayed in a shared screen facility and users communicated with the remote Travel Agent via video and audio links run over a dedicated local area network. The quality of the video images was low, temporal resolution was 4-5 frames per second, this being a common feature of publicly available VMC systems. The audio link provided high quality full duplex audio signals. Full orthographic transcriptions of the face-to-face and DVC dialogues were made from high quality audio recordings.

3.2. Method of analysis:

CGA was applied to the transcripted dialogues by two trained coders. Table 1 shows the full set of Conversational Games found necessary and sufficient to capture the speaker's communicative intentions in coding the Travel Game Dialogues. Definitions are provided, along with examples of the Conversational Games from the face-to-face (Face) and DVC communicative contexts.

Table 1. Seven Types of Conversational Games used in Coding Travel Game Dialogues.

INSTRUCT: Communicates a direct or indirect request for action or instruction.

Examples:
Face: You'll need to take a note of the flight times
DVC: Hang on a minute, no I'll change my mind and go to Michigan

CHECK: Listener checks their own understanding of a previous message or instruction from their conversational partner, by requesting confirmation that the interpretation is correct.

Examples
Face: I'm sorry, which airport did you want to fly to?
DVC: And what was the city?

QUERY-YN: Yes-No question. A request for affirmation or negation regarding new or unmentioned information about some part of the task.

Examples:
Face: Can I move on to Detroit?
DVC: Can I get a connecting flight to Casper?

QUERY-W: An open-answer Wh-question. Requests more than affirmation or negation regarding new information about some part of the task.

Examples:
Face: Where would you like to go from Syracuse?
DVC: Is there anywhere else you would like to go in Montana?

EXPLAIN: Freely offered information regarding the task, not elicited by coparticipant.

Examples:
Face: I am sorry, there isn't actually a connection between those two airports.
DVC: Right, so that will be day 4 before you can fly out of New York

ALIGN: Speaker confirms the listener's understanding of a message or accomplishment of some task, also checks attention, agreement or readiness.

Examples:
Face: I think it said that in the instructions, didn't it?
DVC: So you leave on day 22, is that okay?

DIRECTIVE: Communicates a decision made by the speaker.

Examples:
Face: Right I'll fly to Memphis to start with.
DVC: Okay, I'm going to fly into Portland, Maine.

The following extract gives an example of a dialogue from the face-to-face context of the Travel Game, and shows the application of Conversational Games Analysis to this task. In this extract 'TA' refers to the Travel Agent, and 'C' to the client. Conversational Games are indicated above the text of the dialogue in upper case, and Conversational Moves are shown underneath the text in italics. The start and end of each Game is shown.

Extract 1. Example of coded face-to-face dialogue.
Game 1: QUERY-W
TA: where would you like to go from Syracuse?
Move: Query-w

Game 2: QUERY-YN embedded
C: I can still go, I have to be still in New York

Move: Query-yn

Game 3: EXPLAIN embedded
TA: you have just to stay in New York until 5.30 that day/
Move: Explain

C: 5.30 >
Move: Acknowledge
End Game 3, End Game 2.

Game 4: QUERY-W embedded
<TA: you could just stay in Syracuse until 5.30 and choose to st.. fly out of state then if you wish/
Move: Query-w

C: yeah>
Move: Acknowledge

TA: or you could go to another city in the meantime?
Move: Query-w cont
C: no l think l will stay
Move: Reply-w

TA: right
Move: Acknowledge
End Game 4.

TA: so where would you like to fly then
Move: Query-w (continuation of Game 1)

C: l would like to fly to ehmm, let me see, to Detroit
Move: Reply-w

TA: to Detroit ... uh huh ...
Move: Acknowledge

Game 5: EXPLAIN em
C: it's in Micshigan, Michigan
Move: Explain
TA: to Detroit in Michigan,
Move: Acknowledge
End 5.

Game 6: EXPLAIN embedded
TA: l am sorry there isn't actually a connection between those two airports.
Move: Explain

C: right, ehmm
Move: Acknowledge
End Game 6

The distribution of Initiating Moves appears to be different from the typical pattern of Initiating Moves found in the Map Task. The task is primarily one of seeking and giving information, demonstrated by the large number of Explain Initiating Moves (giving information) and frequent use of open-ended and yes-no type questions (Query-w and Query-yn).

4.1. Results

Only the results of the CGA will be reported here. The findings on comparisons of task-performance, turn-taking procedures and rates of interruptions are reported in Anderson et al., 1996. CGA can either be carried out at the level of the Conversational Games, or at the more detailed level of the Conversational Moves. In this paper the analysis is based upon Conversational Moves which initiate Conversational Games, as inter-judge agreement for coding Conversational Moves has been found to be greater than agreement over categorization of Conversational Games (Carletta et al., 1997; Newlands, 1998). An inter-coder reliability test was conducted which showed an inter-judge agreement of 91.5%. Agreement on the classification of each Conversational Move was calculated giving a kappa of 0.94 (N=177, k=2), indicating that agreement between coders was not due to chance factors alone ($p<0.001$).

The frequency with which each type of Initiating Move occurred in the DVS and face-to-face contexts was calculated, and the standardised frequency scores (per 100 turns of dialogue) were obtained to allow for differences in length of dialogues in the two contexts. The mean standardised frequency of each Initiating Move in the DVC and face-to-face contexts are show in Table 2 below, standard deviations are shown in brackets.

Initiating Moves	DVC	Face-to-face
Explain	30.62 (11.87)	19.82 (7.00)
Query-yn	16.55 (4.57)	10.92 (4.12)
Query-w	12.18 (5.37)	17.05 (7.11)
Check	5.56 (3.36)	8.20 (3.27)
Align	3.68 (2.54)	1.57 (0.95)
Instruct	2.38 (1.74)	3.63 (2.42)
Directive	0.68 (0.63)	1.23 (0.98)

Table 2: Mean standardised frequency of Initiating Moves in DVC and face-to-face contexts.

The data presented in Table 2 shows that some Initiating Moves were used more frequently in the Travel Game than others. Separate analyses of variance (2 way mixed ANOVA) were computed for each category of Initiating Move. Communicative context (DVC vs. face-to-face) was treated as a between group factor, with the role of the participant (Travel Agent vs. client) as a within dialogue repeated measure. The mean standardised frequency of each type of Initiating Move by the Client and the Travel Agent are presented in Table 3 below.

Context	Face-to-face		VMC	
Role	Travel Agent	Client	Travel Agent	Client
Instruct	0.17 (.033)	3.46 (1.22)	0.38 (0.55)	1.99 (1.34)
Directive	0.04 (0.1)	1.19 (1.99)	0.00 (0.00)	0.66 (0.63)
Explain	17.01 (5.99)	2.80 (4.05)	27.64 (9.22)	4.50 (3.99)
Query-yn	5.29 (2.33)	5.63 (3.44)	3.40 (1.88)	12.55 (4.42)
Query-w	13.71 (8.25)	3.72 (2.08)	7.78 (2.3)	4.40 (2.77)
Align	1.05 (0.75)	0.52 (0.66)	3.29 (2.58))	0.39 (0.54)
Check	2.13 (1.24)	6.06 (3.53)	2.75 (2.35)	2.80 (1.75)

Table 3. Mean Initiating Moves by Travel Agent and Client in the VMC and Face-to-face Contexts (Standardised data).

The analyses revealed non-significant main and interaction effects for the Instruct, Directive, Query-w and Check Initiating Moves (p>0.1), but significant main and interaction effects were observed in the frequency of Explain, Align and Query-yn Initiating Moves.

Explain Initiating Moves

The analysis showed that there was a significant main effect of context [F(1,18) = 6.33, p<0.05]. A greater number of Explains were initiated in the DVC context than in face-to-face interactions (means being 15.32 vs. 9.91 respectively). The main effect of role of participant was also significant [F(1,18) = 119.91, p<0.001]; the Travel Agent initiated a greater number of Explain Moves than the client (means 22.17 vs.3.05). The interaction between communicative context and role of participant was also significant [F(1,18) = 7.94, p<0.05]. Further analysis by Simple Effects showed that the Travel Agent initiated more Explain Moves in the DVC context than in the face-to-face context [F(1,18) = 8.82, p<0.01].

Query-yn Initiating Moves

The analysis showed that there was a significant main effect of context [F(1,18) = 8.37, p<0.01]. A greater number of Query-yn were initiated in the DVC context than in face-to-face interactions (means being 8.28 vs. 5.46 respectively). The main effect of role of participant was also significant [F(1,18) = 18.48, p<0.001]; the client initiated a greater number of Query-yn Moves than the Travel Agent (means 9.09 vs. 4.64). The interaction between communicative context and role of participant was also significant [F(1,18) = 15.79, p<0.001]. Simple Effects analysis showed that the client initiated more Query-yn Moves in the DVC context than in the face-to-face context [F(1,18) = 15.69, p<0.001].

Align Initiating Moves.

The analysis showed that there was a significant main effect of context [F(1,18) = 6.03, p<0.05]. A greater number of Aligns were initiated in the DVC context than in face-to-face interactions (means being 3.68 vs. 1.57 respectively). The main effect of role of participant was also significant [F(1,18) = 13.68, p<0.01]; the Travel Agent initiated a greater number of Align Moves than the client (means 2.17 vs. 0.45 Aligns per 100 turns of dialogue). The interaction between communicative context and role of participant was also significant [F(1,18) = 6.57, p<0.05]. Simple Effects analysis showed that the Travel Agent initiated more Align Moves in the DVC context than in the face-to-face context [F(1,18) = 6.94, p<0.05].

These analyses highlight the effect of communicative context and role of participants in the Travel Game. In the DVC dialogues the Travel Agent initiates a greater proportion of Explain and Align Moves, and the Client increases the use of Query-yn Moves in DVC compared to face-to-face context.

5.1. Discussion

The Conversational Games Analysis showed that participants in the two contexts interacted and collaborated in different ways. In the DVC Travel Games the Travel Agent used proportionally more Initiating Moves to elicit feedback from the listener (Align Moves), or to offer information about the task and her activities (Explain Moves) to the Client. At the same time, the client sought more information by asking a greater number of yes-no questions (Query-yn).

In order to determine why these differences occurred, examples of Explain, Align and Query-yn Games were extracted from the dialogues to determine their functions and the types of information they were eliciting or offering. The increased use of Explain Games occurs because the Travel Agent offers more information to the Client concerning the Agents activities and what she was currently attending to. This also occurred in the face-to-face dialogues, but very rarely. For example, the Travel Agent would tell the Client that she was looking up the flight details, or inform the client that they had now arrived at their destination. The following extract demonstrates this usage of Explain Initiating Moves in DVC dialogues.

Extract 2. DVC dialogue
C: Right, can I get a connection to Jacksonville?
Move: Query-yn

TA: I'll just check that for you
Move: Explain

Yes you can
Move: Reply-y

Do you want to go there?
Move: Query-yn

C: Yes.
Move: Reply-y

TA: (pause) Right, you're in Jacksonville
Move: Explain

C: Okay.
Move: Acknowledge

These types of explanations accounted for nearly 42% of Explain Moves initiated by the Travel Agent in the DVC context. This would indicate that, in the DVC dialogues, the Travel Agent spent a considerably amount of time and effort in keeping the Client informed of her activities, or their position in the task. This may be a result of the restricted information provided by the visual channel, and by the low quality of the video signal. In effect, the Travel Agent verbally offered the Client information that would have been available visually in face-to-face interactions.

Examining the functions of Query-yn Moves initiated by the Client in the face-to-face and DVC dialogues, revealed that these questions were used to gain a wide range of information. For example, yes-no questions were asked to gain information about the rules of the Game, or the possibilities of changing the itinerary. However, the majority of the Initiating Moves occurred when the Client asked the Travel Agent if there were connecting flights between two Airports. These questions accounted for more than 78% of all of the Query-yn Moves initiated by the Client in DVC dialogues, but only 40% of Query-yn Moves initiated by the Client in face-to-face interactions. The following extracts show incidences of yes-no questions (Moves are emphasised in bold print) in which the Client asks about the possibility of connecting flights between airports. The extracts are taken from both communicative contexts.

Extract 3. Face-to-face dialogue
GAME 1. QUERY-W
TA: where would you like to go from Salt Lake City
Move: Query-w

C: I will stay there for three days, and then l will fly out of State
Move: reply-wh

TA: Okay
Move: Acknowledge

GAME 2. QUERY-YN embedded
C: can l, is there, are there flights to Seattle from there?
Move: Query-yn

TA: I'll check that.
Move: Explain

Extract 4. DVCC dialogue
TA: Right, you've now arrived in Grand Rapids, in Michigan
Move : Explain (Ends previous Game)

GAME 1. QUERY-YN
C: Can l move on to Detroit?
Move: Query-yn

TA: I'll just check.
Move : Explain

No, you can't fly to Detroit from Grand Rapids.
Move: Reply-no

C: Right, okay.
Move: Acknowledge. Ends Game 1

GAME 2. QUERY-YN
Uhmm, can l fly to Great Falls?
Move: Query-yn

TA: Ehmm lets see. No, I'm afraid you can't fly to Great Falls either.
Move: Reply-no. Ends Game 2

The examples show typical use of Query-yn Initiating Moves in the two communication contexts. In face-to-face interactions, the Client is often prompted by an open-ended question (Query-w Initiating Move) from the Travel Agent to say where he would like to go next, and many of the Client's Query-yn Moves concerning flight connections then occur within the context of an already initiated Query-w Move. In the DVC dialogues the Travel Agent tends to round off each set of Games by offering explicit information about the Client's progress in the task, this puts the Client in the position of starting off the next stage of the Travel Game. The Client can achieve this most simply (and most explicitly) by asking the Agent if there are connections to a particular airport. Thus, this behaviour does seem to be an indirect affect of the DVC context, but probably depends in part on the Travel Agent's response to the DVC condition.

Searching through the coded dialogues revealed that Align Moves are used in a variety of ways. The following extracts demonstrate some of the ways in which Aligns were used by the Travel Agent to elicit feedback in the face-to-face and DVC contexts. In these extracts the following symbols are used: TA and C indicate the Travel Agent and Client respectively; a short pause is represented by three dots (...).

Extract 5. Examples of Align in Face-to-face
TA: Can you make a note of your decisions as we go along
**M Instruct*

TB: Uhmmm
**M Reply-y*

TA: I think it said that in the instructions
****M Align***

TB: Sure, yeah it did.
**M Reply-y.*

Extract 6. Example of Align in DVC context
TA: Its actually going to be day 28 before you can actually leave Arizona
**M Explain*

TA: Okay?
M Align

As these extracts show, Align Initiating Moves can be quite lengthy (as in extract 5), but sometimes they can be initiated with just a single word (extract 6). Examination of the dialogues showed that these shorter, one word, Aligns occurred more frequently in the DVC dialogues than the face-to-face interactions: 60% of the DVC Travel Agent's Aligns consisted of single words, such as 'okay', 'right', compared to 46% of Aligns in the face-to-face interactions. It is possible that these one word Aligns were being used more frequently in the DVC context to assist the process of grounding; short Align Moves were used instead of gaze to ascertain that the Client had understood the previous contribution. In the face-to-face context participants could see each other clearly, non-verbal forms of establishing mutual understanding were easily accessible, so the need to use verbal alignments was reduced in this context.

Is there any support for these suggested explanations in use of Conversational Games from previous literature? The most relevant paper is by Doherty-Sneddon et al (1997), who examined the structure and content of dialogues from face-to-face and remote spoken contexts as well as several VMC contexts. The findings from Conversational Games Analysis of these contexts showed that people communicate in a more cautious manner when they use an audio-only context; they adopt what Shadbolt (1984, in Doherty-Sneddon 1997) calls a 'low risk' style of communication. This was apparent in the greater use of Align and Check Games in remote spoken interactions, and an increased use of Align Games in a remote computer-mediated (audio-conferencing) context. Doherty-Sneddon et al. concluded that participants interacting in a spoken only context use a greater number of verbal alignments, and a more cautious style of communication.

The results from this study also show an increased use of verbal alignments. In this case, the effect may be due to the poor quality of the visual signals, rather then the total absence of video images. The quality of the visual signals may have been low enough to engender a more cautious style of communication than occurred in the face-to-face context. So the findings reported by Doherty-Sneddon et al. do support the view taken here, that the increased use of verbal alignments could have been due to the quality of the visual signals provided by the DVC system in this study. The other differences in structure and content of the DVC dialogues (increased use of *Explain* and *Query-yn* Initiating Moves) receive no support from previous literature. However, as suggested in the previous discussion these changes may demonstrate some of the different ways in which users adapted to working in a DVC context which affords low quality video images. Travel Agents requested a greater amount of listener feedback (*Align* Moves), and they spent proportionally more time informing the Client of their activities. The Client appears to have responded to this style of interaction by making greater use of simple yes-no questions.

Overall, the findings support the view that in a DVC context where the visual channel provides restricted visual signals, participants may attempt to achieve a greater amount of collaboration through the verbal channel since the quality of the visual channel limits the use of non-verbal communication. This appears to be the common theme linking the differences in proportional use of the Align, Explain and Query-yn Moves in the DVC context. Users of these systems make greater use of the verbal channel to establish common ground, and to ensure that the process of communication proceeds smoothly. These variations in use of pragmatic functions also change the style of communication by users of the DVC system. Users adopt a more explicit style of interaction, offering more information about the task and their activities. This style of communication allows users of DVC to maintain a greater amount of verbal contact with each other than occurred in the face-to-face context, and could be due to the low quality of the video images and could indicate collaborators when the quality of the video images is low.

These findings demonstrate the subtle manner in which people adapt to a video-mediated environment. These adaptations may be difficult to observe when only the surface structure of the dialogues is examined; as the length of turns and number of turns may not be greatly affected. The finer grain pragmatic analysis undertaken in this study therefore has advantages over other ways of determining the impact of communicative contexts upon collaboration and communication.

6. References

Anderson, A.H., Newlands, A., Mullin, J., Fleming, A.M., Doherty-Sneddon, G., and Van der Velden, J. (1996). Impact of video-mediated communication on simulated service encounters. *Interacting With Computers*, 8 (2):193-206.

Anderson, A.H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A.M., 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 K.E. Finn, A.J. Sellen and S.B. Wilbur (Eds.) *Video-Mediated Communication* (pp. 133-172). NJ: Lawrence Erlbaum Associates.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Newlands, A., Doherty-Sneddon, G., and Anderson, A.H. (1997). The reliability of a dialogue structure coding scheme, *Computational Linguistics*, 23 (1):13-32.

Clark, H.H., and Brennan, B.E. (1991). Grounding in communication. In L.B. Resnick, J. Levine and S.D. Teasley (Eds.) *Perspectives on socially shared cognition.* American Psychological Association, Washington.

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

Clark, H.H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition,* 22: 1-39.

Doherty-Sneddon, G., Anderson, A.H., 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.

Heath, C., and Luff, P. (1991). Disembodied conduct: communication through video in a multi-media office environment. *Proceedings of the ACM Conference on CHI 1991*, New Orleans: Louisiana.

Houghton, G. (1986). *The Production of Language in Dialogue: A Computational Model*, Ph.D. 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 P. Morris (Ed.) *Modelling Cognition* (pp. 249-267). John Wiley.

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.

Newlands, A 1998. *The Effects of Computer Mediated Communication on the processes of communication and Collaboration..* Ph.D. thesis, Glasgow University.

Newlands, A., Anderson, A.H., and Mullin, J. (1996). Dialog Structure and Cooperatiave Task Performance in Wwo CSCW Environments. In J.H. Connolly and L Pemberton (Eds.) *Linguistic Concepts and Methods in CSCW* (pp. 41-60). Springer-Verlag. London.

O'Conaill, B., and Whittaker, S. (1997). Characterizing, predicting, and measuring video-mediated communication: a conversational approach. In K.E. Finn, A.J. Sellen and S.B. Wilbur (Eds.) *Video-Mediated Communication* (pp. 107-131). NJ: Lawrence Erlbaum Associates.

Power, R. (1979). The Organisation of Purposeful Dialogues, *Linguistics*, 17: 107-152.

Sellen, A.J. (1995). Remote conversations: The effects of mediating talk with technology. *Human Computer Interaction*, 10: 401-444.

Design Constraints and Representation for Dialogue Management in the Automatic Telephone Operator Scenario

José F. Quesada, J. Gabriel Amores, Gabriela Fernández, José A. Bernal, M. Teresa López

University of Seville (Spain)
jquesada@cica.es

Abstract

This paper describes the dialogue layer of a Spoken Dialogue System prototype. We have developed a methodology for the design of dialogue management systems following Knowledge Engineering principles. A dialogue is treated as an activity that requires knowledge, experience, ability and training. We will also propose a specification language for the representation of the linguistic knowledge involved, as well as a task-oriented inference engine for the control and reasoning about the dialogue. This generic model has been implemented in Delfos, an application developed for the automatic telephone task scenario.

1. Introduction

Our previous research in collaboration with the Speech Technology Division of Telefónica I+D, has focussed on the study of the automatic telephone operator scenario (section 2) within the ATOS project (Alvarez *et al* 97; Fernández & Quesada 99; López & Quesada 99). As a result of this collaboration, we have elaborated a generic methodology for the design and implementation of dialogue management systems. This paper describes the most relevant characteristics of our methodology, and their implementation in the prototype system Delfos.

The methodology we propose is inspired on Knowledge Engineering principles and distinguishes two major levels in the design of systems: a level for knowledge specification and a level for inference and control (section 3). Besides, our methodology divides the specification level into two further levels: the representation of speech acts (section 4) and the representation of dialogue structures (section 3).

2. Design Constraints: A Dialogue System for the Automatic Telephone Task Scenario

This section describes the design constraints of the system. We illustrate a sample conversation (Figure 1) in order to present the functionality we are aiming at (although our corpus is in Spanish, we have translated the conversation into English for expository reasons).

This short conversation will help us to describe the design restrictions we have taken into account:

- **Interaction with the speech recognition system.** Our system is embedded in a Spoken Dialogue System application which takes as input the output of a speech recognition system through the telephone line. Speech recognition errors such as those reported in our sample conversation have been dealt with in previous research work in our group. The natural language processing system Iris (López & Quesada 98) incorporated a number of techniques for the detection and correction of recognition errors during the natural language understanding phase. Nevertheless, the dialogue management system should also be capable of handling recognition errors when configuring the dialogue interaction by using both direct confirmation questions (as in S6) and indirect ones (S7).

- **Task Detection.** Our scenario difers from Task-oriented systems in that the system does not know beforehand the task that the user has in mind. Rather, the user may choose between any of the different functions which have been designed to interact with the PABX. Therefore, the first problem that the system must solve is to figure out which task(s) the user may want to perform.

- **Incomplete functions.** In our scenario, it is common to find situations in which the requested functions are not complete, that is, commands for which the user has not specified all the arguments required to fulfil the task. For instance, in U3 the user is requesting the system to transfer his/her calls, but s/he has not specified the destination number. In this case, the system must be capable of keeping track of the requested information and generating those questions necessary to complete the missing information.

- **Expectations.** The dialogue manager system can benefit from the knowledge of the previous history of the dialogue. In fact, this dialogue history generates answering expectations: in S4 the system is waiting for a confirmation answer either affirmative ("Yes", "OK", ...) or negative ("No", "That's wrong!", ...), in S5 the system is expecting a telephone number from the user. This knowledge (expectations generated from the dialogue history) can be further used both by the speech recognition system and the natural language processing module in order to improve the efficiency and restrict the semantic search in the grammar.

- **Multiple paths in a dialogue.** As we can observe in U3, it is common to find cases where one interaction initiates several functions. The system must be capable of handling all of them, while recalling unfulfilled work at every single point. In our example S4 is asking for direct confirmation of one of the functions. Later on, in S5 the system carries on with the fulfilment of

S1: [Welcome message and management instructions]
U1: I would like to know if I have any message in my mailbox.
S2: You have a new message.
U2: Please, I would like to hear it.
S3: The message is: ...
U3: Now I want to delete it and to transfer my calls.
S4: Shall I delete the last saved message?
U4: Yes, please.
S5: The last message has been deleted. To transfer your calls,
I need a telephone number or an extension.
U5: The number is 123456789
S6: So, do you want me to transfer your calls to the number 132456789?
U6: No, to the number 123456789.
S7: OK, your calls have been transferred to the number 123456789.
Would you like to perform any other function?
U7: No
S8: [Goodbye message]

Figure 1: A sample conversation

the function requested in U3 while indirectly confirming the execution of the first function.

- **Dialogue history and anaphoric references**. Representing the previous dialogue history is also useful in order to deal with discourse phenomena such as anaphoric references. In our sample conversation both answers U2 ("I want to hear it") and U3 ("delete it") refer to "the last message".

Taking into account the design restrictions imposed by the application described, we have developed a general methodology which is described in the next section.

3. A Knowledge Engineering-based Methodology for the Representation and Reasoning in Spoken Dialogue Systems

The level of representation (Hodgson 91) plays a crucial role in the study of the discourse level of language and in the implementation of dialogue management systems. This is evident, for example, in Discourse Representation Theory (Kamp 81; Kamp & Reyle 93; Eijck & Kamp 97). One of the main areas of research has been the development of dynamic representation models of speech acts which are capable of allowing an incremental interpretation of the utterances within a context (Kamp 81; Barwise & Perry 83; Cooper *et al* 99).

On the other hand, those research works which have studied the implementation of dialogue management systems have made evident the need to specify and represent concrete models or dialogue plans in relation to the application domain tasks. Most implementations in the literature (OVIS (Noord *et al* 98), VERBMOBIL (Alexandersson *et al* 98), Philips (Aust *et al* 95), TRAINS (Trains Web Page), TRINDI (Trindi Web Page)) make use of a frame-based representation model and a dialogue management plan-oriented strategy (Schank & Abelson 77).

Thus, the integration of a dialogue management or planning strategy introduces a higher level of control on top of the speech acts, that gives rise to the idea of a dialogue structure representation model.

Bearing in mind the research work described above, as well as the design constraints explained in section 2, we propose a methodology which is based on the following principles (Figure 2 illustrates the architecture of the system Delfos according to these ideas):

- **Unification-based dialogue management**: basically, every speech act obtained from the module (Iris) is represented as a complex feature structure (Rozenberg & Salomaa 97). This facilitates the integration between the dialogue manager module and unification-based grammars (Shieber 86; Kirchner 90).

- **Communication via the CTAC protocol**: This protocol, formally based on the extended Lexical Object Theory (Quesada 98), guarantees an efficient, bidirectional, flexible and transparent communication between the NLP and dialogue management modules (López & Quesada 98).

- **Declarative Specification of the Dialogue Structures**: Delfos incorporates its own language for the specification of dialogue structures. This language profits from plan-based management techniques, thus avoiding those problems associated with dialogue grammars (Aust & Oerder 95). The specification language of Delfos allows, among other things, for the representation of the history of the dialogue, the control of expectations and the treatment of ambiguity.

- **Dialogue Management as an Inference Engine on a Discourse Declarative Model**: The reasoning level can be regarded of as an inference engine (Hodgson 91) specialised on the treatment of dialogue manager systems.

4. Utterance Representation

From a functional perspective, the system receives as input the results obtained from the speech recognition sytem,

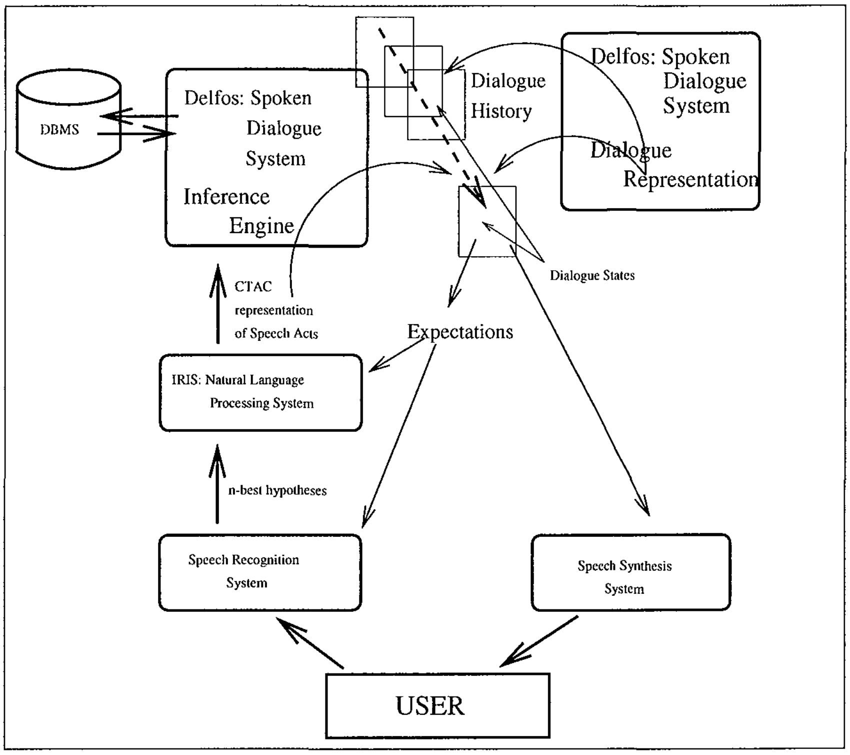

DBMS
Delfos: Spoken Dialogue System
Inference Engine
Dialogue History
Delfos: Spoken Dialogue System
Dialogue Representation
Dialogue States
Expectations
CTAC representation of Speech Acts
IRIS: Natural Language Processing System
n-best hypotheses
Speech Recognition System
Speech Synthesis System
USER

Figure 2: The Delfos Architecture

represented as a set of n-best hypotheses. The NLP system Iris (L6pez & Quesada 98) then carries out the lexical, morphological, grammatical and semantic analysis. It also incorporates a wide range of speech repair strategies (L6pez & Quesada 99).

The NLP system transforms each utterance into a list (one or more) of feature structures according to the CTAC protocol (L6pez & Quesada 98). The CTAC protocol stands for Class, Type, Arg and Contents.

1. **CLASS**: Three classes have been defined for this application: Object, Function and Operation. Object includes any terminal element in the dialogue structure such as Person, Telephone Number, etc. Function describes the set of tasks in the application domain. Finally, Operation will be used for the representation of auxiliary functions such as Confirmation, Cancellation, Help, etc.

2. **TYPE**: This feature specifies the value that each class adopts in a given structure, as described in the **CLASS** section.

3. **ARG**: Some classes may require the presence of one or more arguments. The ARG feature specifies the argument structure of the class. This takes the form of a list in which conjunction, disjunction, and optional operators may appear.

4. **CONTENTS**: This feature represents the particular values associated with each element of the ARG attribute.

As an example, the second part of the interaction U3 above *"transfer my calls"* will be represented as:

$$\begin{bmatrix} CLASS & : & Function \\ TYPE & : & TransferCalls \\ ARG & : & [Name] | [Number] \\ CONTENT & : & \end{bmatrix}$$

For the corpus we are currently using, which contains 60 pieces of dialogues (amounting to more than 1,000 sentences) extracted by a Wizard of Oz, we correctly identified the task and assigned the correct CTAC representation in a 96% of the cases.

5. Dialogue Structures Representation

As mentioned above, the module in charge of the dialogue management is called Delfos. This subsystem incorporates its own language for the specification of dialogue states. These states are triggered by the speech acts resulting from CTAC. In turn, dialogue states may trigger expectations, modify the history of the dialogue, and/or execute actions.

Continuing with the example above, the specification of the dialogue state corresponding to the function *Transfer calls* in Delfos is represented in Figure 3.

Though a complete description of this specification language is beyond the scope of this paper, we will briefly describe some of its components. First, the dialogue state is triggered by a process of unification between the CTAC and the `TriggeringConditions` in the state specification. The `DeclareExpectations` field assigns priority values to other states, and specifies how the result of those states will be incorporated into the history of the dialogue. Thus, the CTAC generated for the interaction U5

```
( StateID:          TRANSFERCALL;
  PriorityLevel:    15;
  TriggeringCondition:
          (CLAS:Function,TYPE:TransferCall);
  DeclareExpectations: {
          Name <= NAME;
          Number <= NUMBER;
  }
  SetExpectations: {
          Confirm <= (CLAS:YesNoEnd,TYPE:Confirm,CONT:Yes);
  }
  ExitState:
          (CLAS:YesNoEnd,TYPE:Confirm,CONT:Yes);
  ActionsExpectations: {
          [Name] => {
             UserPrompt("Por favor, indique el destino del desvo."); }
          [Confirm] => {
             @if (@is-TRANSFERCALL.Name) @then {
                     UserPrompt(@concat("Realmente quiere desviar a ",
                          @is-TRANSFERCALL.Name.CONT,
                          " ?")); }
             @else {
                     UserPrompt(\@concat("Realmente quiere desviar al ",
                         @is-TRANSFERCALL.Number.CONT,
                          " ?")); }
          }
  }
  PostActions: {
          UserPrompt("EJECUCION DEL DESVIO");
  }
)
```

Figure 3: The TRANSFERCALL Dialogue State Specification

(*"The number is 123456789"*, which was understood as *"The number is 132456789"*):

$$\begin{bmatrix} CLASS & : & Object \\ TYPE & : & Number \\ ARG & : & [] \\ CONTENT & : & 132456789 \end{bmatrix}$$

will trigger the NUMBER (a king of destination for a telephone call) dialogue state (Figure 4).

The information obtained by this state will be integrated in the previous dialogue state using the expectation specification `Number <= NUMBER`, thus yielding the representation shown in Figure 5.

With this representation, the system can trigger the direct confirmation question S6.

6. Conclusion

This paper has described a methodology for the design and implementation of Spoken Dialogue Systems which integrates different techniques from Language and Knowledge Engineering. As a result, we have presented an architecture that divides the Dialogue management level into two main components. Firstly, the representation module which distinguishes the representation of speech acts according to the CTAC protocol, and, second, the representation of dialogue states. The system proposed is capable of dealing with all the design constraints previously specified for the task, such as bi-directional interaction with a speech recognition system, control of incomplete functions, manipulation of expectations and multiple paths in a dialogue, and representation of the dialogue history.

7. References

Alexandersson, J., B. Buschbeck-Wolf, T. Fujinami, M. Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz & M. Siegel. 1998. *Dialogue Acts in VERBMOBIL-2*. DFKI Saarbrcken and TU Berlin, Report 226, July 1998.

Alvarez, J., Tapias, D., Crespo, C., Cortázar, I. and Martínez, F. 1997. Development and Evaluation of the ATOS Spontaneous Speech Conversational System. *International Conference on Acoustics, Speech and Signal Processing*, 1139-1142

Aust, H. & M. Oerder. 1995. Dialogue Control in Automatic Inquiry Systems. In Andernach, J. A., S. P. van der Burgt & G. F. van der Hoeven. eds. *Proceedings of the 9th Twente Workshop on Language Technology*. University of Twente, Netherlands.

Aust, H., M. Oerder, F. Seide & V. Steinbiss. 1995. The Philips automatic train timetable information system. *Speech Communication*, 17, 249-262.

```
( StateID:          NUMBER;
  PriorityLevel:   20;
  TriggeringCondition:
                    (CLAS:Object,TYPE:Number);
)
```

Figure 4: The NUMBER Dialogue State Specification

$$\begin{bmatrix} CLASS & : & Function \\ TYPE & : & TransferCalls \\ ARG & : & [Number] \\ CONTENT & : & \\ Number & : & \begin{bmatrix} CLASS & : & Object \\ TYPE & : & Number \\ ARG & : & [] \\ CONTENT & : & 132456789 \end{bmatrix} \end{bmatrix}$$

Figure 5: An unification-based approach to the incremental representation of information states

Barwise, J. & Perry, J. 1983. *Situations and attitudes*. Cambridge, Mass: The MIT Press.

Cooper, R., S. Larsson, C. Matheson, M. Poesio & D. Traum. 1999. *Coding Instructional Dialogue for Information States.* TRINDI (LE4-8314) Deliverable D1.1, February 1999.

Eijck, J. van & Kamp, H.. 1997. Representing Discourse in Context. In van Benthem, J. & Meulen, A. ter. eds. *Handbook of Logic and Language*. Elsevier Science. pp. 179-237.

Fernández, G. & Quesada, J. F. 1999. Delfos: Un Modelo Basado en Unificación para la Representación y el Razonamiento en Sistemas de Gestión de Diálogo. *Procesamiento del Lenguaje Natural*, 67–74.

Hodgson, J.P.E. 1991. *Knowledge Representation and Language in AI.* Chichester, England: Ellis Horwood.

Kamp, H. 1981. A theory of truth and discourse representation. In Groenendijk, J., Jansen, T. & Stockhof, M. eds. *Formal methods in the study of language*. Amsterdam: Mathematical Centre tracts 135.

Kamp, H. & Reyle, U. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Kirchner, C. ed. 1990. *Unification*. San Diego, California: Academic Press Inc.

López, M. T. & Quesada, J. F.. 1998. Spoken Language Parsing Strategies in a Conversational System. In *Proceedings of ECAI'98: XIII European Conference on Artificial Intelligence*, 203-204.

López, M. T. & Quesada, J. F.. 1999. Error Detection and Error Recovery from Speech Recognition: Language Engineering Strategies. *XIV International Congress of Phonetic Sciences*. San Francisco, CA.

Noord, G. van, G. Bourna, R. Koeling & M. J. Nederhof. 1998. Robust Grammatical Analysis for Spoken Dialogue Systems. *Natural Language Engineering*, 1(1), 1-48.

Quesada, J. F. 1998. The Lexical Object Theory: Specification Level. *Grammars*, 1, 57-84.

Rozenberg, G. & A. Salomaa. eds. 1997. *The Handbook of Formal Languages*. Berlin: Springer Verlag.

Schank, R. and R. Abelson. 1977. *Scripts, Plans, Goals and Understanding*. Hillsdale, New Jersey: Lawrence Erlbaum.

Shieber, S.M. 1986. *An Introduction to Unification-based Approaches to Grammar.* CSLI Lecture Notes 4. Stanford, California: Center for the Study of Language and Information.

(Trains Web Page) University of Rochester, Department of Computing Science. *The TRAINS Project: Natural Spoken Dialogue and Interactive Planning.* http://ftp.cs.rochester.edu/u/trains

(Trindi Web Page) TRINDI: Task Oriented Instructional Dialogue. http://www.ling.gu.se/research/projects/trindi

Multi-speaker Utterances and Coordination in Task-oriented Dialogue

Hannes Rieser and Kristina Skuplik

SFB 360 "Situierte Künstliche Kommunikatoren", Projects B3 and C4, Bielefeld University
Postfach 100131, D-33501 Bielefeld, Germany
{Rieser, Skuplik} @lili.uni-bielefeld.de

Abstract

We investigate utterances in task-oriented dialogue which are produced by several agents. Starting with an example where a VP in a directive is initiated by one agent and completed by the addressee, we set out to explain under which conditions coordinations of this type can be successful. The explanatory devices developed are 'action schema' and 'sufficiently informative proposition'. It is argued that propositions embedded in directives must be sufficiently informative for an agent to permit a task-relevant action of his. We suggest that agents' coordination on sufficiently informative propositions be taken as a measure for coordination in dialogue. Using this idea we show how coordination in a full task-oriented dialogue develops. Finally, we demonstrate that agents' "pointwise" coordination can be explained using Asher-Morreau defeasible inference, a version of Asher-Lascarides defeasible practical syllogism as well as additional principles like a Cooperativity Principle and a principle saying 'Make assumed conformity of interests and intentions publicly known!'. If our arguments can be accepted this would first of all entail that theories of dialogue were to be based on coordination and secondly, it would also have repercussions on the notions of proposition, mutuality and common ground.

1. Introduction and Description of Dialogue Example

Although the relevance of dealing with multi-speaker utterances is acknowledged (see e.g. Clark, 1996; Traum, 1999), there is little systematic research on this topic so far. We try to close the gap here, at least to some extent.

Our paper on multi-speaker utterances and coordination is based on a corpus of German task-oriented dialogues (s.a., 1997), in which a Constructor (henceforth Cnst) builds a toy-airplane according to the directives of an Instructor (Inst in the sequel), see Fig. 1. The model on Cnst's side, a toy-airplane, see Fig. 2, has to match in the end its "twin model" on Inst's side. Inst and Cnst are separated by a screen[1].

Intuitively, this can only be achieved, if both agents coordinate. For the purpose at hand we can conceive of agents' coordination in the following way: In order to attain some commonly defined goal, they alternately carry out a sequence of actions. Such a sequence of actions is coordinated iff any of its constituents either reaches the goal or serves as a necessary condition for the next action. The sequence is initiated by the "dominating agent".

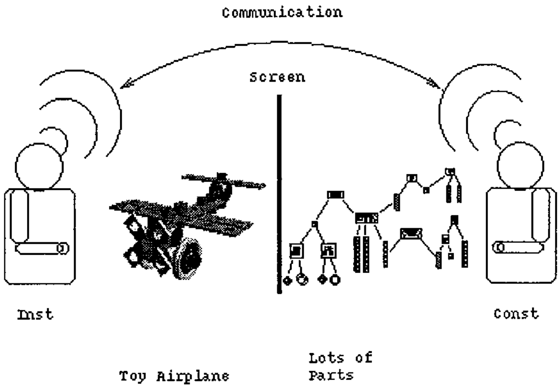


Fig. 1: Setting with Instructor and Constructor separated by a screen

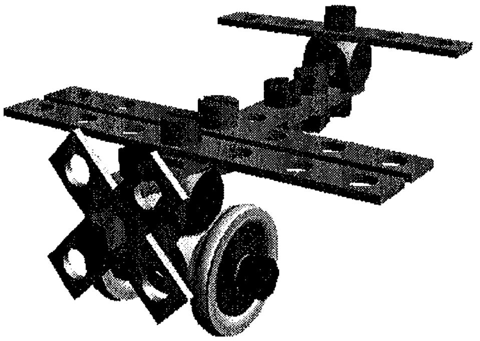


Fig. 2: Toy-airplane ("Baufix"-airplane)

[1] As one of our reviewers rightly observed, the experiments carried out to elicit task-oriented dialogues are in some ways similar to those described in Cohen (1984). However, as easily to be seen, the focus of our research is different from Cohen's. As far as we can tell from the literature, an experimental setting closely matching ours is the tangram scenario used in Clark and Wilkes-Gibbs (1986).

The example of coordinated utterance production below is taken as our point of departure:

(1) Dialogue example for coordinated utterance production

Inst: So, jetzt nimmst du
Well, now you take
Cnst: eine Schraube
a screw.
Inst: eine <-> orangene mit einem Schlitz.
an <-> orange one with a slit
Cnst: Ja.
Yes.
Inst: Und steckst sie dadurch, also
And you put it through there, well
Cnst: Von oben.
From the top.
Inst: Von oben, daß also die drei festgeschraubt werden dann.
From the top, so that the three bars get fixed then.
Cnst: Ja.
Yes.

What happens here can be characterized as follows: The agents' contributions make up two illocutionary acts (Searle & Vanderveken, 1985), roughly one directive introducing a new object (*a screw*) and a second directive demanding a put-through action (*And you put it through there, well*). The first directive serves as a precondition for the second one.

Inst starts demanding the selection of an orange slit bolt from Cnst and some kind of reference act emerges. More precisely, object introduction is at issue, which is not treated as a special type of act in speech act theory. The mechanisms of object introduction are familiar from a somewhat different field, namely the paradigms of Dynamic Semantics and DRT (see Groenendijk & Stokhof, 1991; Kamp & Reyle, 1993). Relying on her anticipation, Cnst continues Inst's locution with *a screw.* However, at the current stage of construction several bolts of different shape and colour are still available (see Fig. 3).

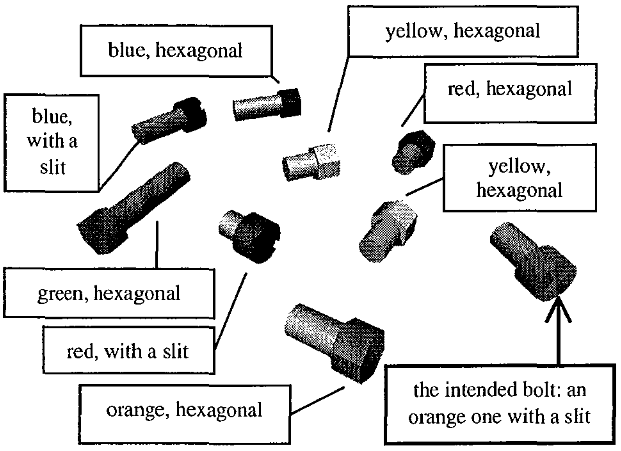


Fig. 3: Available bolts

Hence, the directive *now you take a screw* would not be specific enough to satisfy Inst's intention and so he feels obliged to add *an orange one with a slit.* From Cnst's reply *Yes* Inst may gather that she identified the bolt.

