DiaBruck 2003

Proceedings of the 7th Workshop on the Semantics and Pragmatics of Dialogue

Ivana Kruijff-Korbayová and Claudia Kosny (editors)

4–6 September 2003
Wallerfangen

Organized by
Department of Computational Linguistics,
Saarland University, Saarbrücken, Germany

Preface

DiaBruck 2003 is the seventh in a series of workshops bringing together researchers working on the semantics and pragmatics of dialogues in fields such as artificial intelligence, formal semantics and pragmatics, computational linguistics, philosophy, and psychology. The series were founded by Gerhard Jäger and Anton Benz, who were students at the CIS, University of Munich, at the time of the first meeting, MunDial 1997. The subsequent workshops in the series were Twendial 1998 (University of Twente, The Netherlands), Amstelogue 1999 (University of Amsterdam, The Netherlands), Götalog 2000 (Gothenburg University, Sweden), Bi-Dialog 2001 (Bielefeld University, Germany), and Edilog 2002 (University of Edinburgh, Great Britain).

DiaBruck received 50 submissions. Each was reviewed by at least 2 peers. 19 plus one reserve were accepted for the regular sessions, and 14 submissions for poster presentation. A call for demonstrations and project descriptions attracted 16 more submissions. The 20 full papers and 24 abstracts of accepted posters appear here.

The reviewers had a hard job given the tight schedule resulting from extending the submission deadline. But we tried to uphold the tradition of providing detailed comments, which we hope authors of both accepted and rejected papers find helpful in their further work. Thanks for reviewing go to Anton Benz, Barbara Di Eugenio, Bonnie Webber, Colin Matheson, David Traum, Elena Karagjosova, Hannes Rieser, Henk Zeevat, Jan Alexandersson, Johan Bos, Jonathan Ginzburg, Malte Gabsdil, Manfred Pinkal, Mary McGee Wood, Massimo Poesio, Michael Strube, Robin Cooper, and Stina Ericsson. I would especially like to thank Jonathan Ginzburg who acted as a co-chair in the selection process and helped greatly in the crucial decision making phase.

When organizing a workshop in these series, one important task is choosing a name, that traditionally consists of some form of *dial* or *log* as an affix to part of the name of the host city. Among the options considered but rejected this year were DiaBruecken or Dialuecken, SaarLog or SaarLogue, SaarDial or SaarDialogue, and DiaSaar. DiaBruck was chosen for being simple and disliked by the least voters. More importantly, DiaBruck is one of the alternatives that evoke dialogue and bridges: it will hopefully provide a forum for interaction bridging various layers and strands of dialogue research, thus contributing towards knowledge gain and stimulating exchange of ideas.

What is instrumental for achieving the workshop's goals is not the name but the participants. The papers and abstracts included here promise to provide a good basis for interesting discussions, which we hope the presenters and the audience will enjoy.

Credit for organization goes to Bettina Klingner and Claudia Kosny who devoted much energy and patience to local arrangements. Sabine Burgard and Adriana Davidiescu also helped in the finish. Johan Bos, Colin Matheson and Massimo Poesio provided reusable material from earlier workshops. Frederik Fouvry helped with the proceedings. Special thanks go to Claudia Kosny for professional handling of the website.

Finally, DiaBruck has been made possible by financial support from the Department of Computational Linguistics and Phonetics of the Saarland University, and from the Collaborative Research Centre on Resource-Adaptive Cognitive Processes (SFB 378).

Ivana Kruijff-Korbayová
DiaBruck 2003 chair

August 2003

Programme Committee and Referees

Ivana Kruijff-Korbayová (chair)
Jonathan Ginzburg (co-chair)
Anton Benz
Barbara Di Eugenio
Bonnie Webber
Colin Matheson
David Traum
Elena Karagjosova
Hannes Rieser
Henk Zeevat
Jan Alexandersson
Johan Bos
Malte Gabsdil
Manfred Pinkal
Mary McGee Wood
Massimo Poesio
Michael Strube
Robin Cooper
Stina Ericsson

Invited Speakers

Andreas Herzig, IRIT - Université Paul Sabatier, France
Martin Pickering, The University of Edinburgh, Great Britain
Nicholas Asher, University of Texas at Austin, U.S.A.

Tutorial: Good Practice in Empirically-Based Dialogue Research

David Traum, University of Southern California, U.S.A.
Laurent Romary, LORIA, Nancy, France
Michael Strube, EML, Heidelberg, Germany

Table of Contents

Invited Talks

Contributed Papers

Poster and Demonstration Abstracts

Bias, Tone and Questions in Dialogue

Nicholas Asher

University of Texas at Austin, USA

Ever since the work of Borkin and Linebarger (in the 70s) and Ladusaw (beginning of the 80s), the semantics of questions and their licensing of negative polarity items (NPIs) (*ever, any*) have been the subject of considerable scrutiny in formal semantics. Early on linguists also noticed that questions with NPIs like (1) had a biasing effect:

(1) Does Fred do a damn thing around the house?

Borkin argued that such questions were only acceptable when the speaker evidences a feeling of incredulity or an expectation of a 'no' answer. The observation by and large holds up, and the work on the semantics of questions and NPIs gives us a pretty good explanation of the bias effect. But missing in all of this is the role of intonation which gives intonational prominence to the NPI. Also missing is an account of positive polarity items in questions. And finally what accounts for the bias in framing a polar question negatively rather than positively as in (2a,b)

(2) a. Are you tired?
    b. Aren't you tired?

Further questions that are important for the analysis of dialogue also need to be addressed. How do questions affect the content of a dialogue? Is the bias of the question part of the semantic content or a pragmatic "side effect"? Should an account of dialogue really pay attention to the fine grained differences between

Are you tired? Are you at all tired? Are you somewhat/ a little bit tired? Aren't you tired?

I'll sketch a particular approach to questions in dialogue that comes largely from joint work with Alex Lascarides and attempt to address some of these questions in my talk.

Beliefs, intentions, actions and speech acts

Andreas Herzig and Dominique Longin

IRIT - Université Paul Sabatier, France

We present a logical framework integrating the notions of belief, intention and action that is build on epistemic logic and dynamic logic. Based on that framework we discuss the representation of speech acts as actions that can be characterized by pre- and postconditions, focussing on the evolution of beliefs and intentions when speech acts are performed.

Investigating the interactive-alignment model of dialogue

Martin Pickering

University of Edinburgh, Scotland

The interactive alignment account of language comprehension in dialogue (Pickering & Garrod, Behavioral and Brain Sciences, in press) assumes that the linguistic representations employed by the interlocutors become aligned at many levels, as a result of a largely automatic process. The process greatly simplifies production and comprehension in dialogue. It makes use of a simple interactive inference mechanism, enables the development of local dialogue routines that greatly simplify language processing, and explains the origins of self-monitoring in production. In the course of describing the account, I outline a number of recent experimental studies that provide support for the account.

Negotiative Spoken-Dialogue Interfaces to Databases

Johan Boye and **Mats Wirén**
Voice Technologies
TeliaSonera Sweden
Johan.Boye@teliasonera.com, Mats.Wiren@teliasonera.com

Abstract

The aim of this paper is to develop a principled and empirically motivated approach to robust, negotiative spoken dialogue with databases. Robustness is achieved by limiting the set of representable utterance types. Still, the vast majority of utterances that occur in practice can be handled.[1]

1 Introduction

The need for spoken dialogue with databases is rapidly increasing as more and more non-technical people access information through their PCs, PDAs and mobile phones. The related research area of natural-language (text-based) interfaces to databases has a long tradition, going back at least to around 1970. Still, a kind of culmination of this research occurred already in the 1980s (for an excellent overview, see Androutsopoulos et al. 1995). The high-end systems of this time, for example, TEAM (Grosz et al. 1985), LOQUI (Binot et al. 1991) and CLE/CLARE (Alshawi 1992, Alshawi et al. 1992) used linguistically-based syntactic analysis and powerful intermediary languages for semantic representation, and were able to engage in continuous dialogue involving complex phenomena such as quantification, anaphora and ellipsis. In part, this was possible because the systems were alleviated from the kind of noisy input generated by speech recognition, since text-based input was the only realistic option at the time.

Although some of today's commercial spoken-language information services also provide access to databases, they do not aim at being general database interfaces in the sense of the systems mentioned above. Rather, they provide a limited view of the database which is relevant to a particular task, for example, finding a train trip that fulfills certain constraints. This restricts the user to asking certain types of question that are compatible with the task, whereas other types are not allowed. In many cases, the system also keeps the initiative to itself by treating the user as an answer-supplier.

An advantage of this cautious approach to spoken dialogue with databases is that it allows a very simple semantic representation of user utterances, typically as flat slot–filler structures representing the propositional contents. For example, the utterance "I'd like a two-room apartment on the South Side" could be represented as [*number_of_rooms* = 2, *area* = *south_side*]. Hence, robust methods like phrase-spotting can be used for extracting the meaning even from noisy input.

However, progress in speech recognition and robust parsing during the last decade opens up the possibility of constructing more general spoken-dialogue database interfaces. In this paper, we will investigate *negotiative* spoken-dialogue interfaces to relational databases. "Negotiative" here means that the user is able to discuss and compare several alternative solutions to a problem (Larsson 2002).

[1]The authors would like to thank the Adapt group at TeliaSonera and KTH for discussions and feedback, and the participants at a presentation of this material at Göteborg University for valuable comments. This work was supported by the EU/HLT-funded project NICE (IST-2001-35293).

More specifically, the goal of the paper is to find an empirically based trade-off between robustness and expressiveness in spoken database dialogue: On the one hand, by deliberately restricting the expressiveness of the semantic representation formalism (keeping it suitably coarse-grained), we make it *apt for robust parsing*. On the other hand, the representation must be sufficiently fine-grained to *capture the distinctions actually made in spoken database dialogue*. We present a formalism that meets these goals, as well as a robust parsing algorithm for mapping speech recognizer output to this representation. The approach has been used successfully in Adapt (Bell et al. 2001), a multimodal dialogue system in an apartment domain.

2 Negotiative dialogue

The following piece of dialogue with the Adapt system provides an illustration of the kind of negotiative dialogue we are aiming to handle.

1. **User:** Are there any two-room apartments on the South Side that cost less than two million?
2. **System:** Do you have any other requirements?
3. **User:** A balcony would be nice.
4. **System:** There are seven such apartments. They are shown on the map.
5. **User:** I see ...this apartment on Horn street ...how much does it cost?
6. **System:** The apartment on Horn Street costs 1,900,000 kronor.
7. **User:** Ok ...is there anything cheaper?
8. **System:** The apartment on Blekinge street costs 1,500,000 kronor.
9. **User:** Hmm ...please tell me more about that apartment.

In utterances 1 and 3, the user homes in on the kind of apartment he wants to consider (the *specification* phase). The system briefly presents a number of results (in utterance 4) by indicating their position on the map, after which the user starts asking questions about the individual apartments, comparing their relative merits (utterances 5, 7, 9). This *negotiation* phase is what distinguishes negotiative dialogue systems from simpler systems.

To handle negotiative dialogue, the system must be able to distinguish specification utterances like "I'd like an apartment with a balcony" from utterances seeking information about a particular apartment, like "Does that apartment have a balcony?". The system must also be able to handle references to different objects in the same utterance, like "Is there anything cheaper than that apartment on King's street?".

3 Domain models

From the user's point of view, the purpose of a dialogue as exhibited above is to retrieve information about a set of interrelated objects, such as apartments, prices and addresses. The set of all such objects in the domain, together with their relations, constitutes the *domain model* of the system. From the system's point of view, the goal is then to translate each user utterance into an expression denoting a subset of the domain model (namely, the subset that the user is asking for), and to respond by either presenting that subset or ask the user to change the constraints in case the subset cannot be readily presented.[2]

We will assume that each object in the domain model is typed, and to this end we will assume the existence of a set of *type symbols*, e.g. $\mathsf{apartment}, \mathsf{integer}, \mathsf{money}, \mathsf{street\_name}$ etc., and a set of *type variables* $\mathsf{t}_1, \mathsf{t}_2, \ldots$ ranging over the set of type symbols. Each type symbol *denotes* a set of objects in an obvious way, e.g. $\mathsf{apartment}$ denotes the set of apartments. Both type symbols and type variables will be written with a sans serif font, to distinguish them from symbols denoting individual objects and variables ranging over individual objects, which will be written using an *italicized* font. The expression $\mathsf{t} : x$ is taken to mean the assertion "x is of type t".

Objects are either simple, scalar or structured.

[2] Naturally, this is somewhat idealized, as there are meta-utterances, social utterances, etc. that are not translatable to database queries. Still, 96 % of the utterances in our Adapt corpus correspond to database queries (compare Section 6).

Objects representable as numbers or strings are simple (such as objects of the type money or street_name). Scalar objects are sets of simple objects, whereas structured objects have a number of attributes, analogous to C structures or Java reference objects. Typically, structured objects correspond to real-world phenomena on which the user wants information, such as apartments in a real-estate domain, or flights and trains in a travel planning domain.

We will use the notation $x.a$ to refer to attribute a of object x. For example, an apartment has the attributes $size$, $number\_of\_rooms$, $price$, $street\_name$, $accessories$, etc., with the respective types square_meters, integer, money, street_name, set(accessory), etc. Hence if apartment : x is a true assertion, then so is square_meters:$(x.size)$.

Thus, a (structured) object o_1 might be related to another (simple, scalar or structured) object o_2 by letting o_2 be the value of an attribute of o_1. For instance, an apartment a is related to "King's street" by letting $a.street\_name = Kings\_street$. There is a standard transformation from this kind of domain models into relational database schemes (see e.g. Ullman 1988, p. 45), but domain models can also be represented by other types of databases.

We will further assume that types are arranged in a subtype hierarchy. The type t_1 is a subtype of t_2 (written as $\mathsf{t}_1 \leq \mathsf{t}_2$) if $\mathsf{t}_2 : x$ is a true assertion whenever $\mathsf{t}_1 : x$ is a true assertion.

4 Semantic representation formalism

In this section, we describe expressions called "utterance descriptors", which constitute the semantic representation formalism used internally in the Adapt system.

4.1 Constraints

Constraints express desired values of variables and attributes. The following are all examples of constraints:

- $x.street\_name = King\_street$
- $x.price < 2,000,000$
- $balcony \in x.accessories$
- $x.street\_name \in \{King\_street, Horn\_street\}$

4.2 Set descriptors

Set descriptors are expressions denoting subsets of the domain model. They have the form $?\mathsf{t}:x\ (P)$, where P is a conjunction of constraints in which the variable x occurs free. Such a set descriptor denotes the set of all objects x of type t such that P is a true assertion of x. Thus,

$$?\mathsf{apartment}:x\ (x.area = South\_side \wedge x.number\_of\_rooms = 2)$$

denotes the set of all apartments whose $area$ attribute has the value $South\_side$ and whose $number\_of\_rooms$ attribute has the value 2.

We may also add existentially quantified "placeholder" variables to a set descriptor without changing its semantics. For instance, the set descriptor above is equivalent to:

$$?\mathsf{apartment}:x\ \exists\mathsf{integer}:y\ (x.area = South\_side \wedge x.number\_of\_rooms = y\ \wedge\ y = 2)$$

Thus, set descriptors can also have the form $?\mathsf{t}_1 : x\ \exists\mathsf{t}_2 : y\ (P)$, where P is a conjunction of constraints in which x and y occur.

4.3 Representing context

Utterances may contain explicit or implicit references to other objects than the set of objects sought. For example, when the user says "A balcony would be nice" in utterance 3 of the dialogue fragment of section 2, he further restricts the context (the set of apartments) which was obtained after his first utterance.

Obviously, an utterance cannot be fully interpreted without taking the context into account. Thus the context-independent interpretation of an utterance is a function, mapping a dialogue context (in which the utterance is made) to the final interpretation of the utterance. In our case, a dialogue context is always an object or a set of objects (a subset of the domain model), and the final interpretation is a set descriptor, denoting the set of objects that are compatible with the constraints imposed by the user.

Accordingly, the context-independent interpretation of "A balcony would be nice" is taken to be

$$\lambda S\ ?\mathsf{apartment}:x\ (balcony \in x.accessories\ \wedge\ x \in S)$$

where S is a parameter that can be bound to a subset of the domain model. Thus the expression

above can be paraphrased "I want an apartment from S that has a balcony". The idea is that the ensuing stages of processing within the dialogue interface will infer the set of objects belonging to the context, upon which the functional expression above can be applied to that set, yielding the final answer. In the dialogue example of section 2, S will be bound to the set of apartments obtained after utterance 1.

An utterance may contain more than one implicit reference to the context. For example, "Is there a cheaper apartment?" (utterance 7 of the dialogue fragment of section 2) contains one implicit reference to a set of apartments from which the selection is to be made, and another implicit reference to an apartment with which the comparison is made (i.e. "I want an apartment from S which is cheaper than the apartment y").[3] Hence the representation is:

$$\lambda \mathsf{apartment}{:}y\ \lambda S\ ?\mathsf{apartment}{:}x \\ (x.price < y.price\ \wedge\ x \in S)$$

The contextual reasoning carried out by the Adapt system then amounts to applying this expression first to the apartment mentioned in utterance 6, and then to the set of apartments introduced by utterance 4.

Therefore, we define an *utterance descriptor* to be an expression of the form $\lambda X_1\ \ldots \lambda X_n\ U$, where X_i is either a set variable or a typed variable $\mathtt{t}{:}x$, and where U is a set descriptor in which the variables of $X_1 \ldots X_n$ occur free. Thus, an utterance descriptor is a function taking n arguments (representing the context), returning as result a subset of the domain model.

Yet an example is given by the utterance "How much does the apartment cost", which is represented by

$$\lambda \mathsf{apartment}{:}y\ ?\mathsf{money}{:}x\ (y.price = x)$$

Utterance descriptors can also contain type variables, when sufficient type information is lacking. For instance, "How much does it cost?" would be represented by

$$\lambda \mathtt{t}{:}y\ ?\mathsf{money}{:}x\ (y.price = x)$$

[3]For comparisons with sets of objects, see section 7.

4.4 Minimization and maximization

In many situations one is interested in the (singleton set of the) object which is minimal or maximal in some regard, for example the "biggest apartment" or the "cheapest ticket". To this end, we will further extend the notion of utterance descriptor.

Consider an expression of the form

$$?\mathtt{t}_1{:}x\ \mu \mathtt{t}_2{:}y\ (P)$$

where $\mathtt{t}_2$ is a numerical type and P is a conjunction of constraints in which x and y occur. We will take such an expression to denote the (singleton) set obtained by first constructing the set denoted by $?\mathtt{t}_1{:}x\ (P)$, and then selecting the object whose value for y is minimal. For instance, the utterance "Which is the cheapest apartment?" would by represented as

$$\lambda S\ ?\mathsf{apartment}{:}x\ \mu \mathsf{money}{:}y \\ (x.price = y\ \wedge\ x \in S)$$

When applied to a context set S, the function above returns an expression denoting the singleton set of the apartment in S whose price attribute has the minimal value. There is also a analogous maximization operator M.

5 Robust parsing

The robust parsing algorithm consists of two phases, pattern matching and rewriting. In the latter phase, heuristic rewrite rules are applied to the result of the first phase. When porting the parser to a new domain, one has to rewrite the pattern matcher, whereas the rewriter can remain unaltered.

5.1 Pattern matching phase

In the first phase, a string of words[4] is scanned left-to-right, and a sequence of constraints and meta-constraints, triggered by syntactic patterns, are collected. The constraints will eventually end up in the body of the final utterance descriptor, while the purpose of the meta-constraints is to guide the rewriting phase.

The syntactic patterns can be arbitrarily long, but of course the longer the pattern, the less frequently it will appear in the input (and the more

[4]Currently, 1-best output from the speech recognizer is used.

sensitive it will be to recognition errors, disfluencies etc.). On the other hand, longer syntactic patterns are likely to convey more precise information.

The solution is to try to apply longer patterns before shorter patterns. As an example, reconsider the utterance "I'm looking for an apartment on King's street", and suppose that "apartment on S" (where S is a street), "apartment" and "King's street" are all patterns used in the first phase. If the utterance has been correctly recognized, the first pattern would be triggered. However, the utterance might have been misrecognized as "I'm looking for an apartment of King's street", or the user might have hesitated ("I'm looking for an apartment on ehh King's street"). In both cases the pattern "apartment on S" would fail, so the pattern matching phase would have to fall back on the two separated patterns "apartment" and "King's street", and let the rewriting phase infer the relationship between them.

5.2 Meta-constraints

The pattern matching rules in the pattern matcher associate a sequence of constraints and meta-constraints to each pattern. The most commonly used meta-constraint has the form $obj(\mathtt{t}{:}x)$ which is added when an object x of type t has been mentioned. For instance, in the Adapt parser, the pattern "apartment" would yield

$$obj(\mathsf{apartment}{:}x_1)$$

whereas the pattern "King's street" would yield

$$obj(\mathsf{apartment}{:}x_2), x_2.street = Kings\_street$$

where x_1 and x_2 are variables. The existence of the object apartment : x_2 is inferred, since in the Adapt domain model, streets can only occur in the context of the *street* attribute of the apartment type. If the domain model would include also another type (restaurant, say) that also has an attribute *street*, the pattern could instead yield:

$$obj(\mathtt{t}{:}x_2), x_2.street = Kings\_street$$

where t is a type variable.

The meta-constraint $head\_obj(\mathtt{t}{:}x)$ is a variant of $obj(\mathtt{t}:x)$ that conveys the additional information that x is likely to be the object sought. We will illustrate the use of this and other types of meta-constraints in section 5.4.

5.3 Rewriting phase

In the rewriting phase, a number of heuristic rewrite rules are applied (in a fixed order) to the sequence of constraints and meta-constraints, resulting in a utterance descriptor (after removing all meta-constraints). The most important rules are:

- Unify as many objects as possible.
- Identify the object sought.
- Identify contextual references.
- Resolve ambiguities.

The first rule works as follows: Suppose pattern matching has resulted in:

$$obj(\mathsf{apartment}{:}x_1), obj(\mathtt{t}{:}x_2), x_2.street = Kings\_street$$

Then checking whether the two objects x_1 and x_2 are unifiable amounts to checking whether the types apartment and t are compatible (which they are, as t is a variable), and checking whether an apartment has an attribute *street* (which is true). Therefore the result after applying the rule is

$$obj(\mathsf{apartment}{:}x_1), x_1.street = Kings\_street$$

As for the second rule, the object sought is assumed to be the leftmost $head\_obj$ in the sequence. Failing that, it is assumed to be the leftmost obj. In the sequence above, that means apartment : x_1, which results in:

$$?\mathsf{apartment}{:}x_1\ (obj(\mathsf{apartment}{:}x_1), x_1.street = Kings\_street)$$

No more rewrite rules are applicable on this expression. The final result is obtained by adding the contextual argument S and by removing all meta-constraints:

$$\lambda S\ ?\mathsf{apartment}{:}x_1\ (x_1.street = Kings\_street\ \wedge\ x_1 \in S)$$

5.4 Example

The utterance "I'd like an apartment on Horn Street that is cheaper than the apartment on King's street" exemplifies the use of several rewrite rules in the Adapt system. First of all, "apartment on Horn Street" yields

$$obj(\mathsf{apartment}{:}x_1), x_1.street = Horn\_street$$

The pattern "cheaper" yields the sequence

$$\begin{aligned}&\mathit{head\_obj}(\mathsf{apartment}{:}x_2), \mathit{obj}(\mathsf{apartment}{:}x_3), x_2 \neq x_3,\\ &\quad \mathit{obj}(\mathsf{money}{:}y_2), x_2.z = y_2,\\ &\quad \mathit{obj}(\mathsf{money}{:}y_3), x_3.z = y_3, y_2 < y_3,\\ &\quad \mathit{ambiguous}(z, \{\mathit{price}, \mathit{monthly\_fee}\}, \mathit{default}(\mathit{price}))\end{aligned}$$

which is appended to the first sequence. The *ambiguous* meta-constraint conveys that the variable z is one of the attributes *price* or *monthly_fee*. If no further clues are against, z will be bound to *price*.

Finally, the pattern "apartment on King's Street" appends the sequence

$$\mathit{obj}(\mathsf{apartment}{:}x_4), x_4.\mathit{street} = \mathit{Kings\_street}$$

In the rewriting phase, objects are first unified in a left-to-right order. Thus x_1 and x_2 are unified, but the meta-constraint $x_2 \neq x_3$ prevents unification of x_2 and x_3. Instead, x_3 and x_4 are unified. The ambiguity is resolved (binding z to *price*). Thereafter, the variable x_2 is identified as the main object, and the implicit contextual reference argument S is added.

$$\begin{aligned}&\lambda S\ ?\mathsf{apartment}{:}x_2\ (x_2.\mathit{street} = \mathit{Horn\_street}\\ &\quad x_2 \in S, \mathit{obj}(\mathsf{money}{:}y_2), x_2.\mathit{price} = y_2,\\ &\quad x_2 \neq x_3, x_3.\mathit{street} = \mathit{Kings\_street},\\ &\quad \mathit{obj}(\mathsf{money}{:}y_3), x_3.\mathit{price} = y_3, y_2 < y_3)\end{aligned}$$

Finally, the variable x_3 is identified as a contextual reference. After removing meta-constraints, this results in:

$$\begin{aligned}&\lambda\mathsf{apartment}{:}x_3\ \lambda S\ ?\mathsf{apartment}{:}x_2\\ &\quad (x_2.\mathit{street} = \mathit{Horn\_street}\ \wedge\ x_2 \in S\\ &\quad \wedge\ x_3.\mathit{street} = \mathit{Kings\_street}\ \wedge\ x_2.\mathit{price} < x_3.\mathit{price})\end{aligned}$$

6 Discussion

Given that the goal of this paper is to find an empirically based trade-off between robustness and expressiveness in spoken database dialogue, there are two questions that need to be answered:

1. Is the parsing algorithm robust enough?
2. Is the formalism expressive enough?

For an answer to the first question, we refer to Boye and Wirén (2003). Basically, that paper demonstrates that the parser is robust in the sense of outputting utterance descriptors with a significantly higher degree of accuracy than the strings output by the speech recognizer.

To answer the second question, we need to look at the kinds of utterances that *cannot* be represented by the formalism. To this end, we have studied two corpora with transcriptions of database dialogue. One is from the Adapt apartment-seeking domain (Bell et al. 2000), comprising 1 858 user utterances, and the other is from the SmartSpeak travel-planning domain (Boye et al. 1999), comprising 3 600 user utterances. Both corpora are the results of Wizard-of-Oz data collections used for development of the systems. In both cases the wizard tried to promote user initiative as well as to simulate near-perfect speech understanding.

Below we provide a list of utterance types that are not representable as utterance descriptors, but instances of which are found in at least one of the corpora. The list was obtained by manually checking several hundred utterances from each of the two corpora and, in addition, searching the entire corpora for a variety of constructions judged to be critical.

1. Constructions involving a function of more than one structured object: "How many two- or three-room apartments are there around here?"

2. Complex and–or nesting: "A large one-room apartment or a small two-room apartment."

3. Selection of elements from a complementary set: "Are there any other apartments around Medborgarplatsen that are about 50 square meters big and that are not situated at the ground floor?"

4. Comparatives involving implicit references where the comparison is made with a *set* of objects rather than with a single object. To illustrate, assume that several flight alternatives and their departure times have been up for discussion previously in the dialogue. The user then asks: "Is there a later flight?", requesting a flight which is later than all the

previously mentioned ones.[5]

The most common of these types is (1), which accounts for 0.4 % in the apartment corpus (but does not show up at all in the travel corpus). In none of the other cases do the number of occurrences exceed 0.05 % of a single corpus. We thus conclude that the kinds of utterances that are not representable by our semantic formalism only occur marginally in our corpora.

It is also interesting to consider utterance types that we cannot handle and that *don't* appear in our current corpora. For example, TEAM (Grosz et al. 1985) handles constructions such as "Is the smallest country the least populous" (comparison involving two superlatives) and "For the countries in North America, what are their capitals?" ("each" quantification). Although it may be argued that these particular sentences are not typical of spoken negotiative dialogue, session data from other spoken database interfaces are certainly useful for the purpose of testing the formalism.

We set out by claiming that negotiative dialogue requires that we go beyond flat slot–filler structures. Indeed, even a quick look at the corpora reveals that a substantial part of the utterances *do* require the added expressiveness. Thus, in addition to trivial specification utterances such as "I'd like a two-room apartment on the South Side", one encounters numerous instances like the following that can be represented by our formalism, but not in general by flat slot–filler structures:

1. Specifications such as "I'd like an apartment with a balcony" as opposed to seeking information about a particular aspect of an apartment, like "Does that apartment have a balcony?".

2. Comparative constructions involving explicit references to different objects in the same utterance, such as "I'd like an apartment at Horn street which is cheaper than the apartment at King's street.".

3. Comparatives involving implicit references, such as "Is there anything cheaper?".

4. Superlatives: "The cheapest apartment near Karlaplan."

5. Combinations of a comparative and selection of a minimal element: "When is the next flight?", which can be paraphrased as "Give me the earliest flight that departs after the flight that you just mentioned."

7 Future work

The current Adapt parser assumes certain contextual references to refer to a single object rather than a set of objects (e.g. see the representation of "I want a cheaper apartment" in section 4.3). Obviously, a more general representation would be

$$\lambda S_2\ \lambda S_1\ ?\mathsf{apartment}{:}x\ \forall \mathsf{apartment}{:}y \\ (x.price < y.price\ \wedge\ x \in S_1\ \wedge\ y \in S_2)$$

("I want apartments from S_1 cheaper than all the apartments in S_2"). It would then be the task of the contextual reasoning component to infer the set S_2. In the same vein, a more general form of representing "How much do the apartments cost" would be

$$\lambda S\ ?\mathsf{money}{:}x\ \exists \mathsf{apartment}{:}y\ (y.price = x\ \wedge\ y \in S)$$

(i.e. "I want the prices of the apartments in S"). As is clear from these examples, such an extension, allowing the user to refer to sets of objects, would require the parsing algorithm to infer which variables are bound by universal quantifiers and which are bound by existential quantifiers[6].

This extension can be realized by introducing two new meta-constraints $all\_obj(\mathtt{t} : x)$ and $some\_obj(\mathtt{t} : x)$. The pattern "cheaper" would then yield:

$$head\_obj(\mathsf{apartment}{:}x_2), all\_obj(\mathsf{apartment}{:}x_3), \\ x_2 \neq x_3,\ obj(\mathsf{money}{:}y_2),\ x_2.z = y_2, \\ obj(\mathsf{money}{:}y_3),\ x_3.z = y_3,\ y_2 < y_3, \\ ambiguous(z, \{price, monthly\_fee\}, default(price))$$

signalling that x_3 should be universally quantified. Similarly, "How big" would yield:

$$head\_obj(\mathsf{square\_meters}{:}y), \\ some\_obj(\mathsf{apartment}{:}x), x.size = y$$

[5] To determine whether such a comparison is made with a set of objects or with a single object, it is in general not sufficient to look only at the last utterance. Thus, to handle this, the context-independent representation of the utterance must cater for both possibilities, thereby allowing the contextual analysis to make the final verdict (see further Section 7).

[6] The current approach avoids this issue by allowing references only to individual objects, in which case the difference between universal quantification and existential quantification disappears.

signalling that y should be existentially quantified.

Future work involves implementing and evaluating the need and robustness of this extension.

8 Related work

Approaches to handling spoken dialogue with databases can be largely divided into two types:

1. General-purpose linguistic rules and powerful semantic formalisms: not robust enough; overly expressive.

2. Pattern matching and flat slot–filler lists: robust but not expressive enough.

Several attempts at synthesizing these approaches have been made, either by "robustifying" (1) (for example, van Noord et al. 1999) or by extending (2) with the capability of handling general linguistic rules (Milward and Knight 2001). However, the semantic representations produced are still limited to that of flat slot–filler lists, and as a result they are not suitable for negotiative dialogue.

We have shown empirically that the subclass of questions handled by our approach is a useful one. In a similar vein, Popescu et al. (2003) define a class of "semantically tractable questions" to their PRECISE system. The idea is that each semantically tractable question provably yields a correct answer, whereas questions outside of the defined class are answered with an "error message" which allows the user to reformulate her question. Hence, the PRECISE system is "robust" in a conservative way, since it guarantees that no incorrect answer will ever be produced. On the other hand, PRECISE is text-based and has no dialogue capabilities, and thereby circumvents the problems introduced by speech-recognition errors and underspecified utterances.

9 Conclusion

The aim of this paper has been to develop a principled and empirically motivated approach to robust, negotiative spoken dialogue with databases. Key to our approach is the choice of semantic representation formalism. On the one hand, it is more expressive than today's commonly used, flat slot–filler lists that are limited to representing the propositional contents of utterances. On the other hand, it is still strongly restricted to make it compatible with the kind of robust parsing needed for spoken dialogue. While more empirical investigation is needed, experience with our current corpora indicates that the robustness–expressiveness tradeoff described here is a reasonable first approximation.

References

Alshawi, H. *The Core Language Engine*. The MIT Press, 1992.

Alshawi, H., Carter, D., Crouch, R., Pulman, S., Rayner, M. and Smith, A. CLARE — A Contextual Reasoning and Cooperative Response Framework for the Core Language Engine. Final report, SRI International, 1992.

Androutsopoulos, I., Ritchie, G. and Thanish, P. Natural Language Interfaces to Databases — An Introduction, *Journal of Natural Language Engineering* 1(1), pp. 29–85, 1995.

Bell, L., Boye, J., Gustafson J. and Wirén, M. Modality Convergence in a Multimodal Dialogue System. *Proc. Götalog*, pp. 29–34, 2000.

Bell, L., Boye, J. and Gustafson J. Real-time Handling of Fragmented Utterances. *Proc. NAACL Workshop on Adaptation in Dialogue Systems*, 2001.

Binot, J-L., Debille, L., Sedlock, D. and Vandecapelle D. Natural Language Interfaces: A New Philosophy. *SunExpert Magazine*, pp. 67–73, January 1991.

Boye, J., Wirén, M., Rayner, M., Lewin, I., Carter, D. and Becket, R. Language Processing Strategies and Mixed-Initiative Dialogues. In *Proc. IJCAI workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 1999.

Boye, J. and Wirén, M. Robust Parsing of Utterances in Negotiative Dialogue. *Proc. Eurospeech*, 2003.

Grosz, B., Appelt, D., Martin, P. and Pereira, F. TEAM: An Experiment in the Design of Transportable Natural-Language Interfaces. Technical note 356, SRI International, 1985.

Larsson, S. Issue-based Dialogue Management. Ph.D. Thesis, Göteborg University, ISBN 91-628-5301-5, 2002.

Milward, D. and Knight, S. Improving on Phrase Spotting for Spoken Dialogue Processing. *Proc. WISP*, 2001.

van Noord, G., Bouma, G., Koeling, R. and Nederhof, M-J. Robust Grammatical Analysis for Spoken Dialogue Systems. *Journal of Natural Language Engineering*, 5(1), pp. 45–93, 1999.

Popescu, A., Etzioni, O. and Kautz, H. Towards a Theory of Natural Language Interfaces to Databases. In *Proc. International Conference on Intelligent User Interfaces (IUI-2003)*, Miami, Florida, 2003.

Ullman, J. *Database and Knowledge-base Systems*, Volume I. Computer Science Press, 1988.

Referential domains and the interpretation of referring expressions in interactive conversation

Sarah Brown-Schmidt
Department of Brain and Cognitive Sciences
University of Rochester
sschmidt@bcs.rochester.edu

Michael K. Tanenhaus
Department of Brain and Cognitive Sciences
University of Rochester
mtan@bcs.rochester.edu

Abstract

This paper describes research investigating the on-line production and interpretation of referring expressions during interactive conversation. In particular, we focus on the interactive processes by which interlocutors establish shared referential domains. In a set of interactive, task-based dialogs, we show that referential domains constrain both the form of referring expressions, and their interpretation. We argue that various task-based factors strongly affect the referential domains used by interlocutors, and that understanding the mechanisms of reference interpretation will require a careful analysis of how these factors affect the referential domains used in interactive conversation.

1 Introduction

Although the generation and interpretation of definite reference has played a central role in real-time sentence processing research, little is known about how addressees interpret referring expressions on-line in interactive conversation. Much of the existing literature investigating the real time interpretation of referring expressions in spoken language comprehension focuses on the interpretation of noun phrases such as "the cube" in sentences like "Put the cube in the can". These sentences are embedded in tasks in which participants are instructed to manipulate a set of objects which are placed on a table in front of them. The instructions are typically pre-recorded, and the referential domain for interpreting the referring expressions is assumed to be the entire workspace. The experimental situations are typically non-interactive, in that the subject simply follows instructions, and does not converse with another person. Research in these constrained contexts suggests that addresses use multiple sources of information to restrict the domain of interpretation of referring expressions, including common ground (Hanna, Tanenhaus & Trueswell, in press), verb -based constraints (Altmann & Kamide, 1999), task relevant properties of objects (Chambers, et al. 2002) and contrast implied by use of a scalar adjective (Sedivy, et al. 1999; 2003).

The findings from constrained contexts indicate that pragmatic factors particular to the context in which a reference is uttered, are key to understanding how that reference is interpreted. However, detailed analyses of how these factors might arise during a conversation are less well understood. For example, the referential domain at the beginning of each instruction in a standard task is generally assumed to be the set of experimental items placed in front of the subject. However, in a natural discourse context, it is possible that the referential domain could include other objects in the room, such as the items on shelves, or that the referential domain could include only a subset of the experimental items. While experiments in constrained situations sug-

gest that linguistic and non linguistic factors both act to constrain this initial referential domain, it is not well understood how these factors are used during conversation.

We present data from two experiments that investigated the production and interpretation of referring expressions in an interactive task-based dialog between two naive participants. The first experiment shows that referential domains can be quite restricted and closely aligned between interlocutors. Speakers frequently used referential expressions that would be ambiguous if the domain were less restricted and addressees were not confused by these expressions, indicating that these potential entities were never considered as potential referents. We suggest that these effects result from domains becoming restricted and coordinated because of task-based factors. In the second experiment, we verified this observation, investigating the role of explicitly mentioning the referential domain before the onset of the referring expression. As in the first experiment, we found that when the referential domain was sufficiently restricted, listeners quickly interpreted the referring expressions without interference from other competing referents that were outside the domain.

2 Method

In both experiments, two naïve participants engaged in a referential communication task (Krauss & Weinheimer, 1966) in which they worked together to complete a task. The specific details of each experiment will be described in more detail below. In both tasks, the participants could not see one another, and were working with game pieces on physically separate, but matching workspaces. For both tasks, participants needed to instruct each other to move game pieces in order to successfully complete the task. We did not place restrictions on the way in which the participants spoke to one another. However, the characteristics of the task and the game pieces allowed us to investigate hypotheses about the interpretation of referring expressions, through naturally arising utterances. We employed a version of the visual-world eye-tracking methodology (Tanenhaus, et al., 1995) in which we obtained a record of one subject's eye-fixations with the use of a light-weight head-mounted eye-tracker.

Previous work using the visual-world eye-tracking methodology demonstrates that listener's fixations are closely time-locked to speech input. For example, in a task where a subject is asked to "Put the apple next to the frog", approximately 200ms following the onset of the word "apple", participants are more likely to look at the apple, than other unrelated objects in the scene (such as a can). A related finding was reported by Allopenna, et al. (1998). When participants hear an instruction such as "Click on the cloud" when viewing a computer screen which has pictures of a cloud, a clown, a dog, and a parrot, listeners are equally likely to look at the clown and the cloud upon hearing the onset of "cloud". This effect is due to the fact that "cloud" and "clown" begin with the same sequence of phonemes. When participants hear the disambiguating sounds in "cloud", they reliably look to the correct referent. This effect is commonly referred to as a "cohort effect", and words like "cloud" and "clown" are often referred to as cohort competitors in the spoken word recognition literature.

Our experimental methodology is partially based on the cohort effect. In designing our experiments, we used some game pieces which had pictures on them. We carefully selected easily nameable pairs of pictures that were cohort competitors, such as "cloud" and "clown"). Most of the pictures were selected from a database of normed pictures (Snodgrass & Vanderwart, 1980) and were easily recognized by our participants. Because the task required participants to refer to the game pieces, we expected to observe cohort effects during references to blocks with cohort competitors. Presumably in the Allopenna, et al. (1998) study, all four items pictured on the computer screen were included in the referential domain used by the listener. We predicted that if the referential domain was significantly restricted, that in some cases the target referent and the cohort competitor would be in different referential domains. In these situations, we expected that upon the reference to the target, the proportion of looks to the cohort competitor would be significantly reduced. By tracking the presence of cohort effect, we are able to gauge the size of the referential domain.

3 Experiment 1

In experiment 1, we monitored eye movements as pairs of participants, separated by a curtain, worked together to arrange blocks in matching configurations and confirm those configurations. We reported a more comprehensive analysis of this dataset in Brown-Schmidt, Campana & Tanenhaus (in press). Here we focus on the aspects of the data related to cohort effects and the circumscription of referential domains. During the task, participants placed 56 different blocks over the course of 2.5 hours. All 4 pairs of participants developed idiosyncratic ways of referring to the objects, and also developed strategies for completing the task. A popular strategy, for example, was to finish placing blocks in one area of the workspace before moving on to the next. Additionally, the partners tended to move from one area to an adjacent area, suggesting they had a preference to build off of structure they had already created.

Over the course of the experiment, each pair generated approximately 75 references to blocks with cohort competitors, like cloud and clown. While cohort competitors were only placed 3.5 inches apart, during the course of the conversation we did not observe a cohort effect. Upon hearing the word "cloud", listeners looked primarily at the target referent (the block with a picture of a cloud on it) and were no more likely to look at the clown than at an object in the scene with a completely unrelated name, such a penguin. This observation suggests that during the conversation, the cohort competitors were not included in the referential domain. However, we *did* observe a cohort effect during instructions which were not constrained by the task-related conversation. Periodically, participants needed to remove the eye-tracker to take a break. On one occasion when we put the tracker back on and recalibrated, we tested the calibration by asking the subject to look at different items on the board, using instructions like "Look at cloud, look at the lamb, look at the seal." Here we saw clear cases of the subject initially looking at the cohort competitor (e.g. clown, lamp) before looking at the intended referent (e.g. cloud, lamb). While the cohort effect appears large, the 15 trials of this sort did not give us enough statistical power to replicate a standard cohort effect, but the pattern of fixations and mean differences between cohorts and targets are similar to those found in Allopenna et al. (1998).

These results suggest that listeners can use tightly circumscribed referential domains during reference interpretation imbedded in a dialog. Unlike the studies using more constrained contexts, the referential domain did not include all of the objects in the participants view- in some cases this would be a large number of blocks. Instead, it appears that strategies which partners mutually developed in order to complete the task, facilitated the use of small, task-relevant referential domains. These observations supported the primary result from this experiment which was that speakers tended not to modify a noun phrase, e.g., saying "the red block" rather than "the **vertical** red block" even when there was more than one red block in the scene. The situations under which speakers *did* choose to modify noun phrases was when the second red block was physically close to the intended referent and it fit the task constraints. When an unmodified NP was used, addressees' eye movements were primarily restricted to the intended referent, suggesting that non-linguistic factors guided the interpretation of these linguistically ambiguous references.

4 Experiment 2

In experiment 2, we created conditions where cohort competitors were more or less likely to be in the same referential domain. Conversational partners took turns instructing one another to click on objects on a computer screen as we monitored the eye movements of one partner and the speech of both partners. On each trial, each participant's screen contained an identical set of 14 pictures, separated into two domains which looked like 'islands'. At the beginning of each trial, a picture on one participant's screen became highlighted. This was a cue to tell their partner to click on this object. Participants were encouraged to speak freely in order to perform the task and no restrictions were placed on how they chose to describe any of the objects. Target objects always appeared with a cohort competitor and we ma-

nipulated whether the cohort appeared on the same or different island as the competitor. We predicted that in cases where the speaker specified the location of the target (e.g. "on the top island") *before* the onset of the referring expression, that this would establish that island as the appropriate referential domain. If the cohort competitor were on a different island than the target, and if the speaker chose to specify the location information before the noun phrase, then we predicted the cohort effect would be eliminated.

When the cohort competitor appeared on the same island as the target, we observed a standard cohort effect, replicating previous findings using pre-recorded instructions (Allopenna, et al. 1998). Approximately 52% of the time, participants specified which island the target was on before the onset of the noun phrase. In these constructions, when the cohort was on a different island than the target, the cohort effect was eliminated, suggesting that specification of the referential domain restricts attention to entities within that referential domain.

The results from this experiment suggest that our subject's explicit (and unscripted) establishment of the referential domain successfully constrained the interpretation of a subsequent referring expressions. We also observed that speakers tended to explicitly mention when the referential domain would *change*. On each trial, the speaker referred to two different objects. Half of the time, the second object was on a different island than the first. When the second object switched islands, speakers were more likely to explicitly ground which island the second referent was in. This strategy was likely to be helpful to listeners (we are currently analyzing the data to find out). Additionally, this adds support to the observation from Experiment 1 that participants tend to work on the task in a highly localized manner, only moving to a new area of the workspace when the previous area has been completed. We are interested in exploring whether these tendencies are specific to the kinds of tasks we selected, or are related to more general properties of discourse and expectancy for upcoming reference.

5. Discussion

By combining the cohort competition effect, well documented in the word recognition literature, with a referential communication task, we were able to observe how participants with shared task-goals circumscribed referential domains. We found that referential expressions were interpreted with respect to a restricted referential domain, and that these referential domains were closely aligned between conversational interlocutors. These results replicate and extend previous studies demonstrating that referential domains are constrained by contextual and pragmatic factors (Chambers, et al, 2002; Hanna, et al. in press; Hanna & Tanenhaus, in press). Our results also demonstrate that it is possible to study real-time language processing in interactive conversation with the same precision as is typically achieved in controlled laboratory settings with scripted, pre-recorded language. We expect that a satisfactory understanding of the mechanisms of reference interpretation will require addressing the many factors that affect referential domains during interactive conversation.

Acknowledgement

This material is based upon work supported by the National Institutes of Health under award number NIH HD-27206 to M.K. Tanenhaus.

References

Allopenna, P.D., Magnuson, J.S. & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419-439.

Altmann, G.T.M. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264.

Brown-Schmidt, S., Campana, E. & Tanenhaus, M.K. Real-time reference resolution by naïve participants during a task-based unscripted conversation. To appear in J.C. Trueswell & M.K. Tanenhaus (eds.), World-situated language processing: Bridging the language as product and language as action traditions. (MIT Press).

Chambers, C.G., Tanenhaus, M.K, Eberhard, K.M., Filip, H & Carlson, G.N. (2002). Circumscribing referential domains in real-time sentence comprehension. *Journal of Memory and Language*, 47, 30-49.

Hanna, J.E. & Tanenhaus, M.K. (in press). Effects of task constraints and speaker goals on addressee's referential domains in a collaborative task. *Cognitive Science*.

Hanna, J.E., Tanenhaus, M.K. & Trueswell, J.C. (in press). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*.

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

Sedivy, J.C., Tanenhaus, M.K., Chambers, C., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109-147.

Sedivy, J.C. (2003). Informativity expectations and resolving reference: Some evidence from language processing and development. Paper presented at the City University of New York conference on Human Sentence Processing, 2003.

Snodgrass, J.G., & Vanderwart, M. (1980). *Journal of Experimental Psychology: Human Learning and Memory*, 6:3, 174-215.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (l995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 632-634.

Influence of regional features on Map-Task dialogues

Marina Castagneto
Università di Cagliari - Dipartimento di Linguistica e Stilistica
Via Is Mirrionis 1 - Cagliari - Italy - e-mail: m.castagneto@virgilio.it
and
Giacomo Ferrari
Università del Piemonte Orientale - Department of Human Studies
Via G.Ferraris 116 - 13100 Vercelli - Italy - e-mail: ferrari@lett.unipmn.it

Abstract

In this paper Map Task dialogues collected in different regions of Italy are compared and some differences in the interaction styles are highlighted and discussed, especially under the point of view of *considerateness* and *involvemente*. The impact of the recognised regional features on the model of dialogue is stressed.

1 Introduction

Dialogue modelling has been often carried in the assumption that people interact almost in the same way and according to the same schemata. A difference has been made between task-oriented dialogues (see Allen et al., 2001), where the structure of the task dictates the structure of the dialogue, and casual conversation, where the structure is often unpredictable.
In this paper we will show that even in task oriented dialogues there may be different ways of carrying on co-operation between participants. In particular, there exist different approaches to communication and they are certainly related to the anthropological and cultural background of the dialogue participants.
The background of the participants affects the dialogue in many ways, ranging from turn-taking strategies (overlap) to the use of pronouns, or the problem-solving strategy itself. Here we will focus the interaction style according to Tannen's approach, and we will draw some consequences on the length of dialogue, as well as on the nature of some dialogue moves.
We are not advocating for the design of culture-sensitive dialogue (computational) models, but we want to stress the fact that we still know very few of the different aspects that may more or less deeply affect the way how people interact with one another, and that may make a communicative style look more or less coherent with the partecipants anthropological profiles. [1]

2 A task-oriented dialogue: Map Task and the Italian MapTask

MapTask (Brown et al., 1984) has become a sort of standard for the collection of task-oriented dialogues. The best known and more extensively studied corpus of MapTask dialogues is the one collected first in Glasgow around 1991 (Anderson et al., 1991). Map Task dialogues are unscripted but are elicited according to a well defined task schema and they respond, implicitly, to the assumption that a relatively uniform solution strategy and a uniform communication style would be used by dialogue participants facing a common and neatly identified goal.

A collection of Map Task dialogues has been carried in Italy, since 2000, by a consortium of Universities funded by the Ministry of Education (the project AVIP). The research centres involved in the collection of dialogues have been the Laboratorio di Linguistica of Scuola Normale Superiore, Pisa, the CIRASS (Centro Interdipartimentale di Ricerca per l'Analisi e la Sintesi del Segnale) of the Università Federico II, Naples, the Laboratorio di fonetica of Istituto Universitario Orientale, Naples, the Dipartimento di Elettrotecnica ed Elettronica of Politecnico, Bari, and the Dipartimento di Studi Umanistici of the Università del Piemonte Orientale, Vercelli. Thus dialogues have been collected in Naples, Bari, Pisa and, less extensively, Vercelli. For those who are not familiar with Italian geography, each collection area is located in a different area of Italy, namely: Vercelli (in Piemonte region, North-Western Italy), Bari (regional capital of Puglia region, South-Eastern Italy), Pisa (in Toscana region, Central Western Italy) and Napoli

[1] Section 3 and section 4.2 are totally due to Marina Castagneto, while the rest of the paper has been written in cooperation

(regional capital of Campania region, South-Western Italy).
The main focus of research of AVIP has been the carrying of studies on the regional varieties of Italian. Unlike other languages, the impact of regional features on standard Italian is so strong as to allow researchers to clearly distinguish regional varieties of language with their own characteristics and a wide range of use (Berruto, 1987). Regional features affect, mainly, the phonetic and the prosodic level; thus research has been carried in depth on these aspects of the dialogues.
However, a pragmatic analysis of the dialogues, though superficial, allows the hypothesis that even at this level, regional features affect at least the communicative style.

3 Communicative styles through regional areas

The first striking property of Italian dialogues is the variation in length; there are dialogues of even 25 or 30 minutes, face to relatively average lengths of the Scottish dialogues. In addition the Scottish dialogues seem to be exactly how we expect a map-task dialogue should be at least in four ways:

- as the interlocutors have a shared knowledge and the very same task, no conflict about *face* (Goffman, 1955) arise between them. As their reciprocal roles are previously established, the use of some linguistic forms (ex. the imperative mood) cannot be considered a *face threatening act* (Brown and Levinson, 1987), as it may be in other situations

- Scottish givers and followers are inclined to observe strictly the Grice co-operation principles "Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged", as well as the *four maxims*

- according to the theoretical paradigm of Tannen (1984), we'd say that Scottish dialogues are all in a *high-considerateness style*, that is to say they focus on message rather than on the relationship among the speakers, who show distance between them

- in addition they follow the Rule of Rapport *don't impose* , as proposed by Lakoff(1979). The speakers feel clearly detached from what they say and from each other.

The pragmatic situation of Italian dialogues is far more intricate and not homogeneous at all. There are many dialogues that parallel the style of the Scottish ones, but also a good number of them exhibit uncommon interaction styles, with high rate of overlap and, above all, a number of utterances that cannot be assigned a move label in the traditinal Map Task tag-set.
In a deeper analysis, we can take two dimensions to be relevant: *considerateness* and *involvement*. This last property applies to both the giver-follower and the subject-researcher relation.
However we should not consider the depicted conversational style, the *high-considerateness style*, defined by distance, and its opposite conversational style, the *high-involvement style* (accepting again the terminology of Tannen, 1989) as discrete notions. Rather, we will consider them as the extreme points of a continuous scale about conversational styles on which we could place the sub-corpora of dialogues from Vercelli, Bari, Pisa, Napoli. In order to make an evaluation, we have taken as evidence the metaphors used to transform a Map Task into a personal adventure, the expressions of anxiety and doubt about the results of the experiments, face-threatening acts, negotiation subdialogues even where there is nothing to negotiate (beginning and end of dialogue).

3.1 Vercelli

Thus Vercelli dialogues show a lot of pragmatic similarities with HCRC dialogues, falling into the *high-considerateness* style area, except for a certain degree of anxiety about making a good impression on researchers (so we have a first little step of involvement on our scale). Exactly as in HCRC dialogues, the one who starts the dialogue is the giver (as it is expected to be) and we have a sudden start with a *giver instruct*, coming directly to the point. In the same way the dialogues tend to end abruptly, as in dialogue D06_v:

G196: e quello l'arrivo
and that is the arrival
F197: ah OK <sussurro> va bene
ah OK <whisper> all right
G198: quindi sei arrivata all'arrivo,
so you have come to the arrival,
praticamente
as a matter of fact
F199: OK
OK

As there is no involvement with the subject-matter, the story is not dramatized, so we lack style

figures of speech, metaphors and imageries, narration, i.e. the features considered by Tannen (1989) among the strategies signalling high-involvement. Neither we could find in these dialogues any of the linguistic signals of involvement (in the term of the authors *immediacy*) listed by Wiener and Mehrabian (1968): use of inclusive first person plural pronouns, high frequency of deictic particles, high degree of specificity in denotation, etc. Also, there are very few overlaps, and they are not so much extended: this is a very important parameter to check, maybe the strongest marker of a lack of involvement. The rare cases in which we can find a certain degree of involvement are the ones in which giver and follower were already friends with each other.