Now the preconditions for the second directive are satisfied: The bolt and three overlapping bars (introduced several steps before) are there and Inst may demand of Cnst that she stick the bolt through the hole. Inst underspecifies, thus triggering Cnst's cooperative behaviour for a second time.

In addition some other features casting light on coordination procedures deserve comment:

The two adjacent speech acts are joined by an illocutionary connective *and.*

The anaphora *it* in the second directive refers back to an antecedent generated in a side sequence following the pattern of Clark's "presentation" and "acceptance phase" (see Clark, 1996:227ff). It is run through twice here. The first presentation is *a screw* which is extended by *an orange one with a slit.* This "expansion" is accepted in turn by Cnst. The cooperatively produced antecedent is *a screw, an orange one with a slit.*

Finally, the intended perlocutionary effect of the second speech act is indicated by Inst's *so that the three bars get fixed then,* which acts as a sort of control. Here again we have coordination, since Cnst issues an appropriate reply, stating that the result of her action corresponds to Inst's description.

2. Coordination of Directives' Parameter Values

2.1. Sufficiently Informative Propositions in Directives

In the task at hand, Cnst typically has to take a Baufix-object and join it onto the aggregate built up so far. This can be seen from example (1), where she has to look for a particular bolt and put it through the intended junction shown in Fig. 4. The bolt's head has to go into the position designated as the top area of the aggregate before.

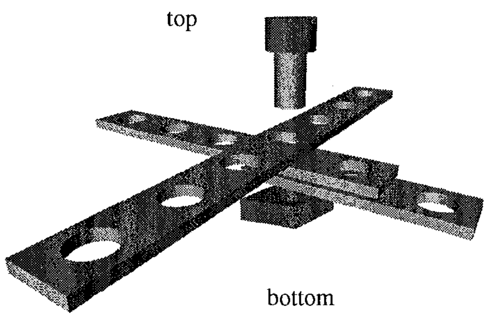


Fig. 4: The intended junction in (1)

Intuitively, we can conceptualize Cnst's task as an action to be carried out. Generalizing to all of Cnst's tasks indicated by Inst's directives we may say that she has to instantiate an action schema. This schema has different parameters to be matched with values in the particular action chosen. The schema is

(2) Action <Agent, Object, Manner, Location, Direction, Instrument, Time, Cause/Reason>

Concerning (1) it is roughly instantiated by

(3) *put-through* <Cnst, *an orange bolt with a slit,* Manner, *there, from the top*, Instrument, *now*, Cause/Reason>

The parameters Cnst, Manner, and Cause/Reason are free. Cnst knows what to do, i.e. which action schema to instantiate, because of Inst's directive. Hence we can say that Inst's directive provides a schema by means of the proposition expressed. The proposition in turn describes the situation Inst wants to see on Cnst's side. Actually, we can represent the put-through-directive in (1), i. e. the speech act, by an illocutionary force marker "Directive" and the list of parameters partially instantiated as in (4):

(4) <Directive, *put-through*<Cnst, *an orange bolt with a slit*, Manner, *there, from the top*, Instrument, *now*, Cause/Reason>>

We call the partially instantiated action schema paired with an illocutionary force marker "proposition embedded in the directive".

The proposition embedded in a directive must be sufficiently informative in order to enable Cnst to carry out the instruction. In other words, a certain number of parameters in (2) has to be paired with appropriate values, which ones depends on the particular action. Roughly, Inst can see his model and hence knows which values must be associated with which parameters. However, what frequently happens is that Inst does not provide sufficient information, e.g. he omits certain parameter-values altogether. This leads to problems on Cnst's side, who cannot carry out an adequate action. Thus a coordination problem becomes manifest. Its solution is that a sequence of agents' contributions is started with the common aim to fix the parameter values needed, cf. the exchanges inexample (1b):

(1b)	
Inst:	Und steckst sie dadurch, also *And you put it through there, well*
Cnst:	Von oben. *From the top.*
Inst:	Von oben, daß also die drei festgeschraubt werden dann. *From the top, so that the three bars get fixed then.*
Cnst:	Ja. *Yes.*

However, coordination on parameter values alone does not explain all cooperative contributions in task-oriented dialogue as example (1a) shows, where Inst had not yet completed the proposition of his directive and yet Cnst intervened:

(1a)	
Inst:	So, jetzt nimmst du *Well, now you take*
Cnst:	eine Schraube *a screw.*

Hence, mechanisms of a different sort have to be added as an explanation as we'll show in chapter 3.

We spot and *inter alia* annotate these exchanges tied up with sufficiently informative propositions in our dialogue data. A preliminary investigation of all our data (Skuplik, 1999) shows that we have at least 126 patterns as in (1a/b). This number comprises only entry sections of coordinative syntax productions. As a consequence, what is captured is coordination at the very first level. Hence, syntax coordination consequential upon entry sections have not been considered, although they frequently occur in the data. We take the number of agents' contributions necessary to yield a sufficiently informative proposition as a measure of coordination in a dialogue.

2.2. Coordination throughout the Dialogue

Following the rationale laid down in the previous chapter we can now demonstrate coordination in one dialogue[2] as follows (see Fig. 5 on the following page): On the horizontal axis we insert the number of sufficiently informative propositions to construct the plane, arranged in a sequence. On the vertical axis we indicate the number of contributions for every proposition. This yields the diagram in Fig. 5 where we have extreme peaks tied up with particular propositions. Intuitively, peaks arise due to

(a) properties of objects (number of surfaces, different holes with/without thread)

(b) complexity of constructions (number of parameters to instantiate as indicated by the action schema (2)),

(c) conceptualization and language variation.

(c) deserves some comment. Inst uses *überschneiden/ intersect* instead of *überlappen/overlap* and *mittig/ aligned, in line* instead of *mittlere/middle.* In addition, Inst applies his reference grid in naming positions, which causes problems for Cnst, who can't use the same grid. For matters of convenience, we use the term "peak-parameters" for (a), (b) and (c). A more satisfactory presentation would have to provide diagrams for the individual peak parameters and use them as a basis to compute the overall coordination diagram. In Fig. 5 the dark line indicates the number of contributions, the light line the number of parameters of sufficiently informative propositions. Observe that Fig. 5 also shows the distribution of peak-parameter (b).

2.3. Preliminary Evaluation: A Note on Discourse Structure and Stable Propositions

It is perhaps worthwhile to comment now upon two emerging problems, one tied up with discourse structure and the other with the notion of proposition to be used in our context of investigation.

[2] We chose dialogue 21 from „Wir bauen jetzt also ein Flugzeug" (s. a., 1997) and annotated it with respect to the following properties: track structure, social coordination, illocutionary force, coordination on sufficiently informative propositions, beginning and end of sufficiently informative propositions.

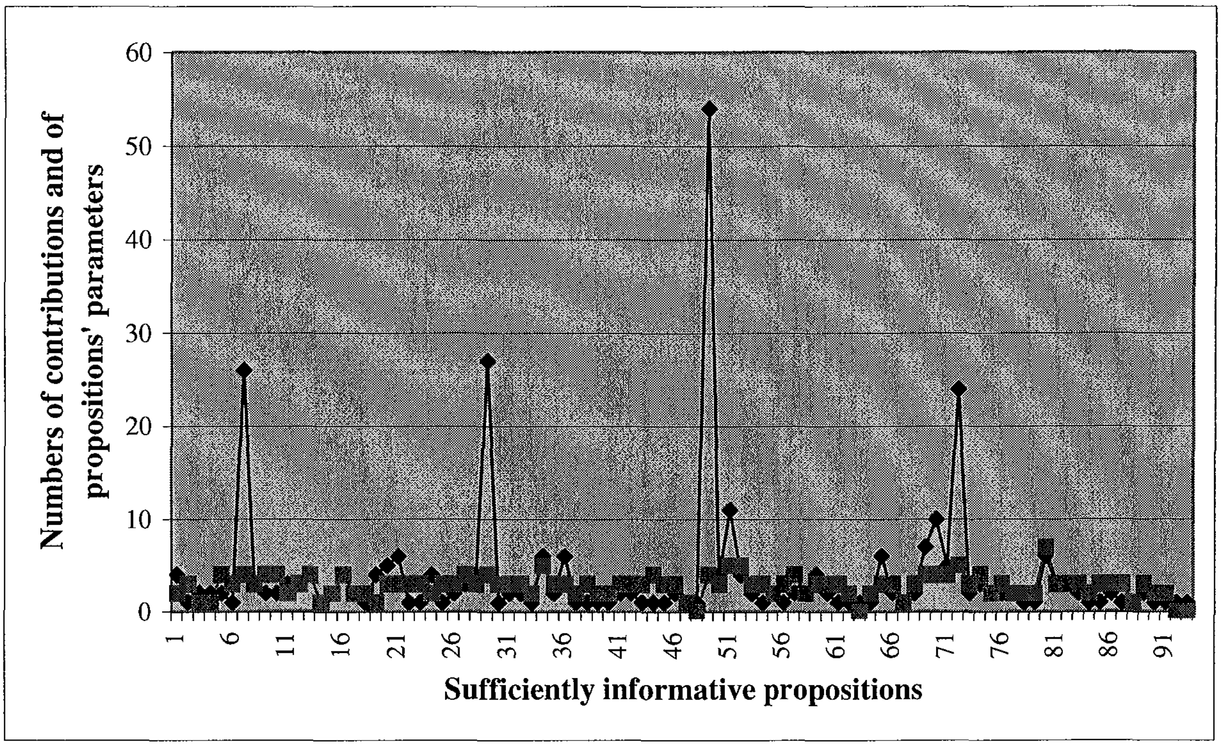


Fig. 5: Coordination throughout the dialogue

Dark lines indicate the number of contributions, light lines the number of sufficiently informative propositions' parameters. We observe the following peak positions: Proposition 7: Five-holes bar goes on top of three-holes bar (26 contributions, 4 parameters). Proposition 29: Three-holes bar (tail) is fixed below five-holes bar of fuselage (27 contributions, 4 parameters). Proposition 49: First wing goes crosswise over fuselage (54 contributions, 4 parameters). Proposition 51: Bolt is put through first wing and fuselage (11 contributions, 5 parameters). Proposition 70: Front-position of blue holes-cube (10 contributions, 4 parameters). Proposition 72: Position of yellow cube in undercarriage (24 contributions, 5 parameters). We did not count 80 of 437 contributions related to social interaction. The mean of contributions per sufficiently informative proposition is 3.5, the mean of parameters being 2.8. Hence, agents need more than one utterance for a directive and more than two contributions in a presentation-acceptance cycle.

The basic tool the agents aim at consists of a sufficiently informative proposition on the side of Inst and a clear reply on the side of Cnst indicating the action carried out. What we encounter instead is a presentation of part of a directive by Inst, followed by a "sub-dialogue" of contributions until the directive's proposition contains sufficient information. We could also say that the underspecified proposition is updated until a level of sufficiency is reached. The mechanism described captures cases where the information provided is clear but insufficient. In example (1) we have the following instantiation of the action schema after Inst's *put-through*-directive:

(5) *put-through* <Cnst, *it*, Manner, *there*, Direction, Instrument, Time, Cause/Reason>

However, Inst's and Cnst's exchange adds values to free parameters in the following way:

(6) *put-through* <Cnst, *it*, *so that the three bars get fixed*, *there*, *from the top*, Instrument, Time, Cause/Reason>

A different sort of cases also leads to sub-dialogues: These are problems related to the information as presented or perceived, i.e. presented or perceived in a non-satisfactory way. We do not treat these here. They are tied up with the first three levels of H. Clark's action ladder[3] and his notion of track (Clark 1996:147ff and 255).

The second question to discuss is when do the agents get at stable propositions in the dialogue, the preliminary answer being: if Cnst can use it in an appropriate way to bring about the situation demanded and also indicates this and Inst signals approval either explicitly or by default. These stable propositions are good candidates for entering into the common ground.

2.4. Linking up with Example (1)

In order to understand why the agents can achieve perfect syntactic fit in (1), we have to point out the level of coordination already achieved before (1) comes into being. Inst and Cnst agree on the position of the bars and the junction as shown in Fig. 4. In addition, Cnst has learned by past coordination procedures that there is only one pattern of producing a rigid join in aggregates, namely, screwing a bolt into a nut-like object. The latter can be a nut proper or a holes-cube with a thread in a

[3] We do not have action ladder problems up to level 3 frequently in our data.

suitable position. It is this insight which doubtlessly feeds her anticipatory potential and reasoning which we capture by practical syllogism (see ch. 3).

3. Coordination Points and How to Detect them

Looking at the coordination process in (1), the following questions seem to be worthwile to discuss:

- How do Inst and Cnst manage to produce a syntactically well-formed directive?
- Which information must be given for Inst to produce (part of) a directive?
- Which information must be given for Cnst to continue Inst's started utterance?
- How is the coordination problem solved?

Finally, with respect to the description of these matters we should ask

- What are the representational tools to be used?

We discuss each of these questions in turn.

3.1. Syntactic Well-formedness

Coordinated productions can show different grammatical characteristics. These may range from close syntactic fit, as in example (1), to mere pragmatic acceptability only loosely based on properties of syntactic form. In this paper our focus is on cases of close syntactic fit.

At first sight, the coordination task of Cnst seems to be simple. Inst is about to produce a VP, the transitive verb being already spelled out. Cnst adds a missing NP, perhaps triggered by a pause of Inst's. The resulting construction is well-formed and so is the construction after Inst's addition. The cooperatively produced utterance hence turns out to be *Well, now you take a screw, an orange one with a slit.* On second sight, things are more complicated.

3.2. Information in Order to Produce a Directive

Inst focusses the particular junction in his airplane (see Fig. 4 above), where a five-holes bar, a three-holes bar and a seven-holes bar are tied together with an orange slit bolt sticked through a hole formed by the overlapping holes-bars. The bolt in turn is fixed by a nut. Also, the bolt head is considered to go into the "on-top"-position, the nut as taking the "beneath"-position.

Inst knows the details of the junction on his side (named J_{Inst} and similarly in the other cases). He intends that Cnst build J_{Cnst}, the "twin" of J_{Inst}. He believes that Cnst has the following "Baufix"-parts at her disposal, introduced in earlier tasks:

- one three-holes bar ($3HB_{Cnst}$),
- one seven-holes bar ($7HB_{Cnst}$),
- one five-holes bar ($5HB_{Cnst}$),
- one orange slit bolt (OSB_{Cnst}),
- one orange nut (ON_{Cnst}).

In addition, he believes that Cnst can identify the hole where the bars should overlap, abbreviated as H_{Cnst}. Inst has the global intention that Cnst take OSB_{Cnst} and put it through the hole H_{Cnst} in order to fix $3HB_{Cnst}$, $7HB_{Cnst}$ and $5HB_{Cnst}$.

Let us make the assumption that Inst's intention distributes over the taking-of-the-bolt-action and the putting-the-bolt-through-the-hole-action in order to achieve the fixing of the bars. What else must we assume concerning Inst? He believes that Cnst has neither selected OSB_{Cnst} nor fixed the bars. He also has the firm belief that uttering an appropriate sequence of speech acts like "Take OSB_{Cnst}!" and "Put OSB_{Cnst} through H_{Cnst}!" will *ceteris paribus* satisfy his intention. Being rational, Inst will intend to produce the said sequence of illocutionary acts, above all the first one concerning OSB_{Cnst}. Things being set up this way, we may assume that Inst maintains the following intentions I_{Inst} in the current situation:

(7) $\exists x\ (I_{Inst}!(\text{now-take}(Cnst,x)) \wedge OSB_{Cnst}(x) \wedge I_{Inst} \ldots.)$

Oversimplifying matters somewhat, we may assume as a *ceteris paribus* rule that things intended are carried out.

Again by *ceteris paribus* Inst will couch the content of the illocutionary act (7) into words by some sort of morpho-syntactic translation procedure. This will give us something like

(8) $\exists x\ !(\text{now-take}(Cnst,x) \ldots \quad \Rightarrow$ *Now you take*

Observe that on both sides of the "⇒" we have incomplete expressions: This nicely matches that Inst is IN the process of issuing a directive's content and IN the process of editing out its morpho-syntactic frame.

Production proceeds by increments, and we know from (1) that after the first increment of Inst's, Cnst comes in. We now turn to the question why she can come in at all and why successfully so.

3.3. Information in Order to Produce a Continuation

Cnst knows that she has $3HB_{Cnst}$, $7HB_{Cnst}$, $5HB_{Cnst}$, OSB_{Cnst}, ON_{Cnst} and that the bars jointly form the hole H_{Cnst}. By anticipation, Cnst intends that the three bars be fixed in the overlap H_{Cnst}. She believes that she hasn't yet brought about the fixing. She also believes that taking some bolt and putting it through H_{Cnst} will serve her purpose. Observe that at the current construction stage Cnst could use various bolts of equal shape (see Fig. 3). Cnst's intention to fix the bars will lead her by some sort of practical reasoning to the intention of immediately taking up a bolt:

(9) $\exists x\ (I_{Cnst}!(\text{now-take}(Cnst,x)) \wedge \text{bolt}(x)).$

Ceteris paribus she will transform the content of her intention into sensomotoric behaviour. With a little effort you can imagine Cnst stretching out her hand for some bolt as intention dictates her. Now we are at the heart of the coordination problem, which, as it turns out, is also a timing problem.

3.4. How is the Coordination Problem Solved?

At this very moment Inst utters *Now you take*. Const maps Inst's utterance onto the following reconstructed intention:

(10) $\exists x\ (I_{Inst}!(\text{now-take}(Cnst,x)) \wedge \text{bolt}(x))$.

Matching her own intention with the reconstructed one, she arrives at the non-matching point $\exists x\ (\text{bolt}(x) \ldots)$. Granted this is a short cut, it could presumably be backed by some procedure using Gricean maxims. Cnst then takes over the reconstructed intention of Inst's and in order to indicate this (that's the main point here), she utters *a screw*.

3.5. What Are the Representational Tools to be Used?

The reasoning of Inst's and Const's can be expressed with practical syllogism as suggested e.g. in Lascarides and Asher (1999:4):

(11) $\mathbf{PS}_{\text{Asher\&Lascarides}}$

$$\begin{array}{ll} \text{(a)} & (I_B\psi \wedge B_B\neg\psi \quad \wedge \\ \text{(b)} & B_B(\phi > \text{eventually}(\psi)) \wedge \text{choice}_B(\phi,\psi)) \\ \hline \text{(c)} & > I_B\phi \end{array}$$

(11) reads as follows: 'If (a) B intends ψ and believes $\neg\psi$ and (b) B believes that he can nonmonotonically infer ψ if ϕ is true, and ϕ is B's choice for achieving ψ, then (c) [nonmonotonically] B intends ϕ.'. All the *ceteris paribus* conditions we use will be expressed by the Asher-Morreau defeasible inference '>'. A > B can be paraphrased as *If A then normally B.*

In order to justify Cnst's utterance as coming into existence, we need a "Cooperativity Principle" satisfying at least the following: 'An agent B is cooperative with agent A if he adopts A's goals.' (Lascarides & Asher, 1999). In addition we stipulate a principle which says 'Make assumed conformity of intentions and interests publicly known!' The Cooperativity Principle will make Cnst adopt the reconstructed intention of Inst's and the second principle will make him utter *a screw.*

The use of some schema of practical syllogism does not in itself seem to need much defense. We have the classical situation for its application: An agent's hypothesis that some means (ϕ) will achieve the derived end (ψ) makes him desire ϕ.

Following a reviewer's suggestion, we now briefly explain firstly, why the Asher-Morreau conditional should be used and secondly, why the Asher-Lascarides version of practical syllogism is a promising tool. Finally, we comment upon the mechanisms leading Inst to produce *Well, now you take* and Cnst to intervene and continue with *a screw,* diagrammatically represented in Fig. 6 on the following page.

The usually encountered conditional in schemas of PS is the material conditional (see e.g. von Wright (1983a, b)). Using the Asher-Morreau conditional '>', one can on the one hand express the fact that agents believe that ϕ will bring about ψ and on the other hand that the connection obtaining between ϕ and ψ is a fairly weak constraint which can be overridden. In order to see this consider the following situation: Let ϕ be tied up with an agent B's production of a speech act and let ψ be the action demanded by its embedded proposition. Then production of ϕ by B could lead to the addressee's intervention which in turn finally could lead to a revision of ϕ, call it ϕ', ultimately resulting in ψ'.

A similar remark can be made with respect to (c): The relation of the premises (a) and (b) to the conclusion (c) is also non-monotonic. Consider the case that something unexpected detracts agent B's typical behaviour resulting in a new intention, i.e. one no longer related to ϕ. Then the conclusion would not be maintained any more.

In general, if one adopts (PS), one is committed to an analysis in terms of intention.

3.6. Parallel Moves, Parallel Situations

Fig. 6 shows the mechanisms necessary to explain cooperative production of utterances. Inst extracts the information ψ_{Inst} out of his model, i.e. the information that OSB_{Inst} put through H_{Inst} fixes the three bars $3HB_{Inst}$, $7HB_{Inst}$ and $5HB_{Inst}$. He desires that the same should be the case on Cnst's side and he assumes in addition that this has not been accomplished yet. He also believes that uttering an appropriate speech act ϕ would lead to the desired result. Hence (a) and (b) of (11) are clearly satisfied.

At the same time Cnst, observing her situation, notices that a bolt put through H_{Cnst} could fix the bars on her side (ψ_{Cnst}). This entails her belief that $\neg\psi_{Cnst}$. She considers ψ_{Cnst} as the outcome of a possible future action. Hence (a) and (b) of (11) also obtain with respect to Cnst.

Now we reflect upon the consequences of the respective conclusions. On Inst's side the consequence of (11) leads to an intention to produce a speech act, whereas on Cnst's side we have an intention to pick up a bolt. Intentions to act typically lead to corresponding acts. On Inst's side the result is the planning and production of a speech act involving the action schema (2) and on Cnst's side the planning and production of a sensomotoric action.

Concurrently with her own sensomotoric action Cnst perceives the spelled out fragment of Inst's illocutionary act *Well, now you take.* Cnst reconstructs Inst's intention as an intention involving a take-something action of Cnst's on the basis of the fragment *now you take.* Now the explanation assumes a Cooperativity Principle guaranteeing that she matches her own action intention directed against grasping a bolt with Inst's assumed action intention involving a taking-of-something act. The Cooperativity Principle is designed to trigger the comparison process. Cnst's matching of intentions will reveal an overlap, the taking-of-something part. Finally, Cnst's planning and production of *a screw* seems to be based on another principle or interactional maxim saying 'make assumed conformity of intentions and interests publicly known'. As before, we have it that planning and production of an act lead to an act planned by default.

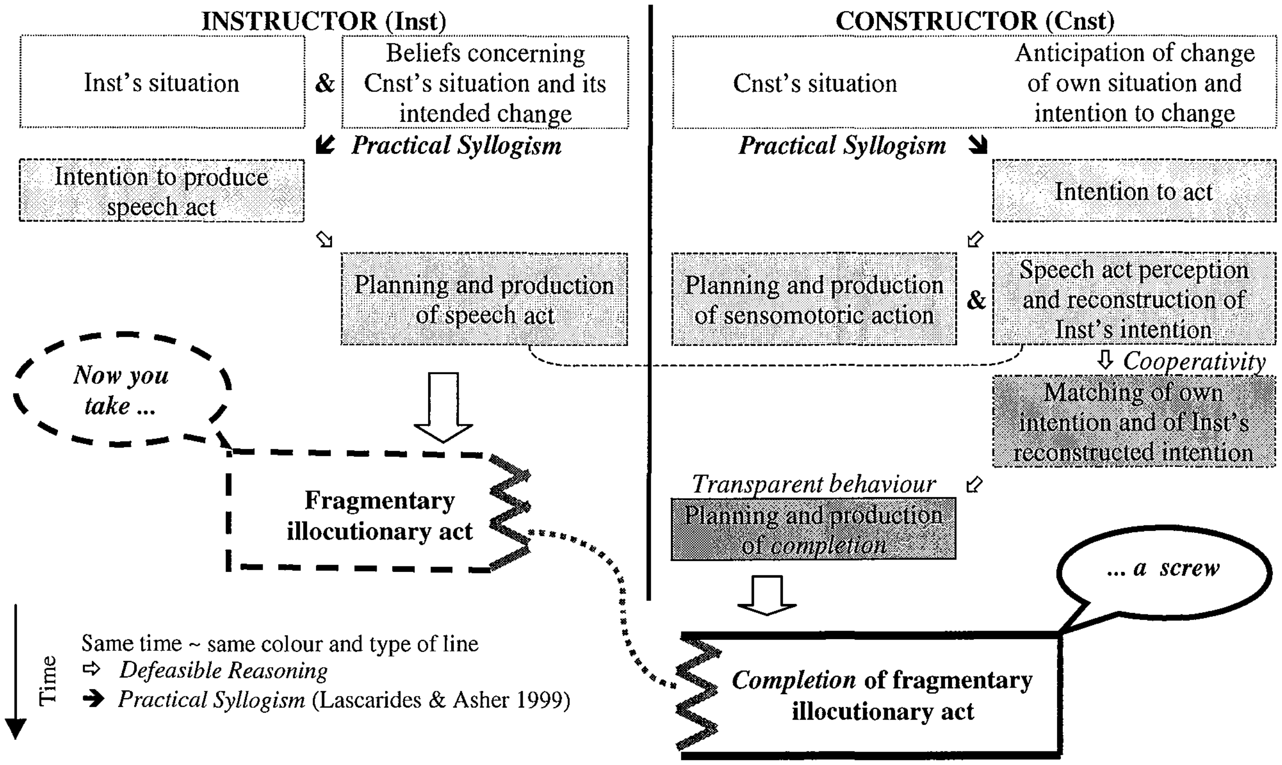

INSTRUCTOR (Inst)
Inst's situation
&
Beliefs concerning Cnst's situation and its intended change
Practical Syllogism
Intention to produce speech act
Planning and production of speech act
Now you take ...
Fragmentary illocutionary act
CONSTRUCTOR (Cnst)
Cnst's situation
Anticipation of change of own situation and intention to change
Practical Syllogism
Intention to act
Planning and production of sensomotoric action
&
Speech act perception and reconstruction of Inst's intention
Cooperativity
Matching of own intention and of Inst's reconstructed intention
Transparent behaviour
Planning and production of completion
... a screw
Completion of fragmentary illocutionary act
Time
Same time ~ same colour and type of line
⇨ Defeasible Reasoning
➔ Practical Syllogism (Lascarides & Asher 1999)

Fig. 6: Explaning Agents' Coordinated Behaviour Using Practical Syllogism and Defeasible Reasoning

4. Conclusion and Outlook

4.1. Explanations Using Practical Syllogism and Various Principles

So far we get at the coordination point in the following way: Inst wants Cnst to act and thinks that a directive will make him do so. Therefore he plans the directive and starts producing it. Cnst anticipates Inst's aim. She reconstructs Inst's intention from his fragmentary utterance, comes to believe in near-identity of her and Inst's intentions and indicates that by completion of the directive. As a result, close fit of utterance structure is achieved.

The formal tool we use for intention-and-belief-based explanation is the Asher-Lascarides schema of Practical Syllogism. However, some effort will have to be put into answering the question which properties the Asher-Morreau conditional integrated into the Asher-Lascarides schema has and why it should be considered superior to other forms of reasoning under uncertainty.

In the context of explanation by defeasible reasoning we also have to look more closely into the principles used, which could not yet be accomplished. Take-over of intention has to rest on a „Cooperativity Principle" the formulation of which can be based on already existing proposals. Spelling out coordination points token-wise can be grafted upon a Principle of "Transparent Behaviour" saying 'Make Conformity of Intentions and Beliefs Public!' for which, however, there is no precedent so far.

4.2. Empirical Investigations

Departing from example (1) we have shown in this study that agents can produce utterances together by some sort of theoretically ill-understood division of labour. In this context, syntactic fit is a remarkable property, but it is, of course, not the whole story: what is needed is pragmatic fit. The pragmatic fit of syntax completions or continuations depends on the agents' coordination with respect to their task. The explanatory concepts like "sufficiently informative propositions", "action schema", principles etc. we need to describe pragmatic fit were developed on the basis of a fine-grained annotation of one task-oriented dialogue, which also provided example (1). However, a preliminary investigation of the whole corpus revealed that similar cooperative productions can be found in the other dialogues too (see ch. 2.1 and Skuplik, 1999). From our data it follows that theories of (task-oriented) dialogue would have to be based on agents' coordination. As a consequence, the concept of coordination would also affect structures being part and parcel of a theory of dialogue. The discussion of stable propositions at the end of ch. 2.3 provides a case in point. Propositions coordinated on should get a privileged position in dialogue theory, since they form the bases of ensuing future verbal and non-verbal acts and finally also determine the felicitous completion of the task at stake. A coordination-based concept of proposition would of course also have considerable impact on theoretical concepts which normally are grounded on some notion of proposition, e.g. mutually believed information or common ground.

Acknowledgement

We want to express our gratitude to Peter Kühnlein, SFB 360, as well as to Guido Heumer and Martin Hoffhenke, AG Wissensbasierte Systeme, Technical Faculty, University of Bielefeld, for providing the pictures for our paper. We are also indebted to two anonymous reviewers, who suggested improvements to our paper and specified profitable research lines.

Bibliography

[Amstelogue'99] s. a. (1999). *Preproceedings of Amstelogue'99.* Workshop on the Semantics and Pragmatics of Dialogue, Univ. of Amsterdam.

André, E., Poesio, M. & Rieser, H. (Eds) (1999). *Deixis, Demonstration and Deictic Belief.* Workshop Proceedings. ESSLLI 99, Univ. of Utrecht.

Asher, N. (1998). *Common Ground, Correction and Coordination.* Department of Philosophy. The Univ. of Texas, MS.

Benz, A. & Jäger, G. (Eds) (1997). *Mundial '97.* Munich Workshop on Formal Semantics and Pragmatics of Dialogue. Proceedings. CIS, Univ. of Munich.

Chierchia, G. (1995). *Dynamics of Meaning.* London: The University of Chicago Press.

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

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. In *Cognition* 22:1-39.

Cohen, Ph. R. (1984). The Pragmatics of Refererring and the Modality of Communication. In *Computational Linguistics,* Vol. 10. No.3:97-147

Dekker, P. (1997). On First Order Information Exchange. In Benz, A. & Jaeger, G. (eds), 21-40.

Groenendijk, J. & Stokhof, M. (1991). Dynamic Predicate Logic. In *Linguistics and Philosophy* 14:39-100.

Heydrich, W. & Rieser, H. (Eds) (1998). *Mutual Knowledge, Common Ground and Public Information.* Proceedings of the ESSLLI'98 Workshop, Univ. of Saarbruecken.

Hulstijn, J. & Nijholt, A. (Eds) (1998). *Twendial 98.* Twente Workshop on Formal Semantics and Pragmatics of Dialogue. Proceedings, Twente Univ. Faculteit Informatica.

Kamp, H. (1990). Prolegomena to a Structural Theory of Belief and Other Attitudes. In Anderson, C.A. & Owens, J. (Eds). *Propositional Attitudes* (pp. 27-90). CSLI Lecture Notes 20.

Kamp, H. & Reyle, U. (1993). *From Discourse to Logic.* Dordrecht: Kluwer.

Lascarides, A. & Asher, N. (1999). Cognitive States, Discourse Structure and the Content of Dialogue. In [Amstelogue'99].

Poesio, M: (1998). Cross-speaker Anaphora and Dialogue Acts. In Heydrich, W. & Rieser, H. (Eds), *Mutual Knowledge, Common Ground and Public Information* (pp. 41-46). Proceedings of the ESSLLI'98 Workshop, Univ. of Saarbruecken.

Rieser, H. (1999). The Structure of Task-oriented Dialogue and the Introduction of New Objects. Invited Speaker's Address. In [Amstelogue'99].

s. a. (1997): "Wir bauen jetzt also ein Flugzeug ...". Konstruieren im Dialog. Cognitive Science Unit 360 "Situated Artificial Communicators", Univ. of Bielefeld.

Schegloff, E. (1979). The relevance of repair to syntax-for-conversation. In Givon, T. (Ed.). *Syntax and Semantics* Vol. 12 (pp. 261-286). New York: Academic Press.

Searle, J. R. & Vanderveken, D. (1985). *Foundations of Illocutionary Logic.* Cambridge: Cambridge University Press.

Skuplik, K. (1999). *Satzkooperationen: Definition und empirische Untersuchung.* Report 1999/03 of the Cognitive Science Unit 360 "Situated Artificial Communicators", Univ. of Bielefeld.

Traum, D. R. (1999). 20 Questions on Dialogue Act Taxanomies. In [Amstelogue'99].

von Wright, G. H. (1983a). Practical Inference. In von Wright. *Practical Reason* (pp. 1-17). Oxford: Blackwell.

von Wright, G. H. (1983ba). On So-called Practical Inference. In von Wright. *Practical Reason* (pp. 18-34). Oxford: Blackwell.

Decision Problems in Pragmatics

Robert van Rooy

Institute for Language, Logic, and Computation
University of Amsterdam,
Nieuwe Doelenstraat 15, 1012 CP Amsterdam,
the Netherlands
vanrooy@hum.uva.nl

Abstract

Decision theory is used to define a notion of 'relevance' in terms of decision problems. Decision problems are also used to explain why attitude attributions are made. Assuming that belief attributions are made to explain unexpected actions, and that assertions have to be relevant, it is shown that potentially ambiguous, or underspecified, *de re* belief attributions can be disambiguated by context. In the last part of the paper I show how decision problems can be used to derive conversational implicatures.

1. Introduction

Where semantics is the study of how to determine the meaning of a complex expression in a systematic way from the meaning of its parts, pragmatics can be seen as the study of whether we could use certain expressions *appropriately* or *relevantly* in certain circumstances or not, and *why* this is the case. Whether the use of a certain linguistic expression is appropriate and/or relevant or not depends obviously on the attitudes of the agents involved in the conversation. It is widely recognized that the relevance, or appropriateness, of the usage of linguistic expressions depends crucially on the *common ground*, what is *presupposed* among the participants of the conversation. But it should be equally clear that the appropriateness depends also on the *beliefs* and *preferences* of the agents involved and on their rationality. This suggests that pragmatics should be based on a theory of rationality, i.e. *decision theory*. In some related work (Van Rooy; 1999, 2000) I use decision theory to propose a measure of *pragmatic relevance* of questions, and use this to determine which question is actually expressed by an interrogative sentence. In this paper I want to show how decision theoretical tools are useful (i) to define a measure of pragmatic *relevance* of *assertions* that can be used to determine what is actually expressed by declaratively used sentences; (ii) to use this notion of relevance in particular for the analysis of *de re* belief attributions, on the assumption that belief and desire attributions are made to *explain behavior*; and (iii) to account for certain *conversational implicatures*.

2. Decision Theory

Decision theory is a theory that tells us which action an agent will or should perform given his beliefs and his desires. It is normally assumed that the beliefs of an agent can be represented by a *probability function*, a function from events to real numbers that add up to one. To represent the preferences, or desires, of the agent, on the other hand, we need not only a probability function, but also an additional function that assigns to each action a payoff given an event, a *utility function*. A *decision situation* can then (in the finite case) typically be represented by a *decision table*, which identifies the conditional gain associated with every possible combination of the acts under consideration and the possible mutually exclusive events with their probabilities of occurrence:

				Actions		
World	Probability	a_1	a_2	a_3	a_4	a_5
e_1	1/9	5	1	4	2	7
e_2	2/9	3	3	5	6	1
e_3	4/9	1	3	0	3	3
e_4	2/9	6	2	3	5	2

Once we have both the probability and the utility function around, the most sensible way to determine which action the agent should perform is by *maximizing the expected value*. The expected value of each action is determined by multiplying the conditional values for each action/event combination by the probability of that event, and summing these products for each act. The expected utility of action a_1, for instance, is $(5 \times 1/9) + (3 \times 2/9) + (1 \times 4/9) + (6 \times 2/9) = 27/9 = 3$. The action which should be chosen is a_4, because it is the action which maximizes the expected utility, $36/9 = 4$.

Assuming that the relevant possible events are indeed exclusive, we can say that the probability functions take *worlds* as arguments. If we assume that the utility of performing action a in world w is $U(w, a)$, we can say that the expected utility of action a, $EU(a)$, with respect to probability function P is

$$EU(a) \quad = \quad \sum_{w} P(w) \times U(w, a).$$

The decision situation of the agent as described above is also known as a *decision problem*, and can in general be modeled as a triple, $\langle P, U, A\rangle$, containing (i) the agent's probability function, P, (ii) her utility function, U, and (iii) the alternative actions she considers, A. Given this, we might now define a function, O, from decision problems to sets of actions, such that $O(\langle P, U, A\rangle)$ gives us the set of actions in A which have maximal expected utility with respect to the decision problem. If we denote the decision problem of the above example by DP, for instance, $O(DP)$ will denote $\{a_4\}$. In this paper I will use this function, O, to determine whether or not an attitude attribution was relevant when used to explain an action.

3. The relevance of assertions

It is well known that Grice's (1989) famous maxims of conversation should not be taken as arbitrary conventions among language users, but rather describe means for conducting rational co-operative exchanges. This means that these maxims should be explained by a more general theory of cooperation among rational agents. We might say that Grice's supermaxim of cooperation between S and cospeaker T can be implemented within our framework by saying that T takes over the goals of speaker S. More in particular, T wants to help S to determine what would be the best action for her to perform among the alternative actions she considers. Let us now assume that S's decision problem is common knowledge among S and T. It turns out that in that case we might determine the *utility of assertions* in terms of the relevant decision problem in a very natural way.

If we assume that $EU(C, a)$ denotes the utility of action a after conditionalization of the initial probability function P with C, $\sum_w P(w/C) \times U(w, a)$, we can denote the value of a decision problem after proposition C is accepted by $max_i EU(C, a_i)$. In terms of this notion we can determine the value, or *relevance*, of the assertion C. Referring to a^0 as the action that has the highest expected utility according to the original decision problem, $\langle P, U, A\rangle$, we can determine the *utility* of the *assertion* C, $U(C)$, as follows:[1]

$$U(C) \quad = \quad max_i EU(C, a_i) - EU(C, a^0)$$

This value can obviously never be negative, and will be equal to zero if accepting the new proposition will not lead to a different action that the agent will perform. Now we might say that assertion C is *relevant* just in case the value of $U(C)$ is strictly higher than zero, $U(C) > 0$, or if the set of optimal actions of the new decision problem after the acceptance of C will not be the same as before the acceptance, $O(\langle P, U, A\rangle) \neq O(\langle P^*_C, U, A\rangle)$. This notion of relevance might be used to determine what is actually said by declaratively used sentences. In particular, it might be used to select the relevant contextual parameter needed to determine what is expressed by belief attributions.

4. Attitude attributions as explanations

It is common wisdom that a lot of belief and desire attributions are made to explain the behavior of an agent. But the question is what this explanation consists in. It is normally assumed that the conclusion of a practical reasoning is an action. But how does this conclusion follow? By the traditional account of explanation, the given belief and desire, together with a law given by theory, results via modus ponens to the explanandum. But it seems very unlikely that a psychological theory will ever have laws like "If an agent wants A and believes B, he (probably) will perform action a". A more natural assumption would be to assume that the background theory is the theory of rational action, i.e. *decision theory*. This theory prescribes that the action that the agent should perform is the action which has the highest utility. But on the assumption that only those acts are done that have the greatest utility, the attitude attributions won't be very illuminating either. The given belief and desire attributions do not give enough information to assure that the action to be explained has the greatest utility. Can we give a better account of what the contributions of belief and desire ascriptions are to explain the given action? I think we can, if we adopt van Fraassen's (1977) theory of explanation.

4.1. Van Fraassen's theory of explanation

To have an explanation of a fact is to have a theory which is acceptable and which explains that fact. According to the covering-law-model of explanation, primarily associated with Hempel, the explanandum is explained by a covering law given by the theory which either entails the explanandum, or gives it a high probability (in case of statistical theories). In general, the information given (the covering law) must give good grounds for believing that the phenomena did, or does, indeed occur. However, it was soon realized that 'giving good grounds for believing' is neither a *sufficient* (asymmetry of explanation) nor a *necessary* condition for a good explanation. That a barometer falls when storm is coming does indicate (and so gives good grounds for believing), but does not explain the fact that a storm is coming. In the same way, the length of the shadow indicates (together with the elevation of the sun, and the principles of optics) but does not explain the height of the tower. This is known as the *asymmetry problem*, because storm *does* intuitively explain the falling of the barometer, and similarly the height of the tower intuitively *does* explain the length of the shadow. Epistemic relevance is not the same as causal relevance. Giving good grounds for believing is also *not* a *necessary* condition for explanation, because a doctor can in some circumstances explain why his patient died from the disease called 'paresis' by saying '*Because he had latent syphilis which was left untreated*', although only a low percentage of such cases die of their illness.

Van Fraassen (1977) wants to give a theory of explanation that can account for the facts of the asymmetry of explanation and for the fact that the given explanation doesn't have to give the explanandum a high probability. To account for the asymmetry problem, he argues that the explanation only gives the *salient factor* in the part of the causal net that leads up to the explanandum. What the salient factor is depends of course on the assumed background knowledge of the hearer, his orientation and his interests. His interests and background knowledge determine the appropriateness of the explanation given. The explanation must be informative for the hearer about that part of the causal net where he was interested in.

To account for the fact that an explanation need not even give a necessary condition for the explanandum, van Fraassen argues that an explanation is an answer to a *why*-question, '*Why P_k?*', where P_k is the explanandum. The underlying structure of such a question is according to him '*Why P_k in contrast to $P_l, P_2, \ldots$?*' The set of those P_i's, i.e. X, is called the *contrast class*.[2] The answer to such a

[1] It is useful sometimes to think of a^0 as a *mixed* action of *indecision*.

[2] See Rooth (1984) for a very similar use of contrast classes in his theory of focus.

question must give relevant information (it must be informative about the causal factor the questioner is interested in) that P_k in contrast to other members of X is true. That reducing explanation to answering *why*-questions can account for the paresis example is explained by van Fraassen in the following way:

> If a mother asks why her eldest son, a pillar of the community, mayor of his town, and best beloved of all her sons, has this dread disease, we answer: because he had latent untreated syphilis. But if that question is asked about this same person, immediately after a discussion of the fact that everyone in his country club has a history of untreated syphilis, there is no answer. The reason for the difference is that in the first case the contrast-class is the mother's sons, and in the second, the members of the country club, contracting paresis. (van Fraassen, 1977)

So, a *why*-question can be identified with the triple $\langle P_k, X, R\rangle$, where P_k is the *explanandum* and the *topic* of the question, X the *contrast class*, and R the appropriate relation. Both X and R are context dependent. The presuppositions of $\langle P_k, X, R\rangle$ are that (a) P_k is true, (b) every P_j in X is false if $j \neq k$, and (c) there is at least one true proposition A that bears R to $\langle P_k, X\rangle$.

The normal form of an answer to a *why*-question is

(*) P_k in contrast to the rest of X because A

By the answer it is claimed that (1) P_k is true, (2) no member of X other than P_k is true, (3) A is true, and (4) A bears R to $\langle P_k, X\rangle$, i.e A is a reason. How is 'A bears relation R to $\langle P_k, X\rangle$' to be explained? Van Fraassen is not very clear about this, but I think that a natural answer can be given when we make the reasonable assumption that *why*-questions are typically asked when an *unexpected* or *abnormal* event occurred. In the next section of this paper we will concentrate ourselves only on *why*-questions related to belief and desire attributions.

4.2. Attitude attributions answer *why*-Questions

It seems natural to assume that belief and desire attributions are normally given to explain behavior, and that explanations are basically very context-dependent answers to *why*-questions. Taken together, this means that attitude attributions are usually given to answer questions like '*Why did the agent do* <u>*that*</u>*?*' Assuming van Fraassen's theory of *why*-questions, the problem is to determine both the *contrast class* and the relevant *relation R*. To solve this problem, I will make the further (but natural) assumption that *why*-questions are normally asked only when an *unexpected*, or *abnormal*, event occurred. In our case this means an unexpected action, b, performed by the agent. Action b was unexpected because the one who asked the *why*-question, Q, expected another action, a; he assumed that the attitude state of the agent should be represented by something like a decision problem, say $\langle P, U, A\rangle$, and that the optimal action of this decision problem, the unique element of $O(\langle P, U, A\rangle)$, is not b, but a.

Now that we have related van Fraassen's theory of *why*-questions for belief and desire attributions with decision problems, we can determine what the contrast class and the relevance relation should be. The contrast class can simply be thought of as the set of alternative actions of the decision problem. Determining the relevant relation, i.e. determining when an answer counts as an explanation, however, is a bit more complicated. I have assumed that when Q asks a question like '*Why did the agent do* <u>*that*</u>*?*', Q represented the attitude state of the agent by a decision problem like $\langle P, U, A\rangle$. Now I will make the further assumption that when someone answers the question by making an attitude attribution, he will know (more or less) how Q represented the attitude state of the agent. When Q accepts the attitude attribution given as an answer to the *why*-question, the effect of the attitude-attribution will be that Q changes his representation of the decision problem of the agent from $\langle P, U, A\rangle$ to $\langle P', U', A\rangle$. Assuming that $O(\langle P, U, A\rangle) = \{b\}$, but that the agent actually performed action a, I propose that the attitude attribution counts as an explanation exactly when $O(\langle P', U', A\rangle) = \{a\}$. So, I propose that an attitude attribution like *John believes C, and he desires D* stands short for a normal answer to a *why*-question like 'John performed a in contrast to the rest of A, because he believes C and desires D'.

But how will attitude attributions change Q's representation of the agent's decision problem? Here the difficult problem of belief and preference *revision* comes in, obviously. Belief and desire attributions change Q's representation of the agent's attitude state differently. It seems natural to propose that a belief attribution of the form *John believes C* effects the change from the attitude state $\langle P, U, A\rangle$ to the state $\langle P^*_C, U, A\rangle$, where P^*_C represents the posterior probability function resulting from the *revision* of the prior probability function P with C. Normally this change is just belief revision by conditionalization, but as explained by Gärdenfors (1988), among others, sometimes belief change should be accounted for in a more complicated way. The effect of desire attributions is more difficult to determine, because the desirability, or expected utility, of an action depends on both the probability function *and* the utility function.[3] In this paper I won't discuss the problem how to account for preference-change, although this problem has recently been taken up by Hansson (1995).

In this section I have argued that attitude attributions typically are relevant in a certain conversational situation when they help to explain an unexpected action performed by an agent. In the following section I will argue that this notion of relevance might be used to select the relevant contextual parameter needed to determine what is expressed by a *de re* belief attribution in case Quine's famous double vision problem arises.

[3] Notice that by using Savage's decision theory, I implicitly assume that *actions* are desirable, and not *propositions*. You might protest and say that propositions can be desirable, too, and that to account for desire attributions it is more natural to assume that just like *belief*, also *desire* denotes a relation between agents and propositions. If so, you might adopt Jeffrey's (1965) theory of decision.

5. Relevance and Quine's double vision problem

5.1. Context-dependence of *de re* belief attributions

Consider Quine's (1956) Ralph who, one evening, sees a man with a brown hat whose suspicious behavior leads Ralph to believe that the man is a spy. On another occasion, Ralph sees the same man at the beach, but he does not recognize him as the same man; and the thought that the man he sees at the beach is a spy does not even occur to him. Intuitively, we can attribute his beliefs by saying (1a) and (1b):

(1) a. Ralph believes of the man with the brown hat that he is a spy.

b. Ralph doesn't believe of the man he saw at the beach that he is a spy.

As it turns out, however, the man with the hat who is later seen at the beach happens to be Ortcutt. So we seem to be allowed to infer (2a) from (1a), and (2b) from (2a):

(2) a. Ralph believes of Ortcutt that he is a spy.

b. Ralph doesn't believe of Ortcutt that he is a spy.

Now, does Ralph believe that Ortcutt is a spy or not? Or better, how can we account for the beliefs attributed to Ralph that seem to be about Ortcutt without concluding that Ralph is irrational?

A natural reply to Quine's Ortcutt problem would be to demand that a *de re* attribution can be truly made only if the agent *knows* the object that the belief is about. According to this picture, *de re* attributions can be truly made only if the possibility of mistaken identity does not exist. But this means that neither (2a) nor (2b) can be true. Both of them will be false because Ralph doesn't know that a single individual, Ortcutt, is the source of the two relevant bodies of information; he knows the identity of neither the man with the brown hat nor of the man seen at the beach.

This reply, however, gives rise to problematic predictions. Suppose we tell only one half of the story. One evening, Ralph sees a man with a brown hat who behaves suspiciously and who, he comes to believe, is a spy. Ralph has never heard the name *Ortcutt*, but in fact it is the person named *Ortcutt* who is the suspiciously-behaving man with the brown hat whom Ralph has seen earlier. In these circumstances the belief attributions (1a) and (2a) seem to be appropriate and true, although Ralph has no discriminating knowledge about Ortcutt.

This, then, raises the question how we should account for the fact that Ralph might have two beliefs *about* the same individual that are (apparently) mutually inconsistent.

One reaction would be to assume that Ralph really believes in propositions that are mutually inconsistent, and thus that his belief state itself is mutually inconsistent. Another way to response to Quine's problem about Ortcutt, the response that I am going to adopt, would be to deny with Quine (1956) that Ralph's beliefs really are (internally) inconsistent. The most straightforward way to go about this in possible worlds semantics is to follow Kaplan (1969) and to judge a *de re* belief attribution like (2a) as being true iff Ralph has a representation in mind (i) that was actually caused by Ortcutt, and (ii) whose instantiation is a spy in each of Ralph's belief-worlds. In this way, two *de re* belief attributions like (2a) and (2b) need no longer lead to a contradiction, because there might be two representations that were actually caused by Ortcutt, but do not denote the same individuals in all worlds compatible with what Ralph believes.

But if in some of Ralph's belief worlds the actual Ortcutt has *two* representatives, or counterparts, which one do we refer to by a belief attribution like (2a)? According to van Fraassen (1979), and Stalnaker (1988), among others, which representative or counterpart we refer to depends not so much on the belief state of the agent itself, as on the *conversational situation* in which the belief attribution is made.

Notice that when we assume that it depends on context which representation is relevant for the analysis of the *de re* belief attribution, there might be conversational situations where we might truly say that Ralph *doesn't* believe of Ortcutt that he is a pillar of society, which indeed seems to be in agreements with the facts. If I had only given half of the story, and only told you that Ralph saw a man with a brown hat who behaves suspiciously, the belief attribution (2a) seems to be true.[4] In the next subsection I will discuss Stalnaker's proposal how context helps to select the meant interpretation of *de re* belief attributions.

5.2. How context selects interpretation

Assertions are analyzed with respect to a context. Such a context should contain at least the information that is presupposed by the members of the conversation. For the assertion to be appropriate, it has to be the case that what is said by the assertion should be *consistent* and *informative* with respect to what is presupposed in this context. Sometimes this constraint helps to explain why what is asserted is *inappropriate* in its context of interpretation, but the more interesting case is when it helps to determine what is actually expressed by the sentence being used that allows for several interpretations. If what is expressed by an utterance violates the constraint on one of its interpretations, but does not on the other, then the constraint suggests that the meant interpretation of the sentence was the latter one (cf. Stalnaker, 1978).

Not only whole sentences should be interpreted with respect to contexts, but embedded clauses too. The contexts with respect to which these embedded clauses should be interpreted, however, need not be the same as the ones with respect to which whole sentences need to be evaluated. It's standardly assumed, for instance, that the context of interpretation of the second conjunct of a conjunction, or the consequent of an indicative conditional, contains *more* information than the context with respect to which the whole sentence is interpreted. Sometimes an embedded clause has to be interpreted with respect to a context that not only is not identical with the context of interpretation of the embedding sentence, but also does *not* contain *more* informa-

[4] See the first chapter of van Rooy (1997) for a possible way of working this out by making use of a counterpart theory.

tion than the latter. Typical cases are contexts of interpretation of embedded sentences of *attitude attributions*. If we want to account for the intuition that the utterance '*Ralph believes that Mary stopped smoking*' can be appropriately uttered in a context where it is assumed that Mary never smoked, and where it has just been asserted that Ralph believes that Mary used to smoke, the context of interpretation of the embedded sentence has to be incompatible with the context of interpretation of the embedding sentence. Stalnaker (1988) proposes that the context of interpretation of the embedded clause of a belief attribution to Ralph should be the context representing what is commonly assumed to be believed by Ralph.[5]

This context now fulfills the same role as the main context fulfills for whole sentences: what is expressed by the embedded clause has to be consistent and informative with respect to it. Also, just like the main context can be used to help selecting what is expressed by an assertion due to the acceptability constraints, Stalnaker (1988) argues that these constraints on interpretation sometimes might also help to select what is expressed by the embedded clause of a belief attribution.

Consider, for instance, the previously discussed example (2a) of a *de re* belief attribution, repeated here as (3):

(3) Ralph believes that Ortcutt is a spy.

Although, as we have argued, the attribution (3) is in principle ambiguous, depending on which of Ralph's representations of Ortcutt the speaker has in mind, the context of interpretation of the embedded clause suggests only one particular reading in some situations. If it is presupposed, for instance, that Ralph believes that the man he saw at the beach is not a spy, but that no such thing is presupposed about what Ralph believes concerning the man with the brown hat, what is presupposed about Ralph's beliefs triggers the interpretation of (3) where 'Ortcutt' is interpreted as 'the man with the brown hat'.

But now suppose that the context is different such that what is expressed by the embedded clause of the belief attribution is on both interpretations consistent and informative with respect to what is presupposed to be believed by Ralph. Would that mean that the embedded sentence is, thus, always ambiguous between the two readings? Of course not! Although Stalnaker's (1978) proposed method for 'disambiguating' utterances is related to a notion of 'relevance', there are plenty of cases in which relevance determines the right interpretation, but where the notion of relevance cannot be reduced to mere consistency or informativity. For those cases the relevant notion of *relevance* is closer to the notion we discussed earlier in this paper. In the next section we will discuss such an example related to *de re* belief attributions.

5.3. Ortcutt and the English Gentleman

Consider the following case:

The famous former Polish-American politician Paderewski is found dead, murdered with a knife in his hotel room. Among the other guests in the hotel are Pierre, his French servant, and Bernard J. Ortcutt. Ortcutt is like Paderewski a former American politician, but his career was crushed by Paderewski, his political and personal enemy, some years ago.

Detective Ralph is asked to investigate the murder. He was told soon that Paderewski had an enemy for life, and so is acquainted with Ortcutt in this way. Because *we* know about the personal relation between Paderewski and Ortcutt, we expect Ralph to arrest Ortcutt for the murder. But then something unexpected happens: Ralph does not arrest Ortcutt, but Pierre. I ask you *Why?* You tell me that Ortcutt pretended to be a noble English gentleman during the investigation, that Ralph never found out that 'Ortcutt' was his true name, but that he was told that Pierre would inherit Paderewski's golden watch in case the latter would die. This latter fact gives Ralph reason to believe that Pierre had a motive for murdering Paderewski. Moreover, you explain me that Ralph thinks that Pierre also had the opportunity to do so by saying

(4) Ralph believes that both Ortcutt and Pierre had a key to Paderewski's door.

In this example Ralph is acquainted with Ortcutt in two different ways: as the personal enemy of Paderewski who's career was crushed, and as the noble English gentleman. The story makes clear that Ralph does not realize that the two 'individuals' are really one and the same. So, just like in Quine's (1956) example, the *de re* belief attribution allows for different interpretations. Notice that in this case the right interpretation of the belief attribution cannot by saying that only on one of the interpretations the attribution will be consistent and informative with respect to what is presupposed to be believed by Ralph. The example does not rule out, for instance, that Ralph has a belief that could be expressed by the *de dicto* attribution '*Ralph believes that Paderewski's personal enemy had a key to his door.*' Still, our intuition is that (4) should be interpreted such that the name 'Ortcutt' refers to the noble English gentleman in the belief-worlds of Ralph. The reason is, or so I would argue, that only under this interpretation the belief attribution would be *relevant*.[6]

The reasoning goes as follows: We expect Ralph to arrest Ortcutt, but in fact he arrests Pierre. This makes two actions relevant in the decision problem we attribute to Ralph. But we had a decision problem ourselves, too: should we bet on the hypothesis that says that Ralph will arrest Ortcutt, h, or on the hypothesis saying that Ralph will arrest Pierre, h'. On the basis of our conception of what Ralph believes, or will find out, we assume that Ralph believes (or will come to believe) that it is more likely that Ortcutt murdered Paderewski, and thus that he will arrest Ortcutt.

[5] So, if C represents what is presupposed about the actual world, and if $Dox(r,w)$ represents what Ralph believes in world w, what is presupposed to be believed by Ralph in context C is going to be $\bigcup\{Dox(r,w)|\ w \in C\}$. If a belief attribution to Ralph is made in context C, the embedded clause of the belief attribution has to be interpreted with respect to context $\bigcup\{Dox(r,w)|\ w \in C\}$.

[6] I don't want to claim, though, that my notion of relevance can do the whole job of selection of right contetual parameter.

But this means that according to our prior beliefs, represented by our prior probability function, proposition h has a higher probability than proposition h'. Because utilities are irrelevant here, this also means that the 'act' which has maximal expected utility is to go for hypothesis h. Now we learn that not proposition h, but proposition h' is true, and we want an explanation for this. The explanation is given partly by the belief attribution, and for the explanation to be sufficient it has to be the case that after accepting the attribution the posterior probability of proposition h' will be higher than the posterior probability of proposition h. But this can only be the case if after the update of the information state that we took to be Ralph's belief state with what is expressed by the embedded clause of (4), the most 'useful' act for Ralph to do would be arresting Pierre, instead of arresting Ortcutt. This, however, will only be the case when we interpret 'Ortcutt' as 'the noble English gentleman', and not when we interpret the name as 'Paderewski lifelong personal enemy'. But this also means that only under the first interpretation the belief attribution as a whole is going to explain for *us* the unexpected behavior of Ralph. In other worlds, only under one of the two interpretations the belief attribution is going to be *relevant*. And this is the reason that I, as a hearer, will conclude that the relevant interpretation used for 'Ortcutt' is going to be '*the noble English gentleman*'.

6. Implicatures

6.1. Two-way interaction context and utterance

In pragmatic theories the notion of context plays two roles: (i) it should contain enough information about the conversational situation to determine what is expressed by a context dependent utterance, and (ii) it should contain enough information about the presuppositions, beliefs and desires of the participants of the conversation to determine whether what is said by the speaker is appropriate or not. The central idea behind Stalnakerian pragmatics and recent theories of discourse is that there is a single notion of context that plays both of these roles, and that both kinds of information modeled by this single context change during a conversation in an interactive way. In the previous sections I have argued that (what is believed to be) the decision problem of a particular agent is a crucial contextual parameter to determine in decision-theoretic terms the utility of assertions, and that this measure can be used to determine what is expressed by declaratively used sentences. Thus the attitudes of the participants of a conversation can partly determine the content of what is expressed.

Central to Stalnakerian pragmatics is not only that context helps to determine what is expressed, but also that we might learn something (by means of *accommodation*) about the context due to the sentence being used. This two-way interaction holds in particular for relevance and decision problems; (i) the decision problem determines which of the possible interpretations of the sentence is most relevant, and thus which interpretation was probably meant, and (ii) a hearer who hears a sentence can sometimes only make the sentence relevant when he assumes that the context, i.e. the relevant decision problem, has specific properties, and thus concludes, by means of accommodation, that the context/decision problem indeed has those properties. The perhaps most familiar way in which a context is being accommodated to make the sentence being used appropriate in its context of interpretation is to account for *presuppositional* inferences. These inferences are normally triggered by particular expressions that for more or less *conventional* reasons can be interpreted appropriately only in particular kinds of contexts.[7] In the next subsection, however, I will consider cases where the context is being accommodated for a more general reason: without this accommodation what is said by the sentence would not be *appropriate* in its context of interpretation. Not being appropriate in its context of interpretation traditionally means that one of the Gricean maxims has been violated. In the next sections I will argue that for several cases that are normally treated separately, we can reduce the inappropriateness to the claim that what has been communicated violates the assumption that communicative acts should be *relevant.*

6.2. Conversational implicatures

In this section I want to show how we can account for particularized conversational implicatures by making use of decision theory. Particularized conversational implicatures arise, according to Grice (1989), whenever we may infer more from a sentence than its truth-conditional content, due to the particular conversational situation. But what, then, are the relevant contextual parameters? In this section I propose that the *decision problem* of one of the relevant agents is a prime example here.

Consider the following case: John is desperately in love with Mary. His acquaintance Bill gives a party tonight. John thinks that this party might well be his last chance of meeting Mary. He doesn't know, however, whether Mary will come, and is thus not sure whether he should come himself. Bill is well aware of John's state of mind, and John is aware of this. Pete, not John's best friend, knows about John's wondering whether he should go to the party tonight, but is not aware of the fact that his decision depends on whether Mary will come. Pete's own decision to come or not depends on whether John will come. In the midst of a discussion with Bill and John about the party, Pete asks Bill somewhat rudely whether he thinks John will come. Bill answers *Well, Mary will come.* Pete understands, and decides not to go to the party himself.

In the above example, the assertion that Mary will come to the party will intuitively *conversationally implicate* that John will come, too. This implicature can be explained formally by making use of decision problems, and assumed relevance of the assertion/answer.

In the example John's decision problem, i.e. whether to go to the party or not, was common knowledge among John, Bill and Pete. John and Bill, but not Pete, also know what kind of information would 'resolve' John's decision problem. That is, John and Bill, but not Pete, know that John will come to the party if he learns that Mary will. Pete assumes, however, that Bill's answer to his question will be relevant, and thus assumes that Bill's answer will be an indication whether John will come. Pete finds it more likely

[7] For this reason, these inferences are sometimes called *conventional implicatures*.

that the answer indicates that John will come than not. Pete concludes from Bill's assertion two things: (i) the missing premise that John will go to the party if he learns that Mary will, and (ii) that John will go to the party. The latter conclusion follows straightforwardly from the above assumptions, because he knows that John is one of the participants of the conversation who accepts Bill's answer. Notice that the real thing that John has to find out is what the missing premise is that would make Bill's answer relevant.

In our above example, John's decision problem can be thought of as the *yes/no*-question '*Will John come to the party?*', where the answers correspond one-to-one to the actions John is considering. Although Bill is not explicitly answering this question, the implicature can be derived by assuming that Bill's response implicitly answers the question. Many cases of so-called particularized conversational implicatures are like this, or so I claim. Consider, for instance, the example explicitly discussed by Grice (1989) in his William James lectures where A is standing by an obviously immobilized car and it is approached by B, after which the following exchange takes place:

(5) a. A: I am out of petrol.

b. B: There is a garage around the corner.

Grice notes that because B's remark can only be relevant in case the garage is open, A can conclude that this is something conversationally implicated by B. What is going on in this example, intuitively, is that B's remark answers the question '*Where can I get petrol?*' that was implicitly asked by A's assertion. This implicitly asked question, in turn, corresponds to A's decision problem: '*Where should I go to get some petrol?*' It is clear between A and B what A's implicit question and decision problem is, and it is also clear that B's remark is meant to resolve these issues. Because B's reaction can only resolve these issues when the garage is open, A understands that this is conversationally implied by B; otherwise the relevance of his assertion would be 0, i.e. his information would be pointless.[8]

The implicatures discussed in this section are inferences that arise from the assumption that what the speaker says is relevant, i.e. that he observes the Gricean maxims. The apparent violation is restored by accommodation of context. But Grice (1989) argued that implicatures also arise when the violation is genuine, where the maxims are *flouted* and the semantic meaning is really *irrelevant*. I have nothing to say about these kinds of implicatures.

6.3. Quantity impliatures

From a sentence of the form (6a) we intuitively infer that (6b) is not true:

(6) a. Max ate *some* of the cookies.

b. Max ate *all* of the cookies.

Grice (1989) argued that this inference should not be accounted for in terms of semantic entailment, because the inference is cancellable. Instead, he argued that the negation of (6b) is only conversationally implicated, and suggested that this could be accounted for in terms of his first maxim of *Quantity*:

Ceteris Paribus, make your contribution as informative as required for the current purposes of the exchange

Horn (1972) and Gazdar (1979) called implicatures that should be explained in terms of this maxim *scalar*, or *Quantity*, implicatures, and proposed a formal implementation. In their formal implementations, however, both neglected the phrase 'as required for the current purposes of the exchange'. Both assumed that the relevant propositions in the scale are just contextually given, and that the ordering relation is explained completely in terms of entailment. Horn (1972), for instance, argues that (6a) implicates the negation of (6b) due to the first maxim of Quantity, because (6b) is higher on the scale than (6a), which in turn is explained by the entailment (assuming that there are some cookies) from (6b) to (6a).

Hirschberg (1985) convincingly argued, however, that the scales relevant for Quantity implicatures cannot always be ordered by entailment, for it cannot explain why Wendy's answer (7b) to question (7a) of her potential future employer gives rise to the inference that (7c) is not true:

(7) a. Do you speak Portugese?

b. My husband does.

c. Wendy speaks Portugese.

This gives rise to the problem how propositions (or answers) could be ordered such that not only the standard Quantity implicatures could be explained, but also the ones discussed by Hirschberg.[9]

In section 3 of this paper I showed how we could determine the relevance of assertions in a *quantitative* way.[10] It is clear that once we quantitatively measure the value of propositions, this measure immediately gives rise to a (total) *ordering* relation. The propositions that have a higher value are considered to be more relevant. Notice that this ordering relation orders not only propositions that stand in

[8] In chapter 7 of his yet unpublished book, P. Parikh (ms) also accounts for particular conversational implicatures in terms of relevance defined in terms of decision problems. Still, there is an important difference between my analysis and his. My explanation of the implicature is from assumed relevance of assertion to context in which it is relevant. This explanation is very Gricean in nature; if the context would have been different, the Gricean maxim of relevance would have been violated. Parikh, on the other hand, argues that the implicature follows without a threat of violation of the maxim of relevance. So, for Parikh an implicature doesn't seem to be an inference from use of the sentence to context. I must admit, though, that I don't fully understand what is supposed to trigger the implicature on Parikh's analysis.

[9] And also Grice's own well known example of professor U's evaluation of Mr. X by saying *Mr. X has excellent handwriting* implicating that Mr. X is not suited for the philosophy job.

[10] I should note, though, that there are other quantitative ways of measuring the relevance, or usefulness, of assertions. See van Rooy (ms) for discussion.

an entailment relation to each other, and thus could in principle overcome the limitation Hirschberg pointed to. Indeed, I propose that our quantitative ordering relation is the relevant one for these so-called scalar implicatures.

Notice that the induced quantitative ordering relation between propositions depends on the relevant decision problem. Although it is not always obvious what the relevant decision problem is, sometimes it is pretty clear. In the exchange between Wendy and her potential future employer, for instance, the relevant decision problem is the one of the employer whether he should hire Wendy for the job, or not. The question makes clear that it would be useful for the job to speak Portugese, and if you don't, that you can call on the services of someone who does, and virtually gratis. So, Wendy's answer is relevant, although not as relevant as the answer *Yes* would have been. That's why we can conclude via Grice's maxim of Quantity that Wendy doesn't speak Portugese herself.

Notice that there exists an important difference between my above analyses of conversational and quantity implicatures, respectively. In distinction with the former case, for quantity implicatures we did not conclude from the utterance being used to specific properties the context must have to make the utterance relevant.

The above sketched analysis of Quantity implicatures closely resembles the Decision Theoretic one recently proposed by Merin (1997). On first look it might seem, however, that Merin does not determine the relevance of assertions in terms of decision problems. He rather accounts for them in terms of the *argumentative force* these assertions have for the participants in an argumentative dialogue. But first appearance is sometimes misleading. To account for his notion of 'argumentative force' in a formal way, Merin adopts the assumption that the two participants of a dialogue always argue for two mutually exclusive hypotheses. These two hypotheses form a *dichotomy*, and such a dichotomy is of course nothing else than a bi-partition, a decision problem which of the two hypotheses to accept. Thus, we might say that also Merin's analysis of Quantity implicatures makes implicit use of decision problems.

7. Conclusion and Outlook

In this paper I have argued that we could use decision theoretic tools to make some pragmatic inferences precise. In particular, I have shown how to define a notion of *relevance* in terms of the decision problem that one of the participants of the dialogue faces, and how this notion helps to determine what is expressed by *de re* belief attributions on the assumption that attitude attributions are in general made to explain some unexpected behavior of a relevant agent. Furthermore, I have sketched how some kinds of Gricean implicatures could be accounted for in terms of our notion of relevance, too. In van Rooy (1999, 2000) I used a similar analysis to account for pragmatic features related to questions. It remains to be seen whether my rather sketchy remarks related to implicatures can be turned into a predictive formal theory. Another main future concern will be to see in how far the work presented in this paper fits into the more general game-theoretical analysis of the semantics/pragmatics interface of Dekker & van Rooy (1999).

8. References

Dekker, P. & R. van Rooy (1999), "Optimality Theory and Game Theory: Some parallels", In: H. de Hoop & H. de Swart (eds.), *OTS^2; Papers on Optimality Theoretic Semantics*, OTS, Utrecht.

Fraassen, B. van (1977), "The pragmatics of explanation", *American Philosophical Quarterly*, 14, pp. 143-50.

Fraassen, B. van, (1979), "Propositional attitudes in weak pragmatics", *Studia Logica*, 76, pp. 365-374.

Gärdenfors, P. (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemics States*, MIT Press, Cambridge.

Grice, H.P. (1989), *Studies in the Way of Words*, Harvard University Press.

Gazdar, G. (1979), *Pragmatics*, London: Academic Press.

Hansson, S.O. (1995), "Change in preference", *Theory and Decision*, 38, pp. 1-28.

Hirschberg, J. (1985), *A theory of scalar implicature*, Ph.D. thesis, UPenn.

Horn. L. (1972), *The semantics of logical operators in English*, Ph.D. thesis, Yale University.

Jeffrey, R. (1965), *The Logic of Decision*, McGraw-Hill, New York, University of Chicago Press, Chicago.

Kaplan, D. (1969), "Quantifying in", In: D. Davidson and J. Hintikka (eds.), *Words and Objections, Essays on the work of W.V. Quine*, Dordrecht, pp. 178-214.

Merin, A. (1997), "Information, relevance, and social decisionmaking", In: L. Moss, J. Ginzburg, M. de Rijke (eds.), *Logic, Language, and Computation, Vol. 2*, Stanford.

Parikh, P. (ms), *The Use of Language*, manuscript, Stanford.

Quine, W.V. (1956), "Quantifiers and propositional attitudes", *The Journal of Philosophy*, 53, pp. 177-187.

Rooth, M. (1984), *Association with Focus*, Ph.D. thesis, University of Massachusetts, Amherst.

Rooy, R. van (1997), *Attitudes and Changing Contexts*, Ph.D. thesis, University of Stuttgart.

Rooy, R. van (1999), "Questioning to resolve decision problems", In: P. Dekker (ed.), *Proceedings of the Twelth Amsterdam Colloquium*, ILLC, Amsterdam.

Rooy, R. van (2000), "The pragmatics of mention-some readings of questions", manuscript, Amsterdam.

Rooy, R. van (ms), "Comparing Questions and Answers: A bit of Logic, a bit of Language, and some bits of Information", manuscript, Amsterdam.

Savage, L. (1954), *The Foundations of Statistics*, New York: Wiley.

Stalnaker, R. (1978), "Assertion", In: P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics*, pp. 315-332.

Stalnaker, R. (1988), "Belief attribution and context", In: R. Grimm and D. Merrill (eds.) , *Contents of Thought*, Tuscon, University of Arizona Press.

Acknowledgements

This paper was written as part of the 'Sources and Streams of Information' project, sponsored by the Dutch Organization for Scientific Research (NWO). It was written mainly during a visit at MIT. I would like to thank Maria Aloni and Robert Stalnaker for discussion.

The Nature of Common Ground Units: an Empirical Analysis Using Map Task Dialogues

Lesley Stirling*, Ilana Mushin*†, Janet Fletcher*, Roger Wales†

*University of Melbourne, Melbourne 3010, Australia
{l.stirling, i.mushin, j.fletcher}@linguistics.unimelb.edu.au
†La Trobe University, Melbourne 3083, Australia
r.wales@latrobe.edu.au

Abstract

This paper evaluates the meso-level 'Common Ground Units' (CGUs) proposed by Nakatani & Traum (1999) by applying this category to four dialogues from the Australian map task corpus. We first consider two formal characteristics of CGU boundaries: associations with following turn boundaries and associations with Intermediate and Intonational Phrases. We then give a profile of CGU initiating, grounding and final elements in terms of DAMSL dialogue act codes. We discuss some of the problems which arise in applying the category of CGU and conclude by proposing some parameters for consideration in a typology of CGUs.

1. Introduction

In this paper we consider the nature of 'Common Ground Units' (CGUs) as part of higher-level dialogue structure, by examining CGUs in four dialogues from the MAP TASK section of the Australian National Database of Spoken Language (ANDOSL). The work is part of a larger on-going study of the relationship between dialogue structure and prosodic structure, where we consider dialogue structure at different levels (micro-, meso- and macro-) and investigate whether and how dialogue segmentation is associated with various correlates of prosodic structure.

The ANDOSL Map Task (Millar et al., 1994) is closely modelled on the HCRC Map Task (Anderson et al., 1991). Participants worked in pairs, each with a map in front of them that the other could not see. One participant (the 'instruction-giver' IG) had a route marked on their map and was required by the task to instruct the other (the 'instruction-follower' IF) in drawing the correct route onto their own map. The maps differed to some degree in presence, position and names of landmarks. The four dialogues considered for this study were from two mixed-gender pairs, one pair who previously knew each other well and one pair who had never met before. Each pair produced one dialogue with the female as IG and one with the male.

'Grounding' is the process whereby information contributed by participants in a communicative interaction is mutually acknowledged as having entered the 'common ground', or shared knowledge of the participants (Clark & Marshall, 1981; Clark & Schaefer, 1989; Traum, 1994). This process takes place by virtue of a contribution being proposed by one participant and then evidence being given by the other that they have perceived and understood it: the evidence may be as minimal as 'proceeding as usual', may consist in a head-nod or other non-verbal cue, or may be an overt verbal acknowledgement or response. It is a dynamic process which can be modelled as a series of collaborative negotiations interspersed with 'moments' of grounding leading to change in the pragmatic and semantic status of the information considered to that point, as 'grounded' pragmatic and semantic information is added to the participants' assumed common ground (cf. Poesio & Traum, 1997).

Recently attention has been focused on the way in which the grounding process may interact with higher-level dialogue structure. Nakatani & Traum (1999) proposed a new coding scheme that applies simplified principles of grounding theory at the 'meso-' level of dialogue structure. 'Common Ground Units' (CGU's) are defined which they hypothesise might function as the minimal units for even higher-level dialogue units based on intentional/informational structure.

The status of CGUs is still under discussion (e.g. Core et al., 1999). In this paper we first give a formal profile of CGUs by examining how CGU beginning and end points correspond with measures of discontinuity in the speech signal such as turn boundaries and intonational phrase boundaries. We then give a functional profile of CGUs by examining the mapping between CGUs and dialogue acts labelled according to a version of DRI/DAMSL. These issues should contribute to our understanding the nature of CGUs, their status as dialogue units and their relationship to other kinds of dialogue structure.

2. Method

The dialogues were digitised for analysis at 22 kHz using Entropic's ESPS / Waves + speech analysis software running on a Sun Workstation in the Phonetics Laboratory of the University of Melbourne. A complete orthographic transcription of the dialogues was carried out.

As part of the larger study described above, a range of prosodic characteristics of the dialogues had been independently coded. Turn start and endpoints were labelled; in cases of speaker overlap, it was still possible to clearly mark a turn beginning and a turn end because the original ANDOSL recordings were dual channel files. The speech data were also annotated for Break Indices 3 and 4 (corresponding to intermediate and intonational phrase boundaries, respectively) according to the ToBI (Tones and Break Indices) prosodic transcription conventions for Australian English detailed in Fletcher and Harrington (1996).

The dialogues had also been independently coded for dialogue act using the 'Switchboard' version of DAMSL (SWBD-DAMSL) described in Jurafsky et al. (1997). Stirling et al. (Forthcoming) give more details and discussion of this coding. Since a major goal of the project was precisely to investigate associations between prosodic characteristics and functional discourse categories, we coded for discourse categories independently from dividing the speech signal into prosodic units: for example, dialogue act coding did not presume prior division of the speech into intonational units (cf. the discussion in Zollo & Core (1999)).

2.1. CGU coding

We followed Nakatani & Traum (1999) in our identification of CGUs in the four dialogues. Their CGUs are similar, but not identical, to Clark & Schaeffer's (1989) 'Contributions' and to the 'Discourse units' defined in Traum & Hinkelman (1992), Traum (1994). Like these other units of grounding, a CGU is considered to consist of all the linguistic material involved in achieving grounding of an initial contribution of information by one participant. Thus the prototypical CGU in our dialogues was an initial contribution by either leader or follower in the map task, and a response to this by the other participant which indicated that they had heard and understood this contribution (though not necessarily agreed with it). An example is given in (1).

(1) IG: just cross that river
IF: yes

Note that there is no requirement on CGUs to be informationally or intentionally coherent: providing that the information is grounded (or groundable) in the same way, it is considered to belong to the same CGU. Thus the initial contribution to the CGU may be complex in consisting of several dialogue acts of the same or different kinds, as long as they can all be grounded by the same response.

CGUs may also be complex in containing repair or clarification sequences which need to be negotiated prior to grounding being possible. While Clark & Schaeffer treated such sequences as embedded contributions, and therefore modelled grounding units as highly recursive tree structures, Nakatani & Traum avoid coding such complexity in terms of embedded CGUs, and we have followed this practice. An example of a CGU containing a repair sequence is given in (2).[1]

(2) IG: now come UNDER the Brownwood S~~
IF: [Branded Steers yep]
IG: [Steers just] about catching his tail
IF: right

Like earlier writers, however, Nakatani & Traum do allow for overlapping CGUs. This is a function of the fact that an utterance may simultaneously be used to ground (and conclude) an already open CGU, and to make a contribution which initiates a new CGU and itself requires grounding. An example is given in (3).

(3) IG: now from that cross can you see the Galah Open-cut Mine?
IF: I've got a Dingo Open-cut Mine *[grounds first CGU / initiates second* CGU*]*
IG: right

CGUs as defined by Nakatani & Traum have several additional interesting characteristics.

First, CGUs can be discontinuous if the same information treated in one grounding negotiation is revisited later (for further confirmation or repair, for example). We adopted Nakatani & Traum's heuristic of allowing later mentions to be included in a previous CGU only if no more than three CGUs intervened. Both in deciding whether to treat an interaction as a repair sequence within a CGU and in identifying discontinuous CGUs we found that it was a nontrivial matter in the map task dialogues to distinguish between negotiation over grounding of a contribution and negotiation at a deeper level of agreement or understanding. Some of the dialogues we examined did contain rather complex discussions of information concerning the position of landmarks and the direction of route segments being described by the leader; in some cases it was difficult to decide whether to handle these discussions as part of a discontinuous grounding negotiation or at a larger level of structure (such as Nakatani & Traum's IUs). We took a relatively conservative approach in these decisions and ended up with few instances of discontinuous CGUs.

Second, CGUs need not segment the speech signal exhaustively, since speech material which does not contribute to the grounding of information may be omitted from the coding of CGUs. This includes such material as clear false starts and other disfluencies, 'self-talk', and 'phatic' communion such as the establishment of the channel at the beginning of an interaction.

Material which is excluded from CGUs in this way can be distinguished from instances where one of the participants attempts to begin a grounding negotiation which is subsequently cancelled or dropped before the information is grounded: following Nakatani & Traum, abandonned CGUs of this kind were coded but starred, and in fact are excluded from the quantitative results presented below. Abandonned CGUs included cases where one participant began a contribution, which was then interrupted either by the other participant or by the speaker herself, and which were not resumed at least within the three CGU limit mentioned above.

One further issue requires elaboration. This concerns the criteria used for determining the end point for an open CGU and for recognising the beginning of a new CGU.

First note that grounding may consist in nonverbal signals such as head nods, smiles or laughter, or in the most minimal case simply continued attention. It is usually not possible to identify such signals in audio data, although we did in a very few cases count audible laughter as grounding a CGU, and in one case the context made it clear that a non-verbal signal must have been present to allow the dialogue to proceed. In all other cases, we expected the grounding element to be an utterance of some kind.

Second, Clark & Schaeffer argued that every utterance (indeed, every signal – p. 266) is a contribution which requires grounding, including minimal acknowledgements of other contributions. It was assumed that the kind of grounding evidence required in such cases would be the most minimal possible: continued attention and/or moving on to the next relevant contribution. Nakatani & Traum, following others such as Traum & Allen (1992), adopted an approach where minimal acknowledgements made without any new information being introduced were simply included in the CGU they grounded without themselves setting up a new CGU: this makes for considerably less complexity in the coding. We followed this approach.

[1] Square bracketting indicates overlap between utterances.

We also adopted the perhaps more controversial approach of including minimal answers to questions, especially positive answers to check questions, just in the CGU initiated by the question. We will return to this point later.

Third, we noted above that in some cases CGUs overlap, with the same utterance acting as a response and grounding element for one CGU and as the initiating element of the next. In some cases it is difficult to decide on what should be included in which CGU in such cases. We followed the principle that if we could distinguish a grounding element from a continuation which provided new information, we did so, and included just the grounding element in the first CGU. If this was not possible, we included just the first utterance of the new information chunk in both CGUs, on the principle that after the first utterance the former CGU was clearly grounded.

As indicated above, Nakatani & Traum advocate including in a CGU any material produced in the negotiation over grounding including questions designed to achieve clarity or understanding by the ultimate grounder. We followed this practice by making a distinction between clarifying questions which did not ask for extra information and those which did; the former were coded as part of the previous CGU and the latter were treated as both grounding the previous CGU and initiating a new CGU.

Finally, note that, as a number of other authors such as Clark & Schaffer (1989), Nakatani & Traum (1999) have pointed out, initiating contributions for CGUs need not be complete illocutionary acts. It was quite common in our dialogues for participants to cooperatively divide up complex material into small chunks, for example negotiating the grounding of a reference to a landmark or to the source for a route before going on to deal with more complex actions concerning it. An example is given in (4). We treated these units as separate CGUs, thus effectively treating virtually all 'backchannels' as grounding elements marking a CGU boundary, as long as the expression being grounded was in some way informationally complete. However note that this decision fails to address the issue of grounding by non-verbal signals such as head-nods, which may also occur throughout an utterance. (For a discussion of this issue see Core et al. (1999).)

(4) CGU1 IG: now from there
CGU1 IF: mhm
CGU2 IG: underneath that you should have
Consumer Trader Fair
is it there?
IF: yes I've got that
IG: oh right

Two coders independently labelled the dialogues for CGUs then resolved their differences to produce a consensual version.[2] CGUs were numbered and their beginning and end points were entered into a separate xwaves label tier for each participant in the dialogue: coding CGUs in xwaves label tiers in this way fails to adequately reflect the potential and actual complexity of some units, since although overlaps can be coded (by including material within the boundaries of two CGUs), discontinuities can be marked (by numbering the discontinuous chunks the same), and excluded material can in principle be indicated, the result is a complex set of codes which is difficult to compile for analysis. Nevertheless, we found it to be a useful way to associate the boundaries of CGUs with other discourse and prosodic phenomena.

While various divisions have been made of the internal structure of grounding units and the subfunctions of the elements within the unit (for example see Clark & Schaeffer (1989), Traum (1994)), we will be most concerned here with characteristics of the initiating element, the grounding element, and the final element.

3. Results

The total number of grounded CGUs in the corpus was 412. In what follows we give a profile of these grounded CGUs from a formal perspective, in considering turn taking and prosodic features of their boundaries, and from a functional perspective, in considering their dialogue act profile. In this way we can contribute to an evaluation of CGUs as defined above in terms of their validity as discourse units and the way in which they interact with other discourse phenomena.

In the map task domain the Instruction-Giver has a privileged position in terms of the information available to him/her. This is reflected in the fact that while leader and follower produce roughly the same number of turns – in fact, leaders produced 308 turns overall while followers produced 335 turns – the average turn length was shorter for the follower in each dialogue (averages overall were 6 seconds for leader turns (range 4 – 11.85 seconds) and 1.8 seconds for follower turns (range 1.38 – 2.67 seconds)). While it is virtually always the case that the grounding element in a CGU is produced by the participant who did not produce the initiating unit, deriving initiative data about grounding is complicated by the fact that in 13% of cases the final unit in a CGU was not the grounding element. Taking this into account, we did in fact observe a pattern in which leaders produced relatively more contributions to be grounded and followers produced relatively more grounding units: overall, 64% of the CGUs were grounded by the follower, and 36% by the leader, with little variation across dialogues.

3.1. Formal profile of CGUs: turn taking and prosodic characteristics

It is well-accepted that aspects of prosodic structure are often associated with discourse segment boundaries (for example see Grosz & Hirschberg, 1992; Nakatani et al., 1995; Hirschberg & Nakatani, 1996; Swerts, 1997). However most studies have focused on the interaction between various correlates of prosodic structure and discourse segmentation at the micro- (e.g. dialogue act) level. The question of how to calculate prosodic associations with larger level discourse units, especially when they may not be temporally sequential, is nontrivial and is an ongoing concern of our project. In this study we make a start by measuring the association between the end

[2] Intercoder reliability was not a focus of this study, however it is worth noting both that it is difficult to devise an appropriate measure of this for CGU coding, and that other work has suggested that CGU coding in general does not exhibit high degrees of intercoder agreement (cf. Core et al., 1999).

of the final utterance in the CGU and turn boundaries and Break Index.[3]

42% of grounded CGUs overall were followed by a turn boundary, that is, a change of speaker (range 26%-50% across the four dialogues). More interestingly, CGUs followed by turn boundaries tended to be grounded by IFs (76%, as compared with 64% of CGUs grounded by IFs overall, as mentioned above). This is because turn-bounded CGUs tended to be grounded by simple acknowledgements. These results are summarised in Table 1.[4]

CGUs followed by turn boundaries	Total for 4 dialogues
Grounded by IF, no turn boundary	131 (54.8%)
Grounded by IG, no turn boundary	108 (45.2%)
Total no turn boundary	239 (58%)
Grounded by IF, with turn boundary	132 (76.3%)
Grounded by IG, with turn boundary	41 (23.7%)
Total with turn boundary	173 (42%)
Total CGUs	412 (100%)

Table 1: Associations between endpoints of grounded CGUs, turn boundaries, and initiation of grounding

There was a much stronger association between CGU endpoints and Break Indices. 79.6% of CGU final boundaries coincided with a BI 4, or Intonational Phrase.[5] This is consistent with other unpublished work which has shown that association with BI 3 or 4 is a good indicator not just of dialogue act boundaries but also of the boundaries of larger-level units. These results are summarised in Table 2.[6]

BI value for final unit of CGU	Total for 4 dialogues
BI 4	328 (79.6%)
BI 3	62 (15%)
Neither BI 3 nor 4	22 (5.4%)
Total grounded CGUs	412 (100%)

Table 2: Associations between endpoints of grounded CGUs and Break Indices

The 22 cases where the CGU boundary did not coincide with a major prosodic unit were those where coders felt it was possible to make a functional division between an element which grounded the previous CGU (usually a yes/no or discourse particle such as *Okay, Allright*), but where speakers did not make an intonational break between this element and the following material.

We also found that a third of the time (in 33% of cases) the grounding units in CGUs overlapped with the contribution being grounded – in other words, the new speaker didn't wait until the previous speaker had finished their contribution before grounding it.

3.2. Functional profile of CGUs: dialogue act types

Recently, considerable interest has been expressed about the relationship between micro-level dialogue act coding such as DRI/DAMSL and grounding phenomena. For example Core et al. (1999) noted as questions for further consideration the relationship between backward-looking functions and grounding and the question of whether grounding units such as CGUs should be composed of dialogue acts as their minimal units; Nakatani & Traum (1999) on the other hand assume that both are separately coded from minimal units such as intonational phrases. Poesio & Traum (1998) define different kinds of conversational acts including both core speech acts and grounding acts, and divide the backward-looking functions of DRI/DAMSL between the two, with Understanding BFs as grounding acts. Zollo & Core (1999) describe a method of automatically extracting grounding features and identifying grounding units once a dialogue has been coded with the DRI backward- and forward-looking tags.