3.2 Bari

Bari dialogues are still in a *high-considerateness* style area, as they lack the most peculiar parameters of *high-involvement* style too: there are very few deictic particles used as demonstratives or first person pronoun with inclusive value, while the language-level is still formal, without dialectal and slang words to show *camaraderie* (according to Lakoff, 1979). Yet, in Bari dialogues we can recognize a certain amount of anxiety and uncertainty about the fulfilment of the task, betraying a first step of involvement on our scale. In most of them, when the participants come to the arrival, they psychologically refuse to stop, starting again the whole map-task in order to check if everything is OK. The extreme case is represented by dialogue D02_b, in which the task is accomplished four times: after the first time the follower asks the giver to do it again in order to look for the landmarks she was not able to find yet (F080), so the giver repeats everything. Then the follower, not yet satisfied, declares that now it's her turn to try (F229), and they perform the map task once again. But the game is not over yet. At the very end the follower asks to her giver "do you want me to do it again?" and the giver answers "as you like" (F369-G370).
Another piece of evidence for involvement in Bari dialogues is that the reciprocal attitude between givers and followers is not properly neutral, but there is a little amount of aggressiveness showing in some face threatening acts, with or without redressive actions (as chuckles). Being aggressive is just a possible way to express that you care. In particular, in these dialogues the follower tends to be more aggressive than in the rest of the dialogues in the Italian corpus, as if he couldn't accept the temporary inferiority given by his role. So he pushes his giver to obtain all the information lacking to him without saying anything in turn on the difference of landmarks between the two maps. He simply doesn't change his behaviour until he obtains a "global view" taking the lead over his giver. That's why most of Bari dialogues are started by the follower, and we find very often (false) questions like "so?", "and then?", "now what?", a sort of lightly aggressive way to push the giver to offer new instructions and information. All D01_b dialogue is built up in this way, as in:

G003: il tuo punto di pa+
your point of st+
F004: allora dai Ivo
go on, Ivo
G005: volevo sapere il tuo punto di
I wanted to know if your starting
partenza coincide con il mio,
point corresponds with mine
cioè
F006: c'è scritto "aia" al mio punto di
there is "farmyard" written on
partenza
my point of departure
G007: okay
okay
F008: quindi
so what

Moreover, and this is crucial, in Vercelli and Bari dialogues there is no emotional involvement at all with what dialogue acts expresses.

3.3 Pisa

The presence of features of emotional involvement allows us to consider Pisa dialogues already in *high-involvement style* dialogue area. In dialogue A03_p, for example, the follower behaves as if the landmarks were real places, as if he went there living an adventure. Asked to go toward a river (but he has no river on his map) he answers with a new question on whether he could use boats (as he has got boats on his map); then he wants to know, when hungry, if there is a pastry shop somewhere, and so on.

3.4 Naples

The last group of dialogues is from Naples. These dialogues are in a *high-involvement* conversational style, and, indeed, they represent a prototypical *high-involvement style*, with all its set of features. Beyond the identification with the text, here we find a strong "metamessage" of rapport between the communicators, who thereby experience that they share communicative conventions and inhabit

the same world of discourse" (see Tannen, 1989). Here both the beginnings and the ends of dialogues tend to be negotiated, as in dialogue C04_n:

G001: allora tu stai a partenza, giusto?
so you are at "start", are you?
F002: io sto a partenza, mo' dobbiamo
yes, I am at "start", now we have
trovare l'arrivo
to find the arrival

The negotiation in the beginning of a dialogue is not aimed at solving the task, obviously, but at establishing a communicative relationship between the giver and the follower. This is the way in which the same dialogue ends:

G329: prima calcola un pollice,
first calculate a thumb,
schiattaci l'arrivo e hai finito
squash the arrival down and you can put an end on it
F330: diciamo che dovrebbe essere aspe'
(let's say it should be...wait
G331: mamma mia! (risata)
goodness gracious! (laugh)
F332: aspe'
wait
G333: è stata una battaglia
it has been a real battle
F334: (risata) è stata è stato bello però!
(laugh) and still it has been great!
aspetta, quindi diciamo che un pollice
wait, a thumb should be
dovrebbe essere
...let's say
G335: noi abbiamo finito, tu sei arrivato
game's over, you arrived
F336: ah! (risata)
ah! (laugh)
G337: sano e salvo
safe and sound
F338: e su per giù, okay, a posto?
more or less, okay, it's alright?
G339: allora che dobbiamo fare mo'?
What else should we do?
Abbiamo finito!
we've arrived!
F340: noi penso che abbiamo finito oo
I suppose we've arrived...
ho trovatooo penso che ho trovato
I found the.. yes I suppose I found
l'arrivo si penso
the arrival I suppose

This dialogue ending section lasts twelve moves, it has a lot of (false) interrogative and exclamatory sentences, there is a repeated use of inclusive first person plural pronoun, a structural metaphor ("it has been a battle" -the very same metaphor discussed by Lakoff and Johnson, 1980), an idiomatic syntagm ("safe and sound") and so on. Moreover, in Naples dialogues there are frequent comments and narrations built up on the names of landmarks with a certain degree of imagination. see, for example, the dialogue C04_n:

F292: l'ho accerchiato il leone
I've surrounded him, the lion
G293: hai accerchiato il leone, bravo!
You've surrounded the lion,
sparalo (risata)
congratulation! Shoot him (laugh)
F294: si (risata) m'ha morso/
sure (laugh) he bit me/
m'ha mors 'o leone
he bit me, the lion
G295: sparagli
shoot at him

The position of the aforesaid sub-corpora of dialogues on this scale is justified by the presence or absence of some linguistic and pragmatic markers or set of markers signalling the greater or smaller proximity or distance from the two polarized opposite styles:

from	*high considerateness*	HCRC samples
		Vercelli
		Bari
		Pisa
to	*high involvement*	Napoli

4 Regional features

It turns out that *high involvement* substantially affects the communicative style and the dialogue structure itself. We assume that these consequencies affect at least three levels, the syntax of utterances, the turn-taking strategies and the dialogue structure itself.

4.1 Syntactic level

As we have remarked in different sections, there are some syntactic features that seems to be typical of the *involvement* style. These are the use of inclusive first person plural pronouns instead of first person singular and, more significantly, imperatives or directives. This is not a grammatical variant of a directive, but just a way of saying *we'll do it together.*

4.2 Turn-taking

In *high-involvement* style dialogues, especially in the Neapolitan ones, givers and followers continuously overlap and interrupt each other; this atti-

tude is not a sign of aggressiveness, but, on the contrary, it's a sign of co-operation. Many researchers consider overlapping sentences in which the interrupting speaker integrates or accomplishes his interlocutor's sentence as a marker of involvement. Tannen (1984) includes them in a so-called "enthusiasm strategy", Jefferson (1973) considers them a type of "collaborative sentences", while for Duncan (1974) they must be classified among the "backchannels behaviours" and Ervin-Tripp (1987) states that "groups which emphasize solidarity and 'positive politeness' are likely to look for evidence of closeness through the production of joint text, proxy completion and simultaneity". Actually this kind of sentences created a hard problem of pragmatic annotation, so we decided to create a new label for this move, [continue], integrating the HCRC system; for ex. (dial.C04_n):
G025: devi scendere # F026 [continue] e devi andare a # sinistra
G025: you have to go down # F026 [continue]: and you have to go towards # left
In case the interrupted speaker regains the conversational turn and leads to an end the sentence, we have to face two possibilities:
1) if the follower makes the right guessing about the intention of the giver, as it is the case of the previous monorhematic sentence "left", we assign the label [acknowledge] to this very last move.
2) if the follower fails in guessing, the giver must correct the target of the textual communication. For this class of cases too we created a new label: [correct].
For other problems of labelling in Italian dialogue, see the Proceedings of API Conference, Naples 13-15 / 2 / 2003 (to be issued on web-site).

4.3 Dialogue structure

From the point of view of the structure of dialogue, the following two features seem to dominate *involvement* dialogues:

- the high number of unnecessary utterances. These, in general, are difficult to relate to the main topic of the dialogue as they are based on some sort of metaphor *over* such topic.

- widely spread negotiation subdialogues, even in those sections, such as the end of the Map Task where there is nothing to negotiate.

4.4 Length of dialogues

The general consequence of the above features is the indefinite temporal lengthening of dialogues. This parameter correlates very well with *involvement*: the more involvement is in a dialogue, the more time it takes. So, we'll not be surprised to discover that the shortest Italian dialogues are Vercelli's, while the longest are in Naples. The average length in Vercelli dialogues is 5min25sec, and it could have been even less if it was not for only one dialogue of 10min15sec. Half dialogues of Vercelli sub-corpus last less than 3min30sec. They are really very short dialogues, even shorter than HCRC's, where the average length in time is 7min45sec. Bari and Pisa dialogues have approximately the same length (Bari dialogues: approx. 11min vs. Pisa dialogues: 10min49sec), so, in this situation, no inference about involvement degree can be drawn. As we expected, Neapolitan dialogues are the longest: the average length is 15min50sec. The longest Italian dialogue (D02_o), lasting 26min., is, of course, Neapolitan.

5 Conclusions

There may be different sociological/anthropological explanations of the distribution of these features. One could hypothesise that the more industrialised a society is (the more it is a *Gesellschaft*, in the terms of sociologist Tönnies), the more we find a *high-considerateness* conversational style of dialogue. On the other hand, the more a society preserves archaic features and it is structured as a collectivity (i.e. *Gemeinschaft*, for Tönnies), the more we meet a *high-involvement* conversational style. Italian data, at least, are in this direction, considering that Vercelli is in the industrialized Northern Italy and Bari, in Puglia, is an industrial enclave in Southern Italy, while Pisa was an independent town since the Middle Ages, a maritime republic and later on a city-state, full of local pride and Naples, due to its history, has had (and it's still having) a very late industrial development.

5.1 Consequencies for computational dialogue models

Whatever the explanation for this distribution, it is a fact that different ways of interaction exist and should, probably, be taken into account in the development of (computational) dialogue models. The above mentioned aspects have the following consequencies:

- there is a large number of utterances whose semantic interpretation is inconsistent with the dialogue domain; even in practical spoken dialogue systems for Italian, there are interaction

during which users introduce a number of personal and private (unnenecessary) accounts in order to reinforce or motivate their queries. A dialogue system aiming at naturality shall at least include the ability of skipping large portions of utterance as not relevant

- the large extension of utterance completions or overlap may cause serious problems to speech recognisers
- a dialogue is more "costly" in terms of time length and "useless" subdialogues. These have no use to the dialogue domain, but have the only impact of "making people feel at ease".

5.2 Future work

In order to clarify the impact of regional features, further research is being carried in the following domains:

- identify different problem-solving strategies and there distribution through the analysed geographic areas
- extend the geographic area to be inspected; all the Mediterranean area is expected to behave in a *high-involvement* style.
- quantify the impact of each of these features on a more extended sample
- verify the actual impact of these features in practical dialogues like those used in transaction systems

References

ALLEN, J.F., BYRON, D.K., DZIKOVSKA, M., FERGUSON, G., GALESCU, L. and STENT, A.(2001): "Toward conversational human-computer interaction", in *AI Magazine*, 22(4):27-37, 2001.

ANDERSON, A., BADER, M., BARD, E., BOYLE, E., DOHERTY, G.M., GARROD, S., ISARD, S., KOWTKO, J., McALLISTER, J., MILLER, J., SOTILLO, C., THOMPSON, H. and WEINERT, R. (1991), "The HCRC Map Task Corpus", in *Language and Speech*, 34 , pp. 351-366.

BERRUTO, G. (1987), *Sociolinguistica dell'Italiano contemporaneo*, NIS-La Nuova Italia, Firenze, 1987.

BROWN, G., ANDERSON, A. , YULE, G. and SHILLOCK, R. (1984), *Teaching talk*, Cambridge: Cambridge University Press.

BROWN, P.; LEVINSON, S.C.(1987) *Politeness. Some universals in language usage*. Cambridge, Cambridge University Press.

DUNCAN, S. (1974), "On the structure of speaker-auditor interaction during speaking turns", in *Language in society* 3 , pp. 161-180.

ERVINTRIPP, S. (1987), "Cross-cultural and developmental sources of pragmatic generalization", in J.Verschueren, M. Bertuccelli-Papi (eds.) *The pragmatic perspective*, Amsetrdam/Philadeplphia: Benjamins, pp. 47-60.

GOFFMAN, E. (1955), "On face work: an analysis of ritual elements in social interaction", in *Psychiatry*18, pp.213-231.

GRICE, H.P. *Logic and conversation*. William James Lectures, Harvard University

JEFFRSON, G. (1973), "A case of precision timing in ordinary conversation: overlapped tag-positioned address terms in closing sequences", in *Semiotica* 9.

LAKOFF, R.(1979) "Stylistic strategies within a grammar of style", in J.Orasanu, M.Slater, L.Adler (eds.) *Language, Sex and Gender*. Annals of the New York Academy of Science, number 327, pp. 53-78.

LAKOFF, G., JOHNSON, M. (1980), *Metaphors we live by*, The University of Chicago, Ill. University of Chicago Pess.

TANNEN, D.(1984), *Conversational Style: Analyzing Talk among Friends*. Ablex, Norwood NJ.

TANNEN, D (1989). *Talking voices. Repetition, dialogue and imagery in conversational discourse*. Cambridge, Cambridge University Press.

TöNNIES, F. (1963), *Comunità e società*. Milano, Comunità.

WIENER, M., MEHRABIAN, A. (1968), *A Language within a Language*, New York, Appleton.

Presuppositions and commitment stores

Francis Corblin
Université Paris-Sorbonne
& Institut Jean-Nicod (CNRS)
Francis.Corblin@paris4.sorbonne.fr

Abstract

This paper revisits the classical question of presupposition projection in a dynamic approach to dialog using Hamblin's "commitment stores" (Hamblin, 1970). It is based on a view of presuppositions as selectionnal restrictions, conceived as constraints on the definition domain of functions. The specific update of commitment stores achieved by the mere use of lexical item having restricted definition domains (presuppositions) is captured by means of layered commitment stores distinguishing *background commitments* and (classical) commitments arising from what is said. The compatibility of this dialogic approach with the dynamic theories of Heim (1983) and van der Sandt (1992) is discussed in the course of the paper.

1 Commitment stores

In Hamblin's (1970) style, a dialog updates the commitment stores (CS) of the speaker and hearer, which means that linguistic expressions should be defined in terms of instructions for updating current commitment stores. CS are represented by Hamblin as lists of propositions, keeping track of what the speaker and hearer got committed to in virtue of what they have said (and in virtue of what they have not raised objections against) in the current dialog.

CS should be distinguished clearly from what will be called here the knowledge-base (KB) of each agent involved in the dialog. This data-base is the whole set of propositions that the agent takes for granted. Of course, there is a link between KB and CS, but this relation is not simple and we do not want to take an *a priori* position on this link. For instance, to end a dialog with p in one CS does not imply that the agent belief is that p is true, and even taking the initiative to update one's own CS with p does not imply that the agent belief is that p is true. Moreover, as Hamblin (1970) himself puts it, assuming that a KB does not contain p and $\neg p$ is a standard assumption, but to assume that a commitment store never contains both would be probably too strong. [1]

CS has the general structure of a blackboard. Both participants can "see" what both CSs show at any time; only one participant's CS at once can be updated; speech acts and linguistic expressions are defined as rules for updating CS (adding/retracting commitments).

The project is to use DRT as the language for representing CS, which means providing a first-order representation of the propositional content enriched with dynamic information about context change potential.[2]

Moreover, the DRS material will be split in two parts: the foreground part and the background part. This difference, which can be intuitively seen as the use of two different colors, for writing the conditions of a DRS, does not trigger any change in the classical mechanism which computes the binding relations and the truthful embeddings of CSs in a Model : all DRS conditions are just DRS conditions and only DRS conditions. The difference between foreground/ background will only play a role in the rules for updating CS. This kind of strategy contrasting

[1] An important point, not considered in this paper, is related to the *effects of when it's said* (Hamblin, 1970 b) : a CS is a list, hence the order in which propositions are entered in a CS is relevant. A given CS can contain, for instance p, and then $\neg p$. A real discussion of such cases being far beyond the scope of this paper; let us just consider that in such cases , if a commitment contradicts a previous one, the previous one is just retracted.

[2] The view that DRS should be conceived as CS is explicitly introduced by Geurts (1995 : 29) : "I assume, therefore, that a DRS is is a partial picture of a commitment slate".

different kinds of semantic information of the same type has been adopted many times in the literature and it seems to me that layered DRT developed in Nijmegen by R. van der Sandt and his colleagues is based on the same idea.[3]

The conception of CS used in this paper will be kept as simple as possible by means of some idealizations. We consider only *light side* commitments, in the terminology of Walton & Krabbe, (1995, pp. 134), which means that we stick to Hamblin idea that commitments are public. We do not consider either the difference between assertion (an explicit claim of A), and *concession* (what B claims without receiving any objection from A). As in Hamblin's original examples, we assume that what B claims without receiving any objection from A becomes part of A's CS. This latter simplification will allow us, to use in most cases identical CS for A and B. Some very schematic working rules can be used for the sake of illustration.

If A uses a declarative sentence *p*, *p'*, the representation of this sentence is entered in A's CS.

If B does not object, *p'* becomes a part of B's CS.

I will not consider in the paper any "syntactic rule" on dialog moves, i.e. rules which try to predict what a given expression, or a given CS configuration, prohibits or imposes for the next moves. I will only consider "semantic rules" i.e. rules defining what a given expression triggers in terms of CS updates for the user of this expression, and I will consider only declarative sentences (for the interpretation of questions in dialog, see Ginzburg (1995).

2 Presuppositions as commitments

Although there are many debates about the nature and behavior of presuppositions, most theories would accept at least the following claims: 1– presuppositions are propositions; 2– they are lexically triggered.

From this, we infer that presuppositions should be entered in CS. The interpretation of 2 is not straightforward. Let us assume that a presupposition of a lexical item *I* is triggered by the use of *I*. This means that any use of *I*, in any logical context (negation, modality, etc.), and for any illocutionary force (assertion, question, etc.) triggers the ascription of *I*'s presupposition to the CS of the user of *I*. Although this view is by no means original it faces two problems:

1. It is too strong: in some contexts, *I* is used, but its potential presuppositions are not projected. Some theories see these cases as "cancellation" cases.

2. It is unmotivated: why should the mere use of a lexical item, irrespective of the logical (e.g. negation, conditional, etc.) context in which it is embedded, commit its user to any propositional content?

I will first try to provide an answer to question (2). The general idea is to consider presuppositions as selectional restrictions, a linguistic phenomenon conceived as analogous to domain restrictions on functions.

2.1 Selectional restrictions

Selectional restriction is a notion introduced by Chomsky (1965), and exemplified by (1) :

(1) I drink something.

It can be described, tentatively, as follows : the lexical expression *drink x* cannot be used for updating any CS, unless the variable *x* is restricted to a domain of individuals such that they satisfy the condition *liquid (x)*.Anyone trying to bypass this rule would not be using the language properly. Consider for instance the potential utterance (2) :

(2) Were you drinking an apple?

It seems to me that the most widespread judgment on (2) is that it violates the rules of English. If someone tries nevertheless, no update stemming from the sentence itself is performed, but the next moves in dialog will try to elucidate this problem about the language supposed to be the common language of the dialog.

Note that the condition *liquid (x)* has any properties of what is called a presupposition : it is propositional, lexically triggered, and stemming from the mere use of the verb *drink* whatever its logical environment is. It gives rise to the same phenomena, and can be captured by the same test than classical examples of presuppositions. (3) will quickly illustrate this:

(3) A : I was not drinking my X.

[3] Corblin (1991) argues that presuppositions should be treated as distinguished conditions of DRSs, which seems to be the same idea.

After (3), *liquid (X)* becomes a part of A's CS. The hearer B can test this commitment against her KB and can find three things in it : 1. *liquid (X);* 2. ¬ *liquid (X);* 3. no information about *liquid (X)*.

Case 1 is the unmarked case: *X* is the known common noun or a known trademark for a liquid.

Case 2 is a case where B must ask clarification about the language used by A : either *drink* or *X* is used by A in a non-standard way. It can trigger for instance moves like (4) which are typical of reactions to presupposition failure:

(4) B : But one cannot DRINK an X, X is not a liquid

Case 3 can be a case where B will learn that *X* is a liquid; at least, she will learn that A uses it as something that A takes for being in A and B's KB.

There is a correspondence between these three cases and the concepts found in the literature which must be clarified. Case 1 reminds van der Sandt (1992) notion of *binding*: the content *liquid (X)* is *found* in some representation of the context. Similarly it is close to the notion of *satisfaction* used in satisfaction theories. Case 2 corresponds to what is called *presupposition failure* in most theories. Case 3 evokes strongly what is called *accommodation*, in most theories using the concept.

2.2 Presuppositions

Having argued that this typical example of selectional restriction can be described in the same terms than presuppositions, we are lead to conclude that presupposition might be nothing else than selection. Selection itself is conceived on the model of domain restrictions on the definition of functions. It is based on the fact that some lexical items contain constraints on their definition domain: the function associated to *drink (x)*, is defined for any *x* satisfying the function *liquid(x)*, and undefined otherwise. Using lexical items with no consideration of their definition domain is just not speaking the language. Conversely, using a lexical item commits to the satisfaction of its domain restrictions.

I will illustrate on some classical examples this view of presupposition as selection:

A. Factive verbs.

(5) I regret that P.

Factive verbs presuppose the truth of their complement *p*. What we said before is that *drink (x)* is defined if *liquid (x)* is true, and undefined otherwise. Similarly *regret (p)* is defined is *p* is true, and undefined otherwise. We will thus analyze *regret (p)* as triggering the background commitment (BC) *p*. One can note that the form of the presupposition is slightly different. For *drink* the BC is a condition on the individual members of the domain (*liquid)*; for *regret* it is just the truth of the propositional argument which is presupposed. This difference might be relevant for delimitating sub-classes, but the similarity is important enough for claiming that we have examples of the same phenomenon.

2. *Manage to*.

(6) John managed to P.

In this case, *p* is not presupposed, but a condition equivalent to: *doing p was difficult for John* is associated to *manage*. The situation is very close to the case of *drink*.

3. Definite descriptions.

(7) The king of France is bald.

It is easy to see *the*, in *the X* as imposing the BC that there is one and only one X.

2.3 Background commitments

We use for the representation of dialog CS designed as classical DRSs distinguishing explicitly BC from the other conditions. We will use in this paper italics for BC.

A simplified CS for (7) is (8) :

(8)[*x [king- of France (x),* bald (x)]]

Although binding is allowed between all commitments, BC are distinguished from foreground commitments by a set of properties which supports the decision to set them apart. In a sense, BC are just the price one has to pay for *using* lexical items, whatever one wants to convey about a model by using these items. A dialog, thus, is not designed for making public the BC of the lexical items she uses, although just by using such items the speaker is committed to them. BC are made public, although the dialog is not set up for making *them* public.

The most salient property associated to this background status is that if *p* is introduced as a

BC, the probability to isolate *p* as an antecedent for a propositional anaphor is very low. Depending on one's own theory of propositional anaphora, one might suggest different reasons for this. It can be seen as a mechanical consequence of the fact that BC are propositions which are most often not encoded under the linguistic form of a clause. If your theory requires that the antecedent of propositional anaphors be clauses, you explain that presuppositions are not accessible to them. If you do not make this assumption and let anaphors work on the semantic representation of the discourse or dialog, you will have to block by an additional stipulation the accessibility of BC. Some examples will illustrate the difficulty to isolate a BC as the antecedent of a propositonal anaphor. In (9), although the sentence is represented as a conjunction (*It was difficult for John to fail* & John failed), it is impossible to interpret a propositional anaphor as taking the sole presupposition as its antecedent :

(9) A: X managed to fail.
B: I cannot believe it.

Most speakers interpret B's sentence as : I cannot believe X failed. [4] This is a case in which the presupposition does not show up in the sentence as a clause. The same is true for *The King of France* case, and the presupposition is not accessible. The most interesting cases for the discussion are factive verbs, since they exhibit the presupposition under the form of a subordinate clause. Consider the contrast (10)/(11) :

(10) X regrets that Y left.
(11) X says that Y left.

If it can be shown that the accessibility of the subordinate clause is significantly lower in (10) than in (11), it would be an argument showing that as such, the BC status of an overtly expressed clause reduces its anaphoric accessibility. I will not pursue the discussion on this for space consideration.

A consequence is that a BC is introduced without being isolated as an accessible topic for the ongoing dialog. A BC, thus, cannot be a *QUD* (questions under discussion), see Ginzburg (1995). In other words, encoding as a BC a given information gives no chance to know more about it in the dialog because it is not an accessible topic. The other side of the coin is that BC encodes normally shared information, that is to say information that the dialog is not designed to manipulate.

Some other distinctive properties can be associated to BC.

- *BC must be contested immediately*. Beaver (2001) insists on this distinctive property of presuppositions, and the present proposal provides a nice context for discussing this point. In dialog, the participants can hold contradictory theses, and a single participant can change her mind in the course of the discussion. If I do not object immediately to your assertion *p*, which becomes part of my commitment, *p* is by no means protected from latter attacks. But it is a common observation that to come back after a while on a *BC p* for adopting explicitly a commitment ¬*p*, is judged unfair if the initiative is taken by the opponent; if taken by the proponent of the BC, it is even worse. I will try to suggest some justification of this based on the very notion of BC.

– *the rejection of a BC by the opponent comes with the cancellation of the utterance containing its trigger*. "Cancellation" means that any effect on the opponent CS update is cancelled. Many BC being existence commitments one might think that this is not a property of BC, but a property of existence commitments. But in the case of selectional restrictions (see the case of *drink)*, the BC is not a BC of existence, and this is also true for some classical examples of presuppositions (e.g. *manage)*. In those cases, my intuition is that, if one rejects the BC, one suspends any update, and waits (or asks) for an elucidation. Suppose for example someone says to you that *John managed to pass the exam*, and that for you, John is the best student of the class. Would you just reject explicitly the BC, keeping the information that *John passed*; or would you suspend any update, thinking for instance that the proponent may be actually speaking of someone else that John (satisfying the BC) and convey no information about *John?* My impression is that when one is in doubt about the capacity of the proponent to

[4] Note that the reading *I cannot believe that it was difficult for him and that he failed* is not accessible either. If it were, one could argue that what happens for the succession accommodated presupposition/assertion is not that different from what happens for two conjoined assertions:
A. It was difficult for him to fail and he failed.
B. I cannot believe it.
Most speakers can interpret B's assertion as : I cannot believe that it was difficult and that he failed, even if they express a preference for taking as antecedent the last expressed proposition.
This shows that bakground commitments cannot be treated as would be a previous assertion.

choose the right word for speaking of a Model, this doubt extends to any part of the utterance.

- for rejecting a BC there are specific linguistic devices which are also used for language mistakes: metalinguistic negation and stress on the "wrong" word. Typical cases are illustrated in the following examples :

(12)But one cannot DRINK an X, one can only EAT an X.

(13)John did not MANAGE to succeed; he succeeded because the exam was very easy for him.

All these properties are coherent with the view of BC advocated in this paper. The rejection of BC, if any, must be immediate because it is unexpected, it would cancel the whole utterance, and it would cast doubts about the language used in the current dialog.

3 The projection problem

The general idea is that any user of a lexical BC trigger is committed to its BC. We expect, then, in general, the BC of a complex sentence to be the conjunctions of its BC. The only resources we have for deriving the so-called "cancellation" cases, are linked to the notion of "user of a lexical item", and to the notion of context of satisfaction of a BC.