While it would be informative to consider the pattern of dialogue act coding across CGUs in some detail, space considerations preclude this here, so we will summarise our findings only.[7] Tables 3 and 4 show the distribution of CGU initial elements and CGU grounding elements between DAMSL forward-looking functions and backward-looking functions.[8]

[3] We initially also considered associations between CGU beginnings and ends and silent pause location, where a silent pause was defined as a break in the acoustic waveform of more than 150ms that was not part of a stop closure phase. However CGU boundaries were only randomly associated with silent pause location, which is consistent with findings reported elsewhere that silent pause location and duration are not reliable indicators of dialogue segment boundaries (cf. Stirling et al.,. Forthcoming) although they have been found to be in monologue (e.g. Swerts, 1997; Nakatani et al., 1995).

[4] In deciding whether the final element of a CGU was followed by a turn boundary we considered only material which had been coded as belonging to some CGU, whether grounded or abandonned.

[5] Endpoints were measured at the righmost boundary of the final word in the CGU, including the ends of utterances performing 'double' function as ending one CGU and beginning the next.

[6] A chi square analysis was done to check on the possible relation between the four dialogues and the sub-categories of BI analysis: the results of this analysis were chi square =4.7; df, 6; significance, 0.5. Thus there is no statistical relation between the BI categories and the different dialogues. As a consequence it is legitimate to pool the data across the four dialogues, which is the form of the data that is reported below.

[7] The full set of SWBD-DAMSL labels can be found in Jurafsky et al. (1997) or Stirling et al (forthcoming); they are very similar to the DRI/DAMSL labels of Allen & Core (1997). Here we mostly refer to the major top level categories of forward-looking and backward-looking function types: 'Statement', 'Information-request' and 'Action-directive' (actually subcategories of 'Influencing-addressee-future-action', 'Agreement', 'Understanding', and 'Answer'.

[8] As noted above, CGUs may be initiated by a turn consisting of more than one dialogue act: here we give figures for the first act in the CGU only. While as we shall see not all grounding elements were final elements in the CGU, at this level of generality

Dialogue act type	Total for 4 dialogues
FF	348 (84.5%)
BF	63 (15.3%)
Other	1 (0.2%)
Total	412 (100%)

Table 3: DAMSL Dialogue act type of initial elements of CGUs

Dialogue act type	Total for 4 dialogues
FF	59 (14.3%)
BF	347 (84.2%)
Other	6 (1.5%)
Total	412 (100%)

Table 4: DAMSL Dialog act type of grounding elements of CGUs

Unsurprisingly, CGUs were most likely to be initiated by FFs – 'Statements', 'Information-requests' or 'Action-directives' - and most likely to be grounded by BFs. 32% of grounding elements (n=132) were simple 'Acknowledgements'. Most of the BFs which initiated CGUs were those with a 'double' function (52% – chiefly extended answers to initiating questions) or were continuations of extended answers. Similarly, most cases of FFs grounding or concluding CGUs were also 'doubles' (71% – statements or questions made in response to an initiator's original contribution). There were 75 cases of 'double' function elements (thus, 150 CGUs or 36% began or ended with a shared element). Table 5 shows the breakdown of these units by dialogue act type.

Dialogue act type	Total for 4 dialogues
Statement	21
Information-request	20
Action-directive	3
Other FF	1
Agreement	6
Understanding	7
Answer	20
Total FF	42
Total BF	33
Total	75

Table 5: Dialogue act type of 'double' function elements

Zollo & Core (1999) identified as problematic for their automatic extraction method the class of cases where the grounding element also initiated a new contribution and therefore would not be coded with a BF: such cases would be missed as grounding elements. From our data this is a problem which would affect a sizeable 18% of CGUs. Had we followed Clark & Schaefer's practice of identifying a new grounding unit for each acknowledgement, grounded for the most part by the occurrence of a next relevant utterance, the proportion of grounding units which were forward-looking functions would have been much higher.

In 54 CGUs (13% of the total) the grounding element was not the final element in the CGU. Upon further analysis these cases were of three main kinds:

(i) The first case, which accounted for 26 of the 54, was an artefact of our decision to include simple responses, especially positive responses to check questions, just within the CGU initiated by the question: in 32% of cases these responses were followed by an acknowledgement by the CGU initiator.

(ii) The second type were cases where the grounding element was an acknowledgement in the form of a repetition. 10 out of 22 (45%) repetition-acknowledgements in the data were themselves followed by an acknowledgement by the CGU initiator.

(iii) Thirdly were cases where a simple 'Acknowledgement' or 'Accept' functioning as grounding element was itself followed by a further 'Acknowledgement' or 'Accept' from the CGU initiator.

Finally, CGUs did not necessarily preserve DAMSL dialogue act boundaries. This occurred in two cases. First, as mentioned in section 2, some dialogue acts were grounded in installments. Second, again because of the overlapping CGUs already discussed, what had been coded as a single dialogue act was in some cases divisible into a grounding unit followed by the initiating unit of a new CGU.

4. Discussion

A number of authors have proposed taxonomical distinctions according to which grounding units can be characterised. Clark & Schaefer (1989) listed different types of presentations in terms of how they corresponded to complete utterances or turns and different types of acceptances in terms of the kind of evidence they provide for grounding. Traum (1994) distinguished different types of 'grounding acts' according to the subfunctions of utterances within a grounding negotiation. Core et al. (1999) include a proposal for distinguishing CGUs in terms of three types of acknowledgement. The results we have presented are less concerned with the internal structure of CGUs but suggest two additional parameters along which CGUs can be distinguished:

1. Turn properties of final unit in CGU

- The final unit may be a complete turn by one speaker (in these cases the CGU normally ends with a BI 4)
- The final unit may be part of a turn, with a later part of the turn functioning to initiate the next CGU (in some cases the CGU then ends in a BI 3 or even less; there is a problem in deciding when you divide a turn into a CGU completion part and a CGU initiation part – sometimes it seems the turn can be divided functionally, but there is no division in intonation unit)
- The final unit may be part of a turn, where the segment cannot be divided into CGU completion and

the figures are essentially the same whether we consider grounding elements or final elements.

CGU initiation elements (these are the 'double' function cases; here the CGU may well end in a BI 4, but by virtue of the fact that it is a whole unit which is doing double duty)

2. Grounding properties of final unit in CGU

- The grounding element is also the final element (the simplest case)
- The grounding element is followed by further material from the initiating speaker, acknowledging / grounding the grounding element

One of the characteristics of our data was the occurrence of complexity in what Clark & Schaefer described as the 'acceptance' phase of a grounding unit: the part where grounding is negotiated, after the initial contribution of information under consideration. In addition to expected complexity where repair or clarification sequences occur, we found that complexes of grounding units or of multiple acknowledgements made by both participants without any new information being introduced occurred quite commonly. Clark & Schaefer (1989), Nakatani & Traum (1999) and Core et al. (1999) all note that it sometimes takes several acknowledgements back and forth by the two speakers to establish common ground sufficient for current purposes: where the requirement on mutual understanding is high, as in task-based interactions of the kind we consider, this may be particularly the case. The following are representative examples.

(5) is a case where there were simply a number of acknowledgements provided by the grounding participant.

(5) IG: now we're going to start you off on tour one
IF: thank you darling
allright we're starting on tour one

(6) illustrates the case where an acknowledgement was repeated after overlap during an initial grounding unit; this occurred very commonly in our data.

(6) IG: uh about an inch below the bottom branch of the
weste[rn tr]ee
IF: [yeh]
yeh

(7) is the case where a simple response by one speaker is followed by a simple acknowledgement by the other, which has a grounding function but perhaps also other interactive functions such as priming the listener for new information to come.

(7) IG: no you're above the Statistics Centre
IF: right
IG: allright

(8) illustrates the case mentioned in section 3.2 where a grounding element in the form of a repetition elicits further acknowledgement.

(8) IG: you head south
IF: um right head south
IG: yeh

And finally (9) is a case where there were multiple acknowledgements by both parties, perhaps reflecting a particularly difficult sequence of negotiations.

(9) IG:not that far
you're only supposed to go about a centimetre
IF: oh
IG:[yeh]
IF: [OK]
right
O[K]
IG: [O]K
IF: allright [so~~]
IG: [um]
IF: only a centimetre OK

5. Conclusion

The notion of grounding is well motivated but how well does it translate into formal definitions of units in dialogue which can be measured for associations with other dialogue phenomena? For instance, examples such as those in the previous section raise questions about the validity of attempting to identify a single 'grounding element' at which point the information under consideration is taken to enter the Common Ground. More generally, they raise the question for us of what is being measured by CGUs as defined here and in Nakatani & Traum (1999). The decision not to open new CGUs for 'simple' responses of various kinds in effect means that we are not looking at grounding per se here, but at something like the subset of 'more contentful' grounding units: this makes the meso-level units defined much easier to work with especially with large amounts of data, and more akin to other meso-level units such as adjacency pairs or games, but arguably misrepresents the grounding profile of the dialogue and its participants.

Nevertheless, we found that coding the map task dialogues for CGUs has enriched our understanding of what is going on in the dialogues and raised questions which further work will pursue. These will include the relation of grounding to higher-level discourse structure, and the development of a metric of complexity of CGUs to correlate with stylistic characteristics of particular dialogues, including degree of perceived dysfluency (evident in particular in dialogues between pairs who were previously unknown to one another).

6. References

Allen, J. & Core, M. (1997). Draft of DAMSL: Dialog Act Markup in Several Layers. Draft contribution for the Discourse Resource Initiative. Also available at: http://www.cs.rochester.edu/trains/research/annotation.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. & Weinert, R. (1991). The HCRC map task corpus. *Language and Speech*, 34, 351--366.

Clark, H. & Marshall, C. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of Discourse understanding* (pp. 10--63). Cambridge: Cambridge University Press.

Clark, H. & Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*,13, 259-294.

Core, M., Ishizaki, M., Moore, J., Nakatani, C., Reithinger, N., Traum, D. & Tutiya, S. (1999). The report of the Third Workshop of the Discourse Resource Initiative, Chiba University and Kazusa Academia Hall.

Fletcher, J. & Harrington, J. (1996). Timing of intonational events in Australian English. In P. McCormack & A. Russell (Eds.), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 611—615).

Grosz, B. & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing*, Banff (pp. 429--432).

Hirschberg, J. & Nakatani, C. (1996). A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of the 34th Annual Meeting of the ACL, Santa Cruz* (pp. 286—293).

Jurafsky, D., Schriberg, L. & Biasca, D. (1997). Switchboard SWBD-DAMSL Shallow-Discourse-Function-Annotation Coder's Manual, Draft 13. Technical Report TR 97-02, Institute for Cognitive Science, University of Colorado at Boulder. Also available at: http://stripe.colorado.edu/~jurafsky/manual.august1.html.

Millar, J., Vonwiller, J., Harrington, J. & Dermody, P. (1994). The Australian National Database of Spoken Language. In *Proc. ICASSP-94* (pp. 197—100).

Nakatani, C., Grosz, B., & Hirschberg, J. (1995). Discourse structure in spoken language: studies on speech corpora. In *Proceedings of the AAAI-95 Spring Symposium on Empirical Methods in Discourse Interpretation and Generation.*

Nakatani, C. & Traum, D. (1999). Coding discourse structure in dialogue (Version 1.0). University of Maryland Institute for Advanced Computer Studies Technical Report UMIACS-TR-99-03.

Poesio, M. & Traum, D. (1997). Conversational acts and discourse situations. *Computational Intelligence*, 13(3), 309--347.

Poesio, M. & Traum, D. (1998). Towards an axiomatizatoin of dialogue acts. In *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues* (pp. 207—222).

Stirling, L., Fletcher, J., Mushin, I. & Wales, R. (Forthcoming.) Representational issues in annotation: using the Australian map task corpus to relate prosody and discourse structure. *Speech Communication*.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America* 101(1), 514--521.

Traum, D. (1994). *A computational theory of grounding in natural language conversation*. PhD thesis, Department of Computer Science, University of Rochester. Also available as TR 545, Department of Computer Science, University of Rochester.

Traum, D. & Allen, J. (1992). A speech acts approach to grounding in conversation. In *Proceedings of the 2nd International Conference on Spoken Language Processing* (ICSLP-92) (pp. 137—140).

Traum, D. & Hinkelman, E. (1992). Conversation acts in task-oriented spoken dilaogue. *Computational Intelligence*, 8(3). Special issue on Non-literal Language.

Zollo, T. & Core. M. (1999). Automatically extracting grounding tags from BF tags. In *Proceedings of the 37th Meeting of the ACL*, Workshop on Standards and Tools for Discourse Tagging, College Park.

The Meaning of *And* in a Formal Theory of Discourse and Dialogue

Isabel Gómez Txurruka

Institute for Logic, Cognition, Language and Information
Villa Asunción
Apartado 220; 20080 Donostia-San Sebastián
Spain

isabel@uts.cc.utexas.edu

Abstract

In this paper we first try to show that, contrary to the Gricean tradition (Grice 1989, Blakemore and Carston 1999), *and* is not semantically vacuous and, contrary to temporal approaches such as Bar-Lev and Palacas 1980, *and* is not temporally loaded. We propose then to consider a commonsense intuition: while sentence juxtaposition can be interpreted either as discourse coordination or subordination, *and* only signals coordination. Segmented DRT or SDRT (Lascarides and Asher 1993, Asher 1993) already includes a notion of coordinating and subordinating Discourse Relations, and we relate the meaning of *and* to this distinction. Similar distinctions playing a crucial role in anaphora resolution have also appeared in AI's recent work—e.g., Grosz and Sidner 1986, Scha and Polanyi 1988, Webber 1991, Polanyi 1998, 1999, and Seville and Ramsay 1999. Since it has not been well defined yet, our approach could also be of interest in that it shows that the distinction is needed for independent reasons.

Thus the account presented here uses SDRT to treat *and* as a discourse marker signaling not a Discourse Relation but a class of Discourse Relations, namely Coordinators. Once the basic semantic contribution of *and* is isolated, effects related to its presence such as changes in temporal structure, DR blocking, and interaction with intonation, are shown to follow from the defeasible architecture set up by the formal discourse theory.

1. Previous Approaches

In *Studies in the way of words* Grice maintains that *and* and sentence juxtaposition have the same semantic content, both being equivalent to logical conjunction—remember that 'p&q' is true if and only if p is true and q is true. In his view, *and* does not add any particular semantic meaning to the interpretation (different from juxtaposition). Other researchers such as Schmerling 1975 and Posner 1978 have also maintained this position. But any proposal of semantic vacuity is faced with the problem of explaining the temporal effects that seem to show up when *and* is inserted in some contexts. Bar-Lev and Palacas' 1980 paper focused on this problem. Consider their crucial examples in (1) and (2):

1. a. Max fell; he broke his arm.
   b. Max fell and he broke his arm.

2. a. Max fell; he slipped on a banana peel.
   b. Max fell, and he slipped on a banana peel.

While in (1) the insertion of *and* does not change the temporal structure of the discourse, in (2) *and* blocks the possibility of inferring that the second event precedes the first. Bar-Lev and Palacas concluded that *and* must have a temporal import. They realized that the incompatibility of *and* with temporal reversal was quite systematic. They realized further that not only temporal succession, but also other temporal relations such as temporal overlapping or inclusion were permitted in the presence of the conjunction, and they concluded that the impossibility of temporal reversal was directly signaled by *and*. Their so-called semantic command captures this idea:

(Semantic command)
Given S' and S", S" is not prior to S' (chronologically or causally).

However, although this proposal makes the right temporal predictions for the examples above, it has an ad hoc flavor, which is confirmed when one realizes that *and* is also involved in discourse meaning variations when temporal structure is not at issue. Bar-Lev and Palacas were already aware of this fact, as illustrated by (3)-(4) from their paper:

3. a. Wars are breaking out all over; Champaign and Urbana have begun having skirmishes.
   b. Wars are breaking out all over and Champaign and Urbana have begun having skirmishes.

4. a. Language is rule-governed; it follows regular patterns.
   b. Language is rule-governed and it follows regular patterns.

Commenting on these examples, they pointed out that "*and* is mutually exclusive with other conjoining relationships, including exemplification, conclusivity, and explanation" (Op. cit.: 143). Our proposal in §3 that *and* blocks directly certain discourse relations is certainly reminiscent of this quote. We will show that, once the idea of incompatibility of *and* with certain discourse relations is developed in a formal theory of discourse, temporal effects can be shown to follow from it and, therefore, no independent temporal meaning associated with *and* is needed.

On the other hand, what Bar-Lev and Palacas did not see is that their proposal has not only unaccounted cases, such as (3) and (4), but also counterexamples. This is shown in Blakemore and Carston 1999. They present the example in (5) among others:

5. A: Did she do all of her education in the States?
   B: No, she did her BA in London and she did her A levels in Leeds.

The second event in (5B) occurs prior to the first event introduced in this sentence. Were *and* to indicate impossibility of temporal reversal, inconsistency would arise.

The same point can be made with an example originally due to Cohen (Cohen, 1971):

6. a. If the old king has died of a heart attack,
   and a republic has been formed,
   then Tom will be quite content.

   b. If the old king has died, and a republic has been formed,
   and the latter event has caused the former,
   then Tom will be quite content.

The events introduced in the antecedent of the conditional in (6a) are understood as forming a resultative discourse and, thus, as being temporally related by succession. Including a third conjunct in the antecedent, as in (6b), produces a revision in the discourse structure—namely, the relation between the dying of the king and the forming of a republic cannot be seen as cause-effect anymore. The temporal relation of succession has been cancelled. The third conjunct explicitly asserts that there is a reversed cause-effect relation and, thus, a reversed temporal relation is inferred. If *and* indicated impossibility of temporal reversal, inconsistency would have been reached again (and, obviously, this is an intuitively consistent piece of discourse).

2. Temporal Defaults Do Not Make the Right Predictions

Therefore, the semantic command has counterexamples. However, note that they could also be thought as exceptions. That is, one could think that the Impossibility of Temporal Reversal (ITR) normally works even though it can be cancelled sometimes. We are going to show that this is not the case. Assume for a contradiction that the semantic command is correct if it is taken as a default, as in (ITR) below (let the symbol '>' be read as 'then normally'):

(ITR) $<\pi \text{ and } \pi'> \; > \neg(e(\pi')<e(\pi))$

This rule, which has been stated using standard notation in SDRT, can be paraphrased as follows: if *and* is the D-marker linking two representations π and π', it normally follows that it is not the case that the event in π' precedes the event in π. Consider the banana peel example again, repeated below for convenience:

7. a. Max fell; he slipped on a banana peel.
   b. Max fell, and he slipped on a banana peel.

Let us try to explain why temporal reversal and explanation are no longer possibilities for interpreting (7b) using the default in (ITR). We assume that the interpreter of (7a) infers nonmonotonically first Explanation and, then, temporal reversal, using rules that can be roughly stated as follows:

$<\tau, \pi, \pi'> \;\&\; \text{Cause}(e(\pi),e(\pi')) > \text{Explanation}(\pi, \pi')$

$\text{Explanation}(\pi,\pi') \rightarrow (t(\pi')<t(\pi))$

The first rule says that, assuming we want to attach π' to π in the discourse model τ, if the event in π' causes the event in π, it normally follows that the Discourse Relation to attach π' to π is Explanation. The second rule introduces a temporal postcondition that follows from (this kind of) Explanation: the event in π' precedes temporally the event in π.

Suppose now that the interpreter infers $(t(\pi')<t(\pi))$ using these defaults. As we assumed that ITR holds, he will also infer $\neg(t(\pi')<t(\pi))$. Both meanings, $(t(\pi')<t(\pi))$ and $\neg(t(\pi')<t(\pi))$, have been inferred only nonmonotonically and thus none can be preferred over the other (see Asher and Morreau 1991). Either no one is concluded or both are. Therefore ITR cannot be used to predict the blocking of temporal reversal. We conclude that the putative meaning of impossibility of temporal reversal works neither as hard rule nor as default.

There is another temporal default that has been suggested in order to (partially) account for the effects of *and*. Blakemore and Carston's 1999 paper suggests that a rule of Temporal Precedence of Events (TPE) is used by default whenever two conjoined clauses containing events are interpreted out of the blue. Using the same notation as before, we can paraphrase TPE as follows:

(TPE) $<\pi \text{ and } \pi'> \; > (e(\pi)<e(\pi'))$

Following Blakemore and Carston, TPE would be responsible for the prevailing temporal readings below:

8. Mary put on her tutu and (she) pruned the apple tree.
9. Bill went to bed and took off his shoes.
10. She rode into the sunset and jumped on her horse.

We do not agree with this analysis. Note that all the examples above include the same entity as agent in both conjuncts. If we change this, as in (11) below, TPE does not hold anymore; temporal overlapping is strongly preferred:

11. Mary put on her tutu and Melissa pruned the apple tree.

Examples such as (11) allow us to conclude that conjoining two events is not enough to trigger TPE. On the other hand, remember that temporal reversal, as in (5), and temporal inclusion, as in (12) below, are also triggered easily in the right contexts:

12. Arantza played football yesterday and (she) broke her knee.

Thus we do not see any conclusive evidence to assume that *and* has a temporal default of temporal precedence associated. Winding up, (a) ITR does not allow to make the right temporal predictions, (b) there are not grounds to assume that *and* has a default of temporal iconicity associated, and (c) data shows that *and* is compatible with any temporal relation (precedence, reversal, overlapping and inclusion). Therefore, we assume that *and* conveys no temporal information. We need to find a new way to explain the following temporal facts: (i) the reason temporal succession is strongly preferred in Blakemore and Carston's examples (8)-(10), and (ii) the reason temporal reversal is blocked in the presence of *and* in the banana peel example in (2).

3. *And* signals Coordinators

We propose to treat *and* as a D-marker. *And* indicates not a particular Discourse Relation (DR), but a class of DRs including Narration, Result, Conditional, Parallel, and List. Let us call this class Coordinators. As a consequence, *and* blocks DRs such as Instance, Elaboration, Reformulation, Background or Explanation that belong to a complementary class that can be called Subordinators. In other words, we include the following axiom in our theory:

(and) $<\boldsymbol{\pi}, \boldsymbol{\pi}'> \& <\boldsymbol{\pi} \text{ and } \boldsymbol{\pi}'> \rightarrow \text{Coordinator}(\boldsymbol{\pi}, \boldsymbol{\pi}')$

This rule intuitively says that, if we are going to attach two representations π and π', and *and* is linking the clauses from which these representations have been built, the DR attaching π' to π must be a Coordinator. Notation must be explained. Boldface letters $\boldsymbol{\pi}, \boldsymbol{\pi}'$... indicate metavariables for π, π'... which stand for speech acts.

Following work on illocutionary acts, SDRT includes a distinction between propositional content and illocutionary force of an utterance. Take, e.g., an utterance of "I'm surprised that Marlow was killed." We distinguish between the propositional content—consisting of a referential expression such as Marlow and a predicate such as to be killed—and the illocutionary force, in this case of surprise. Searl 1968 introduces the term speech act, which includes illocutionary force and propositional content. SDRT assumes that DR predicates take speech acts as arguments and this assumption will prove to be crucial to account for examples such as (14) below.

We also assume that a particular Coordinator, List, is triggered by *and* by default.

(List) $<\boldsymbol{\pi}, \boldsymbol{\pi}'> \& <\boldsymbol{\pi} \text{ and } \boldsymbol{\pi}'> \; > \text{List}(\boldsymbol{\pi}, \boldsymbol{\pi}')$

List is a weak DR in that other Coordinators such as Narration or Result can defeat it (using Specificity given that their antecedent will be more specific; see Asher and Morreau 1991). The DR of List comes with a postcondition requiring a discourse topic for the *and*-segment.[1] The symbol '$\Downarrow$' stands for discourse-dominance. If $\gamma \Downarrow \pi$, then γ is a Discourse Topic (DT) for π. A DT is either explicitly given by the previous discourse or built by a generalization process:

(DT for List)
$\text{List}(\boldsymbol{\pi}, \boldsymbol{\pi}') \rightarrow \exists \gamma [\gamma \Downarrow \boldsymbol{\pi} \wedge \gamma \Downarrow \boldsymbol{\pi}' \wedge \neg(\boldsymbol{\pi} \Downarrow \boldsymbol{\pi}' \vee \boldsymbol{\pi}' \Downarrow \boldsymbol{\pi})]$

We apply this proposal to account for (i) and (ii) above, providing the right temporal structure for our examples. Moreover, we show that it also (iii) generates the correct predictions for the non-temporal cases, and (iv) provides the basis to account for more complex instances. In order to do so, we need to assume a basic hypothesis about the relationship between temporal structure and DRs: the interpreter relies on information sources such as lexical-K, semantic-K and WK to nonmonotonically conclude a particular DR, and it is only afterwards that the temporal relation is inferred. This directionality has been motivated in Lascarides and Asher 1993:

(Directionality)
Temporal Relations are inferred from Discourse Relations.

Our first task is then to explain the preference for temporal succession in the examples (8)-(10) on the basis of our proposal, showing that the assumption that *and* is associated to a default of temporal succession is spurious. Consider Blakemore and Carston's tutu example again, repeated below as (13):

13. Mary put on her tutu and (she) pruned the apple tree.

Temporal simultaneity, inclusion and overlapping are ruled out using WK, since an agent does not normally get involved in two different activities at the same time (unless explicitly indicated). Using (List), the interpreter could assume that no temporal relation holds between these events, since List is triggered by default in the presence of *and.* But he will not conclude List unless there is no other way to build a coherent whole. Thus the interpreter is led to consider temporal succession and temporal reversal. In order for temporal reversal to be triggered, the interpreter should assume Explanation, but this is not possible because Explanation is a Subordinator and thus systematically blocked by *and* (by our hypothesis). Thus the only temporal relation left is temporal succession, which is recovered using the DR of Narration, a Coordinator. If this kind of explanation makes sense, it allows the construction of temporal structure without using a temporal default. The same type of reasoning can be shown to work in example (9) and (10).

3.1. Blocking Discourse Relations

We apply our hypothesis to temporal reversal cases and non-temporal *and*-examples in this section. Remember that these were our second and third tasks. From our view there is no difference w.r.t. how the meaning of *and* affects those examples: they are all produced by what has been called elsewhere a meaning-blocking process (See Alves and Gómez Txurruka 2000). We restrict ourselves to particular DRs in order to show that *and* blocks (at least) Explanation, Instance, Reformulation, Elaboration, and Background. These incompatibilities are formalized as theorems in SDRT (they follow from (and) and our distinction between Coordinators and Subordinators), and it can be shown that the crucial examples are accounted for.

The theorem below says that Explanation is not allowed as a DR linking speech act $\boldsymbol{\pi}'$ to $\boldsymbol{\pi}$:

($\neg$Expl)
$<\boldsymbol{\pi}, \boldsymbol{\pi}'> \& <\boldsymbol{\pi} \text{ and } \boldsymbol{\pi}> \rightarrow \neg\text{Explanation}(\boldsymbol{\pi}, \boldsymbol{\pi}')$

Consider the banana peel example again:

14. a. Max fell; he slipped on a banana peel.
    b. Max fell, and he slipped on a banana peel.

[1] A very similar definition of common topic is given in Asher 1993 for Narration. All Coordinators might need to be associated to a discourse topic.

Explanation and temporal reversal are concluded in the absence of *and* in (14a) using WK. Inserting *and*, as in (14b), does not make much sense, at least out of the blue; two possible readings, a narrative and a time-free list, are both rejected. Let us introduce a context for (14b); we can consider a Cohen-sentence explicitly introducing explanation in a third conjunct:

15. If Max fell, and he slipped in a banana peel, and his slipping explains his falling, we don't need to resort to any esoteric or voodoo related explanation to understand what happened.

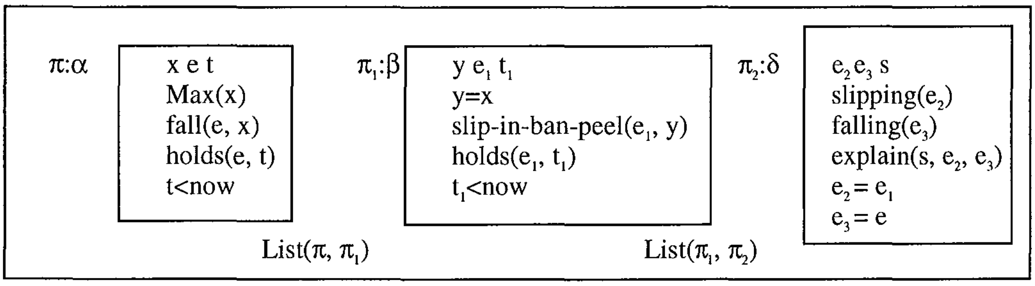


Figure 1: Discourse model for the antecedent in example (15)

The discourse model (informally) derived from the antecedent in (15) is given in Figure (1). Each clause produces a DRS, which comes with a name (α, β...) and an associated speech act (π, π'...). Anaphors are already solved. As we said above, List is triggered by default in the presence of *and*.

The crucial issue we need to address is whether the information derived from the third conjunct is consistent with the assumption that *and* blocks systematically the DR of Explanation. In other words, is the condition explain(e_2, e_3) inconsistent with ¬Explanation(π, π_1)? We defend that they are compatible using the distinction between speech acts and propositional contents drawn above. First note that the predicate "explain" takes events as arguments, while "Explanation", a DR predicate, takes speech acts, and thus they are not directly inconsistent. But inconsistency would arise if we could infer from one to the negation of the other.

Let us first look at the possibility of inferring Explanation(π, π_1) from the condition explains(e, e_1). To see whether this is possible, we need to know how Explanation, and DRs in general, are inferred. SDRT assumes the following general scheme to infer a DR:

<τ, π, π'> & some condition related to LEX-K, SEM-K, or WK > DR(π, π')

That is, suppose we want to attach π' to π in the discourse model τ; the inference of particular discourse relations depends normally on information coming from information sources such as Lexical Knowledge, Semantic Knowledge or World Knowledge.

Suppose now that the interpreter tries to revise the discourse relation of List(π, π_1) in Picture (1) given the new information introduced in π_2, explain(e_2, e_3). He will have to use an instance of a rule that may be roughly stated as follows:

<τ, π, π_1> & explain(e, e_1) > Explanation(π, π_1)

Note that all he gets is that normally Explanation(π, π_1). On the other hand, remember that he is already assuming ¬Explanation(π,π_1), which has been inferred using (¬Expl), and note that (¬Expl) is a hard rule. Thus, given that one rule is hard and the other soft, and that they lead to inconsistency if used together, the default is cancelled.

We also need to go the other way, and see whether it is possible to infer the condition ¬explain(e_2, e_3) from ¬Explanation(π, π_1). The question is whether the information about discourse structure that there is not a DR of Explanation between speech acts π and π_1 supports the information that the main event in π_1 does not explain the main event in π:

(?) ¬Explanation(π, π_1) $\models$ ¬explain(e_2, e_3)

It seems to us that the answer is that we cannot. There seems to be a difference between a speaker using *and* to assert that she is not linking two speech acts with Explanation (level of discourse structure), and a speaker asserting directly that two events are not linked by explanation. The first assertion seems to be weaker than the second. Note also that, if we allowed (?) to hold, we should also allow, by deduction theorem and contraposition, rule in (??):

(??) explains(e_2,e_3) → Explanation(π,π_1)

Rule in (??) is unacceptable since it allows the interpreter to go from a condition holding between events to a DR holding between speech acts without assuming that the interpreter is looking for a DR.

Winding up, Cohen sentences do not constitute a problem for assuming that *and* blocks systematically the DR of Explanation, if we introduce a distinction between speech act and propositional content in our discourse model.

The account proposed here also provides the basis to deal with complex examples involving interaction with anaphora resolution such as the following (Moxey and Sanford 1988; also discussed in Blakemore and Carston 1999):

16. a. Few MPs were at the meeting. They stayed at home and watch it on TV.
    a. Few MPs were at the meeting and they stayed at home and watch it on TV.

As Blakemore and Carston point out "while the pronoun *they* in (a) is best taken to refer to the members that were not at the meeting, this possibility is blocked in the corresponding *and*-conjunction case in (b), leaving only a nonsensical interpretation where the MPs are both at the meeting and have stayed at home." (footnote 5)

The second sentence in (16a) is constructed as an Explanation of the first. It is only in the frame of this Explanation that the pronoun *they* is allowed to refer to a non explicitly introduced referent (some kind of bridging?). When this DR is blocked in the presence of *and* (using (¬Expl)) this way of solving the anaphor is not acceptable anymore.

The same procedure used to explain the blocking of Explanation can be used to account for the blocking of Instance (see Olman 1998 for details about this DR). The following theorem captures this idea:

(¬Inst) $<\pi, \pi> \ \& \ <\pi \text{ and } \pi'> \rightarrow \neg\text{Instance}(\pi, \pi')$

Consider Bar-Lev and Palacas' 1980 example again, repeated below for convenience:

17. a. Wars are breaking out all over; Champaign and Urbana have begun having skirmishes.
    b. Wars are breaking out all over and Champaign and Urbana have begun having skirmishes.

In (17a) the interpreter concludes that the second clause constitutes an Instance of the first clause using roughly the following scheme:

$<\pi, \pi'> \ \& \ \pi$(concepts: all over, wars) &
π'(subconcepts: Champaign and Urbana, skirmishes) $>$ Instance(π, π')

On the other hand, the presence of *and* in (17b) allows to monotonically conclude $\neg\text{Instance}(\pi, \pi')$ using (¬Inst), thus blocking the nonmonotonic rule used in (17a). The interpreter still needs a DR to construct a coherent discourse model for (17b) and uses (List) to do so.

The systematic blocking of Elaboration[2] is assumed in the theorem below:

(¬Elab)
$<\pi, \pi'> \ \& \ <\pi \text{ and } \pi'> \rightarrow \neg\text{Elaboration}(\pi, \pi')$

The example in (18) has been taken from Blakemore and Carston's 1999 paper (originally due to Wilson):

18. a. I had a great meal last week. I went to Burger King.
    b. I had a great meal last week and I went to Burger King.

The relation object-attribute allows the interpreter to conclude Elaboration by default in (18a):

$<\tau, \pi, \pi'> \ \& \ \pi$(object: meal) & π'(attribute: in Burger King) $>$ Elaboration (π, π')

From (this kind of) Elaboration the interpreter concludes a relation of part-of between the relevant events (and temporal inclusion can be recovered from it):

$\text{Elaboration}(\pi, \pi') \rightarrow e(\pi') \subseteq e(\pi)$

On the other hand, the conclusion of Elaboration is blocked in (18b) using (¬Elab). The interpreter needs a DR and uses (List). Concluding List agrees with our intuitions that there are no temporal conditions involved in (18b).

The example in (19) is probably more complex and perhaps more controversial than (18). But I believe that (19b) cannot be used to convey an Elaboration. Note that Elaboration needs the first clause to be discourse topic for the second. This seems to be impossible in the presence of *and.*

19. a. I went to London. I stayed in the Meridian.
    b. I went to London and I stayed in the Meridian.

The relation part-whole indicates nonmonotonically Elaboration in (19a):

$<\tau, \pi, \pi'> \ \& \ \pi$(whole: trip) & π'(part: stay in a hotel) $>$ Elaboration (π, π')

Elaboration comes with a part-of relation between events as postcondition (and again, temporal inclusion is inferred from this relation):

$\text{Elaboration}(\pi, \pi') \rightarrow e(\pi') \subseteq e(\pi)$

On the other hand, our account predicts that Elaboration is blocked in (19b) in the presence of *and,* using (¬Elab).

The presence of *and* seems to coerce the event of going to London: while it is understood in (19a) as a whole trip involving going, staying, and possibly coming back, in (19b) it is interpreted more strictly as a movement verb. The interpreter looks for a stronger DR than List and finds Narration:

$<\pi, \pi'> \ \& \ <\pi \text{ and } \pi'> \ \& \ \pi$(e-going to London(s)) &
π'(e'-staying in a hotel(s)) $>$ Narration(π, π')

From Narration he gets a temporal postcondition:

$\text{Narration}(\pi, \pi') \rightarrow e(\pi) < e'(\pi')$

And also seems to block the DR of Reformulation—which has been considered a subtype of Elaboration in theories such as RST (Mann and Thompson 1986)—as in example (4). Thus Reformulation should also be included among the Subordinators. The following is also a theorem in our theory:

(¬Reform)
$<\pi, \pi'> \ \& \ <\pi \text{ and } \pi'> \rightarrow \neg\text{Reformulation}(\pi, \pi')$

The last Subordinator we consider in this section is Background, which is also predicted to be incompatible with *and* in this account:

(¬Backgr)
$<\pi, \pi'> \ \& \ <\pi \text{ and } \pi'> \rightarrow \neg\text{Background}(\pi, \pi')$

The example in (20) has also been taken from Blakemore and Carston 1999 (originally in Dowty 1986):

20. a. He walked into the room. The director was slumped in her chair.
    b. He walked into the room and the director was slumped in her chair.

[2] See Asher 1993 for details about Elaboration.

Following Asher and Lascarides 1993, the presence of an event and a state in (20a) triggers Background by default:

<τ, π, π'> & π(event) & π'(state) > Background(π, π')

This default is blocked in (20b) using (¬Backgr). Next, the interpreter looks for a Coordinator to attach the representation of the second clause to the representation of the first. Note that (20b) is intuitively understood as a Narration. But Narration is triggered in SDRT only if there are two events and here we have one event and one state. There are probably several ways to solve this puzzle. One of them could be to treat the second clause as a perceptual report of the speaker *seeing* that the director was slumped in her chair. This treatment would give us the two needed events:

<π,π'> & <π and π'> & π(e-walk into the room(x)) & π'(s), where s is taken to be a perceptual report, i.e., x saw s; thus s is raised to an event of seeing s > Narration(π,π')

We conclude that *and* systematically blocks several DRs. These blockings are produced to avoid inconsistency. Two inferred meanings are inconsistent. Given that the meaning of *and* is crucially monotonic, it is able to cancel the other meaning, which had been only nonmonotonically concluded. As a consequence of DR blockings, temporal structure, if involved, can change. However, *and* affects temporal information only in an indirect way. This result agrees with the view about DRs and temporal structure developed in SDRT. More complex *and*-examples should also be taken into account, and we revisit some of Blakemore and Carston's 1999 interesting examples introducing pitch accents in the next section.

3.2. Pitch accents and temporal reversal

Consider again Blakemore and Carston's 1999 example, which was presented in §1 as counterexample of the temporal approach defended by Bar-Lev and Palacas:

21. A: Did she do all of her education in the States?
    B: No, she did her BA in London and she did her A levels in Leeds.

The *and*-sentence in (21B) is a List, explicitly introduced by A's question, and normally concluded using (List). Temporal reversal, i.e., the information that the second event happened before the first, can be inferred at any moment, using WK—even though, contrary to what we will see in the next example, it is not an inference asked by the speaker to be performed. This inference, even if triggered, will not clash with the discourse structure built so far since List is not incompatible with temporal connections being triggered using other information sources. Thus this example is perfectly accommodated by our account. It has been presented only as a first step to account for more complicated examples involving fall-rise pitch accents and temporal reversal. Before presenting these examples let us try to understand a little better the meaning of the fall-rise accent.

Consider the example in (22), from Asher 1998 (originally in Walker 1995). Small capitals indicate fall-rise accent:

22. A: We bought these pajamas in New Orleans for me.
    B: We bought these pajamas in NEW ORLEANS.

Following Asher 1998 this example is a Correction. But the way used to correct is quite remarkable in that the corrected constituent is left out, without being replaced with a different content in the same argument role. The Correction is conveyed as an implicature—an implicature saying that the pajamas were not for A. Intonation seems to be crucial in that it conveys that "what I'm saying is all that can be truly asserted but I encourage you to draw implicatures."

Another well-known example with the same accent is found in Grice 1989:

23 A: Smith doesn't seem to have a girlfriend these days.
   B: He has been paying a lot of visits to New York LATELY.

B's conversational turn could be paraphrased as follows: Peter has been paying a lot of visits to NY lately, and this is all I can assert in relation to your question; but, given the fall-rise intonation, you should draw some implicatures. Thus, B can infer that Peter has a girlfriend in NY as an implicature.

These two examples work in a similar way in the following sense. In both fall-rise intonation seems to convey the following two meanings: (i) what I'm asserting is all I know for a fact, and (ii) you are encouraged to draw implicatures taking as premises what you said and what I replied.

Let us see whether using this informal interpretation of fall-rise intonation and our hypothesis that *and* signals discourse Coordinators we can account for the following example (Larry Horn; also discussed in Blakemore and Carston 1999):

24. A: Did John break the vase?
    B: WELL / the VASE BROKE / and HE dropped it.

An informal interpretation of B's contribution would be along the following lines: B asserts that these two events happened, and she encourages participant A to draw a relevant implicature relative to her question, namely, that John did break the vase. Note that what is apparently puzzling in this example is that A understands that the second event caused the first and that there is temporal reversal. We are now in a position to account for it.

B gives several linguistic clues to help A building this implicature. In particular, besides fall-rise intonation, she uses the discourse marker *well*. The particle *well* signals that what follows is not directly related to the previous conversation turn. In this case, it indicates that what follows is an indirect answer (see Carlson 1994 for details about this discourse particle).

If our hypothesis is right, *and* indicates a Coordinator in this example, and by default, List. A formal representation for the *and*-sentence is given in Figure (2) (assume John(v) is a condition and v an individual D-referent in the DRS for the question posed by A).

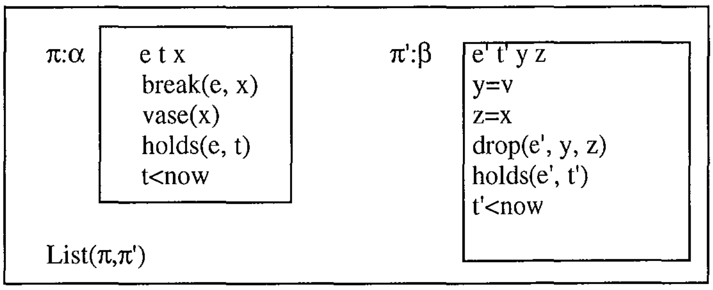


Figure 2: Discourse model for (24B)

Using the linguistic clues—mainly, content, intonation, and the discourse marker *well*—the interpreter extracts the implicature that speaker B probably believes that John did break the vase although she is not in a position to confirm it.

This information means that speaker B believes in fact that the dropping of the vase caused its breaking. Thus, the interpreter could infer from this information that B believes that the event of dropping explains why the vase broke. Note that we were faced with a similar situation when treating the Cohen sentence in (15). We discussed an apparent inconsistency there having to do with the condition explain(e, e_1) being included in a DRS and we concluded that it was not inconsistent with the meaning of *and* (in particular, with ¬Explanation). Similarly in this case, we can recover the same kind of condition (although this time as an inference from an implicature) and one could think that the meaning of *and* is inconsistent with it.

As before, we defend that these two pieces of information are consistent with each other because both predicates take different types of objects as arguments. The DR of Explanation occurs at the level of discourse structure, that is, speaker B asserts that speech acts π and π' are not related by the DR of Explanation in the discourse structure she is built (which is in accordance with our intuitions). On the other hand, the implicature explains(e(π), e'(π')) is not about the discourse structure that speaker B intends to communicate; it is a belief attributed to B about a relation between events in the world. As discussed before, we cannot derive an inconsistency from these two pieces of information.

Although the DR of Explanation is blocked by (¬Expl), the implicature that B believes that the dropping explains the breaking can still be in force because it is not dependent on this DR, it depends on the context and the linguistics clues provided by B. It is similar to the presence of temporal information such as overlapping or inclusion in some contexts. We have seen that these temporal relations are blocked by *and* when they are postconditions of Background and Elaboration respectively. But they are allowed when they follow from a Coordinator such as, for example, Result in example (12). Thus this analysis shows that our hypothesis that *and* blocks the DR of Explanation is compatible with the implicature that the second event explains the first event,

On the other hand, note that, as pointed out in Blakemore and Carston 1999, the order of the conjuncts is meaningful. From our view it is crucial to be able to infer List, since the reversed order would allow the interpreter to normally infer a stronger Coordinator such as Result (we ignore intonation):

25. B': Well, he dropped the vase and it broke.

Assume the interpreter were able to normally infer Result using the following default:

<π,π'> & <π and π'> & Cause(e(π), e'(π')) > Result(π,π')

Note that fall-rise intonation would not make any sense since the implicature would be already given.