3.1 Plugs

The classical theory of presupposition projection set up by Karttunen (1973, p. 178) distinguishes three kinds of contexts: *plugs*, *holes*, and *filters*. Dynamic theories of presupposition projection (the satisfaction theory and the binding-accommodation, see Geurts 1995)) are mainly concerned with filters, and what they have to say about plugs and holes is not very clear. The present proposal, in contrast, provides a straightforward explanation for the existence of plugs.

Plugs are defined by Karttunen as contexts which blocks all the presuppositions of the complement sentence. Typical examples are : *say, mention, tell, ask*.[5] They are precisely contexts that allow taking their lexical trigger as *used by another agent* (reported speech), not by the speaker. In other words, in the context of these verbs, the speaker does not necessarily take herself the responsibility of using these lexical items, but might only be reporting the words *another* speaker used *(John says that the king of France is bald*). Our explanation is straightforward: who uses a lexical item is committed to its BC, and not who *reports* the use of a lexical item by someone else, provided that she makes explicit that she would not use herself this word. The later condition is crucial although we will not go into details here. The fact that this condition is very difficult to satisfy (because after all, a speaker makes use of any lexical item she utters (except in direct reported speech) explains that plugs are not strict barriers against projection of the BC in the speaker CS.

The prediction of the theory is that any other context should be a Karttunen's *hole* i.e. a context in which any BC is made part of the speaker's CS. [6]

3.2 Filters

Filters have the following relevant features: in some constructions (e.g. the antecedent of a conditional), the presence of an expression entailing a presupposition of the consequent, prevents the presupposition to be ascribed to the CS of the speaker. A typical case is (14) :

(14) If France is a monarchy, the king of France is bald.

Gazdar (1979), Heim (1983) and van der Sandt (1992) are well known proposals devoted to explain the existence of filters (see Beaver 2001).

Most properties of context change potential approaches (Heim 1983) transpose nicely in our framework without any ad hoc stipulation: if a trigger is used in the consequent of a conditional, the user is committed to the presupposition in her top-level CS updated by the antecedent. This is a direct consequence of what is a BC : to use a trigger in the consequent means that your top-level CS updated by the antecedent is a context in which the trigger is licensed. If the antecedent update entails the BC, as in (14) this requirement is satisfied and no update is necessary. If not, the speaker is committed to the presupposition satis-

[5] I follow here the presentation of Beaver (2001, p. 54).

[6] We have not enough space here for discussing attitudes verbs (Heim 1992, Geurts 1995). What have been said about verbs reporting speech acts would be a worth trying starting point.

faction in her (toplevel) CS. A comparison to Gazdar (1979) proposal will lead to make this formulation more precise.

Gazdar's treatment of cases like (14) relies crucially on Hintikka's logic of belief, and on the notion of *clausal implicature*:

(15) Clausal implicature of $P \rightarrow Q$: $\neg K(P), \neg K(\neg P), \neg K(Q), \neg K(\neg Q)$. (the speaker has no belief about P)

Gazdar predicts that the presupposition *p* of the consequent is cancelled because if projected, it would contradict the clausal implicature triggered by the use of *if P* (if *P* entails *p*).

But examples like (16) show that this cannot be the right explanation :

(16) Mary is sleeping. If Mary is sleeping, John knows that she is sleeping. Thus John knows that she is sleeping.

The speaker is committed to *P* by the first sentence, but *if P…* is used then, which shows that the use of *if P* is compatible with the commitment to *P*. The key part of Gazdar explanation (i.e. the clausal implicature) seems, in other words, to meet empirical objections.

But such cases force us to strengthen our own proposal. Once admitted that *P* and *if P* are compatible in a CS, we need an additional principle for ensuring that the presupposition of the consequent is not projected at the top-level. I propose that the relevant principle is the Minimal Commitment Principle (MCP) :

(17) MCP : The update of CS by BC is minimal. A potential BC *P* ends up as a BC iff the use of its trigger is not licensed otherwise.

MCP ensures that in (14), since the antecedent of the conditional entails a BC of the consequent, no update of the CS is done.

Van der Sandt's anaphoric treatment of presuppositions would consider (14) as a case in which the presupposition tries to find its antecedent in the *if* clause. In (14) anaphora to the *if* clause fails, and looks for an antecedent at the top-level of the DRS. If (14) is the first sentence of a discourse, no antecedent can be found. Accommodation is then tried at this level. Accommodation is constrained by *contextual acceptability*. I will discuss briefly van der Sandt constraints[7], while running through the example, trying to establish whether they are compatible with the present approach:

(i) informativeness: prevents to accommodate a presupposition which is entailed by the DRS.

MCP seems to have roughly the same effect: if you are already committed to P, any BC entailed by P will trigger no update. Note that in the CS approach, this is a constraint on BC, not on commitments. Nothing prevents the reiteration of an assertion in a dialog, as everyday life shows.

(ii) consistency: prevents to accommodate $\neg P$ if the DRS entails *P*. Although the CS framework allows explicit contradiction in a dialog, the very notion of BC suggests that the retraction of a previous commitment by means of a BC is not a standard strategy.

(iii) no accommodation can be such that some subordinate DRS is either entailed or contradicted by a superordinate DRS.

This constraint seems to be based on the same intuition than Gazdar's clausal implicature (see above). Although van der Sandt gives it explicitly as a constraint on accommodation, he overtly makes it an application of a general principle on discourse coherence which would exclude cases like (16).[8] The MCP, in contrast, will derive (14) without predicting anything about (16).

Let us now consider the treatment of (14) by (i)-(iii). Top-level accommodation is blocked by (iii): one cannot accommodate something that entails a subordinate DRS. *There is a king of France* entails *France is a monarchy*. We try then to accommodate in the antecedent, and this is allowed if *France is a Monarchy* does not entail that *there is a king of France*. It the entailment holds, the informativeness principle (i) is violated.

At first glance, van der Sandt's system makes the same empirical prediction that the MCP for (14), but I see at least one potential problem with this strategy. As I understand the way of treating (14) with (i)-(iii), the theory assumes a strong logical difference between the two propositions. In (14) we need to accept:

There is a king of F. |= F. is a monarchy

---

[7] Van der Sandt (1992, p. 367).

[8] There is no doubt that this is van der Sandt's interpretation of the constraint as shows his illustrative example (63a):*John has a dog. If he has a dog, he has a cat.*
He gives this piece of discourse as unacceptable, although it is for me correct.

If not, accommodation on top-level is allowed. But we need to accept as well:

F is a monarchy $|\neq$ *There is a king of F*.

If we do not, accommodation in the antecedent is not informative.

But this distinction is not plausible if one considers the following couple of examples:

(18) If John has a wife, his marriage is very recent.

(19)If John is married, his wife is French.

The theory, as I understand it, would have to assume that in (18): *to be married* $|=$ *having a wife*, and *to have a wife* $|\neq$ *to be married*. But in (19), the theory would have to assume that *to have a wife* $|=$ *to be married*, and *to be married* $|\neq$ *to have a wife*. It is hard to believe that the processing of these two sentences might be grounded on so strict, subtle and incompatible logical relations between the predicates *to be married* and *to have a wife*.

Our proposal escapes this problem because we do not manipulate logical relations between propositions but constraints on definition domains for the use of lexical items. "To be licensed", in the MCP (17) means than if a proposition is true, it is legitimate to make use of a given lexical item for it. But to the truth of the proposition, we do not associate strict logical inferences involving the lexical item. It is be more in the spirit of our view to see licensing as based on default generic implications like: if X is a monarchy, normally, there is a King(Queen) of X; if X is the King of X, normally, X is a monarchy. We all know real cases in which there are monarchies without sovereign and sovereigns without monarchies.

What we conclude from this discussion is that the MCP derives correctly the "filtering effect" of quantified structures without assuming the "contextual acceptability" constraints needed in van der Sandt's approach, which are not without problems.

3.3 Commitments and Knowledge-Bases

(20) exemplifies a classical problem for presupposition projection theories:

(20) Any woman cherishes her child.

The problem is that this sentence is admissible in a dialog, although none of its users would accept to be committed to *Any woman has a child*. A solution is to let such sentences be interpreted as:

(21) Any woman (having a child) cherishes her child.

But this solution is too strong because it would legitimate, contrary to facts, *any* "intermediate accommodation" like in (22):

(22) Any woman likes her Ferrari.

In the present framework, like in Heim (1983), it is predicted that (20) triggers the BC that any woman has a child. Van der Sandt (1992, p.364) predicts for (20) the interpretation (21) on the basis of a structural constraint on binding/accommodation. The pronoun *her* must be bound by its antecedent *any woman*, and consequently, the accommodation of an antecedent for *the child of x* can only be done in the scope of *any woman*. But the contextual acceptability constraints (i)-(iii), see §3.2, cannot rule out (22).

In a nutshell, although a strictly universal commitment is not projected, a purely accidental and unexpected property like in (22) makes the sentence odd. [9]

We need, it seems, a treatment which commits the user of (20) at least to the BC that "stereotypical women have a child". This would derive the acceptability contrast (20)/(22). Let us stick to the prediction that the basic interpretation of such structures projects a *universal* BC. This is what happens if the discourse is not generic, and is about a restricted set of individuals.[10]

(23) Every boy took his Ferrari (bike)and left.

(23) commits to the BC that every boy (within a particular context-set) had a Ferrari (bike).[11]

Now if the sentence is generic, the KB of the participants knows whether this universal BC is true or not: in the sentence (22), we now that it is not. Let us take this contradiction between the universal commitment mechanically projected by the structure and a proposition of the (shared) KB, a *necessary* condition for triggering the suspension of the universal BC projection. The *sufficient* condition is that the correspondent stereotypical BC be the case, which is true for

[9] See Beaver (2001, §5.6) for a discussion pointing to the relevance of genericity in such examples.

[10] See Beaver (2001).

[11] I am not sure how van der Sandt's approach would accommodate this generic/specific contrast.

bike (in some societies), not for *Ferrari*. The relevant configuration for (20) is thus:

(24) CS potential update:
any W has a C
KB:
¬ (any W has a C)
GEN (W have C)

For (22), although the necessary condition is satisfied, the sufficient condition is not, and the sentence is odd. If both conditions are satisfied, the sentence is interpreted as projecting a weaker BC (stereotypical, not universal). We can explain in this approach why sentences like (20) give the impression that the proponent speaks as if the property were universal (this is the basic interpretation) and why it commits to stereotypes (a typical woman has a child). Moreover we can understand why such sentences are perceived as shortcuts and why these shortcuts although they commit to BC stereotypes are nevertheless so often used. For this, let us compare (20) to the fully explicit (BC-less) version (25):

(25) Every woman having a child cherishes her child.

This sentence is perfectly acceptable and "clean" (deprived of any BC about women and child). It has nevertheless some features which might lead speakers to prefer the "shortened" version. For most speakers the sentence has a redundancy flavor, and the more the property is stereotypical, the more it is perceptible. The reason might be that to restrict X with a stereotypical property of X creates for ordinary discourse the same kind of redundancy than to restrict X with a property strictly entailed by X. The underlying principle might be that by default, one always speaks of stereotypical cases. Compare: *any triangle having three angles, any man having two hands, any French man having a car,….* By choosing this explicit option, then, the speaker treats stereotypes exactly as any property and does not make use of the shared knowledge of the evoked stereotype.

The commitment framework suggests thus to see the so called "intermediate accom-modation" as a pragmatic weakening of the semantic universal BC, relying heavily on KB and stereotypes.

4 Conclusion

The general idea of this paper is that presuppositions are background commitments arising from the use of lexical items involving restrictions on their definition domain. We have tried to give some arguments showing that this idea is worth trying, an the paper has been devoted mainly to illustrate the basic intuitions of this dialogic approach and to point to some issues on which it seems to do better than current dynamic approaches. The next step will be to test this view of presuppositions in a formalized version of the CS framework in order to develop a more systematic comparison with the predictions of these theories.

References

Beaver. D. 2001. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications.

Chomsky. N. 1965 *Aspects of the Theory of Syntax*, MIT Press. Cambridge.

Corblin. F. 1991) Presupposition and discourse context, *Unpublished paper, CNRS Paris 7*.

Gazdar. G. 1979 *Pragmatics. Implicature, Presupposition, and Logical Form*. Academic Press, New York.

Geurts. B. 1995 *Presupposing*, Thesis, University of Stuttgart.

Ginzburg. J. 1995. Resolving questions. *Linguistics and Philosophy* **18**.

Hamblin. C.L. 1970. *Fallacies*, Methuen, London.

Hamblin. C.L. 1970b. The effect of when it's said. *Theoria*. **3 :249-263.**

Heim. I. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of* ***Semantics*** **9: 183-221.**

Heim. I. 1983. On the projection problem for presuppositions. *Proceeding of the 2nd West Coast Conference on Formal Linguistics, Stanford Linguistic Association*. 114-125.

Karttunen. L. 1973. Presuppositions of compound sentences. *Linguistic Inquiry*. **4: 169-193**.

van der Sandt. R. 1992. Presupposition Projection as Anaphora Resolution. *Journal of Semantics* **9: 333-377.**

Walton, D.and Krabbe. E. 1995. *Commitment in dialogue : basic concepts of interpersonal reasoning*, State University of N.Y. Press, Albany.

Combining the Practical Syllogism and Planning in Dialogue

Günther Görz, Alexander Huber, Bernd Ludwig, Peter Reiss
University of Erlangen-Nuremberg
Computer Science Institute

Abstract

We present a dialogue model which relates discourse relations and intentions by reasoning about pragmatic capabilities of dialogue particpants. For that purpose, planning approaches for discourse and application domain are employed. In this way, a computationally tractable version of the pratical syllogism is devised.

1 Rational Dialogues

Our goal is to build dialogue systems for rational interaction. Users can interact with the system in a given (ideally open) domain by conducting spoken dialogues. In principle, it should be possible to augment them by other forms of multi-modal interaction like gestures or the selection of items from a menu on a screen. Interactions are called "rational" because we want to apply rationality principles (at the knowledge representation level) to optimally select appropriate communicative actions. We assume that the satisfaction of user goals within the thematic framework of a particular application domain is to be achieved with the help of a dialogue system proper in cooperation with a technical application which we also call the "domain problem solver". Such a technical application can be an information or reservation system, a system for controlling certain devices, etc.

For dialogue modelling, we will follow a plan-based approach which has its roots in natural language processing and Artificial Intelligence. It provides the means to conduct task- or goal-oriented dialogues which are focussed on accomplishing concrete tasks as mentioned in the introduction. We claim that only a general planning approach enables cooperative response behaviour (pragmatic adequateness, overanswering) and the ability for negotiation. For the reasoning part, i.e. knowledge representation and inference for the interpretation of dialogue as well as for planning to satisfy user goals in the application domain, we insist on a clear commitment to a computational logic framework, in particular description logics. Of course, humans act incoherently and even inconsistently, and common sense reasoning can only to a certain extent be understood in terms of logic, but we are convinced that a coherent and consistent rational reconstruction is the best we can do about it. Such a constructive perspective has the advantage of enabling us to begin with a well understood framework for knowledge representation and reasoning upon which we can attempt to build rule systems for still idealized, but more realistic patterns of argumentation in specific domains. We believe that there is a potential to succeed in a variety of prevailingly instrumentalized contexts as it is the case with technical applications or in forensic argumentation.

Taking these claims serious, obviously a variety of complex issues must be addressed within the framework of our dialogue system, as, e.g., intention recognition, cooperativeness, grounding, or sharing plans – to name just a few important ones. For that, we refer to other publications of our group, e.g., (Görz et al., 2002; Bücher et al.,

2002; Ludwig et al., 2002). In this paper, we are adressing only one specific problem:

2 Choices in Rational Dialogues

In a series of papers, Asher and Lascarides introduced Segmented Discourse Representation Theory (SDRT) as an extension to Kamp's DRT and used it to model dialogue by combining compositional semantics, discourse structure and information about the participant's intentional states. The reason for their approach is to overcome inferential problems encountered in AI approaches to plan recognition in dialogue systems (Asher and Lascarides, 1997). A discourse is represented in the form of a SDRS (Segmented Discourse Representation Structure) which is a recursive structure of DRSs, i.e. semantic representations of linguistic expressions on the clause level, connected by rhetorical relations. Examples for such relations are explanation, contrast, or continuation. For discourse analysis, the authors propose to construct two different SDRSs, one for each discourse participant, in order to take their different cognitive states – we prefer to talk about epistemic states – into account. As a formal means for reasoning from the epistemic states of participants to what they say and vice versa, Asher and Lascarides introduce a version of Aristotle's Practical Syllogism: it "states that normally, people intend to do things that they believe help them achieve their goals"[1]. Under the rationality assumption for dialogues we propose for the multi-agent framework and the applications we work on within it, we will use a tighter, i.e. monotonic version of the Practical Syllogism by dropping "normally": If (a) participant B wants ψ and believes $\neg\psi$, and (b) B believes he can infer ψ from his knowledge base augmented by ϕ, and moreover ϕ is B's *choice* for achieving ψ, then (c) B intends ϕ.

There are two reasons for tightening up the Practical Syllogism into a monotonic version. The first is a technical, albeit quite important one: As already mentioned, in order to achieve a running system we committed ourselves to a particular computational logic framework, description logics, where the decidability – and furthermore an efficient implementation – of the inference problem is the foremost issue. Non-monotonicity in general would mean to lose decidability. The second reason is that in any specific application situation in fact the choice of an appropriate action is monotonic – either it is available or not. If not, a normality assumption is of little help. This decision can be modelled by a Reiter-style default rule, but it is a priori with respect to the application timepoint of the Practical Syllogism: Either counterevidence to the assumption that the chosen action is applicable exists in the actual situation, or it doesn't. So we have two distinct phases: First we have to check for counterevidence; if it exists, the *choice* cannot be executed and the Practical Syllogism is not applicable. If there is no counterevidence, we perform a monotonic derivation. Technically, the first step is represented in Abdallah's FIL calculus (Abdallah, 1995); it allows to express a certain situation description (model) with alternative model extensions where within each extended model we can infer monotonically. The monotonic version of Practical Syllogism now reads as follows:

$$\begin{array}{ll} \text{(a)} & (\mathcal{W}_B(\psi) \wedge \mathcal{B}_B(\neg\psi) \wedge \\ \text{(b)} & \mathcal{B}_B((\phi \rightarrow \psi) \wedge choice_B(\phi, \psi)) \\ \text{(c)} & \rightarrow \mathcal{I}_B(\phi) \end{array}$$

With the assumptions of Grice's maxims of cooperation and sincerity, the Practical Syllogism plays an essential part in the reasoning behind how responses to questions are interpreted in dialogue.

In Asher's and Lascarides' original version the choice operator is left open – there is just an existence assumption. To fill this gap, we propose a constructive approach: What we will argue for in the following is that we need to recur to planning in the application domain to provide a precise meaning for the *choice* operator in the rule above. The operationalization of the choice operator in terms of planning will exhibit the respective options in the actual search space, which applies to discourse as well as to domain operations, and furthermore provide an effective decision procedure. Dealing with both kinds of operations in a uniform way depends on how they are modelled in our system: in the underlying conceptual hierarchy both are represented formally in the same way.

[1] (Asher and Lascarides, 1997), p. 16 of the preprint version

Without any doubt, plans are an important source about discourse structure, as has been demonstrated by Grosz' and Sidner's studies as well as other AI work on dialogue (cf. also (Brietzmann and Görz, 1982)). We agree with Asher and Lascarides in their criticism that there is no direct way to structure dialogues satisfactorily by a global planning approach, because there isn't a one to one mapping between dialogue and plan structures. Our approach[2] as well as theirs is built upon the conviction that discourse interpretation requires intention recognition, and therefore a semantic-based theory of discourse structure is needed to assess when exploiting plans is appropriate. We have to distinguish between a domain level plan dealing with actions and a discourse level plan dealing with speech acts[3]. What we need is to predict automatically when the intentional structure of the discourse is isomorphic to the commonsense plan, and when it isn't; this requires a formal representation of semantics. Otherwise, because there is no strict isomorphism between both structures in general, wrong predictions about possible antecedents for anaphors in utterances that continue the dialogue may result.

The thesis proposed in this paper is that the key to a solution to this prediction problem lies in grounding the *choice* operator in the semantics of the resp. application domain. With respect to discourse structure, the role of action planning on the domain level is not to provide a global discourse segmentation, but rather to assign a precise meaning in terms of domain semantics to the *choice* operator in each situation where it is applied. The intended (perlocutionary) effect of a speech act constitutes a planning goal in the domain; it is determined further by the options to act in a particular cooperation context, e.g. by conjunctive subgoals. Whether there exists a precedence relation between subgoals in the conjunctive case can immediately be taken from the task plan in which possible dependencies are represented based on domain knowledge, resulting in a temporal order for the execution of subtasks. This matches well with Asher's and Lascarides' observation that a transition to the epistemic level, i.e. from structural relations between actions in commonsense plans to the level of knowledge, cannot provide a general solution, because the order in which facts are known by the dialogue participants usually doesn't matter. Only on the level of domain knowledge it can be decided whether dependencies between subgoals exist and in which order the corresponding tasks have to be executed.

In the following sections we will point out how this claim can be implemented in the context of an example domain that can be formalized with (classical) planning languages – in our case PDDL (Ghallab et al., 1998). PDDL is decidable and therefore computationally tractable, and there are several efficient planners implemented whose output – as discussed below – delivers the pragmatic options which are subject to the *choice* operator.

In our view, discourse segmentation is a consequence from planning actions in a setup where partners may exchange information about common goals. From a perspective complementing Sadek's work (Bretier and Sadek, 1996) on why partners have common goals at all, we elaborate an effective model of dialogues that explains the analysis and generation side of rational dialogues in human-computer interaction. In an application domain, there are several sources for a *choice* to be made by a dialogue participant:

- The derivation of a common goal from a natural language contribution to a dialogue.
- There may be more than one unique plan to achieve the common goal.
- It is possible as well that no plan can be found. In this case, *choice* is between relevant alternative contexts (cf. (Sperber and Wilson, 1995)) that can be computed effectively taking certain decision criteria into account (e.g. the usefulness of a alternative with respect to the initial intention of the speaker).
- During execution of a plan, the dialogue participant may run into trouble when assumptions that are vital for the plan to be carried out completely are violated.

Throughout the remainder of this paper, we try to illustrate our approach with the help of examples

[2] as described e.g. in (Görz et al., 2002; Bücher et al., 2002; Ludwig et al., 2002)

[3] (Asher and Lascarides, 1997) p. 5 of the preprint version

taken from a prototypical domain we have implemented: we use a model train to implement cooperative user interfaces to complex technical systems. In our scenario, a coffee machine is controlled by a PC as well as a robot that can load and unload cups from railway waggons and position it underneath the spout of the coffee machine that fills them with coffee. The cups are transported to destinations specified by the user.

3 Handling Underspecification in User Utterances

Our goal is to use a dialogue system for spoken language to control technical application systems. Let us consider the scenario described above. In very simple cases, of course there is no need for planning at all – the choice of the appropriate action is obvious. E.g., if only one switch exists, and this switch has only a set of predefinded states, the utterance "set the switch to position 1" (where 'position 1' is one of the defined states) has only one sensible corresponding system command.

But usually, the environment is more complex – there exists more than one switch and railtrack, several trains are moving with different speeds, etc. Now, the same command may intend different actions at the application level, depending on (a) the current state the application system is in, (b) possible user preferences, and (c) the fact that other requests may still be in the queue for processing. So, there is a 1:n-mapping between an utterance and possible commands to the technical application system, which is a kind of incomplete knowledge for the dialogue manager.

Because the user is free to let his descriptions for domain operations underspecified, utterances like "faster please" are possible. Here, not only the device is not specified (it could be one of several trains or even the robot), but also the degree of speed. The challenge for the dialogue system is to react in a sensible and cooperative manner. One way to resolve the underspecified items would be to initiate a clarification dialogue. But the system should always be as cooperative as possible, and before further inquiry, user preferences should be used as an additional source of knowledge, as well as the current state of the application environment. Taking these sources of knowledge into account, it is possible to determine a precise command for broad range of utterances[4].

We believe that planning can be used to determine the user's intentions and to choose the system action he probably wanted, taking into account the application state and his preferences.

Every plan has an initial state, a goal state and a set of possible actions. These actions are defined as plan operators, where each operator has its required initial state and one or more effects. When trying to find a plan, the initial states of the operators are matched with a representation of the current state of the application and of the current user with his (general) preferences. The intended application state is encoded in the user's utterance, but depends also on his preferences.

So, in the domain model, plan operators for different application states have to be defined. The set of (initial and goal) states is limited by the capabilities of the technical system that is addressed. The user preferences are encoded in the definitions of the plan operators, too. So it is clear that only underspecifications that are (directly or indirectly) handeled in the definitions of the plan operators can be filled. One simple example for encoding user preferences in a plan operator would be:

```
(:action faster
  :parameters  (?engine1 ?user)
  :precondition (not (maxspeed engine1))
  :effect (and
            (when (likesFast ?user)
              (maxspeed ?engine1))
            (when (likesSlow ?user)
              (mediumspeed ?engine1))))
```

Now, the command "faster please" leads to an application state where the train is running with maximum speed, if the user likes it fast. If the user's preference wrt. speed is "slow", the train will only move with medium speed. In this example, the application status is not regarded. But the operator definition can be expanded. For example, the current speed can be increased by two steps if the user likes it fast and by one step otherwise. The speed cannot be increased, if the train already runs at maximum speed.

[4] But there will still be cases where a set of actions of the technical system corresponds to one particular utterance. In those cases, it depends on the confi guration of the dialogue system, wheter a clarifi cation dialogue is initiated or e.g. a randomly choosed action is performed.

4 How is the Discourse Related to the Real World Situation?

The type of discourse situations we are considering here in a multi-agent framework is characterized by the following prominent factors: First, dialogue participants behave according to rational principles of conversation and action. Second, dialogues follow a certain rationale which in turn is determined by the intention to elaborate and execute joint plans (cf. (Chu-Carroll and Carberry, 1995; Chu-Carroll and Carberry, 1996; Carberry and Lambert, 1999)). As a consequence, dialogues are considered to be a means of exchanging information and requiring other dialogue participants to execute certain tasks defined in the application domain.

In the light of these guidelines of dialogue analysis (and generation), choices in a dialogue are determined by options while planning and executing plans for joint tasks. Furthermore, these options are constrained by the capabilities of a dialogue participant to act in his or her environment.

Options can appear on two levels in rationale dialogues: first, when reasoning about the effects of a speech act with respect to a model of interaction, and, second, when reasoning about the content of a speech act. Reasoning about content has many aspects: syntax, semantics and pragmatics not the only but the most prominent ones. The focus of this paper is on pragmatics and its effects on reasoning about the state of interaction.

(Carberry, 1990; Lambert, 1993; Lesh et al., 1999), among others, have pointed out that reasoning about plans is a key to understand dialogues; nevertheless, their work has – to the best knowledge of the authors – never been integrated with the work on planning problem solvers in AI in order to achieve an implementation of their discourse models that can reason efficiently and can be configured in a simple fashion to new applications. This paper tries to gap the brigde between linguistic theory and theoretical and practical work on planning by exploiting the expressivity of PDDL for the definition of application (and discourse) domains. In order to make planning applicable to dialogue processing, the following issues have to be addressed:

Planning must be bidirectional. The problem here is that PDDL operators must be applied for plan recognition as well as for plan generation. As we model collaboration in the sense of (Chu-Carroll and Carberry, 1996), application domain operators must be defined for each perspective in the modelled collaboration. In the applications we consider, normally user and system play complementary role: the user wants the system to perform some task and the systems tries to find and execute a plan that fulfils the task. As the task is to be defined in terms of a planning goal, natural language processing must construct it from utterances. To do that, we determine whether the semantic representation of the utterance talks about objects, states or actions. In the case of states and objects, a plan has to be computed if the information given in the utterance is not entailed in the current application situation; in the case of actions, we have to compute their possible effects by applying plan operators in a forward fashion. Of course, in order to avoid infinite recursion this process has to stop after a finite (small) number of iterations and therefore limits the system's capabilities to foresee the consequences implied by the user utterance. By planning in a forward direction, the system tries to get an imagination of what the user wants to happen. Ambiguities arise when the applied operators have conditional effects. In such a case, a decision procedure has to be applied that tries to get a good guess of what continuation would have the most positive and less risky effects on the user's intentions.

Maxims of conversation in the sense of (Grice, 1969) control planning. The need for a decision procedure is a consequence of the system to follow these general principles that of conversation. In our current implementation, there is no reasoning about these principles; therefore, the decision procedure always tries to fulfill user requests as fast as possible. In the worst case ambiguities cannot be resolved, and clarification is requested from the user. Otherwise, the completion of the task would have to be cancelled.

Planning in dialogues must be interactive. This is also due to the fact that planning is always interleaved with plan execution and there is no way to guarantee that each action in a plan can be ex-

ecuted as expected. Differences between planned and observed states motivate the need for replanning in order to fulfill obligations to satisfy user requests. In this way, handling conflicts can be reduced to the problem of controlling planning that was discussed above.

5 Examples for Computing *Choice*

The behaviour of our system, as it has been described in general terms up to now, is discussed in several examples showing when options come up and how they are handled.

In our example domain, for the sake of an intuitive example, we consider different reactions of the hearer to the speaker requesting *"Please bring me a cup of coffee!"*

In the first case, the setup is as follows: The robot – in station C – has one cup stored, an engine is positioned in station B, a waggon at a siding, whereas the coffee machine is in station C. Now, the speaker's request has to be translated into a planning goal. Obviously, the utterance is underspecified with respect to the destination. Here, two levels of *choice* have to be considered. On the content level, there are several options for destinations (all stations in our case); on the level of controlling the execution of a joint plan, there is a *choice* between deciding autonomously for an option, or to clarify this issue in interaction with the other dialogue participant:

- *You are in station A. I will bring the coffee to station A.*
- *Where do you want the coffee to be delivered? Is station A ok?*

As one can observe immediately, there are numerous options for speech acts and text generation when the need for clarification is verbalized.

```
(:init (at engine1 stationB)
       (at waggon1 siding)
       (cup-state cup1 empty)
       (stored-on cup1 stack1)
       (coffee-machine-state cm1 off)
       (at cm1 stationC)
       (at cup1 stationC))
(:goal (at cup1 stationA))
```

The presentation of the computed options are intended to show that the major issue to be addressed is underspecification, not nonmonotonicity. As far as discourse planning is concerned, our approach is to find (disjunctive) alternatives on the basis of observed facts instead of finding contradictions to default assumptions. This observation leads to our claim that a monotonic practical syllogism is sufficient.

In the example, to resolve the underspecification, a decision is made in favour of station A as the destination. As the translation of the request shows, for the interpretation of an utterance contextual information is taken into account as well as the content of the utterance. We have configured the IPP Planner (Koehler et al., 1997) with plan operators for the functionality of the model train domain. The domain plans shown in this paper are always computed by the IPP Planner. IPP finds the following solution for the request:

```
0: switch-on cm1
   put-below-spout cup1 cm1 stack1
   compute-route stationB siding engine1
1: go engine1 stationB siding
   fill-cup cm1 cup1
2: connect waggon1 engine1 siding
   compute-route siding stationC engine1
3: go engine1 siding stationC
4: put-on-waggon engine1 waggon1 cup1
                 CM1 stationC
   compute-route stationC stationA
                 engine1
5: go engine1 stationC stationA
```

In the next setup, there are two cups: somebody put a cup (*cup2*) underneath the spout and prevents the plan from being executed[5] These constraints are added to the initial situation:

```
(:init (at cm1 stationC)
       (at cup1 stationC)
       (underneath-spout cup2 cm1)
       (at cup2 stationC)
       (cup-state cup2 full))
```

Observe that additional constraints in the application domain influence the course of the dialogue. Starting in the same way as above, now a conflict of the plan with external events has to be handled. Again, *choice* is between options on the application as well as on the control level:

[5] In general, the case of implicit conflicts must be handled: A user may give a precise command that could be satisfied directly by the system. But its execution may conflict with a former wish or preference (e.g. delivering the coffee with train 1 may block the transport way connection for train 2). So, an execution of a wish may or may not have side effects that may be desired, irrelevant, or should be avoided.

- *There is a cup underneath the spout. Your request is cancelled.*
- *There is a cup underneath the spout. It's currently being filled. Do you want this cup?*
- *There is a cup underneath the spout. It will be removed immediately. Is it ok for you to wait a few seconds or do you prefer to cancel your request?*

How are conflicts related to the practical syllogism? Again, we claim that they give no argument for non-monotonicity. In the sense of Abdallah's partial logic (see (Abdallah, 1995)), in a logical reconstruction of what was expected to hold this situation and what was observed, of course a logical contradiction is entailed. But only up to the point when – as it comes out now – the wrong option was choosen, or – in Abdallah's words – our "justification" knowledge was not correct. This means, we just have to leave the wrong path from history to future and follow the right one to remain in a "monotonic world".

In the last setup, we have to handle the following situation: There are two cups in station C, one of them underneath the spout and filled, the other one on the robot's stack. In contrast to the last example however, the user's request for a cup of coffee was interpreted to be underspecified as far as the selection of a cup is concerned. As a consequence, the goal contains a disjunction of all the known cups as possible candidates for satisfying the user request.

```
(:goal (or (at cup1 stationA)
           (at cup2 stationA)))
```

Now, with this underspecification in the goal, a plan is found that may satisfy the user's request. The output of the planner below indicates in step 6 that there are other options that could be an answer to the request as well.

```
0: compute-route stationB siding engine1
1: go engine1 stationB siding
2: connect waggon1 engine1 siding
   compute-route siding stationC engine1
3: go engine1 siding stationC
4: put-on-waggon engine1 waggon1 cup2 cm1
                 stationC
   compute-route stationC stationA
                 engine1
5: go engine1 stationC stationA
6: GOAL-REACHED
```

So, correct and useful utterances in the spirit of (Webber, 1986) include:

- *I will bring you cup2. It's filled already.*
- *Is cup2 ok for you? Or do you prefer cup1? It would take me a bit longer, however.*
- *Do you want cup1 oder cup2?*

In each of the example cases discussed above, a particular discourse relation is assigned to each option available. As a consequence, the user's reaction determines how the structure of the dialogue will develop in further steps to come: Depending on the user's *choice* for a proposed option, the current discourse and application situation have to be updated and modified differently. New constraints that result from this update process, may in turn influence the *choices* available on discourse and application level.

6 Conclusions

We argue that the *choice* operator leads to a dialogue model that distinguishes between discourse and application domain as choices may become neccessary between options in the application (how to satisfy a request?) and in the discourse (how to communicate options?). Another consequence is that dialogues are guided by two levels of control for analysis as well as for generation: first, the computation and selection of options in the discourse (for explaining at least semantic, syntactic and discourse pragmatic ambiguities) and second the computation and selection of options in the application depending on decisions made in the discourse domain. Effective selection of options implies the availability of a decision procedure for this task (see (Carletta, 1992)). It must be able to handle ambiguities as well as unsatisfiable intentions. In our opinion, a (computationally tractable) approach requires different algorithms for reasoning and deciding due to the different nature of the tasks to be solved.

In our example application, we try to integrate reasoning and decision procedures in a complete dialogue system for spoken langugae processing. On the basis of the approach outlined in this paper, the knowledge for making the control levels work,

can be set up for various applications by defining the functionality of the application domain in terms of a set of PDDL plan operators.

There is a lot of related work on the implementation of dialogue systems. The most important one seems to be the TRAINS system and its successors as described in (Allen et al., 2001). In our view, a distinguishing feature of our work is the focus on data-driven adaptability to new domains – an issue not discussed in very much detail in (Allen et al., 2001). (Yates et al., 2003) describe a system that responds to user requests for handling a VCR. As our system, it computes plans to satisfying user intentions. However, the paper does not explain if options – as discussed above – can be computed and how they are integrated in a dialogue. We find that our hybrid approach covers a broader range of dialogues.

References

A. N. Abdallah. 1995. *The Logic of Partial Information*. New York.

J. F. Allen et al. 2001. Towards Conversational Human-Computer Interaction. *AI Magazine*, 22(4): 27–38.

N. Asher, A. Lascarides. 1998. Questions in Dialogue. *Linguistics and Philosophy*, 21(4): 237–309. Preprint version at: http://www.utexas.edu/cola/depts/philosophy/faculty/asher/main.html.

P. Bretier, D. Sadek. 1996. A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction. In: Müller, J. P. et al. (ed.): *Intelligent Agents III – Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)*, Lecture Notes in Artificial Intelligence: 189–203. Heidelberg.

A. Brietzmann, G. Görz. 1982. Pragmatics in Speech Understanding – Revisited. In: J. Horecky (ed.): *COLING 82 – Proceedings of the Ninth International Conference on Computational Linguistics*: 49–54. Amsterdam.

K. Bücher, G. Görz, B. Ludwig. 2002. Corega Tabs: Incremental Semantic Composition. In: Görz, G. et al. (ed.): *KI-2002 Workshop on Applications of Description Logics CEUR Proceedings*. Aachen.

S. Carberry. 1990. *Plan Recognition in Natural Language Dialogue*. MIT Press.

S. Carberry, L. Lambert. 1999. A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics*, 25(1): 1–53.

J. Carletta. 1992. *Risk-Taking and Recovery in Task-Oriented Dialogue*. University of Edinburgh.

J. Chu-Carroll, S. Carberry. 1995. Generating information-sharing subdialogues in expert-user consultation. In: *Proceedings of IJCAI 1995*: 1243–1250.

J. Chu-Carroll, S. Carberry. 1996. Conflict detection and resolution in collaborative planning. In: *Intelligent Agents: Agent Theories, Architectures, and Languages*: 111–126. Berlin.

M. Ghallab, A. Howe, C. Knoblock, D. McDermott, A. Ram, M. Veloso, D. Weld, D. Wilkins. 1998. *PDDL – The Planning Domain Definition Language*. AIPS-98 Planning Committee.

G. Görz, K. Bücher, Y. Forkl, M. Klarner, B. Ludwig. 2002. Speech Dialogue Systems – A "Pragmatics-First" Approach to Rational Interaction. In: Menzel, W. (ed.): *Natural Language Processing between Linguistic Inquiry and System Engineering*. Festschrift für Walther von Hahn (in print). Preprint version at: http://www8.informatik.uni-erlangen.de/IMMD8/staff/Goerz/papersgg/vHahnColl_2002.ps.gz. Hamburg.

H. P. Grice. 1969. Utterer's Meaning and Intention. In: *Philosophical Review, Vol. 78*: 147–177.

J. Koehler, B. Nebel, J. Hoffmann, Y. Dimopoulos. 1997. Extending Planning Graphs to an ADL Subset. In: *Proceedings ECP-97*: 273–285. Berlin.

L. Lambert. 1993. *Recognizing Complex Discourse Acts: A Tripartite Plan-Based Model of Dialogue*. University of Delaware.

N. Lesh, C. Rich and C. L. Sidner. 1999. Using Plan Recognition in Human-Computer Collaboration. In: *Proceedings of the 7th International Conference on User Modelling*: 23–32. Banff.

B. Ludwig, K. Bücher, G. Görz. 2002. Corega Tabs: Mapping Semantics onto Pragmatics. In: Görz, G. et al. (ed.): *KI-2002 Workshop on Applications of Description Logics CEUR Proceedings*. Aachen.

D. Sperber, D. Wilson. 1995. *Relevance – Communication and Cognition*. Oxford.

B. L. Webber. 1986. Questions, answers and responses: Interacting with knowledge-base systems. In: Brodie, M. (ed.): *On Knowledge Base Management Systems*: 366–402. New York.

A. Yates, O. Etzioni, D. Weld. 2003. A Reliable Natural Language Interface to Household Appliances. *Proceedings International Conference on Intelligent User Interfaces*: 189–196. Miami, Florida.

Roles in Dialogue

Joris Hulstijn
Institute of Information and Computing Sciences
Utrecht University
jorish@cs.uu.nl

Abstract

This paper investigates the use of social and participant roles in the definition of dialogue games. We present a formal representation which can be used in the specification of role related requirements for interaction between between artificial agents, and for the specification of human-computer interfaces.

1 Introduction

Dialogue participants have some reason to engage in a dialogue: they are executing some social activity, bound by conventional rules. Often the linguistic realisation of a social activity activity is standardised, and has turned into a dialogue type or genre, such as information exchange or negotiation. See Walton and Krabbe (1995) for a categorisation. The conventional rules of a genre may be expressed as a dialogue game. This approach has been applied in linguistics, e.g. (Mann, 1988; Carletta et al., 1997) and in research on multi-agent communication and argumentation (Walton and Krabbe, 1995; Reed, 1998; Kraus et al., 1998; McBurney and Parsons, 2002).

A dialogue game is a set of rules which determine for each participant what dialogue moves are allowed in a given dialogue context. The dialogue context contains the setting, the participants, the dialogue history and the apparent information states of the participants, including commitments and preferences. Update rules specify the expected effect of a move on the apparent information states of the participants. Other rules determine under what circumstances a dialogue can be initiated and terminated and what the apparent purpose of the participants is in initiating or taking up a dialogue game.

Crucial in the description of a dialogue game are the roles of the participants. A source of inspiration on roles is Goffman (1959; 1981). Some roles have direct linguistic relevance. For example, the role of addressee is needed to determine the meaning of the pronoun 'you'. It is well known that speakers adapt their way of speaking to the role of the addressee (Ladegaard, 1995). Like stereotypes, roles generate expectations, and based on a role the participants of a dialogue are assigned the obligations and permissions that make up the dialogue game rules (Traum and Allen, 1994). However, to our knowledge a systematic way of expressing role requirements in dialogue does not exist.

In this paper we want to further investigate the use of roles for understanding the regularities of both human and artificial dialogue. We present a crude formalisation which can be used in the specification of requirements for interaction between artificial agents, and for the specification of human-computer interfaces. The paper is organised as follows. Section 2 lists three examples of organisation models in dialogue. Section 3 defines the concepts of agents, groups and roles. Section 4 elaborates on the formalisation of role related requirements, while section 5 reconsiders the examples of section 2.

2 Roles in Dialogue

Roles in dialogue can be distinguished by their temporal scope. There are roughly three timescales. The participants fulfilling a *dynamic role*, such as speaker or addressee, may alternate repeatedly during a single dialogue. A formalisation of dynamic roles requires a way of expressing the role allocation change. An example is the turn taking mechanism (Sacks et al., 1974) to allocate the roles of speaker, addressee and overhearer. The allocation of *participant roles* on the other hand, such as an expert or novice in an information seeking dialogue, remains stable during a single stretch of interaction, or encounter. Such roles constitute the current dialogue game. Finally, *social roles* or role relations like teacher and pupil extend beyond single encounters. Their scope is determined by the social activity or situation. Social roles are presupposed by some moves in a dialogue game. On the other hand repeated interactions of a particular type help shape social relationships. The assignment of roles is often ambiguous, and is determined in a joint process. Linguistic cues like politeness help to indicate the current roles.

In each case the function of roles is both prescriptive and descriptive. Roles define permissions and obligations, but also trigger expectations and assumptions. In this respect roles are like stereotypes. The following examples illustrate the use of dynamic, participant and social roles, respectively.

Example 1. Consider a classroom situation, in which the teacher is giving a lecture, while the students are attentive (t_1). The teacher is the speaker and the students are the addressees. Then the teacher asks Bill a question (t_2). Now Bill is the addressee. The other students are still considered participants, because the question is intended to be instructive for them too. Such side participants are called auditors. By contrast, if the headmaster would interrupt the teacher by entering the room and asking a question (t_3), the students would not be participants but overhearers. In this case the students are recognised and allowed to be present by the speaker, so they are called bystanders. Unauthorised overhearers are called eavesdroppers (Bell, 1984; Clark, 1996).

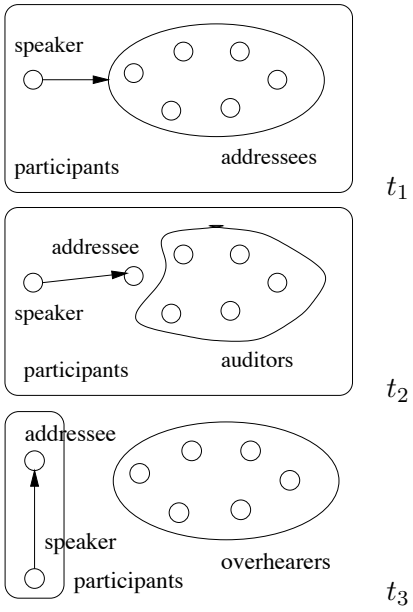


Figure 1: Dialogue roles of example 1.

Example 2. Consider the dialogue genre of cooperative information exchange (Hulstijn, 2000) with the roles of expert and novice. First, an expert knows more about the topic of the particular exchange than the novice. This difference in expertise or 'information potential' is the main reason for engaging in an exchange of information in the first place. Second, in case of a conflict between information from the expert and the novice, both agents are likely to prefer the information of the expert. Third, because the expert knows more and knowledge is valuable, the expert is likely to be of a higher status. This influences the wording and syntax of the utterances in the exchange, the turn-taking and the initiative handling. For example, the novice is less likely to interrupt the expert.

Example 3. Consider a teacher and a pupil. The setting is educational; the teacher is supposed to teach the pupil. There is a power relation, partly based on age and expertise, and partly on deferred authority of the school. For example, the teacher may discipline the pupil, assign homework, set exams and give grades. The setting allows different kinds of encounter, so there is a repertoire of several dialogue games (information seeking, examination, reproach), including also more general dialogue games like discussing the weather.

3 Organisation Structure

We introduce some concepts to give a formal description of organisation structures. An organisation consists of a set of *agents* structured by *groups*, *roles* and *role relations*. The theory is inspired by Ferber and Gutknecht (1998) and modified after other research in multi-agent systems (Carmo and Pacheco, 2003; Wooldridge et al., 2000). A discussion of the origin of roles in the social, psychological and linguistic literature is beyond the scope of this paper. In general, social requirements are defined on roles instead of agents for two reasons. Firstly, agents are autonomous, although they are restricted by responsibilities and obligations (Castelfranchi, 1998). They can for example decide when and how to formulate their responses. Secondly, the social activities that determine role requirements are relatively stable, whereas the allocation of agents to roles may alter quickly.

In this paper the organisational concepts are interpreted as follows.

A *group* is a set of agents that share a group characteristic. For example, agents speaking a particular language, or agents that are mutually connected by a communication channel form a group.

A *role* is a set of related constraints put on an agent by its place in the organisation. A role defines a set of constraints on the expertise, capabilities, responsibilities, goals, obligations and permissions of the agent. A role is always related to some organisational objective or activity. For example, the role of chairman only makes sense during a meeting. Agents may only fulfill a role provided they are *qualified*, i.e., possess the required minimal properties to fulfill a role.

One role can be fulfilled by several agents. Consider for example several postmen in a district. On the other hand, one agent can fulfill several roles. For example, Mintzberg (1979) identifies various managerial roles: resource allocator, disturbance handler, progress monitor, disseminator of information, leader, etc. A *position* is a collection of roles commonly fulfilled by a single agent. The roles for a position must not interfere. For example, a member of a program committee should not review his own or his students' papers. Organisations must be designed in such a way that such conflicts will not occur. This is called *separation of duty* (Sandhu et al., 1996).

Role relations, also known as *dependencies* (Malone and Crowston, 1994) or *channels* (Dretske, 1981) are simultaneous constraints on two different agents, based on their respective roles. Role relations coordinate behaviour. Examples are authority relations (employer-employee), task-based dependencies (consumer-producer) and communication channels (sender-receiver). In computer science, interaction is specified by protocols, often in the shape of (timed) automata. An example is the contract net protocol (Smith, 1980). Below we use dialogue games to express interaction constraints.

It is possible to define group, role and role relation in terms of each other. Groups can be defined by the role of a group member. A role can be seen as a role relation directed towards an abstract entity like the organisation. Such a reference also indicates the scope of the role requirements. According to Bill Mann (p.c.) role relations are prior to roles. Compare for example the doctor role in a doctor-patient relation, which is different from the doctor role in a doctor-nurse relation.

4 Formal Requirements

To allow formal specification, we use predicate logic to define an organisational structure. Consider models $M = \langle D, I \rangle$ where the domain $D = A \cup R \cup G \cup Ch \cup E \cup T$ consists of agents A, roles R, groups G, channels Ch, other entities E and time points T. Since we want to quantify over roles, groups and channels but avoid higher order logic, we use specific predicates R, G and Ch and write $\mathrm{R}(a, r)$ whenever agent a enacts role r, $\mathrm{G}(a, g)$ when agent a is a member of group g and $\mathrm{Ch}(a, b, ch)$ when channel ch is established between agents a and b. Whenever an entity e is crucial to the definition of a role we write $\mathrm{R}(a, r, e)$. We allow nesting of groups and roles. So a group may itself play a role at a higher level of abstraction. To deal with dynamic roles, predicates may be indexed by moments $t_1, t_2, ..$ ordered on a linear scale. More elaborate temporal logics can be used when needed. We use θ to denote assignments of objects to variables.

Organisation structures can be depicted graphically, as in figure 1. Large ellipses represent groups of participants. Small circles represent agents, annotated with roles. Currently instantiated role relations are depicted by an arrow, pointing from the initiator of the establishment of the role relation to the other participant(s).

4.1 Role Definitions

Associated with a role is a set of requirements. Think of these as the constraints put on an agent's behaviour, by virtue of the role. Such requirements may be expressed by formulas of the following form.

$$\begin{aligned} \forall x\, r(\mathrm{R}(x,r) \rightarrow \varphi_r(x)) \quad & \text{qualification} \\ \forall x\, r(\mathrm{R}(x,r) \rightarrow \mathrm{O}_x\varphi(x)) \quad & \text{specification} \\ \forall x\, r(\mathrm{R}(x,r) \rightarrow \mathrm{P}_x\varphi(x)) & \end{aligned}$$

A role definition consists of a set of such formulas. If predicate logic does not suffice, the requirements must be expressed in another logic. We use versions of BDI agent logics (Hindriks et al., 1999; Wooldridge et al., 2000) extended with a deontic logic. Role definitions consist of two parts. The *qualification* restricts the assignment of agents to a role. If an agent does not comply, it is unsuitable to fulfill the role in the first place. Qualifications include expertise, capabilities and motivation. Once assigned, we can use the qualifications of a role as a source of expectations.

The *specification* defines the actual responsibilities, obligations and permissions of a role. If an agent fails to achieve a responsibility, it is not immediately taken from the role. In computer science obligations and permissions are usually expressed in a table or by declarative policy rules that are directly enforced by the operating system. This conflicts with agent autonomy. Moreover, if policies conflict, agents and system designers need a form of 'contrary to duty' reasoning. In such cases an explicit deontic logic with operators O (obligation) and P (permission) becomes useful. The specification requirements can therefore be thought of as embedded in the deontic operators O or P. Instead of $\mathrm{R}(x,r) \rightarrow \mathrm{O}_x\varphi(x)$ it might be more appropriate to use specific conditional obligations $\mathrm{O}_x(\varphi(x) \mid \mathrm{R}(x,r))$ (van der Torre and Tan, 1999).

provided by agent	required by role
knowledge	expertise
capabilities	permissions
goals	responsibilities
practical reasoning rules	interaction constraints

Table 1: Role-based requirements

4.2 Role Requirements

Consider table 1. On the right some kinds of role related requirements are displayed; on the left the corresponding agent characteristics.

Given a role, other agents expect certain *expertise* and *competence*. Expertise can be expressed in a modal epistemic logic with operators like K and B. For example, at school a pupil should know that Paris is the capital of France. But we also want to specify as yet unknown expertise. For example, the chairman of a meeting is required to know who will be present. Actually, one needs quite an elaborate logic to express such constraints. In this case we choose a version of Groenendijk and Stokhof's (1996) semantics of questions combined with a modal operator K^{wh} that embeds issues, i.e. the content of a question, rather than propositions.

$$\begin{aligned} \forall x(\mathrm{R}(x,\mathrm{pupil}) \rightarrow \mathrm{K}_x\mathrm{capital}(\mathrm{France},\mathrm{Paris})) \\ \forall xy(\mathrm{meeting}(y) \wedge \mathrm{R}(x,\mathrm{chair},y) \rightarrow \\ \mathrm{K}^{wh}_x ?z(\mathrm{present}(y,z))) \end{aligned}$$

We found that requirements on expertise generally need to be formulated in terms of knowledge-wh, so as potential answers to questions, rather than as the customary knowledge-that.

Competence (knowledge-how) can be captured by a list of capabilities, i.e., the actions an agent is in principle capable to perform. The deontic counterpart of a capability is a permission. The reasoning capabilities of an agent can be expressed using *practical reasoning rules*, of the form $Cond \Rightarrow Action$, to indicate which actions the agent should perform based under what circumstances (Hindriks et al., 1999).

Task-based roles often contain *responsibilities*. Responsibilities are those intentions the agent is required to pursue by the nature of its role. Using the extended logic suggested above, responsibilities are of the form $\mathrm{R}(x,r) \rightarrow \mathrm{O}_x\mathrm{I}_x\varphi$. Like for goals, one can distinguish *maintenance re-*

sponsibilities, to maintain some property, from *achievement responsibilities* to reach some non-actual state of affairs.

It is difficult to separate responsibilities from plain obligations. One difference seems to be that responsibilities, like commitments, are voluntarily accepted, whereas obligations are imposed. In case of a violation, the violator is liable. A typical sanction is banishment from the community. Failure of a responsibility is less dramatic; only when the agent did not try and apparently did not have the intention, this may be reason for a sanction. Another difference concerns the time scale. The scheduling of an achievement responsibility is left to the discretion of the individual agent; achievement obligations are usually immediate, or have a fixed deadline. The distinction between goals and responsibilities is not always clear either. An agent may identify so much with its role, that its responsibilities become individual goals. This is called embracement by Goffman (1981).