Another example due to Blakemore and Carston, which also makes use of a complex marked intonation, is repeated below. Speaker B asserts that an event of John breaking his leg and an event of John tripping on a Persian rug both happened. However, given the context, participant A understands that the second event caused the first, and thus it explains how John broke his leg. Is this example consistent with our assumption that *and* signals Coordinators and thus Explanation is not possible?

26. A: Bob wants me to get rid of these mats. He says he trips over them all the time. Still, I don't suppose he'll break his neck.

B: Well, I don't know. JOHN / broke his LEG / and HE / tripped on a PERsian RUG.

We think so. Let us first look closer to the informational structure. Focus/Background structure on the predicates indicate that x broke his y and x tripped on z are Background, that is, discourse-old information. Moreover, contrastive topic accent on John signals that this topic is contrasting with an already introduced entity.

In A's conversational turn, A is assuming in her comment that people do not break anything by tripping over mats. This assumption is precisely the target of B's contribution, as signaled by the information structure. B disagrees with it. She begins her turn with the D-marker *well*, indicating that what comes is relevant to the discourse topic (acceptance of DT, in Carlson's 1994 terms), although only indirectly—implicatures are needed.

Applying our hypothesis about the meaning of *and*, the interpreter normally infers that B is communicating List(π, π'). He uses contextual information to infer the implicature that the event in the second conjunct explains the event in the first. The interpreter thus is faced with the following pieces of information:

Implicature:	explains(e(π), e'(π'))
Nonmonotonic inference:	List(π,π')
Hard inference:	¬Explanation(π,π')

And as defended before they are consistent. Therefore, this example shows that a conjoined utterance can inherit causal meaning (and temporal reversal) from context using appropriate intonation. Our hypothesis about the meaning of *and* has proven to be able to handle this kind of example, together with assumptions about the distinction between speech acts and events in the world.[3]

Acknowledgments

I would like to warmly thank Nicholas Asher and Ana Alves who helped me to start thinking about the meaning of *and* and provided many helpful lines of discussion while I was working with them in Austin. Thanks also to the members of the ILCLI seminar on NL Semantics and Pragmatics at Donostia who made relevant comments and contributed to the development of this paper. I am very grateful to the Department of Education, Research and University of the Basque Country for generous support through *Beca Post-Doctoral en el Extranjero* and *Beca de Incorporación de Doctores/as a la CAV*.

References

Alves, Ana and Isabel Txurruka, 2000: The meaning of *same* in anaphoric temporal adverbials. To appear in Bras, M. and L. Vieu (eds.), *Semantic and Pragmatic Issues in Discourse and Dialogue*. Elsevier.

Asher, Nicholas, 1993*: Reference to abstract objects in discourse*. Kluwer Academic Press.

Asher, Nicholas and Alex Lascarides, 1998: The Semantics and Pragmatics of Presupposition. Draft.

Asher, Nicholas and Michael Morreau, 1991: Common sense entailment: a modal theory of nonmonotonic reasoning. In *Proceedings of the 12th International Joint conference on AI*. Sidney, Australia.

Blakemore, Diane and Robyn Carston, 1999: The Pragmatics of *and*-Conjunctions: the Non-Narrative Cases. Draft.

Carlson, Lauri, 1994: Well *in dialogue games*. Amsterdam/Philadelphia: Benjamins Publishing Company.

Cohen, L.J., 1971: The logical particles of natural language. In Y. Bar-Hillel (ed.), *Pragmatics of natural languages*. 50-68. Dordrecht: Reidel.

Dowty. D., 1986: The effect of aspectual class on the temporal structure of discourse. *Linguistics and Philosophy* 9, 37-61.

Grice, H.P., 1989: *Studies in the way of words*. Cambridge, Massachusetts: Harvard University Press.

Grosz, Barbara and Candy Sidner, 1986: Attention, intention and the structure of discourse. *Computational Linguistics* 12: 175-204.

Lascarides, Alex and Nicholas Asher, 1993: Temporal Interpretation, Discourse Relations and Commonsense Entailment. In *Linguistics and Philosophy* 14.

Mann, Bill and Sandra Thompson, 1986: Relational propositions in discourse. Technical Report RR-83-115, Information Sciences Institute, CA: Marina del Rey.

Moxey, L. and A. Sanford, 1988: Quantifiers and Focus. *Journal of Semantics* 5, 189-206.

Olman, Lynda, 1998: Evidence for iconicity: The instance relation in informational exposition. University of Texas at Austin: MA thesis.

Polanyi, Livia, 1998: A formal model of the structure of discourse. *Journal of Pragmatics* 12: 601-638.

Polanyi, Livia, 1999: Linguistic Discourse Structure. Draft.

Posner, R., 1978: Semantics and Pragmatics of sentence connectives in natural language. F. Kiefer, J. Searle (eds.), *Speech Acts and Pragmatics*. Amsterdam: North Holland.

Schmerling, S., 1975: Asymmetric Conjunction and Rules of Conversation. In P. Cole, J.L. Morgan (eds.), *Syntax and Semantics*, 3: *Speech Acts*, 211-232. New York: Academic Press.

Searle, John, 1968: *Speech Acts*. Cambridge University Press.

Seville, Helen and Allan Ramsay, 1999: Reference-Based Discourse Structure for Reference Resolution. Manchester, England: Center for Computational Linguistics. Draft.

Walker, Marylyn, 1995: Corrections, AAAI Symposium on Computational Implicature, CA: Stanford.

Webber, Bonnie Lynn, 1991: Structure and Ostension in the Interpretation of Discourse Deixis. *Language and Cognitive Processes* 6(2): 197-135.

[3] A more substantial distinction should probably be drawn. DRs such as List seem to operate on illocutionaty forces. We said that illocutionary forces are part of speech acts and, as such, they are included in our speech acts constants. On the other hand, DRs such as Narration or Result seem to take something closer to propositional contents as arguments.

Discourse Particles as Speech Act Markers

Henk Zeevat

Computational Linguistics, FdG
Spuistraat 134, 1012VB Amsterdam, NL
henk.zeevat@hum.uva.nl

Abstract

A number of discourse particles are analysed to explain the role they can play in marking speech acts. The analysis uses an optimality theoretic reconstruction of presupposition theory.

1. Introduction

When one tries to further develop Stalnaker's ideas (Stalnaker, 1978) on the conditions for pragmatically correct assertion (informativity and consistency with respect to the common ground between speaker and hearer), it is natural to come up with conditions like the following[1].

(1) a. it is not common ground that the speaker believes A.
b. it is not common ground that the speaker believes that not A.
c. it is not common ground that the hearer believes A.
d. it is not common ground that the hearer believes that not A.

In all these cases, the assertion is improper, or non-standard. In the first case there is little to no effect that the speaker can hope to gain by what she has said: it cannot be a proposal to eliminate possibilities from the common ground. In the second case the speaker is self-correcting, and so faces an inconsistency in her own beliefs as represented in the common ground. In the third case, the speaker is also doing something that is not an assertion in Stalnaker's sense: she is at best assenting to a assertion by the hearer. In the fourth case as well, the speaker is correcting the hearer rather than asserting something.

These theoretical speculations are confirmed by looking at dutch or german sentences that realise such non-standard assertions: they invariably contain discourse particles, like *toch* (*doch*), *inderdaad* (*tatsaechlich*), *immers* (*ja*), *wel* (*doch*). The following examples bear this out. (a) can be a self-correction, (b) an assent to the hearer, (c) a reiteration, (d) a hearer correction.

(2) a. Peter is toch thuis.
a'. Peter ist doch zuhause.
a". Peter is at home (after all?).
b. Peter is inderdaad thuis.
b'. Peter ist tatsaechlich zuhause.
b". Peter is indeed at home.
c. Peter is immers thuis.
c'. Peter ist ja zuhause.
c". As you know, Peter is at home.
d. Peter is wel thuis.
d'. Peter ist doch zuhause.
d". Peter IS at home.

It is important to make the following observations. In contexts for (2a) in which the common ground contains the speaker's opinion that Peter is not home omitting the *toch* makes the utterance infelicitous. Likewise (2b) without the *inderdaad* is infelictous if, according to the common ground, it is already the hearer's opinion that Peter is at home. (2c) without *immers* is infelicitous if it is common ground that Peter is home and xx-1d without *wel* is infelicitous if the hearer has just said that Peter is not at home. This is indeed just what follows from Stalnaker's views on assertion. The particles seem to have the power to make otherwise infelicitous assertions into specialised non-standard assertions that have other goals than standard asssertions, like correcting opinions expressed earlier on or reconfirming established opinions.

An initial hypothesis might be that the particles are in the language just to mark the non-standard character of certain speech acts. But this hypothesis is easily refuted. If this were so, it would not be possible to combine all four particles as in (3), which, though not easy to contextualise, is nevertheless perfectly acceptable Dutch.

(3) Peter is toch immers inderdaad wel thuis.

It follows minimally that the particles do not mark a particular combination of speaker and hearer commitments to the truth or falsity of the proposition, because that combination would be inconsistent. The hypothesis also has to go when one considers the full uses of the particles in question in Dutch or German, as we will later on. And finally, it turns out that although the particles may indicate a combination of speaker and hearer commitments, they also allow other interpretations.

This raises two questions. First, how is it possible that the particles can mark deviant speech acts, i.e. one would like to have an account of their use from which it follows that they can sometimes mark a hearer or speaker commitment? Second, can these insights be used to improve the recognition of the user intention in dialogue systems. In addition, the function of these particles is unclear and any elucidation is welcome.

This paper gives an experimental account of these four particles in terms of an extended presupposition theory and manages to explain the uses quoted in this introduction. It follows that there is a potential use of the particles in future dialogue systems, i.e. the ones that have a capacity for presupposition treatment. Section 2 introduces the presuppo-

[1] For a full discussion of these conditions see (Zeevat, 1997)

sitional treatment of particles and sections 3, 4 and 5 apply the treatment to the four particles in question.

2. The proper treatment of the particle *too*

Kripke's notes on presupposition (Kripke, s.d.) started a new period in the study of presupposition where the analogy with anaphora became more and more prominent. The two most successful accounts are (Heim, 1983) and (Van der Sandt, 1992). Yet, in terms of Kripke's original example these theories do not perform very well at all.

Kripke is puzzled by the example (4).

(4) John will have dinner in New York too.

The traditional theories predict that this sentence presupposes (5) which for (4) is a mere triviality.

(5) Someone other than John will have dinner in New York.

After all, New York is a vast city where millions have dinner every night. If this were the presupposition, the *too* would not give us extra information about the context. It would also be the case that we can always add a *too* to the sentence *John has dinner in New York.* Both of these predictions are wrong: *too* is infelicitous if the common ground does not entail that another person has dinner in New York and it gives us the information that the common ground has this property. Kripke's suggestion is that *too* tells us that the context and not the world contains another person who has dinner in New York and tha the *too* is anaphoric to this part of the linguistic context.

Both Heim's and Van der Sandt's theories contain a resolution mechanism that can pick up the antecedent in the context (in that case the *too* does not give new information). But they also allow the presupposition to be accommodated. In that case, we get precisely the prediction that Kripke criticises, i.e. the requirement of an unidentified other person who has dinner in New York. The theories should rule out accommodation for *too*, but do not have the means to do that. In this way, the theories also predict that *too* can be freely added to our example, without truth-value change or infelicity.

There are some other aspects of *too* in which it is different from standard presupposition triggers, like factive verbs, definite descriptions and lexical preuppositions. The first is that *too* itself does not seem to give information. The following example of Heim brings this out. Two kids are secretly phoning each other after bedtime without the permission or knowledge of their parents.

(6) A: My parents think I am in bed.
B: My parents think I am in bed too.

In one of the interpretations of the utterance by B, the *too* belongs to the complement of the belief sentence. Yet, B's parents know nothing about A being in bed or not. The example also illustrates another problem with *too*. *Too* (and other particles) take antecedents that are not available according to Heim or Van der Sandt. The antecedent *A is in bed* in (6) is not entailed under the operator *B's parents think* and neither is it accessible according to the Discourse Representation Theory in which Van der Sandt's theory is couched. The last property of *too* that is unexplained by the two theories is that its occurrence is obligatory in the sense that in most of the utterances in which it occurs it cannot be omitted without resulting infelicity.

My proposal (Zeevat, 2000) is to (a) liberalise the set of allowed antecedents for presupposition triggers to the veridical contexts and to (b) assume a generation constraints. (c) Embedding the theory within a form of Bidirectional Optimality Theory then allows an explanation of the absence of accommodation for *too* and other presupposition triggers. I will sketch the three steps.

Veridical contexts were proposed by (Giannakidou, 1998) as a characterisation of the contexts that do not license negative polarity items[2] and include beliefs, dreams, suggestions, possibilities and iterations of these. Properly inaccessible antecedents (and negative polarity items) must be in the scope of at least one non-veridical operator. (7)shows some of the possibilities with *too*.

(7) A. Maybe John will go to Paris.
B. I will go there too.
John suggested that Mary left and Bill said Susan did too.

There are some limitations to the antecedents *too* can take, as illustrated by (8)which some people do not like.

(8) John dreamt that Bill is Paris and Tom will go there too.

The English *indeed* is more liberal and (9) illustrates the wider range of antecedents it can take. I do not know why *too* is less liberal than other particles in this respect.

(9) John dreamt that he passed the exam and indeed he passed.
John thinks that Mary hates him and Bill said that she does indeed.

Generation constraints are defeasible constraints that the human generator tries to optimally satisfy when generating a sentence from a characterisation of the semantics. The generation constraint needed for *too* is **ParseOther**, a principle that forces the marking of the presence of another entity of the same type in the context. *Too* marks the presence of another element of the same type, like *also*, *another* or *adifferent*. It is possible to defend the view that this is all that we have to say about the semantics of *too* and that its function provides the explanation of its lack of semantic content.

A similar principle is **ParseOld**, a principle that forces the marking of material that is not new to the context as old material. *Indeed* is one of the linguistic elements that carries out this job, other are pronouns and definite descriptions.

In a bidirectional optimality theoretic framework we can combine the above generation principles with (Blutner, 2000)'s reconstruction of Van der Sandt's presupposition theory by two interpretation principles: **DoNotAccommodate** and **Strength**. The first principle, ranked above

[2] Giannakidou's notion is more restricted and omits suggestions and *maybe*-environments that in some languages allow certain negative polarity items.

the other, militates against accommodations, the second one selects the strongest reading from among the different readings that come out of the accommodation possibilities. In the resulting system, the following principle (Blutner's Law) can be derived.

(10) If a presupposing expression has simple non-presupposing alternatives, it does not accommodate.

The motivation is simple: with a common ground that requires accommodation, a speaker will always select the non-accommodating alternative because it does not lead to a violation of **DoNotAccommodate**. (In the particular version of bidirectional optimality theory advocated by Blutner interpretation constraints are scored together with generation constraints in both directions.)

The predictions that our theory makes for *too* are non-accommodation (this does not rule out a fair amount of partial resolution), the availability of all veridical antecedents, and obligatory occurrence when the veridical context contains another element of the same type. Non-accommodation is a consequence of existence of the simple expression alternative where *too* is omitted. The lack of semantic content is responsible for the possibility of veridical antecedents: it does not matter where the antecedent comes from because it does not need to exist locally.

In these respects, *too* contrasts sharply with a trigger like *regret*. First of all *regret* does not have simple expression alternatives, which means that it allows accommodation. Second, its presupposition makes a strong semantic contribution: it identifies the fact to which the subject has her emotional reaction. This fact must at least be a belief of the subject for the subject to have an emotional reaction to it. Therefore, only real facts and beliefs of the subject can be antecedents and other veridical antecedents are ruled out. The strongest requirement arises when the antecedent identifies a participant, a cause or a precondition of the event described by the clause that contains the trigger (pronouns or definite descriptions). Here the only antecedents are proper constituents of the context of the trigger.

The specification of a trigger is exhausted by a statemnt of its presupposition and its semantic contribution. The overlap between presupposition and semantics filters away unwanted veridical antecedents. Accommodation or not is controlled by the inventory of the language.

For further details I refer to (Zeevat, 2000).

3. *Inderdaad* and *Immers*

My hypothesis about *inderdaad* (*tatsaechlich, indeed*) is that it is just a presupposition inducer, in this case presupposing the positive version of the sentence to which it attaches. As such, it is an old marker and the generation constraint **ParseOld** is responsible for its obligatory occurrence. It takes veridical antecedents, because it does not contribute to the semantics of the clause. It does not accommodate, because as a particle it has simple expression alternatives.

What does this predict about the speech acts in which it occurs? Basically, it says that the hearer, or the speaker or both can have an old opinion that the sentence is true. But it is not necessarily the opinion of one or both of the conversational partners, since the antecedent can also be the opinion of a third party or even weaker, the content of a dream, a suggestion etc. A dialogue system can conclude from an occurrence of *inderdaad* that what is said is already present and it is only the presupposition resolution itself that forces the selection of a speech act of reconfirmation, when resolution is to the speaker or the common ground. It can be the speech act of assenting if the resolution is to a hearer opinion that is not shared. Absence of *inderdaad* when no other old-marker is present, can lead to the conclusion that we have a proper assertion and not a reconfirmation or assent.

The same holds for an occurrence in a question.

(11) Is Harry inderdaad thuis?
Is Harry indeed at home.

(11) presupposes that Harry is at home. In imperatives, it can only presuppose the imperative itself (or the desirability of the course of action).

If we look at a sample of actual uses[3] the hypothesis is largely confirmed, except for an antiquated use as a synonym for *feitelijk* (in fact). This older use is important, because *inderdaad* seems to imply that the new information is better than what we had before. This is either because *inderdaad* retains some properties of *feitelijk*[4] or it is a pragmatic implication of *reconfirmation* or *assent* as such. If *inderdaad* does not add semantical content, the purpose of reconfirmation or assent can only be that new evidence has been found. There is a also subtle distinction between an assent with a single *inderdaad* and one with *ja* (yes) or a nod of the head. If *inderdaad* is used, the speaker claims to have better information than the other speaker whose assertion she assents to. We could capture the distinction by claiming that a sentence with *inderdaad* must still be informational in the sense of Stalnaker, in indicating that the speaker believed it not as a result of what the interlocutor asserted, but already before that. If we supply our reconfirmation or assent with an assertion containing *inderdaad*, the new information can only be the elimination of an existing uncertainty.

Immers is like *inderdaad* in presupposing the truth of the clause to which it attaches, but it is quite different at the same time. *Immers* makes a quite clear semantic contribution. It turns the clause into a reason for accepting what was said just before. Now reasons why something is the case must be the case too in order to be reasons. That John dreamt he was in Spain, or that Charles has suggested so are not reasons why John is away from home. That is why *immers* in simple clauses only takes proper antecedents and no non-entailed veridical contexts. It also does not bring the effect of the new and better view that we noticed with *inderdaad* and we would not expect that since *immers* contributes to the semantic content of the clause.

[3] I used a net-version of Multatuli's Max Havelaar, a classic dutch novel.

[4] *Feitelijk* is not a presupposition trigger, though it can indicate another point of view on the issue at hand. Its analysis is not straightforward.

Like *inderdaad*, it is obligatory. If the statement is already common ground, *immers* is needed to mark the fact that we are dealing with old information. This leads to the following curious fact. *Omdat* like its English counterpart *because* is a presupposition trigger. This gives Dutch two ways of expressing the sentence (12a).

(12) a. He did not come because he is in Paris.
b. Hij kwam niet omdat hij immers in Parijs is.
c. Hij kwam niet omdat hij in Parijs is.

(12b) is obligatory resolved to the common ground. (12c) is obligatory accommodated, because, if it were old information, *immers* must appear. *Omdat* without *immers* is a presupposition trigger that is marked for obligatory accommodation, comparable to a complement of *regret* that has new intonation, or perhaps also indefinite NPs.

Formally, *immersA* has two presuppositions, the one we discussed and the current last sentence. It asserts that the first presupposition is a reason why the second one holds.

Looking at our data, one finds complete confirmation, although there cases where the causal connection is not very clear. *Immers* is not a high frequency item unlike its german equivalent which has quite a number of other uses next to the one discussed here. Questions and imperatives with *immers* are not possible and my analysis explains why.

The occurrence of *immers* in a user utterance is a reliable indication for assuming that the user is not making a normal assertion, but assumes both that the material is already established and relevant at the current point in the dialogue.

4. *Wel*

The marker *wel* in the uses we are focussing on is the typical marker of a correction to a negative utterance made by the other party. It is accented in that case and the most likely explanation is that *wel* is entering in a contrast relation with the negation in the corrected sentence.

(13) A: Jan is niet thuis. (Jan is not at home)
B: Jan is WEL thuis. (Jan IS at home)

In corrections to non-negated sentences, accented *niet* takes over this role.

(14) A: Jan is thuis.
B: Jan is NIET thuis.

But it is not clear there is an element here with which *niet* constrasts. Nevertheless, the relation of contrast with the corrected sentence is so strong that the correct explanation is probably that the whole sentence bears contrast, with everything except *niet* deaccented as old material.

But there are many other uses of *wel*. Typical is the use in a concession:

(15) A: Jan kwam het boek toch gisteren terugbrengen.
A:John was going tot the return the book yesterday, wasn't he?
B: Jan kwam wel, maar hij had het boek niet bij zich.
B:John came allright but he did not have the book.
B1: Jan kwam niet, maar hij heeft het boek wel teruggegeven.
B1: Jan did not come, but he gave the book back allright.

Here the *wel*-clause marks the part where the speaker agrees with the other speaker. But this can be reversed, as in B1. The A. sentence invokes a context in which the plan that Jan was bringing the book yesterday is assumed and evidence is available that the plan has not been carried out. Another case is (16).

(16) A: So they came?
B: Jan WEL, maar Marie NIET.
(Jan did, but Marie did not)

(17) So they did not come?
Jan WEL maar Marie NIET.

Almost idiomatic are the combinations with modal verbs.

(18) Het moest wel.
It had to be.
implies: I/we did not want to but I/we had no choice.

(19) Het zal wel beter gaan in het voorjaar.
It will probably be better in spring
context: now not ok

(20) Het lijkt wel of je nooit meer thuis bent.
It would appear that you are never at home anymore.
presupposes falsity

(21) John shows Mary his new dog.
M: Het lijkt wel een varken.
M: It looks like a pig.
(presupposes it is not one)

(22) Kom je WEL? (presupposes the opposite)
Kom je wel? (expresses doubt)

(23) Wil je wel?
DO you want?
expresses doubt

Quite generally, we seem to be able to say that *wel* p presupposes $\neg p$. In concessive phrases, the presupposition can disappear and the main function is the contrast with the negation in the other half of the pair.

The accented uses require overt negations to contrast with, either within a concessive pair or outside one. In the last case, the negated clause coincides with *wel*'s antecedent.

The explanation of *wel*'s appearance in a sentence must be two-fold. We need a principle that inserts it in a concessive pair, if the concession is built around a positive and negative element, but the generation of concessive constructions does not concern us in this paper. The other occurrences are due to the **ParseOld** principle we discussed before.

Wel takes veridical antecedents, as shown in (24).

(24) Karel droomde dat hij niet voor zijn examen zou slagen, maar hij haalde het WEL.
Karel dreamt he would not pass his exam, but he passed it allright.
Piet zei dat Marie niet zou komen, maar ze kwam WEL.
Piet said that Marie would not come, but she did.

The use of *wel* can help in identifying the dialogue move the speaker is making. It is helpful in identifying corrections, though it must be distinguished from concessive uses and from other presupposing uses.

5. Toch

This is by far the most complicated of the four particles that are the protagonists of this paper. Compare the examples in (25), based on clauses meaning: *he is in Amsterdam* or *come to Amsterdam*

(25) a. Laten we hem vrijdag opzoeken. Hij is dan toch in Amsterdam.
a'. Let us visit him on Friday. He is then in Amsterdam anyway.
b. Hij is toch in AmStErDaM?
b'. He is in Amsterdam, isn't he.
c. Hij is TOCH in Amsterdam.
c'. He is in Amsterdam after all.
d. Is hij TOCH in Amsterdam?
d'. Is he in Amsterdam after all? (We thought he would not be)
e. Kom toch naar Amsterdam. (exhortation)
e'. Come to Amsterdam. (you know you'll like it).
f. Kom TOCH naar Amsterdam.
f'. Come to Amsterdam, (although I see why you do not want it).

The emphatic uses of *TOCH* are pretty straightforward. They indicate that the speaker presupposes the negation of the statement or question she is making. In the case of the imperatives, it is the opposite plan or the desire not too that is presupposed. But the non-emphatic uses are difficult to accommodate in this scheme.

Example (25b.) is the most involved. Often it is treated as a question (a confirmation question) but the form is of an assertion and the intonation is not that of a normal question. Also the facial expression appropriate to its utterance indicates that it is really an assertion uttered expressing surprise at the content, like the assertion in (26).

(26) Hij is in AmStErDaM?

The surprise indicates that the speaker believes to know that what he says is false, in (26). It is a reaction to information that "he" would be in Amsterdam. What the *toch* does in (25b) is to invert these speaker assumptions: the speaker now believes that "he" is in Amsterdam and reacts to information to the contrary.

We could perhaps say that *toch* resolves to the negation of the statement made by the interlocutor. But then after resolution we have assertion with the expression of surprise, which is quite different: the speaker is not surprised that "he" is in Amsterdam, she is surprised that "he" is not. It would seem that this indicates that the *toch* here resolves to the positive information that "he" would be in Amsterdam and -because that rules out surprise at the positive information- the surprise is caused by something else, nl. the information supplied by the interlocutor.

If we look at (25a) this confirms that pattern.

The *toch* here is a device of reminding the interlocutor of some old information and it is functioning not unlike *immers* which could take the place of *toch* in this context. In fact, there are dutch speakers who never use *immers* and always use unaccented *toch* instead.

Uses of unaccented *toch* in questions seem to be impossible. In imperatives, it softens the appeal made on the interlocutor. It does not seem to be impossible to understand this as presupposing a similar desire in the interlocutor. Again the opposite of the accented *TOCH* which presupposes a contrary attitude to the action ordered in the imperative.

In my corpus, by far most uses of *toch* are proconcessives, i.e. single word concessives (like isolated *though* in English) that can be paraphrased by full although-sentences whose content is given by the context. This is a weakening of what we find in (25.c) which seems naturally characterised by presupposing the negation of the clause. Though concessive sentences provide reasons for thinking that the main-clause is false, they do not (cannot) provide the information that the negation is true. It is possible to bring them closer by the notion of a suggestion. The contextually given concessive material can be taken as a suggestion that the clause is false and this would be an appropriate veridical antecedent. Alternatively, we should start from the notion of a reason to be false and let (accented) *TOCH* presuppose a reason for the clause to be false. I prefer the first alternative, since the second alternative makes the integration of the unaccented uses even more problematic than they are already.

What can we make of *toch* in our presuppositional theory? I am not very sure. I would like to say the following. *Toch* is just an old-marker without a preference for positive or negative antecedents. If the antecedent of *toch* has the same polarity as the current clause, no accent is provided by the speaker because there is no contrast between the clause and the recovered presupposition. If the antecedent has opposite polarity, accent results from the recovery of the antecedent. The accent would just be the result of the existence of an alternative in the speaker's mind, here created by the speaker's awareness that she is old-marking a clause for the prior occurrence of a negated version of the clause. I do not have a fully worked out accenting theory from which this accenting pattern would follow, but such a theory is needed. The alternative is that we have a tonal distinction between two lexical items *toch* and *TOCH* with different semantic properties. But this runs against the following argument that I owe to Manfred Bierwisch (p.c.). It would then be completely incomprehensible how it can be that Dutch and German have almost exactly the same *toch/doch* and *TOCH/DOCH* and the same for other accented and deaccented particles.

In other respects, *toch* seems to follow the pattern of the other particles discussed in this paper. It takes veridical antecedents as in (27), it makes no contribution to the content of the clause and it cannot be omitted (but sometimes replaced) where it occurs.

(27) Jan droomde dat hij was gezakt voor het examen, maar hij had het TOCH gehaald.
Jan dreamt he failed the exam, but he passed.

Toch is useful for future dialogue systems as an indicator of corrections when it is accented and when the corrected element can be found in the common ground.

6. Conclusion and Further Research

My first encounter with particles occurred in half-way the eighties when I was working on pronoun resolution. Hypotheses about discourse and dialogue structure can have dramatic consequences for the correct resolutions. It was then —as it is now— difficult to recognise discourse and dialogue structure and in our system we did not even have the resources to reconstruct speaker plans. Particles seemed a way out: in German they are extremely frequent and together with tense shifts and topic they seemed to offer a heuristics that would make our recognition of the discourse and dialogue structure better.

This did not work because particles are not very well understood: many meanings are normally distinguished and few of the meanings seem to be very relevant for the discourse grammarian. The *anyway*, the "pop-marker" of classical discourse grammar is almost an isolated case. And *anyway* is not a pop-marker at all. It marks that what is said in the current clause does not depend on the issue of the last clause or paragraph. The discourse function of closing of a topic is derived from this more primary function.

It is much the same I believe with the particles I have focussed on in this paper. Their function can be clarified to a large extent by analysing them as as presupposition triggers with a number of special properties. It follows that they have certain discourse functions, but those functions are not their primary function. As I hope to have shown in this paper, a reduction to presupposition makes it feasible to use certain particles for the recognition of the speech act the user is making.

There is a considerable class of particles that can be analysed as presupposition triggers. For *again*, I refer to (Kamp & Rossdeutscher, 1994). Next to *again* we find *still* and *already*. Our four old-markers should also include *instead* and perhaps *dan*. As presupposition triggers, they have overwhelming similarities, like the avoiding of accommodation and a strong preference for partial resolution. The dividing line is the question of semantic contribution. The temporal particles clearly sit with *immers* in requiring proper antecedents.

The implementation of the current approach to particles is not much more involved than the general approach to presupposition and anaphora resolution in e.g. Johan Bos's DORIS system, an approach that could clearly be integrated in logic based dialogue systems. The main but unimportant difference is that a larger class of contexts needs to be searched to take care of nonveridical antecedents. A difference —really an advantage— is that the generation constraints also allow inferences about the absence of certain antecedents. The most serious obstacle to a full implementation is the difficulty of doing partial resolution, but this is a difficulty shared with any computational treatment of presupposition. A good discussion of the task for German *wieder* can be found in (Kamp & Rossdeutscher, 1994).

Future research will have to determine what other discourse particles can be captured in the presuppositional analysis.

7. References

R. Blutner. 2000. Some Aspects of Optimality Theory in Interpretation. *Journal of Semantics*, to appear.

Anastasia Giannakidou. 1998. *Polarity Sensitivity as (Non)Veridical Dependency.* John Benjamins.

Irene Heim. 1983. On the projection problem for presuppositions. In WCCFL, 2: 114-26

H. Kamp and A. Rossdeutscher. 1994. Remarks on Lexical Structure and DRS Construction. *Theoretical Linguistics*, 20: 97–164.

S. Kripke. 1964. Presupposition. MS.

Rob van der Sandt. 1992. Presupposition projection as anaphora resolution. In *Journal of Semantics* 9: 333–77.

R. C. Stalnaker. 1978. Assertion. In Peter Cole, editor. *Syntax and Semantics 9: Pragmatics.* New York. pages 315 – 32.

Henk Zeevat. 1997. The Common Ground as a Dialogue Parameter. In G. Jaeger and T. Benz (eds.) Proceedings Mundial 97, CIS, Universitaet Muenchen.

Henk Zeevat. 2000. Explaining presupposition triggers. submitted to ESSLLI 99 book.

Chains and the Common Ground

Anton Benz

Humboldt Universität Berlin
DFG–Project Dialogsemantik
Prenzlauer Promenade 149–152, 13189 Berlin
toni.benz@german.hu-berlin.de

Abstract

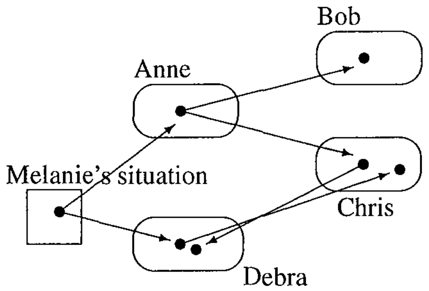
In this paper we provide a description of how the iterated specific use of an indefinite NP can lead to the establishment of referential chains across dialogues and dialogue participants. We describe how they introduce discourse referents, how they are related to the common ground, and how this common ground can be represented by the dialogue participants. Of central concern is the methodological part. We combine methods known from dynamic semantics/DRT on the one side, and theories for multi-agent systems on the other. The last part provides us with a natural, and non–ad hoc model for mutual information, and the interpretation of dialogue acts.

1. Introduction

This is an investigation into the pragmatics of chains in dialogue which are established through sequences of specific uses of indefinite descriptions by different speakers, which are linked to one another, and which are related to the same object.

We can assume that basically each use of an indefinite NP introduces a new discourse referent into the knowledge base of the hearer. We may use here a DRT-like mechanism (Kamp & Reyle 1993; v. Eijck & Kamp 1997) which describes the way a hearer interprets an assertion by the speaker. What is of special interest in the case of the described chains, is that they build a connection between different dialogues, and therefore between different dialogue participants.

(1) Two passenger, Anna and Debra, observe how a Doberman bites a young girl, Melanie. The next day Anna meets Bob and Chris. They sit together, and she tells them that yesterday she saw how *a* young *girl* was bitten by a Doberman. Some weeks later, Chris meets Debra, and they come to talk about dangerous dogs. Debra tells him: "Last week, I witnessed how such a dog bit *a* little *girl*." Chris: "Oh, really! Anne told me that she too saw how a Doberman bit *a girl*."



Here, we have two dialogues, one between Anna, Chris and Bob, the other between Debra and Chris. One and the same object, Melanie, is the source for a branching chain. For subsequent dialogues, it will be necessary for the involved persons to keep track with whom they share which referent.

The problems here are closely related to the phenomena handled in the theory on *First Order Information Exchange* developed by *P. Dekker* (Dekker 1997). He starts out with examples like

(2) A: Yesterday, a man ran into my office, who inquired after the secretary's office.

B: Was he wearing a purple jogging suit?

A: If it was Arnold he was, and if it was somebody else he was not.

He observes that A's answer sounds strange, even if we assume that there was more than one person coming into the office, one of them Arnold. Dekker claims here that

> All natural language terms (definite and indefinite noun phrases alike), are assumed to relate to specific subjects in the information state of a speaker. Indefinite noun phrases which set up discourse referents in a felicitous way, must refer to specific subjects in the information state of the speaker, although they may provide no clue so as to which of his own subjects a speaker refers to. (Dekker & van Rooy 1999)

Dekker and van Rooy developed this approach further to handle belief attributions. The meaning of a discourse like "*Melanie is a 12 years old girl. Chris believes that some dog bit her.*" can be described by a DRS like:

$$\begin{array}{|c|}\hline x,y,yx,yz\\ \hline \text{Melanie}(x)\\ \text{twelve-years-old-girl}(x)\\ \text{Chris}(y)\\ y: \begin{array}{|c|}\hline x,z\\ \hline \text{dog}(z)\\ \text{bit}(z,x)\\ \hline\end{array}\\ x \text{ is the source of } yx\\ \\ \hline\end{array}$$

This framework can be developed straight forward to be able to describe the building of chains across dialogues and dialogue participants. We will do this in a framework of *Multi–Agent Systems*, see (Fagin e.al. 1995). I.e. we will describe the dialogues and the updates of knowledge bases of the participants as games. This has the advantage that we can exploit standard techniques to define the information an agent has in a certain dialogue situation in a possible worlds framework, and we get the usual definition of mutual knowledge. For technical reasons we will develop the theory as a *possibility approach*, see (Gerbrandy & Groeneveld 1997). One effect will be that the *source*–relation, which is a primitive relation in the theory of Dekker and van Rooy, is defined through the rules of the dialogue games.

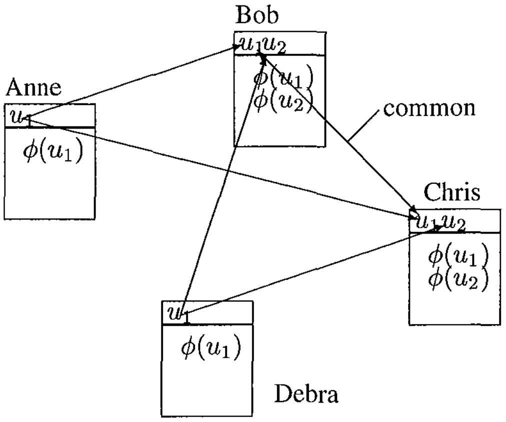
2. Definite Reference and the Common Ground

The relation between established chains and the use of definite descriptions is of special interest, because it forces us to investigate how discourse referents are connected to the *common ground.*

It has become usual to identify the common ground with what is *mutually known* by the dialogue participants. The relation between the referential use of definite descriptions and mutual knowledge has been extensively studied in (Clark & Marshall 1981). For a *visual situation* use, it can be shown that the referential use of a definite description def $x.\varphi(x)$ is successful if the object referred to is the only one for which it is common knowledge that it has the property φ, see (Benz 1999).

The referent of a definite description is an object in the real source situation but this situation is normally not known to the discourse participants. That the anaphoric referential use of a definite is sensitive to the *common* discourse referents can be seen in examples like

(3) At 7:00 am Anna and Debra see how a Doberman bites the young girl Melanie. Anna must leave Debra with the girl. Therefore she can't see that the dog again attacks and bites another girl, Stefanie, some minutes later. Then (1) Anna meets Bob and Chris and tells them that she has seen how a Doberman attacked a young girl. The next day, (2) Debra meets Bob, and she tells him that the dog attacked also another young girl. Later, (3) she meets also Chris and tells him the same. Chris, who does not know that Bob knows already the whole story, (4) meets Bob again and says to him: *"The young girl* was not the only one who was attacked by the dangerous Doberman."



The use of *the young girl* by Chris is felicitous although both of them know that there have been two young girls who were attacked by the Doberman. Only one of them is available through a common discourse referent.

3. Dialogue as Multi–Agent System

For our presentation of multi–agent systems we follow (Fagin e.al. 1995). It consists of a set of *global states*, a set of possible *dialogue acts*, a *transition operation* τ, which models the effect of performing an action, and a function P which tells us, which actions can be performed in a given situation.

A global dialogue state consists of the *local* states of the participants DP $= \{1, \ldots, m\}$, and the state of the environment. Essentially, our states will contain the same information as the pictures provided above. They are represented by tuples $\langle \mathcal{M}, \dashrightarrow, D_1, \ldots, D_m \rangle$. Here, $\mathcal{M}$ is a first order model which describes the situation talked about, D_a is a simple DRS extended with information (1) about the dialogue acts where the participant a was involved, and (2) about the real objects he has observed. $\dashrightarrow$ is a relation between objects and subjects, or subjects and subjects, where a subject is a pair $\langle a, u \rangle$ of a participant a and a discourse referent u. We write u^a for $\langle a, u \rangle$. If a new discourse referent is introduced into a DRS, then $\dashrightarrow$ will connect this referent to it's *source*. Now we are able to say what is a possible sequence of dialogues in our model. We identify them with sequences $G = \langle s_0, \mathtt{act}_0, \ldots, s_n \rangle$, where all s_i are global states, s_0 is a state where $\dashrightarrow^{s_0}$ and all $D_a^{s_0}$s are empty. Further, $\mathtt{act}_i \in P(s_i)$, i.e. it must be a *possible* dialogue act in situation s_i, and $s_{i+1} = \tau(\mathtt{act}_i, s_i)$ for all $i < n$. We denote the set of all such possible sequences of dialogues by $\mathcal{G}$.

We allow for three actions an agent can perform: $\mathtt{send}(a, H, D, l)$, $\mathtt{get}(a, H, D)$ and $\mathtt{observe}(H, D, l)$. $\mathtt{send}(a, H, D, l)$ is the actions which represents an assertion of speaker a with co–present addressees H. D is a DRS, which is the result of translating the speaker's utterance into a DRS by standard techniques known from (Kamp & Reyle 1993). l is a function which relates the discourse referents in $\mathcal{U}_D$ to the subjects in D_a. If the action $\mathtt{get}(a, H, D)$ is performed on the local state of an agent $b \in$ DP, then it means that he is an addressee $b \in H$ of an assertion with content D and speaker a. $\mathtt{observe}(H, D, l)$ means that he is a member of a (co–present) group H which observes some fact represented by a DRS D, where l is an injective function which relates the referents in $\mathcal{U}_D$ to real objects in the universe $|\mathcal{M}|$. These actions can be performed as parts of *joined* actions. They can be identified with sequences $(\mathtt{act}_a)_{a \in \mathrm{DP}}$, where $\mathtt{act}_a$ is one of the three (local) actions defined before. We allow for two sorts of joined actions: Either the sequence has the form $\mathtt{act}_a = \mathtt{observe}(H, D, l)$ with fixed H, D, l for $a \in H$, and $\mathtt{act}_a = \bot$ for $a \notin H$. Or, it has the form $\mathtt{act}_a = \mathtt{send}(b, D, H, l)$ for $a = b$, $\mathtt{act}_a = \mathtt{get}(H, D, l)$ for $a \in H$, and $\mathtt{act}_a = \bot$ for $a \notin H \cup \{b\}$, D, H fixed.

What are the effects of performing a joined action $(\mathtt{act}_a)_{a \in \mathrm{DP}}$? If it represents an assertion by a speaker b, then $\mathtt{send}(b, H, D, l)$ should not change the state of b except that he remembers that he has performed this action, i.e. we assume that D_b has a component Act_{D_b} such that for the new state D'_b we get $Act_{D'_b} =$

$Act_{D_b}{}^\wedge\langle\mathtt{send}(b, H, D, l)\rangle$. The act $\mathtt{get}(b, H, D)$ should result in a merge of D_a and D for $a \in H$, and in an extension of Act_{D_a} to $Act_{D_a}{}^\wedge\langle\mathtt{get}(b, H, D)\rangle$. There is some freedom in defining this merge. We may assume that it introduces for each referent in D a new referent into $\mathcal{U}_{D_b}$, and adds the conditions of Con_D to Con_{D_b} where the old variables are replaced accordingly. $\rightarrow$ belongs to the environment, and an assertion of the form $\mathtt{send}(b, H, D, l)$ has the effect that new chains are added to $\rightarrow$. In order to be able to start in our examples with *empty* representations, we consider also acts of joined observations. For $a \in H$, an observation $\mathtt{observe}(H, D, l)$ should have the effect that D is merged to his old information D_a in such a way that new discourse referents are introduced for objects which he has not jet observed. We assume that D_a has a fourth component Obs_{D_a} which is an injective function relating referents in $\mathcal{U}_{D_a}$ to objects in $|\mathcal{M}|$, i.e. a remembers which objects he has observed.

We don't go into further details here. Table 1 shows the relevant part of a global state for Example 3 which results after Ann's assertion (1). It translates into a DRS $D = \langle\{u_1, u_2\}, \{\mathrm{Doberman}(u_1), \mathrm{young{-}girl}(u_2), \mathrm{bit}(u_1, u_2)\}\rangle$. H_0 is the group $\{An, De\}$, H_1 the group $\{An, Bo, Ch\}$. l links u_1 to the Doberman (*Dob*), u_2 to Melanie (*Mel*). In addition, we assumed here that Chris has already some information represented in his DRS.

To get a full description of our system we have to say which joined actions can be performed in some global state s. We assume that an joined observation is always possible, if we have for $\mathtt{observe}(H, D, l)$ that $(\mathcal{M}, l) \models D$. If the joined actions represents an assertion with $\mathtt{send}(b, H, D, l)$, then it should be a possible action in s, iff $D_b^s \trianglelefteq_l D$. $D \trianglelefteq_l D'$ holds between DRSes D, D', iff l is a function from $\mathcal{U}_{D'}$ to $\mathcal{U}_D$ such that for all condition $\varphi \in \mathrm{Con}_{D'}$ φ/l is an element of Con_D, where φ/l denotes the formula, where the free variables in φ are replaced by their l–values. This is essentially *Dekker's* condition for the licensing of first order formulas (Dekker 1997). It implies that the speaker can make only *true* assertions. For global states s we denote by $P(s)$ the set of joined actions which can be performed in this situation.