4.3 Interaction Requirements

Interaction constraints are specified by protocols or dialogue game rules. As agents engage in a dialogue, they make a commitment to play by its rules. This creates a temporary community of participants. In this community, the effect of a dialogue game rule on an individual participant can be described using so called discourse obligations (Traum and Allen, 1994). Discourse obligations are conditional on the dialogue context, which includes the latest move. How can we separate interaction requirements into obligations for individual roles? Roughly, there are two cases. If the 'poles' of a role relation, channel or dependency are identified with particular roles, as in the consumer-producer dependency, these roles are used. Otherwise, a dialogue game can be used to specify the kinds of interaction that are allowed to take place. For each dialogue game we identify the roles of initiator and responder (Mann, 1988).

$$\forall xy\, ch\, r_1 r_2(\mathrm{Ch}(x,y,ch) \wedge \mathrm{poles}(ch,r_1,r_2) \leftrightarrow \mathrm{R}(x,r_1) \wedge \mathrm{R}(x,r_2))$$
$$\forall xy\, d(\mathrm{Ch}(x,y,d) \wedge \mathrm{dial\_game}(d) \leftrightarrow \mathrm{R}(x,\mathrm{ini},d) \wedge \mathrm{R}(x,\mathrm{res},d))$$

Interaction can be modelled at several levels (Clark, 1996). Usually, dialogue games are formulated at the level of dialogue acts, having a certain task related function and a semantic content. But there are conventional interaction requirements at the other levels too. At the syntactic level, it is obvious that speakers take the addressee into consideration. At the level below that, of presenting and attending to signals, we may place the dynamic roles of example 1. In multi-party face to face spoken dialogue we may distinguish the roles of speaker, addressee, overhearer, auditor and participant (Bell, 1984). All people within hearing distance are overhearers. An auditor is an overhearer that is ratified as a participant by the speaker.

$$\forall xyz(\mathrm{Ch}(x,y,z) \wedge \mathrm{face\_to\_face}(z) \leftrightarrow \mathrm{R}(x,\mathrm{sp},z) \wedge \mathrm{R}(y,\mathrm{ad},z))$$
$$\forall xzu(\mathrm{R}(x,\mathrm{sp},z) \wedge \mathrm{hear}(u,x) \leftrightarrow \mathrm{R}(u,\mathrm{oh},z))$$
$$\forall xzu(\mathrm{R}(u,\mathrm{oh},z) \wedge \mathrm{R}(x,\mathrm{sp},z) \wedge \mathrm{ratified}(x,u,z) \leftrightarrow \mathrm{R}(u,\mathrm{aud},z))$$
$$\forall xz(\mathrm{R}(x,\mathrm{sp},z) \vee \mathrm{R}(x,\mathrm{ad},z) \vee \mathrm{R}(x,\mathrm{aud},z) \leftrightarrow \mathrm{R}(x,\mathrm{part},z))$$

4.4 Group Requirements

Similar to role related requirements, there are group requirements φ_g that all members of a group g ought to satisfy. For example, all members of a club should have paid the membership fee. This does not mean that all members have actually paid. Violations are possible. On the other hand, having a characteristic may be enough to qualify as a member of a group. For example, hearing the speaker is enough to qualify as an overhearer.

$$\begin{aligned} \forall x\, g(\mathrm{G}(x,g) &\rightarrow \mathrm{O}_x\varphi(x)) \\ \forall x\, g(\mathrm{G}(x,g) &\rightarrow \mathrm{P}_x\varphi(x)) \\ \forall x\, g(\varphi_g(x) &\rightarrow \mathrm{G}(x,g)) \end{aligned}$$

As part of the group characteristics we may require that all pairs of members of a group are connected by some communication channel. On the other hand, given a single communication channel, all the agents connected to that channel form a group, e.g., all ships using radio channel 16.

$$\forall xy\, g(\mathrm{G}(x,g) \wedge \mathrm{G}(y,g) \rightarrow \exists ch \mathrm{Ch}(x,y,ch))$$
$$\forall xy\, ch(\mathrm{Ch}(x,y,ch) \rightarrow \exists g(\mathrm{G}(x,g) \wedge \mathrm{G}(y,g)))$$

Interesting group requirements relate to secrets or classified information. By definition, a formula φ is a secret among agents in group h, when no agent outside of the group knows it, and is allowed to know it. A secret can be maintained by the individual obligation of members of the group to keep it a secret.

$$\forall h(\text{secret}(\varphi, h) \leftrightarrow \forall x(\text{K}_x\varphi \rightarrow \text{G}(x, h)))$$
$$\forall xy\, h(\text{G}(x, h) \wedge \neg\text{G}(y, h) \wedge \text{secret}(\varphi, h) \rightarrow \text{O}_x\neg K_y\varphi)$$

An issue, i.e. the content of a question, may also be defined a secret. Issues can specify as yet unknown knowledge, like a password. To maintain the secret, we can specify an obligation that nobody but you is to know your password.

$$\forall h(\text{secret}(?\varphi, h) \leftrightarrow \forall x(\text{K}_x^{wh}?\varphi \rightarrow \text{G}(x, h)))$$
$$\forall x(x \neq \text{you} \rightarrow \text{O}_x\neg\text{K}_x^{wh}?y(\text{passwd}(\text{you}, y)))$$

In order for roles to work, the allocation of agents to roles must be known among all members of the group that use them. To this end, agents in a role are often indicated by external characteristics, such as a uniform, or placement behind a desk. A role r is called *transparent* in group h whenever all members of h know which agent is enacting it.

$$\forall rhx(\text{trans}(r, h) \leftrightarrow (\text{G}(x, h) \rightarrow \text{K}_x?y\text{R}(r, y)))$$

4.5 Exclusion

A single agent in an organisation may perform many different roles at the same time. However, some roles may not be combined. Such roles are called mutually exclusive (Sandhu et al., 1996). Organisations must be defined in such a way that conflicts do not occur. This can be achieved by putting exclusion clauses in the qualification requirements. There are different kinds of exclusion. *Static exclusion* concerns roles that may not be combined. For example, to avoid conflicts of interest a government minister should not also be manager of a large company.

$$\forall x(\text{R}(x, \text{minister}) \rightarrow \neg\text{R}(x, \text{manager}))$$

Dynamic exclusion concerns roles that may be performed by the same agent, as long as they do not concern the same case. For example, a programme committee member may submit a paper to a conference, as long as she does not have to review her own paper. Thus the roles of submitter and reviewer are mutually exclusive, when it concerns the same paper.

$$\forall xe(\text{R}(x, \text{submitter}, e) \rightarrow \neg\text{R}(x, \text{reviewer}, e))$$
$$\forall xe(\text{R}(x, \text{reviewer}, e) \rightarrow \neg\text{R}(x, \text{submitter}, e))$$

Resource-based exclusion derives from the distribution of limited resources like time, tools or energy. If there is one hammer, only one person at a time can use it. A turn-taking model can be implemented using such an exclusive artifact or token. Consider a relay race in athletics. The second runner may not start before the first arrived. Overlaps are difficult to observe, so the rule is implemented using a baton. When the baton is dropped, the rule is violated.

$$\forall xy(\text{R}(x, \text{runner}) \wedge \text{R}(y, \text{runner}) \rightarrow x = y)$$
$$\forall x(\text{R}(x, \text{runner}) \leftrightarrow \text{hold}(x, \text{baton}))$$

An interesting example of resource based exclusion is the speech channel. Unlike written language for example, speech is not persistent. Therefore, an utterance must be produced and processed at the same time. Moreover, overlapping speech signals interfere, which complicates human processing. For this reason, overlaps should be avoided: the speaker role is largely exclusive. On the other hand, a speech channel (attention) needs to be maintained. This takes effort. Therefore, gaps in the use of the channel should also be minimised.

$$\forall xyz(\text{R}(x, \text{sp}, z) \wedge \text{R}(y, \text{sp}, z) \rightarrow x = y)$$
$$\forall xz(\text{R}(x, \text{sp}, z) \rightarrow \text{speak}(x, z))$$

The turn-taking mechanism is a self-organising process that implements a balance between these two opposing global requirements (Sacks et al., 1974). In this sense, turn taking does not differ much from other resource allocation problems. Similar to a baton, the speaker holds the turn. Holding the turn gives the right to speak, but it also triggers an obligation to speak or otherwise to release the turn. Turn taking rules are summarised in table 2.

For each of the following clauses, take
$\forall xyz\, t_1 t_2 t_3\ t_1 \leq t_2 < t_3, \mathrm{tcp}(t_1), \mathrm{tcp}(t_3)$
$\mathrm{R}(x, \mathrm{sp}, z, t_1) \rightarrow \mathrm{P}_x\mathrm{speak}(x, z, t_2)$
$\mathrm{R}(x, \mathrm{sp}, z, t_1) \rightarrow \mathrm{O}_x(\mathrm{speak}(x, z, t_2) \vee \mathrm{release}(x, z, t_2))$
$\mathrm{R}(x, \mathrm{sp}, z, t_1) \wedge \mathrm{R}(y, \mathrm{ad}, z, t_1) \rightarrow \mathrm{P}_x\mathrm{sel\_sp}(x, y, z, t_2)$
$\mathrm{speak}(x, z, t_2) \wedge \neg\exists u(\mathrm{speak}(u, z, t_2)) \rightarrow \mathrm{R}(x, \mathrm{sp}, z, t_3)$
$\mathrm{sel\_sp}(x, y, z, t_2) \rightarrow \mathrm{R}(y, \mathrm{sp}, z, t_3)$
$\mathrm{release}(x, z, t_2) \rightarrow \neg\mathrm{R}(y, \mathrm{sp}, z, t_3)$
$\neg\mathrm{R}(x, \mathrm{sp}, z, t_1) \wedge \mathrm{R}(y, \mathrm{part}, z, t_1) \rightarrow \mathrm{P}_y\mathrm{speak}(y, z, t_2)$

Table 2: Interpretation of turn taking rules

4.6 Global Requirements

A set of role definitions defines an organisation structure. In addition to role definitions we assume formulas expressing facts about individual agents and about the environment. A set of role descriptions with an assignment of agents to roles defines a system. Since formulas may conflict, different extensions, maximal consistent sets of formulas, correspond to different system configurations. Given some constraints on the assignment of agents to roles, we might be able to prove global requirements of a system. Such constraints on the assignment are (i) that agents are qualified, i.e. satisfy the basic requirements that enable them to fulfill the obligations and responsibilities associated with the role, (ii) that agents are motivated, i.e. incorporate (enough) responsibilities associated with the role as a personal goal, (iii) that agents are obedient, i.e. (usually) respect obligations and permissions associated with the role, (iv) that the roles assigned to one agent are non-exclusive, and (v) that roles are transparent.

This line of reasoning can be sketched as follows: a set of role, group and role relation descriptions Γ_{org}, a set of agent descriptions Γ_{ag}, a description of the environment Γ_{env} and an assignment θ of agents to roles that satisfies (i) – (iv) may provide enough structure to prove global system properties Δ. However, because of the non-determinism resulting from agent autonomy, we expect that many global requirements can only be demonstrated using simulation experiments.

$$\Gamma_{org}, \Gamma_{ag}, \Gamma_{env} \models_\theta \Delta$$

5 Applications

In this section we reconsider some aspects of the examples of section 2.

	novice(n)	expert(e) $x, y \in \{n, e\}$
exp.	$\neg\mathrm{K}_n^{wh}?\varphi$	$\mathrm{K}_e^{wh}?\varphi$
resp.	$\mathrm{I}_n\mathrm{K}_n^{wh}?\varphi$	$\mathrm{I}_e\mathrm{K}_n^{wh}?\varphi$
perm.	$\mathrm{ask}(x, y, ?\psi), \mathrm{inform}(x, y, \varphi), \mathrm{ack}(x, y, \varphi)$	
inter.	$\mathrm{latest\_move}(\sigma, \mathrm{ask}(x, y, ?\psi)) \wedge \mathrm{K}_y\chi$ $\wedge\ \mathrm{answer}(\chi, ?\psi, \sigma) \rightarrow \mathrm{O}_y\mathrm{inform}(y, x, \chi)$	
	$\mathrm{latest\_move}(\sigma, \mathrm{ask}(x, y, ?\psi)) \wedge \neg(\mathrm{K}_y\chi$ $\wedge\ \mathrm{answer}(\chi, ?\psi, \sigma)) \rightarrow \mathrm{O}_y\mathrm{ack}(y, x, ?\psi)$	
	$\mathrm{latest\_move}(\sigma, \mathrm{inform}(x, y, \psi)) \wedge \mathrm{K}_y\chi$ $\wedge\ \mathrm{coherent}(\chi, ?\varphi, \sigma[\psi]) \rightarrow \mathrm{O}_y\mathrm{inform}(y, x, \chi)$	
	$\mathrm{latest\_move}(\sigma, \mathrm{inform}(x, y, \psi)) \wedge \neg(\mathrm{K}_y\chi$ $\wedge\ \mathrm{coherent}(\chi, ?\varphi, \sigma[\psi])) \rightarrow \mathrm{O}_y\mathrm{ack}(y, x, \psi)$	

Table 3: Information seeking dialogue game

First consider the allocation mechanism of the dynamic roles of example 1. Table 1 contains an adaptation of Sacks et al. (1974, p 704) turn taking rules. It is meant to illustrate the definitions; not to be of much empirical value. For now, we use predicates indexed with time points, some of which correspond to the so called turn relevance places. We skip over the problem of projecting the next turn relevance place on the basis of what was said. We are developing a simulation algorithm of the turn taking mechanism, using the agent programming language 3APL (Hindriks et al., 1999).

Next consider example 2. Table 3 contains a rough approximation of the role requirements for a cooperative information exchange between expert and novice on a particular topic, specified here by an issue $?\varphi$, in a dialogue context σ. We use dialogue game rules that model initiative-response units. These are based on the latest move. The answer requirement means that χ is consistent, informative, relevant and not over-informative with respect to issue $?\psi$ in dialogue context σ. The coherence requirement means that χ is consistent, informative and relevant with respect to the dialogue issue $?\varphi$ in dialogue context $\sigma[\varphi]$, i.e., σ with ψ added to it (Hulstijn, 2000).

Finally, reconsider the authority relation of example 3 between teacher and pupils that allows the teacher to assign homework and set exam dates, in short, to assign new obligations and responsibilities. In a way, a teacher has meta-privileges and can alter part of the specification of the roles. However, these alterations must take place by means of a dialogue game that ensures that pupils know the new rules and are in a position to comply. Such dialogues are a topic of future research.

6 Conclusions

In order to specify dialogue game rules, we need to be able to express requirements on the social roles, participant roles and dynamic roles of dialogue participants. To this end, we have presented the concepts of agents, roles, groups and role relations to model an organisation structure. We have given a sketch of a way to formalise such an organisation structure, and have presented various examples of its use in the specification of interaction requirements. Further research is concerned with a simulation of the turn taking mechanism, improvement of the representation formalism and application of the logic to a case study from electronic commerce.

Acknowledgements Thanks to Bill Mann for suggesting this topic to me. The representation is developed in collaboration with Jan Broersen, Mehdi Dastani and Leendert van der Torre.

References

A. Bell. 1984. Language style as audience design. *Language in Society*, 13:145–204.

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.

J. Carmo and O. Pacheco. 2003. A role based model for the normative specificatiuon of organized collective agency and agents interaction. *Autonomous Agents and Multi-Agent Systems*, 6:145–184.

C. Castelfranchi. 1998. Modelling social actions for ai agents. *Artificial Intelligence*, 103:157–182.

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

F. Dretske. 1981. *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA.

J. Ferber and O. Gutknecht. 1998. A meta-model for the analysis and design of organizations in multi-agent systems. In *Proceedings ICMAS'98*, 128–135. IEEE Press.

E. Goffman. 1959. *The Presentation of Self in Everyday Life*. Doubleday, Garden City, New York.

E. Goffman. 1981. *Forms of Talk*. University of Pennsylvania Press, Philadelphia.

J. Groenendijk and M. Stokhof. 1996. Questions. In Johan Van Benthem and Alice Ter Meulen, editors, *Handbook of Logic and Language*, 1055–1124. North-Holland, Elsevier.

K. V. Hindriks, F. S. de Boer, W. van der Hoek, and J.-J. M. Meyer. 1999. Agent programming in 3APL. *Autonomous Agents and Multi-Agent Systems*, 2(4):357–401.

J. Hulstijn. 2000. *Dialogue models for inquiry and transaction*. Ph.D. thesis, University of Twente, Enschede.

S. Kraus, K. Sycara, and A. Evenchik. 1998. Reaching agreement through argumentation: a logical model and implementation. *Artificial Intelligence*, 104:1–69.

H. J. Ladegaard. 1995. udience design revisited: persons, roles, and power relations in speech interactions. *Language and Communication*, 15:89–101.

T. Malone and K. Crowston. 1994. The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1).

W. C. Mann. 1988. Dialogue games: Conventions of human interaction. *Argumentation*, 2:511–532.

P. McBurney and S. Parsons. 2002. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334.

H. Mintzberg. 1979. *The structuring of organisations*. Prentice Hall, Englewood Cliffs, N.J.

C. Reed. 1998. Dialogue frames in agent communication. In Proceedings ICMAS'98, 246–253. IEEE Press.

H. Sacks, E.A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organisation of turn-taking for conversation. *Language*, 50:696–735.

R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. 1996. Role-based access control models. *IEEE Computer*, 29(2).

R. G. Smith. 1980. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, 29(12):1104–1113.

D. Traum and J. Allen. 1994. Discourse obligations in dialogue processing. In Proceedings ACL'94, 1–9. Morgan Kaufmann Publishers.

L. van der Torre and Y.-H. Tan. 1999. Contrary-to-duty reasoning with preference-based dyadic obligations. *Annals of Mathematics and Artificial Intelligence*, 27:79–128.

D. N. Walton and E. C. Krabbe. 1995. *Commitment in dialogue: basic concepts of interpersonal reasoning*. State University of New York Press.

M. J. Wooldridge, N. R. Jennings, and D. Kinny. 2000. The Gaia methodology for agent-oriented analysis and design. *Autonomous Agents and Multi-Agent Systems*, 3(3):285–312.

Towards the Elimination of Centering Theory

Rodger Kibble
Department of Computing
Goldsmiths College
University of London
R.Kibble@gold.ac.uk

Abstract

The objective of this paper is to examine whether the rules and constraints of Centering Theory (CT) can be reduced to more general principles of informativity and processing efficiency. We argue that certain OT approaches to discourse anaphora share with CT an oversimplification of the costs incurred by speakers and hearers in collaborative reference resolution, and conclude by outlining a game-theoretic model which aims to take all relevant factors into account, and allows "classic" CT to be derived as a special case.

1 Introduction

Centering Theory (Grosz et al., 1995) is a theory of local discourse structure which has been motivated on the grounds that discourse transitions defined as "coherent" are easier for hearers to process. However, discussions of the cognitive aspects of CT have tended to overlook its implications for the processing load it assumes for the *speaker*: it turns out (Kibble and Power, 2000) the task of planning sequences of clauses and ordering arguments within clauses to maximise coherent transitions soon becomes intractable. And hearers only benefit from "coherent" transitions if they maintain a discourse model enabling them to predict the occurrence of preferred transition types, which itself imposes some processing load.

These considerations suggest that firstly, an adequate account of the phenomena addressed by CT requires both speaker and hearer processes to be modelled in parallel; and secondly, if CT's rules and principles are indeed congruent with reasonable assumptions about processing costs, it may be possible to reduce them to more general "cognitive" principles. This is not a novel idea: Hasida Nagao and Miyata (1995) proposed a game-theoretic account of communication which they claimed "reduces the specifics of anaphora to general properties, such as utility and probability" and recent work by Rob van Rooy (2003) in the same spirit aims to derive the key notions in David Beaver's (2003) OT-based reconstruction of CT from Blutner's (2000) theory of rational language use, bidirectional OT. Beaver (op. cit.) reformulates CT as COT in terms of constraints which are independently motivated in the linguistics literature, and (Buchwald et al., 2002) outline a CT-like model of discourse anaphora in an OT framework (ROT) where the goal is to dispense with notions like "topic" and "[allow] the whole system to emerge from the interactions between constraints that do not depend on special-status elements".

I shall begin with a brief critical overview of these proposals, which will lead me to question whether OT with its ordinally ranked constraints and strict domination is really an appropriate framework for modelling discourse phenomena. This will be followed by an updated exposition of the game-theoretic model first proposed in (Kibble, 2003), and conclude by considering some outstanding empirical problems.

2 Centering Theory

Centering theory (CT) models the interaction of *cohesion* and *salience* in the internal organisation of a text. The basic assumptions of the theory (Grosz et al., 1995) are[1]:

> 1. For each utterance in a discourse there is precisely one entity which is the centre of attention or *center* (**Constraint 1**). The center of utterance U_n is the least oblique argument of U_{n-1} which is mentioned in U_n (**Constraint 3**).
>
> 2. The center is the entity which is most likely to be pronominalised (**Rule 1**).
>
> 3. There is a preference for consecutive utterances within a discourse segment to keep the same entity as the center, and for the Subject to be interpreted as the center and predicted as the center of the following utterance (**Rule 2**).

These principles together account for the general intuitions that in the examples below, a proper name is needed in (c′) to force the subject to be understood as *Betty*, and the variants (c″, c‴) where the pronoun has to be understood as *Betty* are unnatural.

(1) a. Alice likes Betty.

b. She often visits her for tea.

c. Yesterday she baked a cake for her.

c′. Yesterday Betty baked a cake for her.

c″. ?#Yesterday Alice baked a cake for her.

c‴. ?#Yesterday she baked a cake for Alice.

Centering has been seen as a theory of anaphora resolution but it is more accurate to see it as a model of *predictive reference resolution*. In contrast to models where a search for antecedents is triggered by anaphoric expressions, CT assumes that the process of interpreting an utterance identifies some entity as the default referent for the most salient referring expression in the following utterance, thus enabling the hearer to begin constructing a semantic representation in advance. If this position is taken by an RE which is incompatible with this "default referent" there is a disruptive effect, as in (1c‴).

[1]This is only a cursory sketch; readers should turn to the cited papers for fuller accounts.

3 Game Theory and OT

Hasida Nagao and Miyata (1995) present a reversible decision matrix (Table 1) which allows S and H (as players in a communication game) to converge on a mapping between *contents* $\{c_1, c_2\}$ and *messages* $\{m_1, m_2\}$ in contexts where each message potentially refers to either content (e.g., one is a pronoun and one a definite description). Suppose P(c_1) > P(c_2) and U(m_1) > U(m_2) where P is probability and U is utility. Expected Utility is maximised if S chooses m_1, m_2 to convey c_1, c_2 respectively and R interprets m_1 as c_1 and m_2 as c_2:

$$P(c_1)U(m_1) + P(c_2)U(m_2) \; >$$

$$P(c_2)U(m_1) + P(c_1)U(m_2)$$

(2) a. Tom was late.

b. Bob scolded him$_{Tom}$.

c. The man$_{Tom}$ was angry with him$_{Bob}$.

In the analysis of the above example it is assumed that the pronoun *him* has greater utility than the proper name, as a "lighter" expression, and that *Bob* as Subject of (2b.) has greater probability of occurring in (2c.) The claim is that this account "reduces the specifics of anaphora to general properties, such as utility and probability…" But: the analysis seems to rest on specific stipulations about utility (of pronouns) and probability (assumed to correlate with grammatical salience), and assumes that utility of messages is the same for S and R.

Robert van Rooy (2003) follows Hasida et al's approach, aiming to derive the constraints and ranking in Beaver's (2003) OT account of CT from general principles of information theory. The essential components of an OT grammar are: a set of **Constraints** which are ranked, universal and violable; **Gen** which defines all possible outputs for

		Message	
		m_1	m_2
Content	c_1	P(c_1)U(m_1)	P(c_1)U(m_2)
	c_2	P(c_2)U(m_1)	P(c_2)U(m_2)

Table 1: Decision matrix for a communication game

a given input and **Eval** which selects the optimal candidate according to **strict domination**: candidate A is optimal if there is some constraint C_i such that A has fewer violations of C_i than any other candidate and there is no higher-ranked C_j for which A has more violations than some other candidate (McCarthy, 2002). Centering-related constraints in Beaver's COT include the following:

(3)

AGREE	Anaphors agree with their antecedents.
DISJOINT	Co-arguments of a predicate have disjoint reference.
PRO-TOP	The topic is pronominalised.
FAM-DEF	Each definite NP is familiar.
COHERE	The topic of S_n is the same as the topic of S_{n-1}
ALIGN	The topic occurs in subject position.
Rank:	AGREE > DISJOINT> PRO-TOP > FAM-DEF > COHERE > ALIGN

Beaver demonstrates that his model makes the same predictions as the classic BFP algorithm (Brennan et al., 1987), with the claimed advantage of offering a declarative, reversible set of constraints which determine both the optimal linguistic forms given a semantic input, and the optimal interpretation of a given discourse. The model is extended to evaluate the coherence of multi-sentence text, with violations of each constraint summed and optimality determined in the standard way following strict domination. Thus in example (4a-b-c) the interpretation of the (c) sentence is correctly predicted as *she = jane*, *the young woman = mary*.

(4) a. Jane$_i$ likes Mary$_j$.

b. She$_i$ often visits her$_j$ for tea.

c. She$_k$ chats with the young woman for ages.

This can be verified by reading Figure 1 from left to right and at each column, retaining only candidates that have no more marks against them than any other surviving candidate: the arrow $\rightarrow$ marks the winner (optimum) $k = i, l = j$.

However, the constraint ranking given in (3) turns out to be underdetermined by the examples discussed. For instance the same optimum is chosen for (4) if one reads the table from right to left. One example in (Beaver, 2003) where the constraint ranking affects the outcome is the slightly pathological (5) which is admitted to be "difficult to process":

(5) a. Mary$_i$ likes tennis.

b. She$_i$ plays Jim$_j$ quite often.

c. He$_k$ used to play doubles with Mary$_l$.

The problem here is that both BFP and COT interpret the two occurrences of *Mary* as referring to different people (see Figure 2). As an instance of the flexibility which is a hallmark of COT, Beaver derives the correct reading by re-ranking FAM-DEF above PRO-TOP giving the full ranking

$$\text{AG} > \text{DIS} > \text{FAM} > \text{PRO} > \text{COH} > \text{AL}$$

But hard cases make bad law, and in any case this is not the only re-ranking that does the job: as the reader may verify, the preferred reading $k = j,\ l = i$ is also selected by

$$\text{AG} > \text{DIS} > \text{COH} > \text{AL} > \text{FAM} > \text{PRO}$$

To conclude, Beaver succeeds in unifying CT's heterogenous rules and constraints as an elegant set of principles which conspire to identify optimal interpretations, but it remains to be proven that the additional OT apparatus of ordinal ranking and strict domination actually takes us much further.

Rob van Rooy seeks to motivate Beaver's constraints and ranking on information-theoretic grounds: "light" expressions such as pronouns appropriately realise high-probability referents,

Example (4c)	Ag	Dis	Pro	Fam	Coh	Al
→ k = i, l = j				*		
k = j, l = i			*	*		*
k = i, l = i		*		*		

Figure 1: Optimality tableau for example (4c)

Example (5c)	Ag	Dis	Pro	Fam	Coh	Al
k = j, l = i			*			*
→ k = j, l ∉ {i, j}				*	*	
k ∉ {i, j}, l = i			*	*		*
k, l ∉ {i, j}			*	**	*	*

Figure 2: Optimality tableau for example (5c)

while sequences conforming to COHERE and ALIGN minimise violations of PRO-TOP by maximising candidates for pronominalisation and so reducing speaker's effort. However it is questionable whether the facts about extended discourse can be adequately captured in a standard OT framework, which is after all supposed to be a *grammatical* formalism. Specifying constraints in terms of quantifiable "effort" is difficult to reconcile with **strict domination**. Suppose a unit of effort is one *zipf* or ζ. Suppose the cost of violating C_i is $m\zeta$ and the cost of violating the lower-ranked C_j is $n\zeta, m > n$. Clearly, for some x, y, $xm < yn$: the effort incurred by violations of C_j would outweigh the effort saved by observing C_i.

Buchwald et al (2002) propose a bidirectional, speaker-oriented account of discourse anaphora ROT: S computes the optimal output PF for a given LF, and PFs are screened to ensure the LF is "recoverable" for H. This account can be seen to follow mainstream OT in essentially modelling a conflict between *markedness* and *faithfulness*: simpler NP forms are unmarked, according to *STRUCTURE constraints, while "faithfulness" constraints favour fuller NPs supporting feature-checking with PF over the Salience list (RECOVERABILITY), and sequences where grammatical salience in PF_i is aligned with LF_{i-1} (LINEARITY). The "ultimate preference" of the LINEARITY constraints is "for the orders of entities in SaL_{i-1} and Sal_i to align with each other" (op. cit.:41), where SaL_{i-1} and Sal_i represent S's model of the common ground before and after utterance U_i. It is not clear to me how ROT would handle example (2) above, since both LINEARITY and RECOVERABILITY favour interpreting *The man* as *Bob*, which would leave *him* to be interpreted as *Tom*, contrary to the desired reading. Space forbids a fuller discussion of this account in this short paper, but we may note that in common with other approaches we have discussed it only models benefits for H, disregards planning costs for S and the cost to H of maintaining a discourse model, and assumes a common, grammatical salience ranking is available to S and H.

4 A Framework for Centering Games

Kibble (2003) characterises reference resolution as a (timed) game where both players benefit in terms of efficient communication if H correctly guesses which item will be referred to next, or the intended reference of attenuated expressions (anaphors). So it is in S's interest to produce sequences which are as *predictable* as possible and it is in H's interest to learn what patterns S is producing *through repeated plays of the game*. According to Shannon's information theory, applied to NL semantics by e.g. Dretske (1981), predictability/high probability correlate inversely with informativity; so A's task is to *minimise* the information generated by each referring event. In a random sequence each item has equal probability. The more organised the sequence, the more items have better than chance predictability. Increasing levels of optimisation require more planning effort and memory allocation for S, while H can only take

advantage of this if a record is kept of previous utterances (e.g., the previous clause or the entire discourse).

We now sketch a framework for collaborative reference resolution as a non-cooperative game of incomplete information[2], taking account of the considerations outlined above. *Non-cooperative* need not imply competition or conflict, rather that players do not overtly coordinate their strategies, and players have *incomplete information* as they do not know what strategies other players follow, perhaps even after the event. Players seek to maximise utility and minimise costs, choosing their strategies on the basis of expected utility and expectations about other players' strategies. To capture the fact that not only do S and H have conflicting preferences, but that each player needs to balance competing costs and utilities, we model dialogue participants as *multi-agent systems*, and dialogue as a multi-player game involving speaker agents text-planner **P**, RE generator **REG** and hearer agents Reference Resolver **RR** and Discourse Modeller **DM**. An optimal outcome will then be a "Nash equilibrium" (Heap and Varoufakis, 1995) where no player can improve their payoff by unilaterally changing strategy, but all references are successfully resolved.

The processing modules are schematically described below, with very selective lists of possible strategies listed in ascending (partial) order of processing effort and/or memory costs.

Speaker perspective:

1. Planner/Content Determination (P): responsible for organising input propositions into a text structure, which may already be partially ordered according to coherence relations; plan sentences by e.g. choosing verb forms to realise preferred order of arguments. Increased planning effort will increase objective *predictability* of referring events.

Random: Do nothing, resulting in a random order of propositions and arguments within clauses.

Salient-Arg: Promote arguments within a clause according to perceptual salience.

Align: Plan consecutive clauses to align salience rankings.

Global: Plan sequence of clauses to maximise referential continuity, in addition to salience alignment.

2. Realiser (REG): generates appropriate referring expression to denote arguments of predicates.

Null: Do (almost) nothing: $\emptyset$ or personal pronoun

Short-NP: Reduced definite using a basic-level predicate, such as *the dog*

Full-NP: Uniquely identifying definite *the small white poodle in the kennel*

Hearer perspective:

1. Discourse Modeller: maintains a record of entities mentioned in the discourse which will be candidates for anaphora resolution:

$\emptyset$: no record of discourse referents. Only explicit definites will be resolvable.

1-Cl Keep a one-clause buffer of focal referents

Center Model: retain salience-ranked list from previous utterance U_{n-1} marking which was most salient in U_{n-2}; propose default referent for Subj(U_{n+1})

Full DM: fully-structured discourse model including salience-ranked history list, segmented to reflect discourse structure.

2. Reference Resolver: responsible for identifying referent of REs with an entity in the domain.

Default: Resolve to referent predicted by **DM** (if available)

Search-DM: Search the discourse model constructed by **DM**

Search-Dom: Search the domain if no candidate found in the discourse model

Repair: Recovery strategies: reread the text; internally "replay" speech or query the speaker.

It should be clear from the preceding discussion that the more effort **P** expends in constructing coherent plans, the more opportunities there will be for **REG** to use reduced REs, and that **RR**'s success in interpreting these with least effort will depend on the work done by **DM**. We can visualise these modules as a set of interlocking cogwheels: as **P** and **DM** turn in one direction, **REG** and **RR** will turn in the other.

[2] A more detailed account with worked examples can be found in (Kibble, 2003; Kibble, fthc).

Anaphora in the Centering Game

Following Shannon, the quantity of *information generated* by an event (or *surprisal*) is a function of its *probability* in context: $I(e) = \log_2 1/p(e)$. So for instance if all events $e \in E$ are equally likely then for any e, $I(e) = \log_2 1/p(e) = \log_2 \mid E \mid$, or the number of binary decisions needed to reduce E to $\{e\}$. Consider the schematic example *A B C; A C; ? D E*, where A, B, etc are distinct referents:

$$\text{Assume: } p(?=A) > p(?=C) > p(?=B) >$$

$$p(?=D) = p(?=E);$$

$$\text{then } I(A) < I(C) < I(B) \ldots$$

So equating ? to A will have the smallest "inferential cost" for H. S saves effort on Referring Expression Generation (**REG**) by planning sequences to *minimise* information generated by referential events, allowing for use of non-informative or attenuated expressions (pronouns, null string). Resolving anaphors requires **Hearer** to perform non-monotonic inference. The preferred case is where this involves the smallest possible quantity of information. The optimal state is where a referent d is realised as an anaphor in clause C only if there is no other d' s.t. $I(d') \leq I(d)$ and d' is not also realised as an anaphor in C: in other words the "lightest" RE must pick out the most predictable referent. This is quite similar to Centering Theory's Rule 1. And CT's Rule 2 can be reconstructed as follows: assuming that S and H converge on a scheme of text organisation which maximises predictability of reference, we may expect a grammatically-based salience ordering to emerge along with a preference for continuity of topics. It would follow that the rules and constraints stipulated by CT can in principle be eliminated in favour of general principles of rational communication, as we surmised earlier, though we may retain the terminology of CT for descriptive convenience.

We can make this claim more precise: the interaction of the modules listed above defines a 4-dimensional solution space, which is partially mapped out in Figure 3. Each point marked $\mathbf{O}_n$ identifies a partial equilibrium (solution) where

Planning/ Modelling	**Modeller**	1 ∅	2 1-Cl	3 CM
Planner	1. Random	$\mathbf{O}_1$		
	2. Align		$\mathbf{O}_2$	
	3. Global			$\mathbf{O}_3$

Planning/ RE Gen	**Realiser**	1 Null	2 Short-NP	3 Full NP
Planner	1. Random			$\mathbf{O}_4$
	2. Align		$\mathbf{O}_5$	
	3. Global	$\mathbf{O}_6$		

Modelling/ Resolution	**Resolver**	1 Def	2 S-DM	3 S-Dom
Modeller	1 ∅			$\mathbf{O}_7$
	2 1-Cl		$\mathbf{O}_8$	
	3 CM	$\mathbf{O}_9$		

RE Gen Resolution	**Resolver**	1 Def	2 S-DM	3 S-Dom
Realiser	1 Null	$\mathbf{O}_{10}$		
	2 Short-NP		$\mathbf{O}_{11}$	
	3 Full-NP			$\mathbf{O}_{12}$

Figure 3: Aspects of a Referring Game

neither player can unilaterally increase their utility by moving along the relevant axis. In terms of coordinates $\langle P,\ REG,\ DM,\ RR \rangle$, Centering Theory in its classic form assumes that discourse operates at a point close to $\langle 3, 1, 3, 1 \rangle$: speakers and hearers are assumed to be ideally rational cooperative agents who put the maximum effort into high-level tasks of planning and modelling so they they can each minimise their efforts on the low-level tasks of realisation and reference resolution. A dialogue involving uncooperative speakers, who do not plan ahead or align the order of nominal arguments with salience in the common ground, is represented by $\langle 1, 3, 1, 3 \rangle$. An intermediate state where speakers and hearers are prepared to be cooperative but have conflicting claims on their time and attention (*bounded rationality*) might fall somewhere near $\langle 2, 2, 2, 2 \rangle$.

There are two ways of looking at this model.

One way is to see it as a hypothesis about how two rational agents might co-evolve a set of strategies which enable them to communicate efficiently, over repeated plays of a communication game. For instance Agent A might start out by using grammatical salience to minimise the randomness of his referring actions and Agent B would try different salience rankings until she finds one that provides the most reliable predictions about what A will refer to next. Agent B may notice that A is using pronouns and reduced definite NPs, so it is worth her while to keep a record of what has been referred to in order to resolve anaphors. And so on. On the other hand we could see the various solutions as *conventions* in the sense of (Lewis, 1969). A strategy which has successfully solved a coordination problem gives rise to expectations that this strategy will be followed in future and thus becomes a precedent; A and B will follow the strategies they have co-evolved when talking to C rather than beginning a new process of mutual adaptation.

5 Some Empirical Issues

The above reconstruction of CT rests on the assumption that S and H will converge on a salience ranking corresponding to grammatical obliqueness; however this assumption is challenged by results discussed by Oberlander (1998) suggesting that S and H have distinct preferences, with S following a grammatical hierarchy when choosing the topic of consecutive utterances while H seems influenced by thematic roles in forming expectations of upcoming referring events.

(6) a. John gave the book to Bill so he$_{John}$ ...

b. John gave the book to Bill so John ...

c. John gave the book to Bill so he$_{Bill}$ ...

d. John gave the book to Bill so Bill ...

In continuation experiments, speakers prefer type (a) when talking about John and type (d) when talking about Bill, confirming CT's predictions. But in reading-time experiments, type (c) sentences were processed faster than type (a), indicating an effect of *thematic roles*. Since it has so far proved difficult to reliably identify thematic roles in naturally occurring text, the type of evidence that can be brought to bear on this question is rather limited and so I will regard "Oberlander's paradox" as a puzzle for CT rather than a stumbling block.

Assuming the mismatch is real, it points to a deeper assumption: that S and H will both arrive at the same subjective probability assessment of the distribution of referents in a discourse and that both will agree with the *objective* probability which could be determined by corpus analysis. Reasons why S and H may differ in their assessments are firstly that owing to resource constraints, each may only consider a small part of the discourse history and secondly, they may operate at different levels of representation. For instance: if P plans the sequence of referents in U_i according to their salience in the previously generated utterance U_{i-1}, the *syntactic* structure of U_{i-1} is likely to be more accessible to P than its semantic representation. However when DM calculates the expected referents in U_i according to salience in the previously interpreted utterance U_{i-1} it is more likely that the *semantic* representation of U_{i-1} will be retained and salience for H will depend on thematic rather than grammatical roles. We assume therefore that it is more costly for S to retrieve the underlying semantics of the immediately preceding utterance, and for H to reconstruct the syntactic structure. These operations, though costly, are available for repair strategies in case of failure of anaphora resolution. This hypothesis is consistent with the differing preferences of S and H noted by Oberlander (1998), while its formulation relies on the game-theory framework we propose. The question remains however, why H would continue to operate with a salience ranking which seems orthogonal to the ranking actually used by S. What follows is a conjecture: it may be reasonable to assume that H learns a thematically-based salience ranking as long as there is a significant positive correlation between the distributions of grammatical and thematic roles, so that predictions based on thematic roles have a better than chance probability of being correct. Again, evaluation of this conjecture will require progress in the reliable annotation of thematic roles in text.

There is also an assumption underlying CT that S and H share a model of the common ground: S will only make grammatically salient reference to entities which are known to be salient to H. This assumption is undermined by the findings of (Branigan et al., 2003) who conducted experiments using task-based dialogues between pairs of subjects, where the knowledge states of the speaker and addressee were manipulated independently. They found that speakers tended to reproduce a Given-New ordering in their utterances, but only with respect to the speaker's knowledge: items which were New to the hearer would be referenced more saliently than hearer-Given items if they were Given for the speaker. Branigan et al. are cautious about drawing more general lessons from these experiments owing to the relatively simple nature of the hearer's task. However there are two points worth considering in relation to the model outlined above: firstly "salience" has to take account of the visual field and other perceptual sources as well as the discourse history; secondly, this is further evidence that speakers and hearers can have misaligned internal models, and yet communication proceeds. The framework I have outlined allows these mismatches to be modelled, and the study of these kinds of asymmetry may turn out to be more revealing in the long run in the study of human communication than adherence to theories which assume perfect alignment.

6 Acknowledgements

This work has been presented, in various stages of development, at the ITRI Seminar, University of Brighton; the 5th International Workshop on Computational Semantics, University of Tilburg; and the ICCS Seminar, University of Edinburgh. I'm grateful to the participants in these events for much useful feedback.

References

David Beaver. 2003. The optimization of discourse anaphora. *Linguistics and Philosophy*. To appear.

R. Blutner. 2000. Some Aspects of Optimality in Natural Language Interpretation. *Journal of Semantics*, 17:189–216.

Holly Branigan, Janet McLean, and Hannah Reeve. 2003. Something Old, Something New: Addressee Knowledge and the Given-New Contract. In *Proceedings of CogSci 2003*, Boston.

Susan Brennan, Marilyn Walker Friedman, and Carl Pollard. 1987. A centering approach to pronouns. In *Proceedings of 25th ACL*.

Adam Buchwald, Oren Schwartz, Amanda Seidl, and Paul Smolensky. 2002. Recoverability Optimality Theory: Discourse Anaphora in a Bi-directional Framework. In *EDILOG 2002: Proceedings of the sixth workshop on the semantics and pragmatics of dialogue*.

Fred Dretske. 1981. *Knowledge and the Flow of Information*. MIT Press. Reissued by CSLI Publications, Stanford, 1999.

Barbara Grosz, Aravind Joshi, and Scott Weinstein. 1995. Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

K. Hasida, K. Nagao, and T. Miyata. 1995. A game-theoretic account of collaboration on communication. In *Proceedings of the First International Conference on Multi-Agent Systems*.

Shaun P. Hargreaves Heap and Yanis Varoufakis. 1995. *Game Theory: A Critical Introduction*. Routledge, London and New York.

Rodger Kibble and Richard Power. 2000. An integrated framework for text planning and pronominalisation. In *Proceedings 1st International Natural Language Generation Conference*, Mitzpe Ramon, Israel.

Rodger Kibble. 2003. Both sides now: Predictive reference resolution in generation and interpretation. In *Proceedings of the 5th International Workshop on Computational Semantics*, Tilburg.

Rodger Kibble. fthc. Centering Games. MS, in preparation. Department of Computing, Goldsmiths College, University of London.

David Lewis. 1969. *Convention: A Philosophical Study*. Blackwell, Oxford.

John J. McCarthy. 2002. *A Thematic Guide to Optimality Theory*. Cambridge.

John Oberlander. 1998. Do the right thing. *Computational Linguistics*, 24(3):501–507.

Robert van Rooy. 2003. Relevance and Bidirectional OT. In R. Blutner and H. Zeevat, editors, *Pragmatics in Optimality Theory*.

Analysing Bids in Dialogue Macrogame Theory using Discourse Obligations

Jörn Kreutel
SemanticEdge, Berlin
joern.kreutel@semanticedge.com

William C. Mann
SIL
bill_mann@sil.org

Abstract

We focus on a class of utterances that can be analysed as 'explicit game bids' in Dialogue Macrogame Theory (DMT). We propose an update-semantics analysis that uses discourse obligations to model the observable behaviour of dialogue participants and that is able to generate the intentional structure of interactions assumed in DMT. We also claim that our analysis provides the base for reformulating some key aspects of ILLOCUTIONARY FORCE in terms of BIDDING.

1 Overview

In recent years, research in the field of dialogue theory has been concerned with providing formal analyses of a wide range of phenomena. These range from elaborating the basic functioning of questions and answers in dialogue (see (Ginzburg, 1997), (Asher and Lascarides, 1998), (Larsson, 1998), (Bos and Gabsdil, 2000)) to investigations of indirect language use (e.g. (Asher and Lascarides, 2001)). Among the analytic frameworks, two have been highly influential. They are the extension of discourse representation theory known as SDRT (Asher and Lascarides, 2003), and the INFORMATION STATE UPDATE APPROACH proposed by the members of the TRINDI consortium (Traum et al., 1999). Both approaches can be seen as employing a notion of SPEECH ACT (see (Austin, 1962), (Searle, 1969)) for analysing the contributions of dialogue participants (DPs). SDRT focuses on the fact that speech acts in a dialogue context can be seen as expressing RHETORICAL RELATIONS. The TRINDI approach, on the other hand, decomposes speech acts into DIALOGUE ACTS of a higher granularity that may operate on various levels of information structure (Poesio and Traum, 1997).

Even though both frameworks have worked out an abundant repertoire of rhetorical relations and an extensive taxonomy of dialogue acts (Poesio and Traum, 1998), respectively, there is a whole class of utterances which has not been given sufficient attention by the respective theories – in spite of the fact that their analyses allow for insightful generalisations. This class is exemplified by the initial utterances in the following interactions:[1]

(1)
I[1]: Do you know what happened at last night's party?
R[2]: Sorry, I am not interested in gossip.

(2)
I[1]: Have a recommended roll rate for this PTC, if you could copy.
R[2]: All right. Go ahead.

(3)
I[1]: I have a question about the patient in room 101.
R[2]: Please ask the doctor.

(4)
I[1]: May I suggest something to you for solving your money problems?
R[2]: Ok. What do you think?

What is common to these dialogue fragments is that the utterances I[1] in (1)-(4), rather than being 'full-fledged' assertions, questions or suggestions, actually are 'preparatory moves' for these kinds of speech acts. What they *do* – in Austin's terms – is express a *proposal* directed at their addressees to engage in a (sub)dialogue that deals with an assertion, question, or suggestion, the content of which is left partially unspecified. Whether the content will be presented – and whether thus the actual speech act that was meant to be prepared for will be performed or not – will depend on the acceptance or rejection of the proposal on the part of the addressee.

On the other hand, once a proposal has been accepted, both the initiator of the proposal and the addressee will be committed to act accordingly, e.g. in

[1] see http://www-rcf.usc.edu/~billmann/dialogue/ap13d2-view.pdf for (2)

the case of (4) to make a suggestion and to respond to it, respectively. Hence, the following dialogues do not constitute consistent continuations of (4):[2]

(4′)
I[1]: May I suggest something to you for solving your money problems?
R[2]: Ok. What do you think?
I[3]: Sorry, let's talk about something else.

(4″)
I[1]: May I suggest something to you for solving your money problems?
R[2]: Ok. What do you think?
I[3]: How about a smaller flat?
R[4]: Sorry, I don't want to talk about that stuff now.

In this paper, we will propose an analysis of these examples along the lines of DIALOGUE MACROGAME THEORY (DMT, see (Mann, 2002)), in which the initial utterances of the fragments can be classified as 'explicit game bids'. We will provide an account of the DMT notion of BID in an update-semantics framework that employs DISCOURSE OBLIGATIONS (Traum and Allen, 1994) to express predictions and constraints with respect to the expectable behaviour of dialogue participants at each stage of an interaction. This way, we will show that instances of the interactional patterns ('macrogames') that DMT defines in terms of intentional structure can be seen as being the result of a successfully completed bidding process. On the other hand, we will argue that our analysis provides an alternative perspective on the class of examples that is most prominent in speech act related research, i.e. assertions, questions, requests and other types of speech or dialogue acts. We will see below that if we view these actions as bids, we may be able to refine the widely used idea of ILLOCUTIONARY FORCE (see (Austin, 1962)) in a way close to the one proposed in (Habermas, 1981).

2 Explicit Game Bids

Here, we will provide an analysis of the initial utterances in the above examples as explicit game bids in DMT. We will first give a very brief overview of DMT's basic assumptions and will then show how bidding and its effects can be modelled by means of information state update rules using discourse obligations.

2.1 Dialogue Macrogame Theory

Dialogue Macrogame Theory (Mann, 2002)[3] is a descriptive framework for dialogue analysis and annotation. In contrast to other approaches to dialogue that involve the notion of 'game'[4], be it 'language game' or 'conversational game' (e.g. (Eklundh, 1983), (Levinson, 1983), (Lewin, 1998)), DMT does not make any prescriptions with respect to how the MOVEs that constitute a game are actually realised and whether there are 'preferred' or 'dispreferred' sequences of moves for certain games. Instead, DMT defines different types of MACROGAMES exclusively in terms of conditions on the participants' intentions. DMT assumes that each macrogame can be defined by a set of three intentions that consists of an intention of the initiator of the game, an intention of the responder[5] and a joint intention. DMT regards macrogames as conventions of the language of the dialogue, and therefore allows the possibility of cultural variations. As two of the original DMT game definitions in figure 1 show, a game[6] consists of a TYPE, e.g. IS, IR, etc., and a content , the PARAMETER, which is referred to with expressions like 'the information'.

DMT further assumes that both the initialisation and the termination of a macrogame in an actual conversation can be described – metaphorically – in terms of a 'negotiation' process that involves a BID on the part of one DP[7] and a POSITION TAKING on the bid on the part of its responder, which may result in either an ACCEPTANCE or a REJECTION of the bid. As far as game initialisation is concerned, the result of an acceptance of a game bid is, in DMT's terms, the COMMITMENT by each DP to their assigned portions of the intentions that define the game that has been bid.[8] DMT thus provides the distinctions needed to model the consistency of a sequence of utterances in an interaction by means of the DPs' commitments to the intentions that are expressed and accepted. Hence the example (4″) above would be classified as inconsistent because of a mismatch between the intention to which R is committed due to their acceptance of I's game bid in R[2] and their behaviour in R[4].

[2] An inconsistency of this kind may, however, not result in a complete breakdown of the interaction, but will rather trigger repair.

[3] see also http://www-rcf.usc.edu/~billmann/dialogue/dtsite.htm

[4] see (Wittgenstein, 1953) for the origins of the use of this term in language related research

[5] We will use the DMT term 'responder' rather than 'addressee' in the remainder of this paper

[6] we use 'game' as a synonym for 'macrogame'

[7] As far as bids for initialising a game are concerned, the bidder will be the initiator from the game definitions. Bids for termination can be made by any of the DPs.

[8] In addition to macrogames, DMT introduces the notion of UNILATERALS, which are actions on the part of one DP that do not involve collaboration or joint goals, e.g. acknowledgements, exclamations and several kinds of politeness acts. Note that recognition of whether a bid of a macrogame has been made or whether an action was actually a unilateral is sometimes complex. This paper uses simplifying assumptions that all bids are recognised as bids, that all bids are correctly identified with macrogames, and that no unilaterals are present.

Name of Game	Joint Intention	Intention of Initiator (I)	Intention of Responder(R)
Information Seeking (IS)	I knows the information that is sought	I identifies the information that is sought	R provides the information that is sought
Information Offering (IO)	R comes to know the information that is offered	I provides the information that is offered	R identifies the information that I has provided

Figure 1: examples of macrogames

It is evident that the DMT notion of BIDDING provides quite an intuitive account of the examples referred to in the introduction, particularly because it allows us to classify the initial utterances of the dialogue fragments (1)-(4) as EXPLICIT GAME BIDS, which are a subclass of PARTIAL GAME BIDS, i.e. bids that leave either the type of game or its content (the 'parameter') partially unspecified. In DMT, however, all details about the bidding process are left informal, in particular the issue of how to model the notion of the DPs' commitment to an intention and the way it is brought about. The following sections will therefore provide an account of DMT bidding that draws on – and at the same time further refines – established ideas about the functioning of DISCOURSE OBLIGATIONS in dialogue. It further uses Raimo Tuomela's model of JOINT INTENTION FORMATION (Tuomela, 1998).

2.2 Discourse Obligations

The idea to use discourse obligations in dialogue modelling was first raised by David Traum and James Allen in (Traum and Allen, 1994) and was meant to provide an intuitive analysis of examples like the following ones:

(5)
I[1]: Did Pete drive here?
R[2a]: I don't know.
R[2b]: I don't want to talk about that.

Here, a purely intention-based account of interactions cannot explain why dialogue participants respond in situations where they are *unwilling* or *unable* to adopt the speaker's intentions. An obligation-driven approach, in contrast, predicts R's behaviour in R[2a/b] appropriately by assuming that participants in a dialogue are socially *obliged* to respond in some manner, even if their own intentions are not advanced by responding. This does not mean, however, that intentions are discarded as action triggers. What is modified in contrast to standard intention-based approaches is only the process by which an intention is brought about. Instead of assuming that a responder's intention is created by cooperative adoption of an equivalent intention of a speaker, an obligation-driven model makes the following assumption:

OBLS-INTS

An *intention to respond in a particular way* is the result of trading off the fact that there is an *obligation to respond* against one's private intentions, goals, preferences and capabilities. What kind of response will be given in the end clearly depends on this trade-off.

Based on this starting point given in (Traum and Allen, 1994) and subsequent work (e.g. (Poesio and Traum, 1997)) the following sections will provide an analysis of explicit bids that assumes obligations at various levels of information structure in dialogue.

2.3 An obligation-based Model for Explicit Bids

Here, we will provide an analysis of explicit game bids that is based on the use of discourse obligations in a simple information state update model. We will first give an informal outline of how DMT's assumptions about the intentions of I and R can be expressed in this framework and will then discuss the issue of modelling the joint intention distinctive for the macrogames.

2.3.1 Obligations and the Intentions of I and R

As for the effect an explicit game bid has at a given stage of an interaction, we assume, first of all, that the fact that a game bid has been made can be decomposed in the following way:

UPDATE.BASE

1. I has bid a game γ of a certain type
2. I has partially specified the content[9] of γ
3. R is obliged to take a position on I's bidding γ

A bid thus results in both I and R changing their view of the discourse situation such that it records the above events. Leaving aside problems of misunderstanding at any level of linguistic structure, we assume that 1.-3. above can be considered part of the COMMON GROUND of the participants. The leftmost discourse representation structure (DRS)[10] in Figure 2 shows the effect of applying **UPDATE.BASE** for I[1] in example

[9] As this paper provides an 'outsider view' on DMT, we keep on using common terms like 'content' instead of the DMT terms like 'parameter'. In a longer version of the paper, which will be available on a website, we will further define these and other key terms and clarify their scope compared with other people's use of the same or related terms.