We denote by $\mathcal{S}(\mathcal{G})$ the set of all global states which may arise as a possible dialogue situation, i.e. all situations which belong to a $G \in \mathcal{G}$. For multi–agent systems it is usual to identify the *knowledge* of an agent a in a situation s relative to $\mathcal{S}(\mathcal{G})$ with the set of all situation which are *indiscernible* from s. Two situations are indiscernible for an agent a, iff his local states are identical for both situations. This allows us to include the information of agents about the global state, and their information about others into our model. We either may use *Kripke*–structures, see (Fagin e.al. 1995), or develop our theory along the lines of (Gerbrandy & Groeneveld 1997) as a *possibility* approach. Both descriptions provide us with (equivalent) representations $CG_w(H)$ of the *common ground* for a possibility w and a group H. It is as a set of accessible possible dialogue situations and contains all possibilities which are possible according to the knowledge of one participant, possible according to the knowledge a participant can have according to the knowledge of an other participants, etc.

Hence, the general apparatus for multi–agent system provides us with a natural representation of the mutual information of dialogue participants. But in view of our problem to explain the anaphoric referential use of a definite description we need a representation which provides us more directly with information about which subjects with which properties are common. For this reason we introduce the notion of a *common DRS*.

That a DRS is *joined* should mean that it can be embedded into all the DRSes representing the knowledge of the members of the group H in such a way that the images of one referent are all connected to each other via a common source. Hence, D is a joined DRS for a group H and possibility $w \in \mathcal{W}$, iff there is a family of functions $(l_a)_{a \in H}$ such that for all $a \in H$ $D_a^w \trianglelefteq_{l_a} D$, and for all $u \in \mathcal{U}_D$ $\exists x \forall a \in H$ $x \rightarrow_r l_a(u)$, where $\rightarrow_r$ denotes the reflexive closure of $\rightarrow$. In order to restrict the possible size of a joined DRS we add the condition that for all $u, u' \in \mathcal{U}_D$, $u \neq u'$, there is at least one $a \in H$ such that $l_a(u) \neq l_a(u')$. Intuitively, a DRS is *mutual* joined if it is joined, everybody knows that it is joined, everybody knows that everybody knows that it is joined etc. This means that D must be a joined DRS relative to a family $(l_a)_{a \in H}$, and for all $b \in H$ and for all v which are possible for b there exists a family $(l_a^v)_{a \in H}$ such that D is joined in v relative $(l_a^v)_{a \in H}$ and $l_b^v = l_a$. By a simple iteration of this condition we get an intuitive definition of a *common* DRS, $C_w(D, H)$, for a group H in a possibility w. E.g. for the situation described in Table1 we find that the following DRS D is a maximal common DRS for the group $H_1 = \{An, Bo, Ch\}$.

u_1, u_2

Doberman(u_1)
young–girl(u_2)
bit(u_1, u_2)

A detailed examination of the examples introduced above would show that the uniqueness condition for the referential anaphoric use of a definite description is sensitive to the number of discourse referents in the maximal common DRSes.

4. The Representation Problem

The last section provided us with a reasonable description of a *common* DRS. But how can the participants have access to this DRS? The most intuitive way seems to be that they keep track of the discourse referents which have been introduced to each group, and about the properties of those referents. I.e. a participant will not only update his own DRS, if he gets some new information, but he will also update a DRS representing the knowledge of the group which *commonly* got this information. This leads to an extension of the local states. We add for each participant a and for each group $H \subseteq \mathrm{DP}$, where he is a member of this group, representing DRSes $D_{a,H}$. In the same way as in the last section we can describe the update operations connected to the possible local acts $\mathtt{send}(a, H, D, l)$, $\mathtt{get}(a, H, D)$ and $\mathtt{observe}(H, D, l)$ for global states with representations. Together with the function P, which specifies which actions are possible in a certain situation, this

$\rightarrow$	Anne An	Bob Bo	Chris Ch
$Dob \rightarrow u_1^{An}$ $Mel \rightarrow u_2^{An}$ $u_1^{An} \rightarrow u_1^{Bo}$ $u_2^{An} \rightarrow u_2^{Bo}$ $u_1^{An} \rightarrow u_{n+1}^{Ch}$ $u_2^{An} \rightarrow u_{n+2}^{Ch}$	u_1, u_2 Doberman(u_1) young–girl(u_2) bit(u_1, u_2) $\texttt{observe}(H_0, D, l)$ $\texttt{send}(H_1, D, id)$ l	u_1, u_2 Doberman(u_1) young–girl(u_2) bit(u_1, u_2) $\texttt{get}(An, H_1, D)$ $\emptyset$	$u_1, \ldots, u_n,$ u_{n+1}, u_{n+2} Doberman(u_{n+1}) young–girl(u_{n+2}) bit(u_{n+1}, u_{n+2}) $\texttt{get}(An, H_1, D)$ $\emptyset$

Table 1: (Part of) a global dialogue state for Example (**3**).

will lead to a new set of possible dialogues $\mathcal{G}^+$ for the same sequences of actions.

The following figure describes the local state of Bob in Example 3 after his talk with Debra (2). The first column represents his total knowledge about the biting situation, the second his protocol for what he heard in common with Anne and Chris, and the third for what he has in common with Debra.

Bo	{ Bo,An,Ch }	{ Bo,De }

u_1, u_2, u_3
Dob(u_1)
girl(u_2)
bit(u_1, u_2)
girl(u_3)
bit(u_1, u_3)
$u_3 \neq u_2$
	u_1, u_2
	Dob(u_1)
	girl(u_2)
	bit(u_1, u_2)
		u_1, u_2, u_3
		Dob(u_1)
		girl(u_2)
		bit(u_1, u_2)
		girl(u_3)
		bit(u_1, u_3)
		$u_3 \neq u_2$

If we compare this state with the parallel global state in $\mathcal{G}$, then we find that the DRS in the second column is a maximal common DRS. Hence, if Bob meets Chris, then he can apply the uniqueness condition which is connected to the definite *the girl* to this DRS. As Chris will have the same representation for the common DRS, they both will interpret the description as relating to a subject which is *chained* to Melanie.

We can show in general that the DRSes $D_{a,H}$, which are internal representations of agent a for the referents and conditions which are common, are identical for all $a, b \in H$. Furthermore, we can prove — with some effort — that they are always maximal common DRSes for the related dialogue situation in $\mathcal{G}$.

5. Conclusions

We are able to represent the chains that are defined by iterated specific uses of indefinite NPs. The theory of multi–agent systems, which builds the basis of for our model, provides us with natural descriptions of the common ground as an information state representing mutual information. But to be able to explain the referential anaphoric use of a definite description, and especially how to apply it's uniqueness condition, we found that, in fact, it is sensitive to common substructures of the local states of discourse participants. We characterised them as common DRSes and explained how the participants can represent these common DRSes. There are some points in our approach which should be emphasised:

- Specifically used indefinite NPs introduce *free* variables. The interpretation function, which is necessary to define the truth values for the conditions of a DRS, are provided by an external chain relation. They never get existentially bound.
- There are three distinct objects in our model which are possible representations for the linguistic *common ground*: (1) The information states representing common knowledge, (2) the common DRSes, and (3) the internal representations of the common DRSes. The theory about multi–agent systems allows us to describe exactly how common DRSes are related to common knowledge.
- The uniqueness condition connected to an anaphorically used definite description does not contribute anything to the *meaning* of a sentence where it occurs.

6. References

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

A. Benz (1999): *Perspectives and the Referential Use of Definite Descriptions in Dialogue*; ms., Berlin.

H.H. Clark, C.R. Marshall (1981): *Definite Reference and Mutual Knowledge*; in: A.K. Joshi, B. Webber, J. Sag (ed.): *Elements of Discourse Understanding*, Cambridge University Press, Cambridge.

P. Dekker (1997): *On First Order Information Exchange*; in: A. Benz & G. Jäger (ed.), *Proceedings of the Munich Workshop on Formal Semantics and Pragmatics of Dialogue MunDial'97*, CIS Lecture Notes 106, München, pp. 21–39 .

P. Dekker, R. v. Rooy (1999): *Intentional Identity and Information Exchange*; ms, ILLC/University of Amsterdam.

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, p.147–169.

H. Kamp, U. Reyle (1993): *From Discourse to Logic*; Kluwer, Dordrecht.

J. v. Eijck, H. Kamp (1997): *Representing Discourse in Context*; in: J. v. Bentham, A. t. Meulen *Handbook of Logic & Language*, Elsevier, Amsterdam.

System Description: MIDAS

Johan Bos

Computerlinguistik
Universität des Saarlandes
Im Stadtwald, Postfach 151150
66041 Saarbrücken, Germany
bos@coli.uni-sb.de

Abstract

MIDAS is a dialogue system exploring semantic representation and first-order reasoning to model and plan the ongoing dialogue. Technical features of MIDAS are: deep syntactic and semantic analysis (English string input), DRSs as semantic representations, generation from DRSs to English utterances, and the use of several state-of-the-art theorem provers and model builders for inference tasks. Linguistic features of MIDAS are: treatment of several ambiguities (including scope, anaphora, ellipsis, and presupposition), question answering determination, and utterance grounding.

1. Overview of MIDAS

MIDAS (Multiple Inference-based dialogue analysis system) was developed to study the role of automated reasoning in human machine dialogue systems. The overall system structure is based on the Dialogue Move Engine architecture as proposed in the Trindi (Task Oriented Instructional Dialogue, EC Project LE4-8314) and implemented by TrindiKit (Larsson et al., 1999).

On the analysis side, MIDAS uses a left-corner parser, a lexicon consisting of ca. 4000 inflected forms, and a phrase structure rule grammar (for English). The system's utterances are partly pre-canned, but mostly generated directly from the semantic representation, using the same lexicon and a subset of the grammar rules as used for analysis.

The scenario covers the 'route planning service' task: MIDAS, having primary initiative throughout the dialogue, asks the user for a destination, departure, and traveling time, and whether the quickest or the shortest route should be presented.

The semantic framework underlying MIDAS is Discourse Representation Theory (Kamp and Reyle, 1993). Semantic analysis includes resolution of scope, anaphora, ellipsis and presupposition. The information state of the dialogue is modeled by Discourse Representation Structures, extended with machinery to represent dialogue acts, and utterance grounding, similar to that as proposed by Poesio and Traum (1998).

2. Inference Tasks in MIDAS

In general, we would like to equip natural language processing systems (like MIDAS) with a reasoning component to react intelligently on the user's input. This requires coping with the content of the user's contribution, and a deep semantic analysis to deal with inconsistent information, or with more or less information than was requested. In MIDAS, first-order reasoning is used to resolve ambiguities (scope, anaphora, presupposition, ellipsis), and to interpret answers of users to questions posed by the system (see Bos and Gabsdil, this volume).

Most state-of-the-art inference engines don't work on DRSs directly. By using a translation approach from DRT to first-order logic we are able to use a wide variety of off-the-shelf provers. Note that inference problems need not be theorems—they can be satisfiable as well, and we do not know beforehand. A typical example is the check for consistency: ϕ is inconsistent if $\neg\phi$ is a theorem, and consistent if ϕ is satisfiable. A handicap is the undecidability of first-order logic. By using different inference engines (with different strategies) at the same time in a distributed framework, we reduce this gap to a minimum.

3. Automated Reasoning in MIDAS

Automated theorem proving has seen an enormous increase of performance of (especially first-order) inference engines. We argue to farm out the inference tasks to many different provers simultaneously, by combining MIDAS with the distributed MathWeb theorem proving environment (Blackburn et al., 1999; Franke and Kohlhase, 1999), because typically a high number of reasoning (of rather "simple") tasks are generated by MIDAS; and there are significant differences in speed and coverage that state-of-the-art provers offer.

Using MathWeb and the Internet as medium to distribute the inference engines across different machines provides an ideal testbed for studying the role of automated reasoning in computational semantics in general, and in particular for dialogue systems such as MIDAS. The arsenal of inference engines MIDAS currently calls upon include the theorem provers Bliksem, SPASS, Otter, FDPLL, and the model generator MACE.

4. References

Patrick Blackburn, Johan Bos, Michael Kohlhase, and Hans de Nivelle. 1999. Inference and Computational Semantics. In Bunt and Thijsse, editors, *IWCS-3*, Tilburg, NL.

Andreas Franke and Michael Kohlhase. 1999. System description: Mathweb, an agent-based communication layer for distributed automated theorem proving. In *16th International Conference on Automated Deduction CADE-16*.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Studies in Linguistics and Philosophy 42. Kluwer Academic Publishers, Dordrecht/Boston/London.

Staffan Larsson, Peter Bohlin, Johan Bos, and David Traum. 1999. TRINDIKIT 1.0 Manual. Technical Report Deliverable D2.2, Trindi.

Massimo Poesio and David Traum. 1998. Towards an Axiomatization of Dialogue Acts. In J. Hulstijn and A. Nijholt, editors, *Formal Semantics and Pragmatics of Dialogue, Proceedings of Twendial '98*, pages 207–221, Universiteit Twente, Enschede.

5. URLs

MathWeb: `www.mathweb.org`
MIDAS: `www.coli.uni-sb.de/~bos/midas`
TrindiKit: `www.ling.gu.se/research/projects/trindi/trindikit.html`

From Speech Acts to Search Acts:

a Semantic Approach to Speech Acts Recognition

Marc Cavazza

University of Teesside
Borough Road, TS1 3BA, Middlesbrough, United Kingdom
m.o.cavazza@tees.ac.uk

Abstract

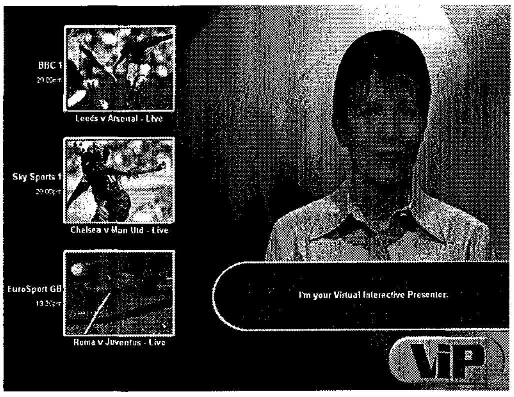
This paper describes the implementation of a human-computer dialogue system based on speech acts. The system has been developed as part of a conversational character for Interactive Television that assists the user in choosing TV programmes from an on-line Electronic Programme Guide (EPG). We have defined a set of specialised speech acts to account for the fact that dialogue actions correspond to the construction of a programme description. These speech acts can be recognised in the course of dialogue by comparing the semantic content of the user utterance with the search filter constructed from the previous dialogue history. The first step consists in parsing the user input to produce a semantic representation. This semantic representation is used to generate a search filter for the EPG. User replies are matched to previous search filters to determine speech acts for acceptance, rejection or refinement of the current selections. These mechanisms are illustrated with example dialogues from the system.

1. Introduction

Speech Acts theory [Austin, 1962] [Searle, 1969] [Récanati, 1979] [Berrendonner, 1981] provides a framework for the implementation of dialogue systems for Information Access applications, as it helps breaking down the information exchange between the user and the system into minimal dialogue units. Recent work in speech acts-based dialogue systems has emphasised the definition of core speech acts [Poesio and Traum, 1997], the specialisation of speech acts according to the task at hand [Busemann et al., 1997] (see also [Lee and Wilks, 1996] [Bunt, 1989]) and the importance of speech acts identification [Traum and Hinkelman, 1992]. The latter point is especially relevant in relation with acceptance and rejection of proposals [Walker, 1994; 1996].

In this paper, we describe the implementation of a dialogue system based on speech acts for Interactive Television. This system is a conversational character, which assists the user in the selection of programmes (Figure 1). The rationale behind the use of dialogue is that it enables users to concentrate on single programme features at each dialogue turn and to refine their selection according to previous results and system's suggestions. During dialogue, the system constructs and updates a filter corresponding to user preferences as these are incrementally refined through dialogue. This filter is used to search the programme database (or Electronic Programme Guide, EPG), which is a hierarchical structure of standard editorial categories defined by broadcasters. More specifically, we will discuss the relations between the semantic content of user utterances, the identification of user speech acts and the subsequent updating of the search filter. Despite the traditional definition of speech acts, Berrendonner [1981] and Recanati [1979] have actually suggested that utterances can qualify as speech acts on the basis of their representational content.

We will first describe the specific set of core speech acts we have defined for our application. We will then show how these task-oriented speech acts can be identified by comparing the semantic contents of the latest user utterance with that of the search filter.



I'm your Virtual Interactive Presenter.

Figure 1. The Virtual Interactive Presenter.

2. Definition of Task-oriented Speech Acts

We have retained most of core speech acts as previously defined e.g. in [Traum and Hinkelman, 1992] such as `yes-or-no-questions`, `request`, `accept` and `wh-questions`. However, we have further specified those speech acts that are connected to the search process, such as `inform` or `reject`. Instead of `inform`, we describe `initial` and `specify` speech acts. Instead of a single rejection speech act, we distinguish between rejections and alternatives and we further refine rejection according to the rejected category. The rationale for using specialised speech acts is that the system's response as well as some dialogue control mechanisms are easier to define in terms of a larger set of speech acts.

Specifications are detected each time the new utterance provides previously non-existing information that is compatible with the information currently available. That is to say, adding a compatible subgenre to a programme genre, adding specific features such as cast or parental rating to a programme description. These do not have to follow a strict top-down refinement.

Rejections are identified each time incompatible features between the current utterance and the search filter are detected. Explicit rejections are introduced by a set of illocutionary speech acts whose surface form most often include negation (e.g., "I don't want to watch that"). The system also allows some explicit rejections in the form of negative comments (e.g., "this is daft"). However, an important form of rejection is constituted by indirect speech acts. Such rejections, as identified by Searle [1975], "constitute a rejection of the proposal, but not in virtue of its meaning". A typical example occurs when the user requests a different movie genre than the one currently considered. Rejections are subsequently typed according to the category or feature rejected.

Alternatives express variants of the current selection. They have to be interpreted contextually. For instance, "can I have another western?" expresses an alternative choice for the instance, which does not reject the subgenre (western). The alternative "can I have another movie?" is slightly more ambiguous, in that it can imply a rejection of the programme subgenre (i.e., the movie genre). Indefinite alternatives ("can I have something else?") cannot be interpreted as such and require additional feedback from the user. Finally, the sentence "Do you have another movie with James Woods?" would reject the top selection, and potentially specify the search, if cast was not already a search feature.

In the remainder of the paper, we concentrate on those speech acts which have a direct impact on database search, mainly rejections, refinements and alternatives. Though traditional speech acts such as greetings, openings and wh-questions are part of the system, we will not discuss them in this paper.

3. From Semantic Analysis to Speech Act Identification

The first step consists in analysing the user utterance to produce a semantic representation. The semantic representation is a feature structure based on semantic categories that correspond to the taxonomic categories of the EPG (e.g. movie, documentary, news...). To this extent, semantic content is more important that logical form, as even the relations that structure these feature representations (e.g. "cast" or "audience") are semantic in nature. The semantic features correspond to the set of editorial categories in the EPG, enhanced with inference mechanisms that can derive categories from connotations (e.g. "entertaining") and infer relevant features (such as deriving the parental rating from the audience). Sentence analysis involves both a syntactic and a semantic step. The syntactic formalism used is a simplified variant of Tree-Adjoining Grammars, whose trees are enriched with semantic features [Cavazza, 1998]. However, the sole purpose of syntax is to ensure propagation of semantic features according to the tree operations carried out during parsing. The system is designed to accept partial parses: in that case the partial semantic structures are aggregated on the basis of feature compatibility. Semantic structures obtained from user input analysis are represented in the various examples illustrating this section as LISP feature structures immediately following the user utterance (see also Figure 2).

These semantic structures are used to instantiate a search filter, which takes the form of a partially instantiated EPG programme record. To this extent, the semantic features in the semantic representation are matched with the Electronic Programme Guide categories. Semantic features that correspond to EPG categories are assembled into a search filter, which can include both positive and negative criteria. This search filter is incrementally refined and updated throughout the dialogue process, to reflect the progressive specification of the user's choice through the course of dialogue.

The next step consists in identifying the user's speech act. In the general instance, speech act identification is a complex process that involves surface forms, semantic content and dialogue pragmatics. Original work on speech acts refers to surface structure for their identification [Searle, 1975], but this is also because in the traditional conception semantic content is denied a role in speech act characterisation. More recent proposals in Computational Linguistics combine surface signals with other heuristics [Hinkelman and Allen, 1989]. Walker [1994; 1996] has specifically studied the recognition of utterance rejection, though not explicitly referring to speech acts, and bases this recognition on content comparison. We thus claim that speech acts, as we have defined them, can be identified by comparing the semantic content of the current utterance with the global filter constructed so far. In other words, the semantic difference between the new filter and the current filter can be used to identify speech acts. A total of 25 rules are used for speech act identification by comparing the semantic contents of the user utterance and the search filter. These rules define speech act recognition in terms of their semantic content differences. For instance, if the latest filter instantiates the same EPG category with a different feature, it consists in a rejection of that category (see examples below). In addition, a distinction is established between speech acts that can be immediately followed by a new EPG search and those who require additional information (characterised as `no_search`). For instance, a speech act rejecting a high-level category without providing a replacement one cannot trigger a new EPG search.

Additional heuristics are used to identify a direct query targeted at a selected programme instance, from the illocutionary nature of the user reply (e.g., "what is its rating?" "who is starring?", etc.), using wh-questions, anaphora and definiteness. These direct speech acts are identified through a set of heuristic rules, which track the explicit mention of a programme feature (e.g., "cast", "parental rating", "starting time") while a specific programme instance is considered (as signalled through pronouns, determiners and deictics). They are not used for filter comparison.

In the next sections, we illustrate different speech acts on sample dialogues (these correspond to actual system runs in its current development status). Each user utterance is followed by its corresponding semantic structure, in the form of a feature structure obtained from the LISP implementation of the dialogue system. The search filter derived from this structure is next, in the form of an attribute-value pair. Finally, the specific speech act recognised from that utterance is presented. Speech acts are represented as tuples, the first item in the list being the speech act category. For the particular `initial` speech acts, which open the dialogue, the tuple includes the initial search filter as well. System replies (labeled "S") have sometimes been edited for the sake of clarity, as the natural language generation component is still under development.

U1: Do you have any action movies?

```
((QUESTION) (EXIST)
 (PROGRAMME ((GENRE MOVIE)
             (SUB_GENRE ACTION) (INDET))))
Filter: ((GENRE MOVIE)
         (SUB_GENRE ACTION))
SA: (INITIAL ((GENRE MOVIE)
              (SUB_GENRE ACTION)) SEARCH)
```

S2: *There are 5 action movies. The first one is: "Raw Deal"*

U3: I'd like a movie I can watch with my kids

```
(((REQUEST)
 (OBJECT ((GENRE MOVIE) (INDET)))
         ((AUDIENCE USER))
((QUESTION)) ((VIEW))
((AUDIENCE CHILDREN)
 (PAR_RATING FAMILY) (POSS)))
Filter: ((GENRE MOVIE)
         (SUB_GENRE ACTION)
         (PAR_RATING FAMILY))
SA: (SPECIFY PAR_RATING FAMILY SEARCH)
```

S4: *Here is "Last Action Hero"*

In this first example, the user opens the dialogue by requesting a specific movie genre. In response to a first selection proposed by the system as a result of the EPG search, the user specifies an additional feature (parental rating), though through an *indirect* speech act (describing the audience). While this utterance is also a rejection of the proposal, the speech act is recognised as a specification, because parental rating had not been previously grounded during dialogue. The system recognises the specification by detecting a new feature in the incoming filter. The system actually rejects the current selection in its subsequent search as it is incompatible with the parental rating introduced.

U5: Can I see a western tonight?

```
 ((QUESTION)
 (SUBJECT ((AUDIENCE USER))
 (CHOICE+ ((VIEW)))
 (PROGRAMME ((SUB_GENRE WESTERN)
(INDET))))
Filter: ((SUB_GENRE WESTERN))
SA: (INITIAL
      ((SUB_GENRE WESTERN))SEARCH)
```

S6: *Here is the western "Unforgiven"*

U7: Can I have a thriller?

```
((QUESTION) (EXIST)
 (PROGRAMME
  ((SUB_GENRE THRILLER)   (INDET))))
Filter: ((SUB_GENRE THRILLER))
SA: (REJECT SUB_GENRE THRILLER SEARCH)
```

In this second example, the user requests a different movie genre from the one already in the filter following his/her opening request. Because the movie genre was previously mentioned in dialogue, this is recognised as a rejection of the movie genre, through an indirect speech act. This is detected by comparing the latest filter with the aggregated filter that was a result of the previous utterance. This comparison explicitly detects a new subgenre, incompatible with the previous one. However, because a new category is provided, the system can trigger a new EPG search that would correspond to the updated filter (this is represented by the `search` feature in the speech act representation).

U8: Do you have any thrillers?

```
((QUESTION) (EXIST)
 (PROGRAMME
  ((SUB_GENRE THRILLER)  (INDET))))
Filter: ((SUB_GENRE THRILLER))
SA: (INITIAL
     (SUB_GENRE THRILLER) SEARCH)
```

S9: *Here is the thriller "the Specialist"*

U10: Can I have another movie with James Woods?

```
((QUESTION) (SUBJECT ((AUDIENCE USER))
 (CHOICE+ ((VIEW))
 (PROGRAMME
  ((GENRE MOVIE) (OTHER-CHOICE+)))
   (INSTR ((CAST WOODS))))
Filter: ((GENRE MOVIE)(CAST WOODS))
SA: (OTHER GENRE SEARCH)
```

This example illustrates the treatment of alternatives (`other` speech acts). Namely, that the alternative rejects the proposed movie instance, which is deleted from the selection, and only retains the most generic feature ("movie") deleting the subcategory as well ("thriller"). The alternative also refines the selection by introducing a specific cast. This is actually taken into account when updating the filter, a process that takes place after speech act recognition. In this particular case, the alternative has enough information for the EPG new search to be directly triggered.

U11: Is there a movie with John Wayne?

```
((QUESTION) (EXIST)
 (PROGRAMME
  ((GENRE MOVIE) (INDET))
   (INSTR ((CAST WAYNE))))
Filter: ((CAST WAYNE))
SA: (INITIAL ((CAST WAYNE)) SEARCH)
```

S12: *Here is the western "Rio Bravo"*

U13: I don't want a western

```
((CHOICE+) (NEGATIVE)
 (AUDIENCE ((AUDIENCE USER))
 (PROGRAMME
  ((SUB_GENRE WESTERN) (INDET))))
Filter: ((SUB_GENRE WESTERN) (NEGATIVE))
SA: (REJECT SUB_GENRE NIL NO_SEARCH)
```

This is a simple and explicit rejection of the movie genre. In this case we find more appropriate that the system returns to the user asking him to specify the subcategory of his choice. This avoids that the system enumerates all possible choices, which can be painful for the user.

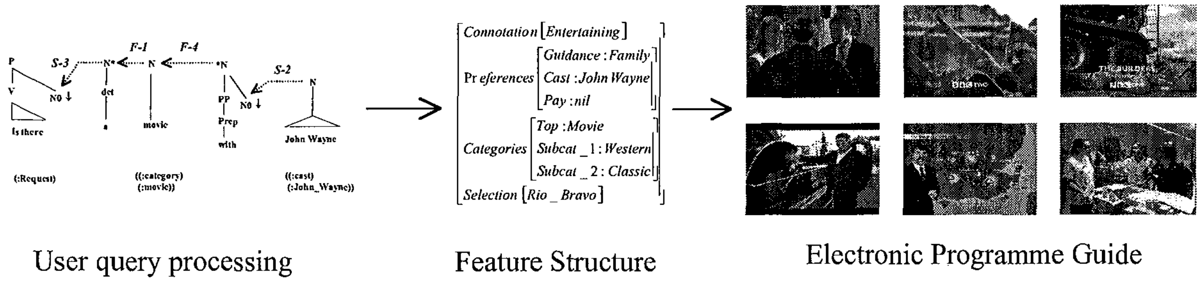


Figure 2. The Natural Language Processing step.

4. Conclusions

We have presented a semantic approach for the recognition and interpretation of speech acts in an Information Access application. This application shares similarities with previous systems [Nagao and Takeuchi, 1994], [Sadek, 1996] [Beskow and McGlashan, 1997], though it proposes a unified treatment of speech act recognition and database search, based on the semantic contents (rather than logical structure) of the dialogue units. This approach appears appropriate for the processing of indirect speech acts as well. The system is however still under development and the above results should be re-assessed in the context of the full EPG database.

Acknowledgements

The Virtual Interactive Presenter is a LINK Broadcast project funded by the DTI. Steve Francis is thanked for Figure 1. EPG contents have been provided by the BBC.

5. References

Austin, J. (1962). *How to Do Things with Words.* Oxford, Oxford University Press.

Berrendonner, A. (1981). Elements de Pragmatique Linguistique. Editions de Minuit, Paris (in French).

Beskow, J., and McGlashan, S. (1997). Olga: A Conversational Agent with Gestures. In: *Proceedings of the IJCAI'97 workshop on Animated Interface Agents - Making them Intelligent,* Nagoya, Japan, August 1997.

Bunt, H.C. (1989). Information dialogues as communicative action in relation to information processing and partner modelling. In: Taylor, M.M., Néel, F. and Bouwhuis, D.G. (Eds.), *The Structure of Multimodal Dialogue,* Amsterdam, North-Holland.

Busemann, S. Declerck, T., Diagne, A., Dini, L., Klein, J. and Schmeier, S. (1997). Natural Language Dialogue Service for Appointment Scheduling Agents. In: *Proceedings of ANLP'97,* Washington DC.

Cavazza, M. (1998). An Integrated TFG Parser with Explicit Tree Typing. In: *Proceedings of the fourth TAG+ workshop,* IRCS, University of Pennsylvania.

Hinkelman, E.A. and Allen, J.F. (1989). Two Constraints on Speech Act Ambiguity. *Proceedings of ACL-89,* pp. 212-219.

Lee, M. and Wilks, Y. (1996). An ascription-based approach to speech acts. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96),* Copenhagen.

Nagao, K. and Takeuchi, A. (1994). Speech Dialogue with Facial Displays: Multimodal Human-Computer Conversation. In: *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics* (ACL'94), pp. 102-109.

Poesio, M. and Traum, D. (1997). Representing Conversation Acts in a Unified Semantic/Pragmatic Framework. In: *Proceedings of the AAAI Fall Symposium on Communicative Actions in Humans and Machine,* Cambridge (MA).

Récanati, F. (1979). La transparence et l'énonciation, Editions du Seuil, Paris (in French).

Sadek, D. (1996). Le dialogue homme-machine: de l'ergonomie des interfaces à l'agent dialoguant intelligent. In: J. Caelen (Ed.), *Nouvelles Interfaces Homme-Machine,* OFTA, Paris: Tec & Doc (in French).

Searle, J. (1969). *Speech Acts: an Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Searle, J.R. (1975). Indirect Spech Acts. In: P. Cole and J.L. Morgan (Eds.), *Syntax and Semantics,* vol. 3: *Speech Acts,* pp. 59-82, New York, Academic Press.

Traum, D. and Hinkelman, E.A. (1992). Conversation Acts in Task-Oriented Spoken Dialogue. *Computational Intelligence,* vol. 8, n. 3.

Walker, M.A. (1996). Inferring Acceptance and Rejection in Dialogue by Default Rules of Inference. Language and Speech, 39-2.

Walker. M.A. (1994). Rejection by Implicature. In Proceedings of the 20th Meeting of the Berkeley Linguistics Society,

Incremental Construction of a Model of the Domain from a Set of Sentences by Means of Contextual Tableaux

Pablo Gervás

Departamento de Inteligencia Artificial,
Universidad Europea CEES,
c/ Tajo s/n,
Villaviciosa de Odón,
28670 Madrid, Spain
pg2@dinar.esi.uem.es

Abstract

During a dialogue, participants build up a model of the domain of predication. This model of the domain is used to interpret subsequent utterances. For instance, a mention of *the man with the limp* relies on locating such a man within the picture of the domain to be interpreted. Quantified constructions like *every man in the room* also rely on this picture of the domain to be interpreted correctly. The present paper concentrates on how this particular ingredient of context can be built up incrementally as a discourse is interpreted. This is done by providing a proof theory that integrates the explicit domain information given by existence statements with the implicit domain information contained in definite terms that give rise to a presupposition of existence. The proof theory is based on Contextual Tableaux, a modified version of semantic tableaux defined over an extended proof language that includes a record of domain information. Additional proof rules are provided for the presupposition of existence of definites, specially designed to account for the defeasible behaviour of presuppositions that originate in negative environments.

1. The Problems

During a dialogue, participants rely on a certain common ground to interpret what is actually said. A crucial ingredient of this common ground is a *model of the domain*: the set of objects that constitute the domain of the dialogue. This model of the domain is built up incrementally from clues contained in the dialogue, and it is then used to interpret subsequent utterances. For instance, a mention of *the man with the limp* relies on locating such a man within the picture of the domain to be interpreted. The present paper concentrates on how this particular ingredient of context can be built up incrementally as a discourse is interpreted.

Explicit information about the domain given by statements such as

(1) *There is a table.*

must be integrated with the information about the domain that one gets implicitly in a sentence like

(2) *The table is set.*

Both (1) and (2) carry information to the effect that the domain under consideration should include a table. In (1) this information is explicit (and communicating this information is the main aim of the sentence itself). This type of sentence is referred from here on as an *existence statement.* Existence statements can be seen as straightforward explicit instructions to add a given object to the picture of the domain. In (2) the information is implicit. In this case the information regarding the existence of a table is referred to as a *presupposition* of the sentence. In particular, as a presupposition of existence resulting from the use of a definite term. There are slight complications to be dealt with when negation is taken into account. If we take sentence

(3) *There isn't a table*

to be true, then sentence (2) can no longer be true. One can assume that sentences about objects that are not in the domain simply cannot be true. This is fine in the given case, but presents problems in trying to interpret sentences of the form

(4) *The table isn't set because there is no table.*

where one is expected to understand a sentence about a table at the same time as one is being told that there are no tables to talk about.

This same problem occurs in sentences like

(5) *Either there isn't a table or the table is not set*

where the different possible states of the world that might be referred to by the sentence involve different combinations of sentences (1) and (2) being true or false. This implies that the different interpretations of the sentence require not only different truth value assignments, but actually different definitions of the domain of predication.

During a dialogue, where each participant has to keep track of new referents introduced by other participants, a common model of the domain is required for all participants to understand the same information from the sequence of utterances. This model is built by each participant using both explicit and implicit information present in the dialogue.

The system developed in this paper models the way in which simple mention of an object as argument of a predicate conveys the 'presupposition' that it exists in the domain and does not need to rely on the presence of an existence predicate. The interpretation of such presuppositions of existence of terms must be addressed within a framework for constructing a logical structure from a sequence of utterances. The alternative views of the domain that appear when interpreting sentences like (5) can be matched with the possible logical interpretations of the sentence in a tree-like structure based on semantic tableaux (Smullyan, 1968).

2. Contextual Tableaux

The formalism presented here relies on a simplified representation language for the discourse under analysis - this representation includes means of representing information about existence and non-existence of terms –, a set of rules for the construction of contextual tableaux for a given discourse, and a definition of the interaction between the (semantic)information made explicit in the discourse and the implicit (pragmatic) information.

2.1. Representation Language

The study is carried out taking as reference *a fragment of natural language* built from declarative sentences and the connectives *if... then*, *and*, *or*, and negation (in its different syntactical forms as determined by the grammar).

The *representation language* used is a zero order logical language L with sorted domain and function constants, with a special predicate $\varepsilon(a)$ to denote existence of an object a on the domain.

A formula of L contains three types of information about the domain being considered: (1) implicit information about what exists (a mention of *the man* conveys the idea that there is a man in the domain), (2) explicit information about what exists (as in *there is a man*), and (3) explicit information about what does not exist (as in *there isn't a man*).

The differences between (1) and (2) affect the interpretation of the discourse after the addition of new sentences (updating) in two ways. On one hand, there are differences in the way implicit and explicit information in the new sentences survives when it is contradictory with previous information. This feature is captured by the model presented in this paper, and it is related to the *presupposition projection problem.* On the other hand, there are differences in the way implicit and explicit information in the representation already built survives when it is contradictory with new explicit updates. Taking this feature into account would require the addition of a Truth Maintenance System to handle the way in which implicit information that had already been interpreted is eliminated when it becomes contradictory with new explicit information. This is not considered in the present system. For this reason, it is not necessary to differentiate between types (1) and (2) of information once it has been interpreted.

In order to make it all simultaneously explicit in the proof theory, contextual tableaux provide an additional step of representation that extends semantic tableaux using two columns[1] for each branch: the first column lists all the terms corresponding to objects that exist (whether their existence is indicated explicitly or implicitly by the formula), the second column lists all the terms corresponding to objects that do not exist. Special domain expansion rules (see table 1) generate this representation using both explicit and implicit information.

$\neg - rules)$

$$\frac{\neg\neg P}{P}$$

$\alpha - rules)$ $\qquad\qquad$ $\beta^* - rules)$

$$\frac{P_1 \wedge P_2}{\begin{array}{c} P_1 \\ P_2 \end{array}} \qquad \frac{P_1 \vee P_2}{\begin{array}{ccc} P_1 & \neg P_1 & P_1 \\ P_2 & P_2 & \neg P_2 \end{array}}$$

$$\frac{\neg(P_1 \to P_2)}{\begin{array}{c} P_1 \\ \neg P_2 \end{array}} \qquad \frac{P_1 \to P_2}{\begin{array}{ccc} \neg P_1 & \neg P_1 & P_1 \\ P_2 & \neg P_2 & P_2 \end{array}}$$

$$\frac{\neg(P_1 \vee P_2)}{\begin{array}{c} \neg P_1 \\ \neg P_2 \end{array}} \qquad \frac{\neg(P_1 \wedge P_2)}{\begin{array}{ccc} \neg P_1 & \neg P_1 & P_1 \\ \neg P_2 & P_2 & \neg P_2 \end{array}}$$

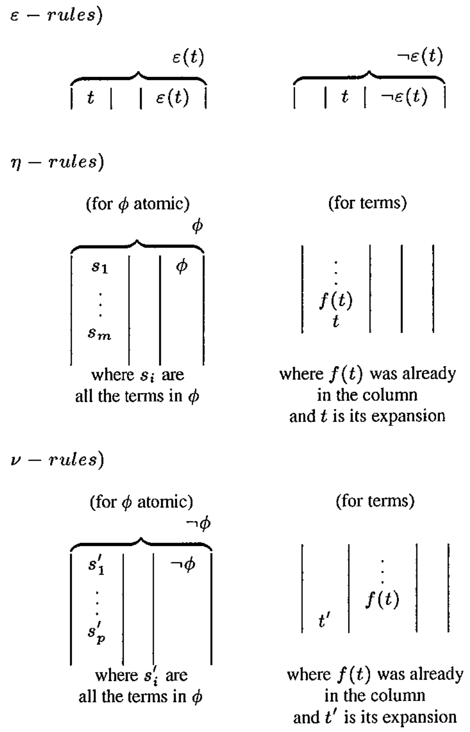


Table 1: Contextual tableaux rules.

2.2. Building Contextual Tableaux

Contextual Tableaux (Gervás, 1995) are a modified version of semantic tableaux (see (Smullyan, 1968) and (Fitting, 1989) for formal details) defined over an extended proof language that includes a record of domain information. Additional proof rules are provided for the presupposition of existence of definites, specially designed to account for the defeasible behaviour of presuppositions that originate in negative environments.

Given the transcription of a natural language sentence into the representation language described, a contextual tableaux representation is constructed[2] by application of the

[1] For ease of reference, an extra column is added to show in each case the formula from which the information was extracted.

[2] The present section includes only the basic elements of the

rules presented in table 1.

The difference between $\eta-$ and $\nu-$ rules accounts for the different behaviour of presupposition under the scope of negation (see (Mercer, 1987)). Domain information resulting from a negated term or formula is marked as t' rather than t, capturing the fact that it is defeasible.

A *contextual tableaux* for a formula X is a tree, whose points are (occurrences) of formulas, which is constructed as follows: (a) formula X is a tableau for X, (b) tableau Γ for X can be extended into another tableau Γ' for X by applying the expansion rules to any of the leaf nodes of Γ according to the following procedure:

- expand using all the rules (mark all additions resulting from $\nu-$ rules, if a term is marked, mark its expansion as well)
- once no more rules are applicable, retract those marked additions whose expansions contribute to the closure of a branch

A *branch of a tableau* Δ can be defined as the path from the root of the tree to one of the leaf nodes. Δ can also be interpreted as the set of atomic propositions found along that path.

For a given branch Δ of a tableau, the set of terms $EX(\Delta)$ is defined as the union of all the Ex columns that appear along the branch, and the set $NEX(\Delta)$ as the union of all the Nex columns that appear along the branch.

Information about the domain information affects the logical interpretation through the following extension of the definition of branch closure.

A branch Δ of a contextual tableau is *closed* if: (1) there are formulas ϕ and $\neg\phi$ in the branch, or (2) $EX(\Delta) \cap NEX(\Delta) \neq \emptyset$. A closed branch represents an inconsistent state of the world. A tableau is *closed* if all its branches are closed.

2.3. The Interaction between the Tableaux and the Domain Notations

The concept of branch closure for contextual tableaux is different from the traditional one because it takes domain information into account when considering branch closure. The procedure may be better understood by constructing the contextual tableau for a sentence such as:

(6) *If there is no king of France then France does not cherish its king.*

The contextual tableaux representation for sentence (6) is given in table 2.

The peculiar behaviour of implicit domain information in negative environments is apparent in this example. In this case, the proposition $\neg C(f, k(f))$ in branch C gives rise to a shorter expansion than it does in branch A. During the construction, a term $k(f)'$ resulting from the expansion of $\neg C(f, k(f))$ closes the branch – there is already a $k(f)$ in the NEX column of the branch. Since the term $k(f)'$ is marked, it is retracted. Any terms resulting from it are also ommited. In this way, the different local contexts of each branch determine different expansions for the same expression.

The EX column of the representation obtained in this way lists explicitly the terms that can be considered as part of the domain for that branch. This listing includes both those terms that had been explicitly stated to exist and those that had simply been presupposed to exist.

3. Comparison with the DRT Approach

The work of (Van der Sandt, 1992) argues that presuppositions – in general terms, including presuppositions of existence – should be read as anaphoric expressions with extra descriptive content. This extra descriptive content allows presuppositions to establish a reference marker in case discourse does not provide one. In this case, the lexical material is accommodated.

The framework provides predictions on the behaviour of presuppositions as determined by the place along the discourse structure where the presupposition is accommodated. These predictions include presuppositions of existence.

The approach of van der Sandt is presented within the framework of DRT (Kamp and Reyle, 1990). This framework provides a representation of discourse markers that can be easily used to model candidate discourse referents (including a binding operation between new markers and markers appeared before). Such a representation provides a structuring of the discourse in terms of nested DRSs over which the constraints can be easily defined as a search path.

In a DRS representation, discourses are divided into two different sets of ingredients: a set of discourse referents (the universe of discourse), and a set of conditions to be satisfied by these referents. Indefinite NPs introduce discourse markers into the universe of the DRS. These markers then serve as referents. Conditions assign properties to the members of the universe of discourse. Anaphoric elements are encoded separately in a DRS. They have to be incorporated into the structure by a process of resolving the anaphoric expressions. Anaphoric constructions constitute an instruction to look for the appropriate referent (a marker satisfying conditions equal or compatible with those of the anaphoric expression) somewhere earlier in the discourse.

This resolution can involve two different processes. If a referent is found earlier in the discourse, anaphoric binding takes place. The corresponding discourse markers are linked by putting the appropriate equations and the conditions associated with the anaphoric expression are transfered to the binding site. If no referent is found both the markers and the conditions are added to the structure at the accommodation site.

On the issue of domains, the DRS translation is evaluated over the domain that its quantifiers create. An initial empty domain is assumed, and objects in the domain are added to it on encountering a quantifier.

Considering that a DRS is actually a representation of the logical structure of a natural language sentence in terms of the natural language connectives that appear in it, there is a certain analogy between DRT's definition of discourse markers of a particular DRS and my definition of a local domain for a tableau branch.

formal system required to present the most elementary ideas.

$$\neg\varepsilon(k(f)) \to \neg C(f,k(f))$$

$\varepsilon(k(f))$ $\neg C(f,k(f))$			$\varepsilon(k(f))$ $C(f,k(f))$			$\neg\varepsilon(k(f))$ $\neg C(f,k(f))$		
$k(f)$		$\varepsilon(k(f))$	$k(f)$		$\varepsilon(k(f))$		$k(f)$	$\neg\varepsilon(k(f))$
f			f			f'		
f'		$\neg C(f,k(f))$	f		$C(f,k(f))$	f'		$\neg C(f,k(f))$
$k(f)'$			$k(f)$					
f'			f					
(Branch A)			(Branch B)			(Branch C)		

Table 2: Contextual Tableaux for (6) *If there is no King of France then France does not cherish its king.*

For a sentence (6) given above, the corresponding DRS is shown in table 3.

<table>
<tr><td><b>x</b></td></tr>
<tr><td><b>France(x)</b>
<table><tr>
<td>¬</td>
<td><table><tr><td>y</td></tr><tr><td>king(y)
poss(x,y)</td></tr></table></td>
<td>→</td>
<td><table><tr><td><b>z</b></td></tr><tr><td><b>king(z)</b>
<b>poss(x,z)</b>
¬ cherish(x,z)</td></tr></table></td>
</tr></table>
</td></tr>
</table>

Table 3: DRS for sentence (6).

The DRS representation does not allow easy representation of the information equivalent to that apparent in the set NEX. The presupposition (that there is a king of France; shown in bold) has had to be accommodated locally because otherwise the constraints would have been violated. This ensures that sentence (6) is (correctly) predicted not to presuppose the existence of a king of France. But this leads to the appearance of two different discourse markers for *the King of France* in the DRS (y and z). One of these discourse markers refers to the king that exists in one of the alternative interpretations of the sentence. The other one does not refer at all, but rather is only used to state the non-existence of the king in the other interpretation of the sentence. The information contained in the DRS is still basically the same as in the contextual tableaux. But the tableau framework makes it explicit what the difference is between these two syntactic occurrences of the condition 'king' applied to discourse markers and has this information available explicitly to be used in inference. Use of this information in inference may result in closure of a branch (as would occur in branch C of the tableaux in table 2 if the existence of such a king is later asserted).

4. Conclusions

Contextual tableaux provide a good formalization of the model of the domain constructed during interpretation of a set of sentences.

Using contextual tableaux, this model of the domain is constructed incrementally, and all the information of previous utterances is automatically stored and used in the interpretation of subsequent ones[3].

The rules that govern contextual tableaux are simple to implement in a practical system. Because they continously refer to the previous context, the complexity of the construction process increases with the size of discourse being represented, a disadvantage shared by all systems that apply thorough semantic analysis. The advantage of using model building is that one need not prove again at each step the consistency of the previous context. In order to improve efficiency over long fragments of discourse, contextual tableaux might be extended with some means of restricting the construction process to obtain a subset of the representation that is relevant to the interpretation of the next sentence in the discourse.

Contextual tableaux may have an interesting role to play in practical dialogue systems that satisfy the following conditions:

- a specific domain is being talked about,
- a simple and precise model of which objects are being talked about in each session is required, and
- the discourses involved do not exceed a reasonable length.

5. References

M. Fitting. 1989. *First Order Logic and Automated Theorem Proving*. Springer Verlag.

P. Gervás. 1995. *Logical Considerations in the Interpretation of Presuppositional Sentences* . Ph.D. thesis, Department of Computing, Imperial College of Science, Technology and Medicine, London.

H. Kamp and U. Reyle. 1990. *From Discourse to Logic*. Kluwer.

R. E. Mercer. 1987. *A Default Logic Approach to the Derivation of Natural Language Presuppositions*. Technical Report 87-35, Department of Computer Science, University of British Columbia, Vancouver, B.C., Canada.

R. Smullyan. 1968. *First Order Logic*. Springer Verlag.

R. A. Van der Sandt. 1992. Presupposition Projection as Anaphora Resolution. *Journal of Semantics*, 9:333–377.

[3] This requires the tableaux to be used as a representational device as well as a decision method. Although this is not an orthodox approach, it is shown to give good results.

Display Acts in Grounding Negotiations

Yasuhiro Katagiri* **and Atsushi Shimojima**†

*ATR Media Integration & Communications Research Laboratories
2–2–2 Hikari Seika Soraku Kyoto, 619-0288 Japan
katagiri@mic.atr.co.jp

†Japan Advanced Institute of Science and Technology
1–1 Asahi Tatsunokuchi Nomi Ishikawa, 923-1292 Japan
ashimoji@jaist.ac.jp

Abstract

Display utterances are those utterances, such as verbatim rephrases or cooperative completions, that exhibit the speakers' construals of their partners' previous utterances. Drawing on both empirical analyses and theoretical considerations, we study the ways they perform the acts of acknowledgment and repair-request. We show that (1) these grounding acts, when performed by display utterances, are *not* autonomous acts whose choices are under the control of the speakers, yet (2) they are instantaneous acts whose types are fixed at the time of utterance. Accordingly, the grounding model proposed in this paper reconciles the non-autonomous nature of display utterances (Clark, 1996) with the necessity for on-line classifications of the grounding acts (Traum, 1994).

1. Introduction

What one says may not be immediately shared in a natural conversation due to various possibilities including communication error and disagreement. The process of grounding (Clark and Schaefer, 1989; Traum, 1994) is a dialogue process in which a piece of information contributed by one speaker becomes established in the common ground between dialogue participants. Each utterance made in a dialogue is considered to perform a particular *grounding act*, depending on the contribution it makes to the overall process. For example, Clark and Schaefer (1989) characterize the fundamental structure of a grounding process as a sequence of two grounding acts, called *presentation* and *acceptance*; Traum (1994) proposes a more fine-grained classification of grounding acts, including *initiation, continuation, acknowledgment*, and *cancel.*

Given this view, it is but a short step to make the following assumption on the nature of grounding acts:

Autonomy: The type of grounding act performed by an utterance is under the control of the speaker.

In this paper, we look at the class of utterances that have been called "displays" (Clark, 1996), and show that the proper treatment of their grounding functions calls for a revision of this autonomy assumption. An utterance is a *display utterance* if it somehow exhibits a part of what the speaker has construed out of a preceding utterance by a different speaker. Therefore, the grounding act attributed to a display utterance is typically an acknowledgment and sometimes a repair-request. Drawing on an analysis of our data on language-prosody interaction in "echoic responses" in real dialogues, we show that the kind of grounding act being performed by a display utterance depends on contextual factors beyond the speaker's access or control, and hence the acts of acknowledgment and repair-request, when performed with display utterances, are not generally autonomous acts.

We then propose a new way of looking at the grounding functions of display utterances, which replaces the autonomy assumption with the following slightly different assumption:

Instantaneity: The type of grounding act performed by an utterance is fixed at the time of the utterance.

Clark and Schaefer (1989) and Clark (1996) emphasize the non-autonomy or jointness of grounding processes; one of the main appeals of the grounding model in Traum (1994) is that it retains the instantaneity of grounding acts in the above sense. The view to be proposed in this paper demonstrates that autonomy and instantaneity are different matters, and giving up one does not entail giving up the other—in other words, we can pursue Traum's program while doing justice to the jointness of certain grounding acts.

2. Correctness of Construals

A quintessential example of a display utterance is an *echoic response*. In the following example, B's response directly displays "on Chestnut Street" as a part of what B has extracted out of A's preceding utterance.

(1) A: Sue's house is on Chestnut Street.
B: On Chestnut Street.

A *collaborative completion* can also be considered as a display:

(2) A: The next meeting will be Tuesday next week
B: at the same time in the same room, right?
A: Yeah.

Unlike the case of an echoic response, the information displayed in B's utterance has never been explicitly stated by A; it is instead something B has inferred from A's somewhat incomplete utterance, as a part of what A had intended to convey.

A display of one's construal may be done rather indirectly, as in the following example of a third-turn repair:

(3) A: Did Mary go to the party?

B: I didn't see her last night.
A: No, I meant John's party on last Thursday.

Here, *B*'s utterance displays her construal that *A*'s preceding question refers to the party last night, and this is why A can ever correct *B*'s construal of the intent of her preceding question.

Our notion of display encompasses what Clark and Schaefer (1989) call "display" and "demonstration"; Traum (1994) discusses display utterances mainly as a way to perform an acknowledgment; furthermore, our notion is almost coextensional with "public display" envisioned by Clark (1996). Thus, display acts have often been at the center of attention in discussions on grounding phenomena. Yet, the exact ways they perform their grounding acts have hardly been incorporated into a systematic theory of grounding.

In fact, a problem occurs if we try to understand the grounding functions of display utterances with a straightforward application of the grounding model of Clark and Schaefer (1989). According to Clark and Schaefer, a speaker *B* is said to *accept* an utterance u by a speaker *A* under the following conditions:

1. *B* gives evidence e' that he believes he understands what *A* means by u.
2. *B* does so on the assumption that, once *A* registers evidence e', he will also believe that *B* understands.

Clark and Schaefer cite display utterances as one of the major ways of performing an acceptance, but if we strictly apply the above conditions, few instances of display can actually perform an acceptance. Take (1) for example. Here, *B* may believe that he understands what *A* means by his utterance, and the timing and prosody of *B*'s echoic response may indeed signal this belief on *B*'s part. Therefore, *B*'s utterance may satisfy condition 1 above.

Yet, it is unlikely that condition 2 holds at the time of this utterance. To *B*, there is always a possibility that his echoic response may be an incorrect repeat of *A*'s original utterance. If that is the case, *B*'s utterance will certainly fail to lead *A* to believe that *B* understands; it may even lead *A* to believe that *B doesn't* understand. Accordingly, unless *B* is sure that he has correctly repeated *A*'s utterance, *B* cannot assume that her utterance will lead *A* to believe that *B* understands. As Clark (1996) shows, however, a display utterance is essentially a part of a joint-construal activity, where a speaker displays her construal of a previous utterance to allow another speaker to check its correctness. Clearly, *B* would not engage in such an activity if *B* were sure that his construal of *A*'s utterance is correct. Condition 2 is therefore hardly satisfied by a display utterance.

The crucial point is that a display utterance always comes with the risk of exhibiting an incorrect construal of a previous utterance, while the correctness of the displayed construal is usually beyond the control of the speaker. (Condition 2 goes against this nature of display utterances by effectively requiring the speaker to know the correctness of her construal.) Now, generally, in order for a display utterance to function as an acknowledgment, the displayed construal must be correct; furthermore, an utterance that exhibits an incorrect construal almost always prompts a repair from the original speaker. This shows that, when display utterances are involved, the act of acknowledgment and the act of repair-request are *not* autonomous acts: the speaker has no absolute control over which act she is performing with her display utterance, inasmuch as she cannot perfectly predict whether she is demonstrating understanding or misunderstanding.

Note that such a radical uncertainty does not exist if the speaker uses a more explicit form of acknowledgment or repair-request. Consider the following modifications of example (1):

(4) *A*: Sue's house is on Chestnut Street.
B: Uh huh. (or *B*: What?)

Unlike a display utterance, the "uh huh" in (4) does not show the content of *B*'s construal of *A*'s utterance; *B* withholds that information, and accordingly gives no concrete clue for *A* to evaluate the correctness of *B*'s construal. Therefore, *A* cannot help but rely on the convention that "uh huh" is uttered only when the speaker understands (or believes to understand) the preceding utterance. For this reason, *B* can assume that his utterance of "uh huh" leads *A* to believe that *B* correctly understands, satisfying condition 2 in Clark and Schaefer's model (1989). Also, precisely because the content of *B*'s construal is withheld, there is no chance that it is revealed to be incorrect. Accordingly, barring exceptional circumstances, *B*'s utterance of "uh huh" never functions as a repair-request. For a parallel reason, *B* can be confident that that her utterance serves as a repair-request rather than an acknowledgment when she utters "What?" In both instances, *B* has control over what type of grounding act she is performing with her own utterance.

3. Integration of Construals

Clearly, whether the construal exhibited in a display utterance is correct is a dominant factor that affects the type of grounding act being performed. But it is certainly not the only factor: the degree in which the speaker is convinced of her construal, often signaled by the timing and prosody of her speech, also seems to be a strong factor.

Consider the echoic response in (1) again. Intuitively, if it were made in a falling tone without any delay, *A* would feel that *B* has integrated the displayed construal ("on Chestnut Street") well in her body of knowledge; *A* would be sure of the success of grounding and perhaps go on to the presentation of the next item of information. On the contrary, if *B*'s response were made in a rising tone with a considerable delay, *A* would doubt that *B* has adequately integrated the information; *A* would be prompted to restate or rephrase the information to supplement the grounding failure. This suggests that, even when confined to the case where the displayed construal is correct, an echoic response shifts its grounding function between an acknowledgment and a repair-request, *depending on* the speaker's integration signaled by the timing and prosody of the utterance.

In fact, our previous studies on the functions of echoic responses (Shimojima et al., 1998; Shimojima et al., 1999) lend empirical supports to this intuition. We conducted a corpus-based observational study and three experiments in the following procedures:

Observational study Instances of echoic responses were extracted from three samples of task-oriented spoken dialogue data, and the correlation between their temporal and prosodic features and the raters' assessments of the speakers' integration were examined.

Experiment 1 Acoustically manipulated speech samples of echoic responses were presented to subjects who were asked to evaluate the speakers' integration.

Experiment 2 Instances of echoic responses extracted from our corpus were presented to subjects who were asked to judge the grounding functions being performed. They were to choose from "acknowledgment" and "request repair."

Experiment 3 The same stimuli were used, while the subjects who were asked to judge the most appropriate response to each instance.

The observational study and Experiment 1 provided an evidence that the prosodic and temporal features of an echoic response indeed signal the degree of the speaker's integration. Specifically, a long delay, a rising boundary tone, a high pitch, or a low tempo indicates a high integration, a short delay, a falling boundary tone, a low pitch, or a high tempo indicates a low integration.

Furthermore, Experiments 2 and 3 showed the correlation between the integration rates associated with echoic responses and the subjects' judgments on their grounding functions: when the integration is high, the subjects tend to take it as an acknowledgment and to choose a response appropriate to an acknowledgment, whereas in the case of a low-integration echoic response, the subjects tend to take it as a repair-request and to choose a repairing response.

Now, the temporal and prosodic signals of the speakers' integration are largely *spontaneous*. Of course, there are cases where a speaker deliberately produces an echoic response with particular prosody in a particular timing, but that must be exceptional, just as a deliberate expression of anger with a face color is exceptional. If so, those temporal and prosodic signals bring in another factor that can break the autonomy of the grounding acts performed with display utterances: even if a speaker ever intends to perform an acknowledgment with an echoic response, its timing and prosody may reveal the low integration on the speaker's part and thus make the utterance function as a repair-request; of course, the opposite is also possible. In either case, the kind of grounding act to be performed by an echoic response is not under the speaker's control.

4. Toward a Dynamic Model for Display Acts

One of the main motivations for Traum (1994) to have developed his own model of grounding is that in the Clark-Schaefer model, it is hard to determine the grounding function of a given utterance "on line," without having to refer to a subsequent development of the dialogue. However, we have just found that the grounding act performed with a display utterance, may it be an acknowledgment or a repair-request, is non-autonomous in two counts. Thus, one may be tempted to think that it cannot be determined on line either. In this view, the grounding act performed by B's response in (1) is fixed only after A responds to it: if A's response is "Yeah, Chestnut Street" (a restatement), it *makes* B's utterance a repair-initiation, whereas if A's response is "And then" (an initiation of new information), it *makes* B's response an acknowledgment.

In the following, we describe a model of display utterances that does justice to the non-autonomy of their grounding functions without being committed to this radical jointness view. We propose redefining the conditions for the grounding acts of acknowledgment and repair-request as follows:

Acknowledgment: In response to A's utterance u, B gives sufficient evidence e that she has correctly identified what A meant by u.

Repair-request: In response to A's utterance u, B gives sufficient evidence e that she has failed to identify what A meant by u.

In both cases, A may not be aware of her identification or lack thereof.

It is helpful to take a layered picture on acts to develop a model that captures the non-autonomous nature of grounding acts. An utterance by A of a certain expression such as "uh huh," "yeah," or "ok" counts as an acknowledgment act by A towards a proposition p under a certain context, e.g., when it is produced as a response to B's utterance of p. This "counts as" relation between acts can be captured by Goldman's *action generation* relation (Goldman, 1970). An act α is said to generate another act β under an appropriate contextual condition C. The same act α, however, may generate a different act β' under a different contextual condition C'. Even though α is in complete control of A, it can generate another act β, which, under a certain context, might not be within her scope of intention.

We characterize a display utterance by the following three components:

(a) the *target* of construal: the dialogue object the display is directed at,

(b) the *content* of construal: what is being displayed, and

(c) the *result* of construal: what cognitive and emotional state one is in.

For the echoic response in (1), the target of B's echoic response is A's utterance of "on Chestnut Street" in the preceding turn, and the content is B's response itself. The result is the integration rate indicated by the prosody that accompanies B's echoic response. Here, the target and the content can be their surface phonological sequences or they can be the semantic contents they express.

The picture we are proposing is (1) a display act α is a lower level autonomous act, characterized by the content and the result components of a display utterance, (2) a display act α generates a grounding act β in a context characterized by the target component, and (3) which grounding

Generating Act (α)		Context	Generated Act (β)
Content	Result	Target	
"uh huh"		following p	acknowledgment
"what?"		following p	repair request
display p	High	following p	acknowledgment
display p'	High	following p	repair
display p	Neutral	following p	acknowledgment
display p'	Neutral	following p	repair request
display p	Low	following p	repair request
display p'	Low	following p	repair request

Table 1: Grounding acts generated by echoic responses.

act is generated is determined instantaneously, but the generated act can be non-autonomous. In the case of echoic responses, a locutionary act of producing an echoic response, characterized by its content and result, generates a certain grounding act under a certain contextual condition, characterized by the target of the echoic response. Table 1 summarizes the types of grounding acts generated out of echoic responses depending on these three parameters. Utterances of typical grounding-oriented expressions are also included in the table for comparison.

In the case of "uh huh" produced after an utterance of p by a partner, it is guaranteed, by the linguistic convention of English, that it serves as evidence that the speaker of "uh huh" believes that she has identified what the previous speaker meant, namely p. Unless further evidence to the contrary is available, it also serves as evidence that the speaker actually identified p. Since this satisfies the condition of an act of acknowledgment, the locutionary act of producing "uh huh" generates an act of acknowledgment. Note that the act of acknowledgment here is both autonomous and instantaneous. Similarly, the locutionary act of producing "what?" generates an autonomous and instantaneous act of repair-request because of the linguistic convention associated with the expression.

An echoic response with high integration prosody provides, if it is a correct echo, two independent pieces of evidence for positive identification, and generates an act of acknowledgment. If it is an incorrect echo, it provides incoherent pieces of evidence. The negative evidence directly obtained from the content defeats the positive evidence, and it becomes an act of repair. An echoic response with neutral prosody, on the other hand, lacks one source of evidence and provides either positive or negative evidence for the identification of the target in the previous turn, and, hence, can generate an act of acknowledgment or an act of repair-request. An echoic response with low integration prosody provides negative evidence for identification, and generates an act of repair-request [1].

The choice of giving up on autonomy in favor of instantaneity for grounding acts might find its support when we consider our daily face-to-face conversations. Most non-linguistic signals are spontaneous displays of one's cognitive and emotional states, which are out of the intentional control of the speakers; people nonetheless invariably exploit these signals to navigate through the course of a conversation. Prosody, which we have found to signal the speaker's integration level in echoic responses, is normally a spontaneous feature of speech. The fact that prosody plays a significant role in grounding by itself shows that grounding acts are non-autonomous.

5. Conclusions

In this paper, we argued that the kind of grounding act being performed by a display utterance depends on two major contextual factors, namely, the correctness of the displayed construal and the degree of the speaker's integration as signaled by the characteristics of her speech. Since both factors are generally beyond the speaker's control, we concluded that the acts of acknowledgment and repair-request, when performed with display utterances, are not autonomous acts.

We then proposed a generation model of the grounding functions of display utterances, which sets up multiple layers of generated grounding acts for a single display utterance. In this model, whether an utterance performs an autonomous grounding act depends on which layer of grounding act the question is directed at.

Under our conceptions of the acts of acknowledgment or repair-request, however, whether a display utterance performs one of these acts is fixed at the time of the utterance, without reference to the subsequent development of the dialogue. In other words, the acts of acknowledgments and repair-requests performed by display utterances retain instantaneity without being autonomous. Thus, this paper suggests a way of reconciling the non-autonomous nature of display utterances (Clark, 1996) with the necessity for on-line classifications of the grounding acts (Traum, 1994).

6. References

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

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

Alvin I. Goldman. 1970. *A theory of human action*. Princeton University Press, Princeton, NJ.

Atsushi Shimojima, Hanae Koiso, Marc Swerts, and Yasuhiro Katagiri. 1998. An informational analysis of echoic responses in dialogue. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, pages 951–956.

Atsushi Shimojima, Yasuhiro Katagiri, Hanae Koiso, and Marc Swerts. 1999. An experimental study on the informational and grounding functions of prosodic features of Japanese echoic responses. In *Proceedings of the ESCA Workshop on Dialogue and Prosody*, pages 187–192.

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

[1] There appears to be asymmetry between direct evidence and indirect evidence. Indirect evidence provided by prosody generally takes precedence over direct evidence, and only negative direct evidence overrides indirect positive evidence, but not vice versa. This is probably caused by the difference between two types of negativities, the lack of identification and misidentification.

Using dialogue information for parsing in a spoken dialogue system

Rob Koeling, SRI Cambridge
koeling@cam.sri.com

May 24, 2000

1. Introduction

Previously we have developed a grammar for the user inputs to a spoken dialogue system for the public transportation information domain. After evaluating the grammar with real world input and optimising the grammar, a number of user utterances were still not analysed correctly. That is, although a better hypothesis was available, the parser favoured some other. In this paper we investigate the hypothesis that (at least some of) these errors can be avoided by taking the context in which some user reply is uttered into consideration. By context we mean dialogue context and especially the question that preceded the user reply. In this project we work on a spoken language system that provides information about the timetable of the Dutch railways. Although the user is allowed to say everything, the system has the initiative in the dialogue and will always confront the user with questions. This makes the dialogues simple and most user answers quite short.

The starting point of our group working on the NLP module of the system is the output of the speech recogniser. We do not have any influence on the performance of the speech recogniser. Our task is to find the best hypothesis within the wordgraph representing all hypotheses suggested by the speech recogniser. A wordgraph is a compact representation of all hypotheses of what the user might have said given by the speech recognizer. To find a path in the wordgraph we begin at the start node and from there we will have to decide which word to choose next. That decision is made on basis of several information sources: the acoustic score assigned by the speech recogniser, syntactic coherence of the hypothesis and the proability given by the *n-gram* language model.

What is important for this paper, is the fact that a wordgraph is only a compact representation of hypotheses of what the user *might* have said. The actually uttered sentence might not be among them. Therefore, it is often not obvious which hypothesis is closest to what was uttered. Robust processing is necessary and all the information sources that are available to guide us through the wordgraph are welcome. So far we have restricted us to the three above mentioned information sources. There is, however, a fourth source of information that, although it is regarded as very important information in human-human dialogues, is not often considered in natural language understanding at this moment: the (linguistic) context in which the sentence was uttered. When we look at some incorrect analyses, we sometimes find that a better hypothesis is available and moreover that there are cases where the favoured hypothesis does not make much sense as an answer to the preceding system question. There may be many reasons why the 'most suitable' (from a semantic point of view) hypothesis is not chosen: just because there is an alternative that is assigned a very high probability by the trigram model (especially in case of short utterances), it might be the case that the actual utterance is not a grammatical one or the actual utterance does not span a path in the wordgraph from start state to final state (i.e. information must be thrown away to choose this hypothesis). But if the system question requests a particular kind of information (e.g. it asks for information about destination station), there is no way to express the preference for a answer that is semantically close to the question. We have previously investigated how to exploit this information using a knowledge based approach, in this paper we will investigate how we can capture the relation between dialogue context and user utterance in a statistical model.

1.1. Suggested approach

In this paper we explore a way to integrate dialogue knowledge with the trigram language model already used in the system. We want to include dialogue knowledge in such a way that it interacts with the information supplied by the previous words (very local information) in a statistical model. We suggest to use a Maximum Entropy model that includes features that model the trigram knowledge, the dialogue knowledge (in this case only the type of the question that preceded the user utterance) and the combination of both knowledge types.

1.2. Other approaches

Another method that might be used to reach the same goal is training different trigram models for every question type. The obvious disadvantage of such a method is the undesirable consequence that you have less data at your disposal to train your language model on. A second disadvantage is very limited flexibility in adapting your model. A second alternative approach might be to define a traditional trigram model and a context model and to combine these models with *interpolation*. A third possible approach is to come up with the n-best solutions given an n-gram model and re-rank the solutions using the context model (Manny Rayner and Price, 1994). The latter two solutions have the disadvantage that it is difficult to account for interaction between the knowledge sources. Similar work as presented here can be found in (Khudanpur and Wu, 1999a; Khudanpur and Wu, 1999b; et al., 1997).

1.3. Why use such heavy machinery?

The use of a sophisticated statistical framework as Maximum Entropy modelling might seem to be an overkill for the problem we want to tackle in this paper. However, in this paper we use a simple example because we only want to illustrate want could be done with this approach. We suggest that this line of work is:

- *Simple.* The MaxEnt framework takes care of the integration of the different knowledge sources. A linguist interested in investigating the effects of incorporating dialogue knowledge in a language model can just concentrate on the interesting knowledge sources instead of how to integrate them into a model. The model described here is created with off-the-shelf software (Dehaspe, 1997).
- *General.* The approach is not limited to the (simple) spoken dialogue system that is used here.
- *Extendible.* More sophisticated dialogue knowledge could be included in the model.
- Not fragmenting available training data.
- *Performing very well.*

However, the traditional problem with MaxEnt models is also encountered here: MaxEnt models are computationally expensive. We will argue later that for the simple example described here this does not have to be a problem, but it might very well cause more problems when the model becomes more complex.

In this paper we want to explain the general line of our approach and why we think spoken dialogue systems in general can benefit from this work. We will concentrate on *how* dialogue information is included and in what way the model that is used as an illustration in this paper can be extended for this or other applications. A more technical account of the approach and comparison with other (simpler) approaches is the subject of a different paper.

2. Maximum Entropy models

Maximum Entropy (MaxEnt) models (Jaynes, 1957) are exponential models that implement the intuition that if there is no evidence to favour one alternative solution above another, both alternatives should be equally likely. In order to accomplish this, as much information as possible about the process you want to model must be collected. This information consists of frequencies of events relevant to the process. The frequencies of relevant events are considered to be properties of the process. When building a model we have to constrain our attention to models with these properties. In most cases the process is only partially described. The MaxEnt framework now demands that from all the models that satisfy these constraints, we choose the model with the flattest probability distribution. This is the model with the highest entropy (given the fact that the constraints are met). When we are looking for a conditional model $P(w|h)$, the MaxEnt solution has the form:

$$P(w|h) = \frac{1}{Z(h)} \cdot e^{\sum_i \lambda_i f_i(h,w)}$$

where $f_i(h,w)$ refers to a (binary valued) feature function that describes a certain event; λ_i is a parameter that indicates how important feature f_i is for the model and $Z(h)$ is a normalisation factor.

In the last few years there has been an increasing interest in applying MaxEnt models for NLP applications (Ratnaparkhi, 1998; Adam L. Berger and DellaPietra, 1996; Rosenfeld, 1994; Ristad, 1998). The attraction of the framework lies in the ease with which different information sources used in the modelling process are combined and the good results that are reported with the use of MaxEnt models. In this paper we want to exploit these advantages, by combining traditional n-gram information sources with information that is not necessarily in the local context. Another strong point of this framework is the fact that general software can easily be applied to a wide range of problems.

3. Trigrams with triggers

3.1. A MaxEnt trigram model

Let us first define the trigram features for the model. We are looking for the probability of w_i, given the fact the we have just seen w_{i-2} and w_{i-1} ($P(w_i|w_{i-2}, w_{i-1})$). In

$$\begin{array}{ll}
P(w_{i-2}, w_{i-1}, w_i) = f(w_{i-2}, w_{i-1}, w_i) & \\
 & \text{if } C(w_{i-2}, w_{i-1}, w_i) \geq K \\
P(w_{i-1}, w_i) = f(w_{i-1}, w_i) & \text{if } C(w_{i-1}, w_i) \geq K \\
P(w_{i-2}, w_i) = f(w_{i-2}, w_i) & \text{if } C(w_{i-2}, w_i) \geq K \\
P(w_i) = f(w_i) & \text{if } C(w_i) \geq K
\end{array}$$

Where C denotes the count function, f the relative frequency and K is some threshold.

Figure 1: Definition of the trigram constraints

figure 1 the constraints are given to define the MaxEnt trigram model. We use the information sources listed in figure 1 above when we have seen enough events in the training material to justify its reliability (i.e when the count of a particular event in the training material is higher than some threshold). What these constraints say is that the probability of some event is the observed relative frequency of this event in the training material. Attention is constrained to models that also assign this probability to these events.

3.2. Adding triggers

A nice feature of the MaxEnt model is the opportunity also to include information from other knowledge sources. Rosenfeld (Rosenfeld, 1996) (who uses his model for topic adaptation) reports very good results on the inclusion of *trigger* features. Trigger features are features that are active if there is some event in the history. For example, if we have a series of words: $h_5, .., h_1, w_i$, we could define a trigger feature that fires when the current word is *w'* and *one of the* preceding 5 words is *h'*:

$$f_j(h_i, w_i) = \begin{cases} 1 & \text{if } w_i = w' \wedge h' \in h_5 ... h_1 \\ 0 & \text{otherwise} \end{cases}$$

The attractive idea about adding trigger features is the fact that *any* type of information can be considered to be included in the model. In a spoken dialogue system, for example, information outside the wordgraph that is parsed can be included. In this application we will explore how to add dialogue state information. The simplest and probably most informative knowledge source will be (in this type of system in which the system asks the questions and the user can't do much more than addressing these questions), the question that preceded the utterance that currently is parsed.

Now we will expand the trigram model by adding features that deal with other information then the last two words. In this paper we will discuss experiments that only use the simplest trigger we could think of: the type of the (system) question that preceded the (user) utterance. The reported results are very promising and suggest that it is useful to look beyond the wordgraph.

3.3. Data

The data we use for training and testing consists of recorded dialogues of people using an early version of the system. The training material consists of a little more than 100.000 question/answer pairs. As mentioned before, people tend to give short answers. This results in an average sentence length of about 3.5 words. The lexicon contains 1650 words. The test corpus consists of 1000 wordgraphs (average size about 50 transitions per wordgraph) with their corresponding system questions.

3.4. Experimental results

Although the grammar was performing quite well already there was still some space for improvement. We calculated what the Word Accuracy (WA), Sentence Accuracy (SA) and Concept Accuracy (CA) would be if we had an oracle giving us the path in the wordgraphs that is as close to the actual uttered sentence as possible. This method is called 'possible' in figure 2. The 'nlp_speech_trigram_keep' method (see (van Noord et al., 1999) for an extensive description of this method) gives us the lower bound of the results. The 'possible' method gives us the upper bound. There is a gap of about 6% between the upper and the lower bound. When we talk bout *error reduction* in what follows, we will refer to the percentage of this gap that is bridged. The numbers for the method '+context' in figure 2 are the result of training a model with the feature setup given in figure 3. Apart from the features we used for the MaxEnt trigram language model, we added 2 features for our '+context' model: one that relates just the type of the previous system question with the current word (the context feature) and one that combines the proper trigram feature with the context feature. Features are selected by using a threshold of 1 for all the features. The resulting feature set consists of 90660 features. The model is trained over 50 iterations. Other feature setups were tested, but no significant improvement could be found.

Feature	History
$(word_{i\_2}/word_{i\_1}, word_i)$	: Previous 2 words
$(word_{i\_2}, word_i)$	: Word 2 to the left
$(word_{i\_1}, word_i)$	: Previous word
$(qtype, word_i)$	: Type of corresponding system question
$(qtype/word_{i\_2}/word_{i-1}, word_i)$	: Previous 2 words and question type

Where Qtype ∈ $whq\_loc, whq\_temp, ynq\_reg,$
ynq_conf, ynq_other

Figure 3: Feature setup for best scoring model

Although, with a error-reduction of 14% (WA); 14% (SA); 23% (CA), the '+context' model performs really well, we have a high computational price to pay. The average cpu-time needed to parse a wordgraph increased enormously. For this particular application this was not really necessary. Because of the fact that we only have to deal with 5 question types, we can have (at most) 5 times as many (question type dependent) trigram probabilities. In this case it will still be possible to compute all the probabilities off-line and simply look them up when needed (as is done for a 'conventional' trigram model). We will run into this problem, however, if we want to benefit from the flexibility that we claimed in section 1. Some serious work is needed to decrease computing time for the general case.

Some observations of the results

In order to get an idea where the better results come from, we have examined in more detail all those parsed wordgraphs where the analysis of the method 'nlp_speech_trigram_keep' deviated from the analysis of the method '+context'. We have mainly focused on those analyses that deviated in Concept Accuracy. Not all the changes improved the performance. In 11 cases we found genuine success. These cases (an example is given in figure 4) show exactly the behaviour that we were looking for. Most of the examples with an improved CA were short utterances, that were represented by rather large wordgraphs.

System Question: (whq_loc) ::
Naar welk station wilt u vanuit Apeldoorn reizen?
To which station do you want to travel from Apeldoorn
Uttered :: Boskoop *(= Dutch city name)*
Trigram :: Half twaalf *(Eleven thirty)*
+Context :: Boskoop

Figure 4: Example of succes

The other interesting class of difference in performance are those examples where it seems that an hypothesis is chosen only because it fits best the type of the corresponding question. We found, however, very few of these examples. The most prominent example is given in figure 5. In this example the user tries to answer the question *and*, at the same time, to correct the information the system supplied in order to check if it was well understood. This is

Method	WA (%)	SA (%)	CA (%)				
			match	prec	recall	CA	
possible	90.4	83.7				90.0	
nlp_speech_trigram_keep	84.453	77.4	82.40	85.76	87.06	83.70	1000
+context	85.259	78.3	84.20	87.21	88.18	85.15	1000

Figure 2: The results of the best setup compared with our baseline method and with the oracle

of course perfectly acceptable in a dialogue, but it causes the problem here. Even while the actual uttered sentence is available in the wordgraph, apparently there is an hypothesis that fits even better a *whq_temp* type of question. Although this problem occurred only once in this test-set, it is clear that this is the behaviour that must be avoided.

System Question: (whq_temp) ::
Op welke dag wilt u vanuit utrecht CS naar hengelo reizen?
(What day do you want to travel from Utrecht CS to Hengelo?)
Uttered :: Op maandag, maar ik wil graag naar venlo
(On monday, but I'd like to go to Venlo)
Trigram :: Op maandag ik wil graag naar venlo
(On Monday, I'd like to go to Venlo)
+Context :: Op maandag morgen venlo
(On Monday morning, Venlo)

Figure 5: Example of bias towards question type

4. Conclusions and future work

In the introduction we pointed at situations in which the parser of an NLP component of a spoken dialogue system might benefit from information outside the wordgraph. As a first step in this directions we have developed a statistical model that combines n-grams with information about the system question that preceded the user utterance. We think the results we have found in section 3. are promising. Only by adding a simple trigger, we managed to improve the performance considerably. Especially the fact that the improvement is not only a matter of concept accuracy, but almost just as strong of word accuracy, implies that there is not just a bias towards semantically better hypotheses. It is even more promising that this approach almost only resulted in positive changes in the analyses. There seems to be space for even more improvement. Therefore, other linguistically motivated triggers could be explored. For this type of application it is probably enough to take only the previous question into account. More complicated spoken dialogue systems might benefit from looking further into the dialogue (Poessio and Mikheev, 1998).

What definitely needs more attention is a method to decrease the computational efforts that are now needed to get the results. Using the current model there will definitely be a lot to gain by computing off-line frequently used probabilities. Another way to reduce computation is by choosing better (i.e. smaller; less redundant) feature sets. One way to do this is selecting only those features that contribute as suggested by (Adam L. Berger and DellaPietra, 1996). However, this is a very expensive selection algorithm. A nice feature of this selection algorithm is that it gives insight in which features are important to describe the process. This would not be very helpful for the trigram features (I would expect that every feature is equally important), but it would be helpful for determining which context-trigger features are important for describing the process.

5. References

Stephen A. DellaPietra Adam L. Berger and Vincent J. DellaPietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1).

Luc Dehaspe. 1997. Maximum entropy modeling with clausal constraints. In *Proceedings of the 7th International Workshop on Inductive Logic Programming*.

Daniel Jurafski et al. 1997. Automatic detection of discourse structure for speech recognition and understanding. In *Proc. IEEE workshop on Speech Recognition and Understanding*.

E.T. Jaynes. 1957. Information theory and statistical mechanics. *Physical Review*, 108:171–190.

Sanjeev Khudanpur and Jun Wu. 1999a. Combining nonlocal, syntactic and n-gram dependencies in language modelling. In *Proceedings of Eurospeech '99*.

Sanjeev Khudanpur and Jun Wu. 1999b. A maximum entropy language model integrating n-grams and topic dependencies for conversational speech recognition. In *Proceedings of ICASSP '99*.

Vassilios Digalakis Manny Rayner, David Carter and Patti Price. 1994. Combining knowledge sources to reorder n-best speech hypothesis lists. In *Proceedings of 2nd ARPA workshop on Human Language Technology*.

Massimo Poessio and Andrei Mikheev. 1998. The predictive power of game structure in dialogue act recognition: Experimental results using maximum entropy estimation. In *Proceedings of ICSLP '98*.

Adwait Ratnaparkhi. 1998. *Maximum Entropy Models for Natural Language Ambiguity Resolution*. Ph.D. thesis, UPenn.

Sven Eric Ristad. 1998. Maximum entropy modelling toolkit. Technical report.

Ronald Rosenfeld. 1994. *Adaptive Statistical Language Modelling: A Maximum Entropy Approach*. Ph.D. thesis, Carnegy Mellon University.

Ronald Rosenfeld. 1996. A maximum entropy approach to adaptive statistical language modelling. *Speech and Language*, 99.

Gertjan van Noord, Gosse Bouma, Rob Koeling, and Mark-Jan Nederhof. 1999. Robust grammatical analysis for spoken dialogue systems. *Journal of Natural Language Engineering*. To appear; 48 pages.

From Manual Text to Instructional Dialogue: an Information State Approach

Staffan Larsson

Department of linguistics, Göteborg University
Box 200-295, Humanisten, SE-405 30 Göteborg, Sweden
sl@ling.gu.se

Abstract

We present preliminary research on the relation between written manuals and instructional dialogue, and outline how a manual can be converted into a format which can be used as domain knowledge by a dialogue system, capable of generating both instructional dialogue and monologue. Starting from a short sample text from a manual, we use the TRINDI information state approach (Traum et al., 1999) to build an experimental dialogue system capable of instructing a user to perform the task. IMDiS, a small experimental implementation based on the GoDiS dialogue system (Bohlin et al., 1999), is presented.

1. Goal of the paper

In this paper, we will present preliminary research on the relation between written manuals and instructional dialogue. We outline how a manual can be converted into a format which can be used as domain knowledge by a dialogue system, capable of generating both instructional dialogue and monologue. Starting from a short sample text from a manual, we use the TRINDI information state approach (Traum et al., 1999) to build an experimental dialogue system capable of instructing a user to perform the task. IMDiS, a small experimental implementation based on the GoDiS dialogue system (Bohlin et al., 1999), is presented. We look at sample monologue and dialogue output and discuss the advantages provided by the dialogue mode in IMDiS. One of the main advantages is that the user can control the dialogue to make the system provide exactly the information needed. Finally, we discuss possible research issues.

We will make two basic assumptions: monologue is a special case of dialogue, and discourse structure corresponds to task structure. These assumptions are by no means original (see e.g. (Grosz and Sidner, 1986)); however, the preliminary work here attempts to combine these assumptions using the TRINDI information state approach to investigate the possibility of generating dialogue (as in built-in automatic assistant) or monologue (as in a traditional written manual) from a single database of domain task plans.

2. IMDiS

IMDiS (Instructional Monologue and Dialogue System) is an adaption of GoDiS to instructional dialogue, and like GoDiS it provides a simple but efficient grounding strategy and facilitates question and task accommodation (Bohlin et al., 1999). In addition, IMDiS can give instructions and the user can request more specific instructions by asking the system how to perform a given instruction. IMDiS can also be made to generate the original text by setting it in "monologue" mode that uses a slightly altered set of dialogue moves and information state update rules, but which still uses the same database and generation facilities as the dialogue mode.

IMDiS is implemented using the TRINDIKIT (Larsson et al., 1999), a toolkit for experimenting with information states and dialogue move engines and for building dialogue systems. 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* updates the information state on the basis of observed dialogue moves and selects appropriate moves to be performed. In this paper we use a formal representation of dialogue information states that has been developed in the TRINDI[1], SDS[2] and INDI[3] projects.

IMDiS has a type of information state similar to that of GoDiS, with the addition of a subfield SHARED.ACTIONS whose value is a stack of actions which the system has instructed the user to perform, but whose performance has not yet been confirmed by the user. The IMDiS information state is shown in Figure 1.

$$\begin{bmatrix} \text{PRIVATE} & : & \begin{bmatrix} \text{PLAN} & : & \text{StackSet(Action)} \\ \text{AGENDA} & : & \text{Stack(Action)} \\ \text{TMP} & : & \text{(same as SHARED)} \end{bmatrix} \\ \text{SHARED} & : & \begin{bmatrix} \text{BEL} & : & \text{Set(Prop)} \\ \text{QUD} & : & \text{StackSet(Question)} \\ \text{ACTIONS} & : & \text{Stack(Action)} \\ \text{LU} & : & \text{Utterance} \end{bmatrix} \end{bmatrix}$$

Figure 1: IMDiS information state type

The main division in the information state is between information which is private to the agent and that which is shared between the dialogue participants. The private part of the information state contains a PLAN field holding 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. The AGENDA field, 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. We have included a field TMP that mirrors the shared fields.

[1] TRINDI (Task Oriented Instructional Dialogue), EC Project LE4-8314, www.ling.gu.se/research/projects/trindi/

[2] SDS (Swedish Dialogue Systems), NUTEK/HSFR Language Technology Project F1472/1997, http://www.ida.liu.se/ nlplab/sds/

[3] INDI (Information Exchange in Dialogue), Riksbankens Jubileumsfond 1997-0134.

This field keeps track of shared information that has not yet been grounded, i.e. confirmed as having been understood by the other dialogue participant. The SHARED field is divided into four subfields. One subfield is a set of propositions which the agent assumes for the sake of the conversation. The second subfield is for a stack of questions under discussion (QUD). These are questions that have been raised and are currently under discussion in the dialogue. The ACTIONS field is a stack of (domain) actions which the user has been instructed to perform but has not yet performed.The LU field contains information about the latest utterance.

The dialogue version uses 9 move types, basically the 6 used in GoDiS (**Ask**, **Answer**, **Inform**, **Repeat**, **ReqRep**, **Greet**, **Quit**) plus instructions to check preconditions (**InstructCheck**), plain instructions (**InstructExec**), and confirmations (**Confirm**). Confirmations are integrated by assuming that the current topmost action in SHARED.ACTIONS has been performed, as seen in the update rule below.

RULE: **integrateUsrConfirm**
CLASS: **integrate**

$$\text{PRE: } \left\{ \begin{array}{l} \text{val\#rec(shared.lu.speaker, usr)} \\ \text{assoc\#rec(shared.lu.moves, confirm, false)} \\ \text{fst\#rec(shared.actions, } A \text{)} \end{array} \right.$$

$$\text{EFF: } \left\{ \begin{array}{l} \text{set\_assoc\#rec(shared.lu.moves, confirm, true)} \\ \text{pop\#rec(shared.actions)} \\ \text{add\#rec(shared.bel, done(} A \text{))} \end{array} \right.$$

Elliptical "how"-questions from the user are interpreted as applying to the currently topmost action in the SHARED.ACTIONS stack.

The monologue mode uses only 3 moves (**InstructExec**, **InstructCheck** and **Inform**). Since there is no user to confirm that actions have been performed, all actions are automatically confirmed using the update rule **autoConfirm**.

RULE: **autoConfirm**
CLASS: **integrate**

$$\text{PRE: } \left\{ \text{fst\#rec(shared.actions, } A \text{)} \right.$$

$$\text{EFF: } \left\{ \begin{array}{l} \text{pop\#rec(shared.actions)} \\ \text{add\#rec(shared.bel, done(} A \text{))} \end{array} \right.$$

3. Manuals and dialogues

The text below is taken from a user manual for the Homecentre, a low end Xerox MultiFunctional Device.

Reinstalling the print head
Caution: Make sure that the green carriage lock lever is STILL moved all the way forward before you reinstall the print head.
1. Line up the hole in the print head with the green post on the printer carriage.
Lower the print head down gently into position.
2. Gently push the green cartridge lock lever up until it snaps into place.
This secures the print head.
3. Close the top cover and reattach the scanner.
4. Press and release the yellow LED button.
The printer will prepare the cartridge for printing.
Note: If the carriage does not move from the center position after you press the cartridge change button, remove and reinstall the print head.

From this text, one can (re)construct a domain plan for reinstalling the print head. Such a plan may be represented as in Figure 2. Note that this is a conditional plan, i.e. it contains branching conditions.

From this plan, IMDiS generates two plans: a monologue plan and a dialogue plan. This is done using the "translation schema" in Figure 3.

The difference between the text plan and the dialogue plan is in the way that conditionals in the domain plan are interpreted. In the monologue plan, they correspond to simply informing the user of the conditional. In dialogue mode, however, the system raises the question whether the condition holds. When the system finds out if the condition holds, it will instruct the user to execute the appropriate guarded action.

In short, here's how conditionals are treated by the system in dialogue mode: When the system has found out what the user's task is, it will load the appropriate dialogue plan into the PRIVATE.PLAN field of the information state. It will then execute the actions in the appropriate order by moving them to the agenda and generating appropriate utterances. When a conditional statement is topmost on the plan, IMDiS will check whether it has been established that the condition holds (by checking the SHARED.BEL field). Since the system has previously asked the user and the user has answered, either the condition or its negation will be in the set of established propositions. If the condition or its negation holds, the conditional will be popped off the plan and replaced by the first or second guarded action (respectively).

4. Monologue and dialogue

In the monologue mode in IMDiS, the control module does not call the input and interpretation modules. The text is output "move by move" as a sequence of utterances from the system[4].

```
S: Reinstalling the print head.
S: Make sure that the green carriage lock
lever is STILL moved all the way forward
before you install the print head.
S: Line up the hole in the print head with
the green post on the printer carriage
```

Compared to the monologue mode, the dialogue mode offers several advantages:

User attention and control The user can direct her attention to the machine and does not have to look at the manual. This means that the user does not have to keep track of the

[4] While perhaps not practically useful, the implementation of a monologue mode in IMDiS is primarilly intended to show how one can construe the claim that monologue is a special case of dialogue.

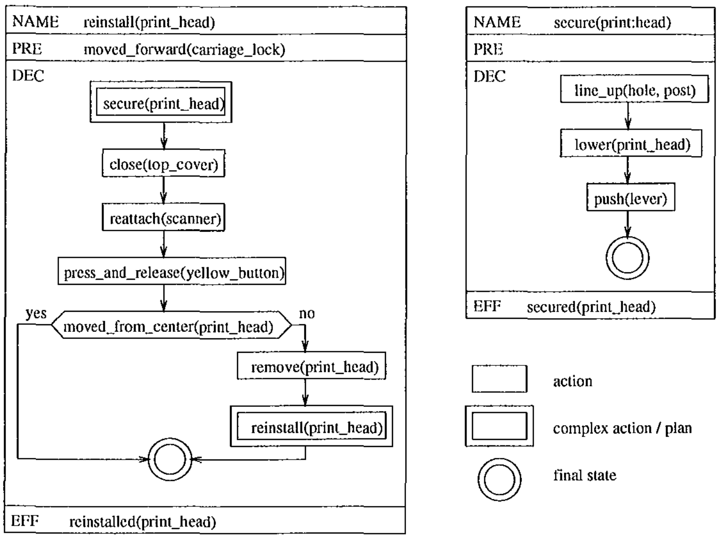


Figure 2: Domain plan

DOMAIN	MONOLOGUE	DIALOGUE
precondition P	**InstructCheck**(P)	**InstructCheck**(P)
action A	**InstructExec**(A)	**InstructExec**(A)
`if_then_else`(C, A, A')	**Inform**(`if_then_else`(C, A, A'))	`findout`(C); `if_then_else`(C, **InstructExec**(A), **InstructExec**(A'))
effect E	**Inform**(E)	**Inform**(E)

Figure 3: Plan conversion table

current point in the manual. Also, the user decides when to move on to the next action, by confirming that the previous action is done.

Avoid irrelevant information When the action to be taken depends on a condition, the system does not give irrelevant information.

```
S: Has the carriage moved from the center
position?
U: yes
S: The print head is now installed
```

Control over level of detail If the user does not know how to perform a substep, she can ask the system for more detailed instructions.

```
S: secure the print head
U: how?
S: Line up the hole in the print head with
the green post on the printer carriage
U: right, ok
S: Lower the print head down gently into
position
U: ok
S: Gently push the green cartridge lock
lever up until it snaps into place
U: ok
S: The print head is now secured
```

On the other hand, if the user already knows how to perform a substep, the system moves on to the next step.

```
S: secure the print head
U: ok, and now?
S: Close the top cover
```

If the user manages to complete the whole action sequence without instructions, she can tell the system this and the system proceeds to the next relevant action.

```
S: secure the print head
U: how?
S: Line up the hole in the print head with
the green post on the printer carriage
U: right, ok
S: Lower the print head down gently into
position
U: ok right, I remember now (pause) the
print head is secured
S: Ok.  Close the top cover
```

Question accommodation The user can provide answers to questions which the system has not yet asked, provided the system is able to accommodate the question. For

```
$S> Press and release the yellow LED button

: private =   agenda =   < instruct_exec(press_and_release(yellow_LED_button)) >
:             plan =     [ raise(move_from_center_position(carriage_head)),
                           if_then(not move_from_center_position(carriage_head),
                                   instruct_exec(remove_and_reinstall(print_head))),
                           inform(reinstalled(print_head)),
                           inform(next(prepare_cartridge_for_printing)) ]
:             tmp = (*surpressed*)
: shared =    bel =      { done(reattach(scanner)),
                           done(close(top_cover)),
                           done(secure(print_head)),
                           done(check(moved_forward(carriage_lock))),
                           task(instruct_exec(reinstall(print_head))) }
:             qud =      < >
:             actions =  < press_and_release(yellow_LED_button) >
:             lu =        (*surpressed*)
```

Figure 4: Sample IMDiS information state, after uttering "Press and release the yellow LED button"

example, the user does not have to wait for the system to ask what task the user wants to perform.

```
S: Hello and welcome to the IMDiS homecentre assistant
U: i want to reinstall the print head
S: Make sure that the green carriage lock
lever is still moved all the way forward
before you install the print head.
```

Grounding If the users does not hear or understand a system utterance, she can ask the system to repeat it.

```
S: Has the carriage moved from the center
position?
U: what ?
S: Has the carriage moved from the center
position?
```

5. Research issues

In building the experimental IMDiS, we have made several simplifications. For example, the problem of NL generation has been side-stepped by using canned text for output. Around 90% of the lexicon is used in both dialogue and monologue mode, while the rest is specific to one mode. It is a research issue to what extent canned text can be used, and how much "real" generation is necessary. Although this is experimental work, it does not seem implausible that useful systems could be constructed fairly easily to the extent that system output can be provided as canned text and that user input is limited in its lexical scope. On a domain level, what needs to be done is to construct domain plans and connect them to the corresponding text output. We make no claims here that this process is easily automated; rather, the idea is that instead of writing a manual (which will, in a sense, encapsulate both domain knowledge and its linguistic realisation), the author constructs the plans and output manually (possibly using a specialised authoring tool).

Also, IMDiS is not capable of referent disambiguation dialogue of the kind common in e.g. the MapTask corpus (Anderson et al., 1991). This type of dialogue would be needed for the system to be able to explain e.g. which component is being referred to and where it is to be found.

So far, we have only explored the extremes of the monologue-dialogue opposition. There are interesting intermediate levels of interactivity, such as dynamically generated text where the content depends on what has previously been related to the user. This is another area of possible future research, where it is likely that higher demands will be put on dynamic language generation.

Although this is not strictly relevant to the monologue-dialogue discussion, we would also like to compare IMDiS to previous instructional dialogue systems such as that described in (Smith and Hipp, 1994).

6. References

A. H. Anderson, M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. 1991. The HCRC Map Task corpus. *Language and Speech*, 34(4):351–366.

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

B. J. Grosz and C. L. Sidner. 1986. Attention, intention, and the structure of discourse. 12(3):175–204.

S. Larsson, P. Bohlin, J. Bos, and D. Traum. 1999. Trindikit 1.0 manual. deliverable D2.2, TRINDI.

R. W. Smith and D. R. Hipp. 1994. *Spoken Natural Language Dialog Systems.* Oxford University Press.

D. Traum, J. Bos, R. Cooper, S. Larsson, I.Lewin, C. Matheson, and M. Poesio. 1999. A model of dialogue moves and information state revision. deliverable D2.1, TRINDI.

Formalizing the Dialogue Move Engine

Peter Ljunglöf

Dept. of Computer Science
Chalmers University of Technology
412 96 Göteborg, Sweden
peb@cs.chalmers.se

Abstract

In this paper we present a calculus for reasoning mathematically about rule-based dialogue systems – so called *dialogue move engines* developed in the TRINDI project. The calculus is similar to term rewriting systems and dynamic logic. It is defined using monads, which are used for describing programming languages, and in functional programming to capture computations with side-effects.

1. Introduction

In this paper we present a calculus for reasoning mathematically about rule-based dialogue systems – so called *dialogue move engines* developed in the TRINDI, SDS and INDI projects[1] (Bohlin et al, 1999; Traum et al, 1999). The calculus is similar to term rewriting systems (Visser and Benaissa, 1998) and dynamic logic (Harel, 1984). It is defined using monads, which are used for describing programming languages, and in functional programming to capture computations with side-effects (Moggi, 1991; Wadler, 1995). In the end we show how the calculus can be used to prove properties of a dialogue system.

1.1. Preliminaries

Since we are only interested in the dialogue manager part of a dialogue system, we assume that there exist good translations between utterances and dialogue moves. Without loss of generality we can then assume that the dialogue participants communicate using dialogue moves.

As a simplification we assume that the dialogue is serial – that the participants make their utterances one after another and that they never interrupt each other. Another simplification is that each utterance can be translated into a time-ordered list of dialogue moves, thus forgetting about overlapping sub-utterances and so on.

In the discussion at the end we will try to argue that these simplifications do not induce severe limitations on the strength of the framework.

1.2. Notational conventions

In this paper we use a lot of terminology taken from programming languages and type theory. For those not familiar with our way of writing things, here are some informal explanations.

We write $a \in A$ to say that the object a is of the type A. The basic type constructors we are going to use are $\times$, $\rightarrow$ and $[]$. Given two types A and B, $A \times B$ is the type of pairs of A and B, $A \rightarrow B$ is the type of functions from A to B, and $[A]$ is the type of lists of type A.

There are some standard operations and predicates on lists which we will use – the *delete* and *add* operations deletes and adds elements to a list, the *append* operation concatenates two lists, and the *member* predicate is a sequential member checking predicate, binding the second argument to each element of the list. We will also use the standard way of using lists to represent backtracking – computations that can fail or return several results – with the empty list representing failure (Wadler, 1985).

2. Defining the dialogue move engine

A dialogue move engine (DME for short) consists of a description of *i*) what the information state (infostate for short) looks like, *ii*) what kinds of dialogue moves there are and *iii*) how they are applied to the infostate, *iv*) a collection of update rules on the infostate, and *v*) an update algorithm which defines how the rules are used to update the infostate.

Parallel with the formalization, we introduce an example DME to illustrate the principles. This is a small subset of the information-seeking DME used in the GoDiS system (Traum et al, 1999), but it is general enough for the purposes of this paper.

2.1. The information state

The *information state* is seen as a representation of an agent's current knowledge, especially the part that change during the dialogue. In this formalism the infostate is a type *IS*. For the example DME we will use a record with the fields shown in table 1 below, where *plan* is a list of things to do in the future, *bel* is a list of beliefs, *qud* is a list of questions currently under discussion, and *lm* is a list of the dialogue moves that the other participant just uttered. We do not further define propositions and questions as the formalization is independent of which notion of proposition or

$$\begin{aligned} plan &\in [Move] \\ bel &\in [Proposition] \\ qud &\in [Question] \\ lm &\in [Move] \end{aligned}$$

Table 1: The information state used in the example

[1] TRINDI (Task Oriented Instructional Dialogue), EC Project LE4-8314. SDS (Swedish Dialogue Systems), NUTEK/HSFR Language Technology Project F1472/1997. INDI (Information Exchange in Dialogue), Riksbankens Jubileumsfond 1997-0134.

r_1 : *integrate_question*	r_2 : *integrate_answer*	r_3 : *answer_question*	s : *select_move*

conditions
member(lm,Q)
is_question(Q)
effects
delete(lm,Q)
add(qud,Q)
	conditions
	member(lm,A)
	is_proposition(A)
	member(qud,Q)
	is_answer_to(Q,A)
	effects
	delete(lm,A)
	delete(qud,Q)
	add(bel,A)
		conditions
		member(qud,Q)
		member(bel,A)
		is_answer_to(Q,A)
		effects
		delete(qud,Q)
		add(plan,inform(A))
			conditions
			member(plan,M)
			effects
			delete(plan,M)
			select(M)

Table 2: The update rules used in the example

question is chosen. In this simple example we have just two kinds of *dialogue moves* – to *ask* a question and to *inform* of a fact, represented as a proposition.

We view a dialogue move as a basic type, so we need to know how to incorporate an utterance, represented as a sequence of moves, into the infostate. This is done with the function $apply \in IS \times [Move] \rightarrow IS$ which updates the infostate with a list of moves. In our example the *apply* function simply adds the moves to the end of the *lm* list.

2.2. Update rules

An *update rule* specifies an update on the infostate (called the effect), which is guarded by a condition – if the condition holds, the effect can be applied. The effect may also have a side effect: it can select one or several dialogue moves to be performed. The conditions and effects are composed by combining basic operations on the elements of the infostate. The update rules of our example are listed in table 2 above. The first two rules interprets the user's last move – if it was a question, it will be added as a question under discussion, and if it was an answer to a question currently under discussion it will be added as a belief. The third rule answers a question under discussion, if the system knows the answer, and the fourth rule selects the first move on the plan to be uttered to the user. Since this last rule selects a dialogue move to be performed, we call it a *selection rule*.

More formally we can say that an update rule is a function that given an infostate, returns either a failure if the condition fails, or the different results of the effect applied to the infostate. This gives us $rule \in Rule$ where $Rule = IS \rightarrow [IS \times [Move]]$.

2.3. The update algorithm

The *update algorithm* defines how these rules should be applied to an infostate; that is, given an infostate, the update algorithm updates the infostate and selects a list of moves to perform. This suggests that the update algorithm is a function $update \in IS \rightarrow IS \times [Move]$.

The naive algorithm is to check the rules in order, and as soon as a rule applies update the infostate accordingly and then repeat the algorithm until there are no rules that apply.

But if we use this naive algorithm on our example rules, all the moves that are in the *plan* will be selected at once – this is maybe not immediate from the definitions, but can be proved using the formalism we will introduce later. Since we want it to just say one thing at each turn, we have to change the algorithm to first apply the rules $r_1 \dots r_3$ until this can no longer be done, and then apply the rule s once, selecting only one move.

With these definitions we can define the *dialogue move engine* to be a function that, given a list of dialogue moves uttered by the user, applies them to the infostate, and then updates the infostate with the update algorithm, selecting new moves to perform during the updating. We now finally have a function $dme \in IS \times [Move] \rightarrow IS \times [Move]$, with the very simple definition $dme = update \circ apply$.

3. A calculus of update rules

In this section we introduce a calculus for the update algorithm and show that this can be used to define the update rules themselves. The calculus is similar to term rewriting systems (Visser and Benaissa, 1998) and dynamic logic (Harel, 1984), with the main exceptions being that the rules also has the ability to communicate to the outer world by selecting dialogue moves to perform, and all the operators are deterministic.

We have two trivial rules and three basic operators that make new rules out of old ones:

- The *identity* rule 1 always succeeds without affecting the infostate and without selecting any moves.
- The *failure* rule 0 always fails.
- The *sequential composition* $r\,;r'$ of two rules first applies r, and if that succeeds, applies r' to the result of r The composition selects all the moves selected by either r or r'. It fails if either r or r' fails. Composition has 1 as an identity and 0 as a zero, which gives the laws $1\,;r = r\,;1 = r$ and $0\,;r = r\,;0 = 0$.
- The *deterministic choice* $r + r'$ first applies r, and only if that fails it applies r'. Choice has 0 as an identity and 1 as a left zero, giving the laws $0 + r = r + 0 = r$ and $1 + r = 1$ (but not necessarily equal to $r + 1$).
- The *repetition* r^* applies r and if that succeeds it executes r^* on the result (concatenating the selected moves). If r fails, it succeeds leaving the infostate unchanged and selecting nothing. The repetition can be unfolded using the other operators: $r^* = (r\,;r^*) + 1$.

$$\begin{aligned}
r_1 &= \exists q \in \mathit{lm}.\, \mathit{is\_question}(q)\,;\, \mathit{delete}(\mathit{lm}, q)\,;\, \mathit{add}(\mathit{qud}, q) \\
r_2 &= \exists a \in \mathit{lm}.\, \exists q \in \mathit{qud}.\, \mathit{is\_proposition}(a)\,;\, \mathit{is\_answer\_to}(q, a)\,;\, \mathit{delete}(\mathit{lm}, a)\,;\, \mathit{delete}(\mathit{qud}, q)\,;\, \mathit{add}(\mathit{bel}, a) \\
r_3 &= \exists q \in \mathit{qud}.\, \exists a \in \mathit{bel}.\, \mathit{is\_answer\_to}(q, a)\,;\, \mathit{delete}(\mathit{qud}, q)\,;\, \mathit{add}(\mathit{plan}, \mathit{inform}(a)) \\
s &= \exists m \in \mathit{plan}.\, \mathit{delete}(\mathit{plan}, m)\,;\, \boldsymbol{\mathit{select}}(m)
\end{aligned}$$

Table 3: Formal definitions of the update rules of the example

With these definitions the update algorithm of our example can be defined as $(r_1 + r_2 + r_3)^*\,; s$. This suggests that the update algorithm is just a very complicated update rule. But this definition of the update algorithm is not completely correct; the type of the *update* function does not correspond to the type of the update rules. The main difference is that the update rules can fail, which the update algorithm is not allowed to. But a correctly defined update algorithm will never fail, which means that the list of results it returns will be non-empty. Then we can use the standard list function *head* $\in [A] \rightarrow A$, which gives the first item in a list, to extract the result we want. This gives for our example the resulting function $\mathit{update} = \mathit{head}((r_1 + r_2 + r_3)^*\,; s)$.

3.1. Defining the update rules

Now it turns out that we can use the calculus to define the update rules themselves. To apply an update rule we first check the conditions, and if they hold we can apply the effects. Both the conditions and the effects are ordered – we apply them in the order they are written. This means that an update rule is just a sequential composition of more basic rules, the individual conditions and effects. There is just one thing that needs to be taken care of – the special *member* predicate which introduces some kind of choice depending on the elements of the first argument. For that purpose we introduce the operator $\exists x \in A.\, r(x)$, where A is a list and $r(x)$ is a rule whenever x is an element in A. The idea is that if $A = [a_1, a_2, \ldots, a_n]$ when the rule is invoked, then $\exists x \in A.\, r(x) = r(a_1) + r(a_2) + \cdots + r(a_n)$.

Another addition is to add the special selection rule ***select***(m), which leaves the infostate unchanged and selects the single move m. With these additions to our calculus, we can define the update rules of our example as in table 3 above. (We still have to give definitions of the basic rules of course).

4. Interpreting the calculus

Monads are standard tools in functional programming for capturing computations with side-effects, and they are also used in denotational semantics for defining programming languages (Wadler, 1995; Moggi, 1991). Here we are going to use them to give a precise definition of our calculus.

4.1. Introducing monads

The standard example of a monad is the type constructor [] which takes any type A and gives back $[A]$, the type of lists of objects of type A. A monad M is a type constructor with two operations: *return* $\in A \rightarrow M(A)$ and *bind* $\in M(A) \times (A \rightarrow M(A)) \rightarrow M(B)$, which also satisfy three identity and associativity laws. Some monads also are equipped with a zero element $0 \in M(A)$, and a plus operation $(+) \in M(A) \times M(A) \rightarrow M(A)$, which in turn satisfy a couple of other laws.

An example of a monad is the state monad $\mathit{SM}(A) = \mathit{IS} \rightarrow \mathit{IS} \times A$, with the definitions $\mathit{bind}(f, k) = \lambda s.\, k(a, s')$ where $(s', a) = f(s)$, and $\mathit{return}(a, s) = (s, a)$. Another example is the monad of lists $[A]$ which is also a monad with zero and plus; where *return* returns a singleton list, 0 is the empty list, and $l + l'$ concatenates the lists l and l'. The bind operation sends each element of the first list to the second function, concatenating the results, and can be defined by cases as $\mathit{bind}([], k) = []$ and $\mathit{bind}(a{:}l, k) = \mathit{append}(k(a), \mathit{bind}(l, k))$.

The list monad is often used to represent backtracking, which we will also do here. We can also combine monads – e.g. combining the two monads above gives us the backtrackable state monad $\mathit{BSM}(A) = \mathit{IS} \rightarrow [\mathit{IS} \times A]$, which is a monad with zero and plus.

4.2. A dialogue monad

The backtrackable state monad *BSM* gives us a way to define the rules and operators of our calculus, since the type *Rule* of update rules is just an instance of $\mathit{BSM}([\mathit{Move}])$. The 0 rule and the (+) operator are exactly the same as 0 and (+) for the monad. The identity and selection rules can be defined as $1 = \mathit{return}([])$ and $\boldsymbol{\mathit{select}}(m) = \mathit{return}([m])$ respecively. Sequentiation simply becomes $r\,; r' = \mathit{append}^M(r, r')$, where append^M is concatenation of lists lifted to the *BSM* monad,[2] and repetition is defined by the unfolding equation $r^* = (r\,; r^*) + 1$. For the $(\exists)$ operator we have to use the fact that the *BSM* monad is a function that takes an infostate and gives a list as result: $\exists x \in f.\, r(x) = \lambda s.\, \mathit{bind}(f(s), r)$.[3]

Apart from giving us a precise definition of the calculus of update rules, it also gives us all the properties of monads, like the associativity and identity laws. These laws are free for us to use when we want to prove statements about, or to rewrite our dialogue move engine to a more efficient one.

[2] The precise definition of the lifting of a function f to a monad is $f^M(r, r') = \mathit{bind}(r, \lambda m.\, \mathit{bind}(r', \lambda m'.\, \mathit{return}(f(m, m'))))$.

[3] Observe that the *bind* used here is the one in the list monad, not in the *BSM* monad. The field label f is seen as the function on the infostate that gives the current value of the field f.

5. Proving properties of the DME

If we have defined a collection of update rules together with an update algorithm, we may want to show that some desirable properties hold for this dialogue system. Most important are to show that the system terminates, always succeeds and always produces some moves to utter to the user, but there are also other interesting properties.

5.1. Termination

To prove that the update algorithm terminates for every given input, we have to show that all repetitions always terminate. For this we can use induction on some parts of the information state. In our example we can do induction on the total length of the *lm* and *qud* lists (to be more precise, $n = 2|lm| + |qud|$), and notice that when any of the rules $r_1 \dots r_3$ is applied, the total length decreases (actually, it's the number n that decreases). This means that $(r_1 + r_2 + r_3)^*$ cannot continue forever, since the *lm* and *qud* lists will finally be empty.

5.2. Non-failure

Since the repetition r^* always succeeds if it terminates, the only thing we have to prove for our example is that the selection rule s always succeeds. In our example case there is a possibility that the *plan* at some point gets empty, so we have to change the update rules in some way – e.g. by replacing s with $s + \textbf{\textit{select}}(m)$ in the algorithm, where m is some default move.

5.3. Productivity

To show that the system is productive we have to show that, whenever it terminates and succeeds, it executes a ***select*** rule. In our example it is easily shown just by looking at the s rule.

We may also want to show that the system does not select too many dialogue moves at the same time (thus giving the user the opportunity to interrupt with e.g. clarifying questions). In our first naive definition of the update algorithm, the system selected all the moves that was on the plan, but in the second algorithm the system selects only one move at the time.

5.4. Other properties

Another interesting property is that the system is efficient. There could be some of the basic conditions or effects that take time to execute (e.g. the *is_answer_to* predicate which probably has to call a theorem prover). We might want to show that the system never calls such a predicate more than, say 5 times. In our example system, the *is_answer_to* condition is called a number of times which in the worst case can be in the order of $|qud| \cdot |bel|$, which in turn means that the system should probably have to be optimized in some way.

A final example of a property of a well-designed dialogue move engine is that two rules are commutative with respect to the choice operator, i.e. $r + r' = r' + r$. The reason why this is a good property is that if a dialogue system is on the form $(r_1 + r_2 + \cdots + r_n)^*$, where each pair of rules r_i, r_j commute, then that system can be implemented asynchronously, executing each rule r_i in parallel.

6. Discussion and future work

In this paper we have introduced a calculus for building and reasoning about dialogue move engines. With the use of a simple example we have defined the basic constructors of the calculus.

In the beginning we introduced some simplifications on the dialogue system, and we will now try to argue that they can be accounted for. The limitations were that the dialogue is serial, and that each utterance can be translated into an ordered list of dialogue moves. But that the participants talk at the same time or interrupt each other can be coded using special arguments to the dialogue moves, as can the overlapping of moves, so these are not real limitations.

If one wants to work with something other than *lists* of moves, e.g. sets or partially ordered collections, one can redefine the (;) and *return* operations in an appropriate manner. (Which means that one uses another monad in place of the list monad). In the same way one can use another definition of the (+) operation, as long as it still obeys the monadic laws – e.g. one may want the choice to look at the current infostate before it decides which of the rules to try.

This is very much work in progress. It remains to show that the framework can be used in real-world problems, where the infostate is much more complicated and there are more than four update rules. One possible research issue would be to see if the framework can be used to model the dialogue behaviour of a system. Possibly the calculus can be used to prove desired properties of the system as a whole – e.g. that it in the end always gives a relevant answer to a question, or that it fulfills given orders.

7. References

P. Bohlin, R. Cooper, E. Engdahl and S. Larsson. 1999. Information states and dialogue move engines. Gothenburg Papers in Computational Linguistics GPCL 99-1. URL *http://www.ling.gu.se/publications/*

D. Harel. 1984. Dynamic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, vol. II.* Reidel.

E. Moggi. 1991. Notions of computation and monads. *Information and Computation*, 93(1).

D. Traum, J. Bos, R. Cooper, S. Larsson, I. Lewin, C. Matheson and M. Poesio. 1999. A model of dialogue moves and information state revision. Trindi Deliverable D2.1. URL *http://www.ling.gu.se/research/projects/trindi/*

P. Wadler. 1985. How to replace failure by a list of successes. In *2nd Symposium on Functional Languages and Computer Architecure.* Lecture Notes in Computer Science LNCS 273. Springer Verlag.

P. Wadler. 1995. Monads for functional programming. In J. Jeuring and E. Meijer, editors, *Advanced Functional Programming*. Lecture Notes in Computer Science LNCS 925. Springer Verlag.

E. Visser and Z. Benaissa. 1998. A core language for rewriting. In C. Kirchner and H. Kirchner, editors, *2nd International Workshop on Rewriting Logic and its Applications.* Electronic Notes in Theoretical Computer Science 15. Elsevier.

Defining Propositional Similarity

Systemizing identification of presuppositional binding

Jennifer Spenader

Computational Linguistics
Department of Linguistics
Stockholm University, 106 91 Sweden
jennifer@ling.su.se

Abstract

When can we say that two propositions in a spoken dialogue are similar enough that the one can function as an antecedent for the other? This study looks at the problem of determining meaning equivalence or meaning relatedness in spoken utterances taken from the London-Lund Corpus of Spoken English. Propositions judged to be equivalent or extremely similar in meaning were categorized according to the type of semantic relationship holding between them. The results here can be used to systemize the identification of presuppositional binding within the anaphoric theory of presupposition.

1. Introduction

Briefly, in the anaphoric theory (van der Sandt, 1992), presuppositions[1] are treated as anaphora. All presuppositional usage falls into one of two categories, binding or accommodation. Either a proposition which can function as an antecedent for the presupposition exists in the previous discourse, in which case binding occurs, or the presupposition must be accommodated into the discourse record. The accommodation process has been the subject of much discussion. The seemingly simpler problem of finding an antecedent for bound presuppositions is, however, deceptively difficult. With most other anaphora, a semantically poor anaphoric element, such as a pronoun, is bound to a semantically rich antecedent. Presuppositions are propositions, which means that determining an antecedent involves identifying a proposition earlier in the discourse that can be said to be sufficiently equivalent to the presupposed proposition to be able to bind with it. In many cases, the antecedent is actually a set of propositions, often related to the presupposition through linguistic relationships and based on world-knowledge, as well as inferential information implicit in the text.

This work attempts to give a more precise characterization of the relationships found between the propositional anaphor and propositional antecedents. It is hoped that a more explicit definition can then be used to automate the identification of binding when working with new examples.

2. Background

If presuppositions are semantically-loaded anaphoric expressions, then examining anaphoric binding in other types of anaphoric relationships should be relevant for identifying propositional binding.

First, what is the nature of the anaphor-antecedent relationship between nouns? Van Deemter (1992) developed a more general definition of anaphora by incorporating many non-traditional anaphoric NP relations that are central in establishing discourse coherence. He points out that referential identity has been the relationship traditionally identified between anaphors and their antecedents, in part because this is the *only* relationship possible for pronominal anaphora. Non-identity relationships are difficult to characterize. Possible relationships include subsumption, which he defines as "subset of a set, part of a quantity, substructure in a given structure" (p. 35) and those generated by relational nouns, (e.g. *the book-the author*, books always have an author). Many of these are sometimes termed bridging anaphora, though definitions differ widely. Van Deemter concentrated on full NP-anaphora but he also mentions anaphoric predicate relations between verbs. He does not, however, go in to how these could be handled.

Non-identity anaphoric relationships are problematic. Kramer & van Deemter (1998) develop a procedure for identifying non-identity relationships, which they term *partial matching,* in their study of definite noun phrases. They associate each discourse referent with a value set, which is the set of characteristics common to the common noun type to which the discourse referent belongs. If the disjunction of the value set of the two discourse referents is non-empty, then we have a partial match. How value sets are determined, and how we determine if the disjunction of two value sets is non-empty, is not elaborated on, which makes it difficult to apply this work to empirical data.

Bos et al. (1995) go further in developing a workable procedure for identifying partial matching by incorporating lexical information into the process. They use qualia structure information (Pustejovsky 1991), which contains information on both lexical relationships and world-knowledge, (an example they give is *bar-barkeeper*). By relating the anaphor to the relevant part of the qualia structure, bridging anaphora can be

[1] Following current usage, all propositions triggered by lexical or syntactic presuppositional triggers are called presuppositions.

resolved. Their work is also limited to definite noun phrases.

One draw-back of Bos et al.'s work is that it has difficulty modeling bridging instances where lexical relationships are not central. Asher & Lascarides (1998) attempt to remedy this by handling bridging by identifying the rhetorical relation that holds between the antecedent and the anaphoric expression. Identifying the rhetorical relationship allows them to compute other semantic information that in turn helps to make the binding relation more clear. They also briefly apply their method to an example of an it-cleft presupposition and an example with the temporal adverbial "again" as a presuppositional trigger. In both cases they discuss the presupposed proposition as an event, and identify a rhetorical relation that helps relate the event to the potential antecedent. Using their method requires that there is already a mechanism for unambiguously determining rhetorical relations.

In summary, work has focused on identifying what types of relationships can license anaphoric binding between full NPs: identity relationships, which are full matches, and several kinds of partial matches. How can we apply the methods found for NP-anaphora to propositions? There are identity propositions, that is when the two propositions are exactly the same. Additionally, we can probably safely allow the applications of some of the laws of logical equivalence found in statement and predicate logic, e.g. De Morgan's laws, laws of commutativity, etc., to also be considered propositional identify. Propositions that are expressed with clear synonyms should also be considered here.

But because even partial matching can license anaphoric binding, we also need to define degrees of similarity between propositions. Relationships such as hyponymy, subsumption, and relational noun inferences have been shown to license anaphoric binding between NPs. One way of applying this information to propositions is by allowing propositions to be considered similar enough for the one to function as an antecedent of the other when they differ in one argument. The relationship between the two arguments must then be one of the relationships defined as partial matching in full NP anaphora. For example, the propositions "The car is broken" should be considered similar enough to "The motor is broken" to allow the first to function as an antecedent for the latter, because of the relationship between *car* and *motor*. This type of relationship could be automated by having access to a lexicon, such as Pustejovsky's generative lexicon, which Bos et al. (1995) argued works with NP-anaphora.

If the relationships discussed here are found in the examples previously judged to be anaphoric binding, then we have the beginning of a method that could be automated to identify presuppositional binding.

3. Method

31 examples previously classified by the author as cases of presuppositional binding for another study were used as data. This earlier study looked at presuppositions with factive verbs, aspectual verbs and it-clefts, so examples are limited to these three trigger types. All examples were taken from the multi-speaker dialogues found in the London-Lund Corpus of Spoken English[2]. For each example, the entire discourse up until the presupposed proposition was examined for a possible antecedent. Examples were categorized according the relationship held between the presupposition and the propositions in the relevant parts of the previous discourse. When necessary, DRT-like representations of the propositions compared are given to make the relationships more clear.

4. Results

Examples of binding could be divided into four main categories that correlated positively with the authors intuitions of the ease of identification of the presupposition as bound. Each category is described below followed by examples. Examples are treated as if all other, non-propositional anaphora have been resolved.

The first two groups were by far the largest, and both deal with propositions where the relationship is one of identity.

1. Sense identity: a proposition/propositions earlier in the discourse have the same meaning as the presupposed proposition.

EXAMPLE 1 >> SHE IS UNUSUAL (1 3 1190)

Speaker A: But at the same time she seems unusual$_{\text{antecedent}}$, doesn't she.
Speaker B: Yes. And everybody notices that **she's unusual**.

This type of binding is most trivial and merely involves matching. There is no need to have access to additional linguistic information that that in the text.

EXAMPLE 2 >> "SOMEONE INVITED ME (= SPEAKER B)" (2 1@1086)

Speaker A:...was the invitation to York for which I did not apply. I was just invited$_{\text{antecedent}}$. (35 LINES LATER)
Speaker A: It was he who **invited me** .

A was invited differs from *someone invited A* only in that the former is passive and the later active. What the two sentences have in common is the fact that speaker A was invited, more explicitly, e.g. [x, A]:[invited(x,A)][3], where the value of x (the invitee) is uninstantiated in the antecedent, as well as in the presupposed proposition.

2. Sense identify by synonyms: Propositions differ only in lexical choice, where the words used were in a

[2] More information on the London-Lund Corpus of Spoken English can be obtained at http://www.hit.uib.no/icame/icame.html

[3] Examples given in DRT-format are simplified to only contain discourse referents and predicates relevant to the example.

well known lexical relationship such as synonymy or hyponymy.

EXAMPLE 3 >>"SOMEONE/SOME ARE ELITISTS" (2 9 630)

Speaker B: Oh no, it's very elitist$_{antecedent}$.
Speaker A: I thought it was the specialists who **are elitists**

This requires access to a lexicon that defines relationships between similar words of different parts-of-speech to identify the connection between "being elitist" and "being an elitist".

EXAMPLE 4 >> "SOMEONE LIFTED IT (=THE CONCEPT OF TRANSFORMATIONS) FROM THEM" (=MATHEMATICIANS) (2 5A 615+C)

Speaker A: I mean every transformation word that I've heard is in at the moment in [dhi] course for mathematics.
Speaker B: That's right. Well, that's where it all comes from$_{antecedent}$.
Speaker A: Yes. And it's {so} fascinating to see the analogy and it's much better in the mathematics than it is in grammar, I think.
Speaker B: But it's us that **lifted it from them**, not vice versa.

Here, B has already pointed out that the concept of transformations originated in mathematics, but seems to feel that A has not really understood that. He therefore reiterates with a more forceful it-cleft construction that linguists took the concept from mathematicians. Understanding that the presupposition is already present requires understanding the similarity between *"it" all comes from (mathematics) = someone lifted "it" from mathematics*, or more simply the synonymy of [X,Y,Z]:[[lifted_from(X,Y,Z)] = [originates(Y,Z)]] .

EXAMPLE 5 >> YOU ARE GOING TO KNOCK OUT AN EXPECTANT MOTHER (1 8 993)

Speaker A It was lethal to expectant mothers$_{antecedent}$ with small children.
(38 more lines)
Speaker A After all, I mean you can't go down and shop if you know that **you're going to knock out an expectant mother**-it was some <violent> streptococcus {that he'd got}

Finding the propositional antecedent in this example involves a kind of recursive process of identifying the relationship between the group of expectant mothers and a single instance of an expectant mother, followed by relating "it was lethal" to "knock out".

3. Non-identity/partial match of arguments:

1) The predicate of the proposition is the same, or a synonym, but one of the arguments to the predicate is different or

2) the predicate and arguments are the same, but the role of each argument differs.

EXAMPLE 6 >> "SOMEONE SAID I WANTED TO SELL OUT" (113 855)

Speaker A: James, it was no good. You didn't tell me to sell out$_{antecedent,}$ it was I who **said I wanted to sell out.**

Here we must in some way characterize [B,A]:[tell_to(B,A,sellout)]=[B,A]:[want(say(A,A,sellout)]. The action of sell-out is common to both, but the difference lies in *tell* and *say* with a volitional operator *want*. Roughly, the expression want(say(X,Y,Z)) can only have the agent of "sell-out" as its subject, whereas the subject of tell can only be the agent of "sell-out" if a reflexive pronoun is used ("I told myself to sell out"). What the two propositions have in common is the relation of *A* as the agent to *sell-out.* Here, the change of polarity does not seem to be a hinder for an anaphoric interpretation.

4. Extended differences: Understanding the relationship requires tracking information over several propositions, and several speakers and then synthesizing this information as well as using extended lexical knowledge, knowledge about the context and world knowledge.

EXAMPLE 7 >> "HE IS ENTHUSIASTIC (HE = PROF. PITT)" (2 1@ 120)

Speaker A: Also Pitt has talked about it a good deal. Professor Pitt here, and he has supported. (SEVERAL LINES) Yes, he has supported you is it with the Cambridge Press. He has supported you quite strongly and we had
(UNTIL LINE 1015)
Speaker A: You could indeed but I should continue also to give Professor Pitt's since I know that
Speaker B: **Know** that **he is enthusiastic**.
Speaker A: Yes, quite. He supported you very strongly.

Being supportive and *being enthusiastic* are not synonyms. However, we can understand them as synonyms in the context of this example because both are used to refer to Prof. Pitt's attitude towards Speaker B, and because the expression of *being supportive* is iterated several times, giving it a stronger import that makes it more possible to consider it synonymous with enthusiasm. Note the gap of almost 900 lines of dialogue from the time Prof. Pitt's support is mentioned until the statement that he is enthusiastic, illustrating the necessity of examining a very large context in order to correctly identify possible cases of binding.

EXAMPLE 8 >> PEOPLE GAVE YOU THINGS BEFORE (MARRIAGE) (210 927)

Speaker A doesn't seem much different except for trying to answer these awful conundrums about what shall we give you - and trying to fit it in (laughs)
Speaker B That's nice.
Speaker C That's the bit Debbie enjoyed enormously.
Speaker B Oh that's much the <cos> that won't, that won't happen again so I should enjoy it very much.
Speaker A Yes, that's true.
Speaker B Cos that stops very rapidly after you get married actually. **People stop giving you things**.

This example illustrates the necessity of tracking the individual parts of a possible propositional antecedent, and the need to use world knowledge. Speaker A's general statement about the difficulties involved in finding gifts is followed by an expression by Speaker C of how enjoyable receiving the gifts are. This "idea" is then referred to with different situational pronouns down until Speaker B's final statement, which according to my intuitions the giving of gifts before marriage was clearly part of the common ground, but pinpointing exactly where and how this is said is very difficult, and beyond the present study.

5. Discussion

Many of the examples studied could be accounted for by identifying simple lexical relationships between the anaphor and the antecedent. Some examples did require relating syntactically different sentences, for example passive sentences related to active counterparts, or ignoring certain types of extra information (such as the "seems" in example 1). Still, all these types of transformations are well understood and could probably be automated.

Typical bridging examples were not found. In the case of propositions this would be a bridging anaphoric relationship between one of the arguments in the antecedent with one of the arguments in the anaphoric proposition. There are three possible reasons for this result. The initial classification of examples as binding or accommodation may have been too strict, and examples with a possible bridging interpretation were placed into the accommodation category. Krahmer & van Deemter (1998) argued that partial match examples, including bridging, should be considered ambiguous between an interpretation as accommodation and an interpretation of binding, meaning that there is a greater chance that individual intuitions play a larger role in the interpretation of examples with bridging. A second possibility is that bridging between arguments in propositions may just be too difficult to take in, because the propositions compared often differ in other ways, e.g. there may be a temporal difference or a change of speaker in the dialogue. At the level of the individual word, the bridging relationship may be clear, but not as a smaller part of a proposition. A third possibility is that bridging is too infrequent to have been found in the limited number of examples studied here.

Not all cases involved a mere matching of synonymous arguments. In particular, group three illustrates some of the possible relationships that can arise when working with full propositions, rather than noun phrases. Here, actual switching of arguments had occurred between the two propositions judged to function in an anaphoric relation. In the data there were only two examples of this type of relationship, but immediately two questions arise. What other kinds of argument switching relationships license anaphoric binding, and where do we draw the line and acknowledge that we are dealing with two different propositions? The examples below give some ideas of what we may have to deal with.

(a) Mike gave Mary a package.
(b) No, Mary gave Mike a package.
(c) No, John gave Mary a package.
(d) No, John gave Mike a package.
(e) No, Mary took a package from Mike.

(ab) and (ac) are sequences similar to those found in the data. In (ab) the actors are the same, but the roles have been switched. In (ac) Mary's receiving a package is static, but the agent of the action changes. Both these examples could be argued to support an anaphoric relationship, and this is partly shown by our ability to use ellipse for c. (e.g. for c': "No, it was John".), so there is definitely an important coherency relation here. On the other hand, (ad) only keeps the action of giving static, but both actors have changed, making the jump too much to support a binding interpretation. (ae) keeps the actors and the initial and the resulting state static, but changes the import of the action. In terms of very basic concepts of transfer of ownership, (a) and (e) are very similar, but my intuitions on whether or not it would really license binding are not very clear here.

Another difficulty unique to propositional anaphora is that parts of the presupposed proposition may have been introduced into the discourse at different times, by different speakers, as illustrated by example 8. Tracking these, and then combing them to make the antecedent proposition requires a good understanding of the discourse.

6. Conclusions

Identifying propositional binding is easy for examples where lexical relationships known to license NP-anaphoric binding can be used. It is however difficult when arguments differ, or are switched, or when the antecedent proposition is a synthesis of several statements. Future work should look at more empirical data. Also, more work could be done working with idealized examples of the relationships found in group three and four.

References

Asher, N. & Lascarides, A. (1998). Bridging. Journal of Semantics,15 (1), 83--113.

Bos, J., Buitelaar, P. & Mineur, A. (1995). Bridging as Coercive Accommodation. In Proceedings of CLNLP'95, Workshop on Computational Logic for Natural Language Processing, Edinburgh.

van Deemter, K. (1992). Towards a Generalization of Anaphora. Journal of Semantics, 9(1), 27--51.

Krahmer, E. & van Deemter, K. (1998). On the Interpretation of Anaphoric Noun Phrases: Towards a Full Understanding of Partial Matches, *Journal of Semantics,* 15(4), 355--392.

Pustejovsky, J. (1991). The Generative Lexicon. Computational Linguistics, 17 (4), 409-441.

van der Sandt, R. A. (1992). Presupposition projection as Anaphora Resolution. Journal of Semantics, 9(4), 333--377.

Author index