[10] see (Kamp and Reyle, 1993) for the basics of DISCOURSE REPRESENTATION THEORY

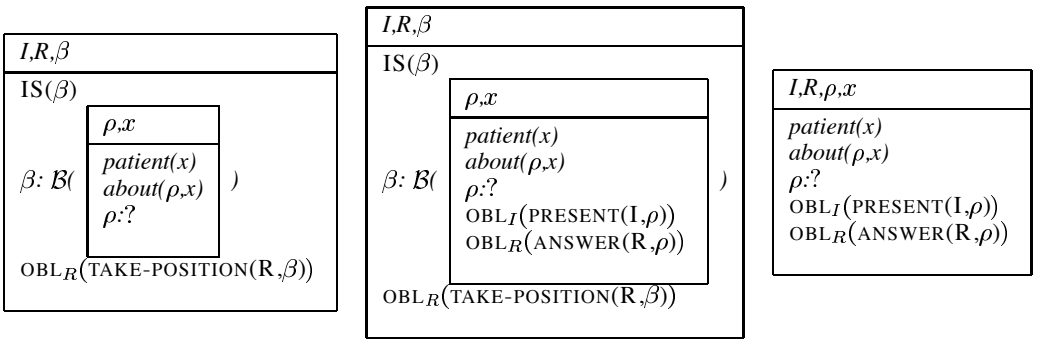


Figure 2: Effect of the Update Rules for Example 3

(3) above,[11] where the root DRS stands for the common ground and the embedded DRS (to which we will refer as K_β) stands for the DRS that holds the actual content of the bid. Note that whereas the fact that a bid has been made belongs to the DPs' common ground, K_β, which expresses *what is actually being bid*, is not grounded, but is in the scope of a modal operator, represented here as $\mathcal{B}$. Adhering to proposals in (Kreutel and Matheson, 2002), we assume that this modality can, in a first approximation, be defined as the initiator's belief that the bid may be accepted by the responder.[12]

Given the knowledge about the type of game that has been bid, the identification of the initiator and the responder, and finally about the (partial) content of the bid, it is possible to introduce obligations via applying a further set of update rules that result in the content of the bid being extended with:

UPDATE.OBLS

1. An obligation on the initiator to specify the content of the game enough so that it can be pursued by R and I.[13]
2. An obligation on the responder that specifies a response specific to the type of game that has been bid

We propose to model the obligation on the initiator as the obligation to actually *present* the content the game shall deal with. As far as the obligations on the responder (R) is concerned, we assume that for IO and IS they can be spelled out as the obligations to ADDRESS the assertive content of the information offer (e.g. by questioning or accepting the content), and to provide an ANSWER that resolves the content of the request for information, respectively.[14] As the obligations will be added to K_β, they will not be considered as belonging to the DPs' common ground, hence they will be pending until the bid has been accepted. In figure 2, the DRS in the middle shows the result of applying **UPDATE.OBLS** on top of the initial state, given these assumptions about the obligations associated with the IS macrogame.

The effect of accepting a bid can, finally, be summarised as follows and is represented by the rightmost DRS in figure 2 for a presumed positive response of R, e.g. *ok*, to the initial utterance I[1] in the example dialogue (3):[15]

ACCEPT-BID

Acceptance of a bid results in the content of the bid[16] being added to the common ground of the DPs.

Having the content of a bid, as proposed, actually include the obligations on I and R that are specific to the game that has been bid provides an intuitive account of the fact that position taking on the bid involves considering the consequences that may result from an acceptance of it on the part of R. From the judgement on R's willingness, capacity, and – as in the case of (3) their being entitled to satisfy the obligation – may then result a positive or negative response. On the other hand, as far as R's intention according to the original DMT definitions is concerned, it can be assumed to be triggered by the game's specific obligation on R following a positive outcome of trading-off the obligation against

[11] At this stage of our project, we will not commit to any of the established update-semantics frameworks for dialogue – be it SDRT (Asher and Lascarides, 2003) or the TRINDI framework (Poesio and Traum, 1997) – and will therefore not provide a spelled out truth conditional semantics for the employed constructs. Nevertheless, as we make use of established expressive means, we assume that a full formalisation will in principle be possible in either of the models mentioned.

[12] Note further that K_β contains a CATAPHORIC discourse referent, ρ, which cannot be resolved at the depicted states in the interaction, but which requires a continuation of the dialogue comprising I's specification of what it exactly is that shall be subject of the information seeking game.

[13] hence to provide a referent for the cataphoric discourse referents in the example structures

[14] The other games can be spelled out in comparable ways.

[15] the MERGING of the content of K_β with the one of the root DRS could be formally modelled following the proposals in (Poesio and Traum, 1997)

[16] i.e. K_β in the example analysis in figure 2

their private mental state, as it is assumed in the standard obligation-driven approach according to the rule **OBLS-INTS** above. In the same way we can assume that the initiator's intention to perform their part in a game is triggered by their knowledge of I's obligations associated with that game. One can assume that intention formation for I will normally take place before I starts the bidding process. However, I will be *obliged* to act accordingly only after the bid has been accepted by R.

Note that this analysis of explicit game bids instantiates the update rule proposed for OFFER in (Poesio and Traum, 1998), in so far as the latter introduce a conditional that formulates the dependency of obligation introduction on the responder's acceptance of the offer.[17] Poesio and Traum, however, describe offering in the generic sense of 'offering an action' and only assume an obligation on the initiator of the offer. Our notion of game bid, in contrast, models the specific 'offer of a dialogue act', the update effect of which comprises not only a further obligation on the responder, but also results – as we will outline in the following section – in a joint obligation on both DPs.

2.3.2 Obligations and the Joint Intention

In the previous section, we reformulated DMT's assumptions about the private intentions of the initiator and the responder of a game drawing upon the EVOCATIVE[18] effect of a dialogue move, in which the raising of an obligation can be seen. Here, DMT's informal notion of a DP's 'commitment to an intention' could be modelled in terms of the DP being obliged to perform an action (or a course of actions) that will result in the satisfaction of the respective intention. As far as the formation of the joint intention associated with a game and the corresponding obligations on I and R are concerned, on the other hand, we will propose an analysis that is based on the EXPRESSIVE content of an utterance and that instantiates the generic model of joint intention formation developed by Raimo Tuomela (Tuomela, 1998); such joint intentions strongly resemble the ones formulated so far for macrogames.

Tuomela's model follows a 'bulletin board view' of the intention formation process, which is seen to be triggered by an initiator's public proposal to engage in joint action. Whether both DPs become committed to the appropriate joint intention will then depend on sufficient uptake on the part of the addressees of this proposal and on all participants' knowledge about each other's uptake. Tuomela remarks that this knowledge is based on 'communication', hence it is the initiator's and the responder's effectively conveying their willingness to engage in the joint action that is the base for assuming that their actions will actually be guided by a joint intention. On the other hand, from the public expression of one's agreement with a proposal for joint action there will further result the obligation on the participants to execute the joined action agreed upon, once the intention formation process has been successfully completed.

In the update framework outlined in the previous section, Tuomela's model of intention formation can be instantiated by assuming that

JOINT-INTS

1. Game bids express the initiator's intention to participate in a joint action that is meant to bring about the accomplishment of the 'joint goal' associated with a game in DMT, which may be an action, outcome, or plan
2. The acceptance of a bid by a responder R expresses the latter's intention to participate in that joint action
3. Given R's acceptance of a bid, we can infer that a joint intention exists to perform the joint action proposed in 1. And that
4. There is an obligation on I and R to adhere to the content of that joint intention

Note that the joint obligations introduced by **UPDATE.OBLS**, on the one hand, and the 'individual' ones inferred via **JOINT-INTS** as outlined in the previous section, on the other hand, can be seen as expressing different levels of COOPERATION: hypothetically, the DPs may act in a way that is compliant with the individual obligations on I and R according to the game's definitions and yet not comply with the joint obligation – even though in bidding and accepting a bid they *pretend* to do so.[19] Under this assumption, whereas only the DPs' adherence to the joint goal of a macrogame can be seen as 'full cooperation', their mere fulfilling the individual obligations of acting and responding to each other in a certain way is characterised by a lower level of cooperation. Here, even though the conversation may be driven forward in a regular and consistent manner, it may eventually not result in a successful macrogame.

We consider it as a strength of the obligation-based approach that it provides a unique expressive means in

[17] (McRoy and Hirst, 1995), on the other hand, require the assumption of additional 'linguistic expectation rules for adjacency pairs' in order to motivate the actual 'telling' that normally follows a 'pre-telling' action like the ones in (1)-(4). With respect to this, the attractiveness of using obligations lies particularly in the fact that they can be intuitively associated – in our case – with the act of bidding and do not require the assumption of further expressive means to reconstruct the rule-governed behaviour of DPs.

[18] our use of the terms EVOCATIVE and EXPRESSIVE follows the proposals in (Allwood, 1995)

[19] This case shows that our framework, apart from allowing discrete degrees of cooperation, is able to touch on SINCERITY, a connection which we will not develop here.

terms of which varying degrees of cooperation, which have been described, e.g., in (Allwood et al., 2000), can be formulated. Following proposals in (Kreutel and Matheson, 2001), we further see the possibility to apply our model in an analysis of Habermas' (Habermas, 1981) concept of STRATEGIC ACTION, which can be seen as an instance of an interaction type that is characterised by a mismatch between an overt cooperation in actions and responses, on the one hand, and an underlying conflict of the DPs' intentions, on the other.

3 Towards a Generalisation: Bids and 'Speech Act Offers'

In the previous sections, we have demonstrated that the structure of interactions described by Dialogue Macrogame Theory can be formulated using update rules for information states and discourse obligations – both individual and shared ones – as expressive means. With respect to the linguistic data, our model provides an intuitive account of the class of utterances that we called 'explicit game bids' and that were exemplified by the examples (1)-(4) in the introduction. In particular, our use of obligations predicts why cases like those in (4′) and (4″) are inconsistent continuations of a dialogue initiated by an explicit bid.[20] As far as the scope of our model of bidding is concerned, we claim that our analysis covers not only explicit game bids like the ones discussed, but also the possibly much more frequent cases of utterances, for which I[1] in the following dialogue provides an example:

	I[1]:	Has the patient in room 101 recovered
(3′)		from the surgery?
	R[2]:	Please ask the doctor.

These cases are actually the ones which traditionally have been constituting the scope of speech act theory and subsequent dialogue research, and which have been investigated with respect to what a DP *does* when uttering something like I[1] above. Given our analyses in this paper, it seems appropriate to us is to assume that I[1] in (3′) expresses a bid directed at R to engage in an IS macrogame. At the same time, I provides the content of the game which in the case of an explicit bid as in (3) was said to be partially unspecified or not specified at all. However, both as far as the obligations on R are concerned and with respect to its consequences for the process of joint intention formation the update effect of I[1] above will be equivalent to the update effect of an explicit game bid.

[20] Given a modest effort, we believe that this development could be extended to the other five game management acts of DMT, i.e. ACCEPT-BID, REJECT-BID, BID-TERMINATION, ACCEPT-TERNINATION and REJECT-TERMINATION.

Given these findings, we see the possibility to generalise our analysis of explicit game bids towards a coverage of other types of speech acts or dialogue acts. From this extended perspective, bidding can be analysed as constituting a major part of what has been described as the ILLOCUTIONARY FORCE of those acts: the act of ASKing, for example, can be seen as the act of bidding the opening of a (sub)dialogue – or macrogame – in which the DPs deal with what is actually being asked. However, rather than being responded to with an answer, in a given context – e.g. as in (3′) above – an ASK act may equally elicit a response that makes the action of asking appear problematic in this context. It is the notion of bidding that provides the background for seeing dialogue acts, hence, as actions of a speaker directed at a responder who is required to take position on what the speaker is doing in order for the action to have an effect on the continuation of the dialogue.

This particular characteristic of actions in dialogue has been described by Jürgen Habermas(1981) as the raising of VALIDITY CLAIMS that need to be evaluated by the responder with respect to the three dimensions of propositional TRUTH, expressive TRUTHFULNESS[21] and normative RIGHTNESS. In Habermas' terms, an action like I[1] in the above example can be conceived of as a SPEECH ACT OFFER which will be either accepted or turned down by the responder, where a rejection may follow (at least) one of these dimensions for argumentation, as exemplified by R[2a]-R[2c]:

	I[1]:	Has the patient in room 101 recovered from the surgery?
(3″)	R[2a]:	He hasn't had any surgery so far.
	R[2b]:	Haven't you just seen him?
	R[2c]:	Please ask the doctor.

As these dialogues show, the claims for truth and truthfulness include the claim of satisfaction of semantic presuppositions and the claim that I's primary interest is finding out something about the state of health of some patient, respectively. R[2a] and R[2b] then demonstrate how these claims can be rejected, given an appropriate context. R[2c], on the other hand, provides an example for a rejection of the rightness claim involved in the performance of I[1]. In the proposed formalisation of DMT bidding outlined in the previous sections of this paper, a response like R[2c] can be modelled by assuming that R will reason over the obligations on I and R that would result from an acceptance of I[1]. Here, R would be obliged to provide an answer to I's question, which may result in a conflict with the rights and obligations R actually has towards I in the given circumstances, hence with R's and I's social ROLES. In consequence, R's reasoning provides

[21] which resembles SINCERITY mentioned in a footnote in section 2.3.2

the grounds for rejecting I[1] by questioning the rightness of I's question.

Our analysis of (3″) is meant to demonstrate that Habermas' notion of speech act offers and the DMT concept of bidding are based on a highly similar view of the nature of actions in dialogue, the adequate description of which requires an interactive model involving an 'initiator' and a 'responder' – rather than an 'agent' and a 'patient'. In particular, a reformulation of bidding in terms of obligations can be considered a promising approach towards a more formal account of Habermas' intuitions about the rightness claim, which (Habermas, 1981) exemplifies using a dialogue initiated by a request.[22] We acknowledge that for an informal notion of 'rightness' the universality of this claim seems to be difficult to grasp,[23] particularly for actions like assertions or questions. Following our model of DMT, however, the introduction of specific obligations on the participants in a dialogue can be seen as a universal property of a wide range of actions[24]. On this base, 'rightness' can be formalised as a matching between the obligations associated with an action and the actual roles of the DPs. [25] How 'roles' need to be conceived of and how the nature of this matching needs to be thought of – e.g. whether it can be described along the lines of formal accounts of PRESUPPOSITION (e.g. (van der Sandt, 1992)) – is an important subject for further research.

4 Conclusions

In this paper, we have proposed a set of update rules that allow us to generate the structure of interactions described by Dialogue Macrogame Theory. A central role in our model is given to the use of discourse obligations, which are claimed to have a determinant effect on the DPs' acting in several respects. On the one hand, any game bid introduces an obligation on the responder to take position on the bid, hence to express either their acceptance or rejection of the bid. The content of a bid, on the other hand, can be seen as comprising further obligations on the initiator and responder of the game that depend on what type of game is being bid. Once a bid has been accepted, these obligations will be added to the common ground of the discourse participants and are prescriptive for their subsequent actions.[26] Any way of acting that is not compliant with these obligations will result in an inconsistent dialogue flow. Further, we have shown that a joint obligation on the DPs to contribute to the achievement of the goal of a macrogame can be introduced via an augmented variant of Tuomela's model of joint intention formation in our update framework. The different types of obligations could further be interpreted as being associated with different levels of cooperation in dialogue.

Finally, we have demonstrated that our account of bidding not only covers explicit game bids, exemplified by the dialogues (1)-(4), but can be generalised and applied to those instances of speech acts or dialogue acts which have formed a main focus of investigation, so far. We could argue that our notion of game bid and Habermas' concept of speech act offer can be seen as closely related to each other and that, hence, our formalisation of bidding could serve as a starting point for clarifying unresolved issues in Habermas' model of validity claims. The proposed analysis of bidding thus provides the base of a universal account that allows us to reformulate some key aspects of illocutionary force in terms of the framework provided by Dialogue Macrogame Theory.

[22] *Please bring me a glass of water* directed by a Professor towards one of their students

[23] (Goldkuhl, 2000), e.g., remarks that a definition of 'rightness' in terms of 'adherence to social norms' has difficulties in capturing other aspects like 'appropriateness', and he therefore proposes to extend Habermas' original repertoire of three validity claims.

[24] In DMT terms, all actions apart from unilaterals

[25] This amounts to seeing 'concordance to social norms' as a moral conformity to what a society regards as appropriate, clean, holy, noble, etc. Roles can be seen as the objectivation of these ideas at various levels of social organisation. It is, hence, conceivable that role-based and moral-based attributions of rightness use all and only the same axioms, and only differ in the starting points

[26] These obligations thus can be seen as giving rise to what in the framework of (Habermas, 1981) is called 'commitments relevant for the continuation of an interaction' which result from accepting a 'speech act offer'.

References

Jens Allwood, David Traum, and Kristina Jokinen. 2000. Cooperation, dialogue and ethics. *International Journal of Human-Computer Studies*, 53.

Jens Allwood. 1995. An activity based approach to pragmatics. *Gothenburg Papers in Theoretical Linguistics*, 76.

Nicholas Asher and Alex Lascarides. 1998. Questions in dialogue. *Linguistics and Philosophy*, 23(3).

Nicholas Asher and Alex Lascarides. 2001. Indirect speech acts. *Synthese*, 128.

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

John L. Austin. 1962. *How to do Things with Words*. Oxford.

Guido Boella, Rossana Damiano, Leonardo Lesmo, and Liliana Ardissono. 1999. Conversational cooperation: the leading role of intentions. In *Amstelog 99, the 3rd Workshop on the Semantics and Pragmatics of Dialogue*. University of Amsterdam.

Johan Bos and Malte Gabsdil. 2000. First-order inference and the interpretation of questions and answers. In *Götalog 2000, the 4th Workshop on the Semantics and Pragmatics of Dialogue, University of Göteborg*.

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. 1997. Querying and assertions in dialogue. Draft paper.

Göran Goldkuhl. 2000. The validity of validity claims: An inquiry into communicative rationality. In *LAP 2000, the Fifth International Workshop on the Language Action Perspective on Communication Modelling*.

Jürgen Habermas. 1981. *Theorie des kommunikativen Handelns*. Frankfurt a.M.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory, 2 Vols*. Dordrecht.

Jörn Kreutel and Colin Matheson. 2001. Cooperation and strategic acting in discussion scenarios. In *Proceedings of the Workshop on Coordination and Action*. 13th European Summer School in Logic, Language and Information, Helsinki.

Jörn Kreutel and Colin Matheson. 2002. From dialogue acts to dialogue act offers: Building discourse structure as an argumentative process. In *Edilog 2002, the 6th workshop on the Semantics and Pragmatics of Dialogue*. Bos, Johan and Foster, Mary Ellen and Matheson, Colin.

Staffan Larsson. 1998. Questions under discussion and dialogue moves. In *Twente Workshop on Language Technology*.

Stephen C Levinson. 1983. *Pragmatics*. Cambridge University Press.

Ian Lewin. 1998. The autoroute dialogue demonstrator: Reconfigurable architectures for poken dialogue understanding. Technical report, prepared by sri international cambridge, Computer Science Research Centre for the UK Defence Evaluation and Research Agency, Malvern.

William Mann. 2002. Dialogue analysis for diverse situations. In *Edilog 2002, the 6th workshop on the Semantics and Pragmatics of Dialogue*. Bos, Johan and Foster, Mary Ellen and Matheson, Colin.

Susan Weber McRoy and Graeme Hirst. 1995. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21(4).

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

Massimo Poesio and David Traum. 1998. Towards an axiomatisation of dialogue acts. In *Twente Workshop on Language Technology*.

John Searle. 1969. *Speech Acts*. Cambridge University Press.

David Traum and James Allen. 1994. Discourse obligations in dialogue processing. In *32nd Annual Meeting of the Association for Computational Linguistics*, pages 1–8.

David Traum, Johan Bos, Robin Cooper, Staffan Larsson, Ian Lewin, Colin Matheson, and Massimo Poesio. 1999. A model for dialogue moves and information state revision. TRINDI Deliverable 2.1, University of Göteborg.

Raimo Tuomela. 1998. Collective and joint intention. In *Selected Proceedings of the Rosseli Foundation conference 'Cognitive Theory of Social Action' held in Torino*.

Rob van der Sandt. 1992. Presupposition as anaphora resolution. *Journal of Semantics*, 9.

Ludwig Wittgenstein. 1953. *Philosophical Investigations*. The Macmillan Company, New York.

Referring to Objects Through Sub-Contexts in Multimodal Human-Computer Interaction

Frédéric Landragin and **Laurent Romary**
LORIA, campus scientifique – BP 239
F-54506 Vandœuvre-lès-Nancy CEDEX – FRANCE
{Frederic.Landragin,Laurent.Romary@loria.fr}

Abstract

There is no one-to-one relation between referential terms and types of access to the referents (referring modes) in multimodal human-computer interaction. We propose a classification of referring modes implying sub-contexts. We describe in detail the nature of these contextual subsets and we show their importance for each type of referring action. Then we define a relation between terms and modes, and we deduce a list of disambiguation principles for the computation of referential terms in order to identify the correct referring mode, the correct sub-context, and the correct referent.

1 Introduction

Many linguistic, pragmatic, and philosophical works deal with referring phenomena. For example, see (Karmiloff-Smith, 1979) or (Corblin, 1987) for a systematic linguistic approach to French referring phenomena, (Sperber and Wilson, 1995) or (Reboul and Moeschler, 1998) for a pragmatic approach, and (Récanati, 1993) for a philosophical work. Two points of interest are the linguistic form of the referring expression (the 'referential term') and the interpretation process (the 'referring mode'). Proper names, definite descriptions like "the red triangle," indexicals like "that," complex demonstratives like "this red triangle," are examples of referential terms. Direct and indirect access to the referent, specific and generic interpretation ("a triangle has three sides") are examples of referring modes.

From a computational point of view, no definitive link can be established between the referential term and the referring mode. The same referential term can be interpreted in several ways, the ambiguity lying in the multiple possible choices of referring mode (Reboul and Moeschler, 1998). In order to resolve the reference, a dialogue system has to be aware of the possible ambiguities (Beun and Cremers, 1998) and has to choose the relevant hypothesis.

In this paper we focus on the nature of these ambiguities, taking as a guide the notion of 'reference domain.' A reference domain is a structured sub-context in which the reference occurs. For example, "this red triangle" associated to a pointing gesture refers to a particular triangle in a domain including several red triangles. The contrast conveyed by the demonstrative determiner is then justified by the presence in the domain of at least another, not-focused, red triangle. The notion of reference domain is useful, first to exploit all the components of the referential term (determiner, category, modifiers, spatial information), second to take implicit attentional phenomena into account. For example, if the user refers to "these triangles" when pointing out two triangles in a salient group of three similar triangles (see the Figure 1), reference domains may be useful to identify the ambiguity between referring to two or three triangles. As the visual scene includes two strong perceptual groups, one on the left and one

on the right, the contrast conveyed by the demonstrative can apply between these two groups, or between the pointed triangles and the third one. The reference domain corresponds to the whole visual scene in the first case, and to the left group in the second case. Then, if the user refers to "the two circles" without any pointing gesture, the reference domain corresponding to the left group will be useful to understand the referring intention, that is "the two circles in the same attentional space."

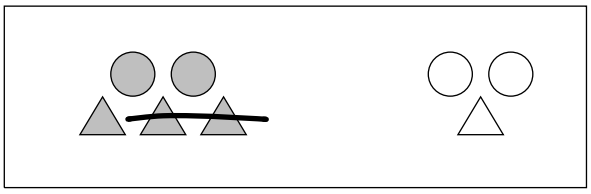


Figure 1: The use of sub-contexts.

As opposed to approaches like the one of (Kehler, 2000), we consider that simple algorithms are not sufficient for a multimodal system to identify the referents. As we see with our example, the gesture does not always give the referents, and the components of the verbal expression are not sufficient to distinguish them. But the combination of these ostensive clues with infered contextual considerations does. We focus here on the access to the referents through reference domains. Using a multimodal corpus (Wolff et al., 1998), we explore the possibilities of referring modes through reference domains. Based on this empirical data, we describe the foundations for a computational model of reference resolution, ie., a model of identification of the correct mode, domain, and referent. Since we used the corpus to collect referring phenomena and to build our model, we cannot validate it using the same corpus. Thus, we present a multimodal dialogue system which is currently under construction, and we show how such a system can validate our approach of referring.

2 Referential Terms and Referring Modes

To a referential term corresponds a preferential use. An indefinite noun phrase is generally used to introduce a new referent; a definite is an indicator to the necessity of identifying a particular referent (Corblin, 1987). A pure demonstrative is generally used together with a pointing gesture. But all of these referential terms have other uses. Reboul in (Moeschler and Reboul, 1994) proposes the following referring modes: direct reference, indirect reference, demonstrative reference, deictic reference and anaphorical reference. Proper names constitute the preferential direct referring mode, and demonstratives the preferential demonstrative referring mode. The problem is that a same noun phrase can be used for several referring modes. For example, the demonstrative noun phrase "this object" can be used for a demonstrative reference that implies an ostensive gesture, or for an anaphorical reference, the antecedent being a previous noun phrase like "the blue triangle." Indeed, demonstratives just as definites can be used for anaphorical purposes.

Thus, there is no one-to-one relation between referential terms and uses. To interpret them, we need to consider the context, which includes perceptual information, previous referring actions (and their results), and various world knowledge. With this linguistic, gestural, and visual information, the context is heterogeneous. Moreover, its scope can be enormous.

To face these problems, we have to consider sub-contexts. Our hypothesis is that each referring action occurs in a reference domain. This contextual subset is generally implicit and has to be identified by the system. It corresponds to a model of user's attention, user's memory, and conversation's area. The utterance's components allow to extract the referents from this subset and to prepare the interpretation of a future reference. For example, the referential term "the red triangle" includes two properties that must be discriminative in a reference domain that must include one or more "not-red triangles." A further referential term like "the other triangles" may be interpreted in the same domain, denoting a continuity in the reference sequence.

Reference domains can come from visual perception, language or gesture, or can be linked to the dialogue history or the task's constraints. Visual domains may come from perceptual grouping, for example to model focus spaces (see (Beun and Cremers, 1998)). When a particular colour is

Mode	**Mechanism details** including the nature–linguistics or multimodal–of the referential expression		**Examples** with singular, quantifier, plural, numeral adjective
new-ref	The referential expression does not refer but can be the antecedent of a future anaphor. No coreferent pointing gesture.		"Create **a square**," "**some squares**," "**squares**," "**two squares**" (all possibilities).
ext-any-ref	The activated reference domain must be more reduced than the whole ontological class of objects. It can be:	delimited by a coreferent pointing gesture	"Delete **a square**," "**some squares**," "**squares**," "**two squares**" with a gesture delimiting a set of squares (all possibilities).
		delimited by a previous referential term	"Select the squares and the triangles" followed by "delete **a square**," "**some squares**" (interpreted as "delete some of the selected squares"), "**squares**," "**two squares**" (all possibilities).
		unprecised (in which case we consider the whole visual context)	"Delete **two squares**" interpreted as "delete two of the visible squares," and eventually as "delete two of the visually salient squares" (all possibilities).
ind-par-ref	It is the pointing gesture that forces the choice of the referent. This case constitutes a deviance from classical theories like the one of (Corblin, 1987). Nevertheless, we found it in the corpus of (Wolff et al., 1998).		"Delete **a square**" with a gesture pointing out a particular square, "**some squares**," "**squares**," "**two squares**" with gestures pointing out particular squares (all possibilities).
gen-ref	Reference to a class of objects. No coreferent pointing gesture.		"**A square** has four sides," "**squares** have four sides," "**two triangles** with a common side make a quadrilateral" (quantifiers are impossible).

Figure 2: Indefinite noun phrases.

focused (ie., activated in the short-term memory, for example after "the red triangle" and "the red circle"), a reference domain based on the similarity of objects considering their colour is built on. Some domains may come from the user's gesture (see (Landragin, 2002)), others from the task's constraints. All of them are structured in the same way. They include a grouping factor ('being in the same referring expression,' 'being in the same perceptual group'), and one or more partitions of elements. A partition gives information about possible decompositions of the domain (see (Salmon-Alt, 2001)). Each partition is characterized by a differentiation criterion, which represents a particular point of view on the domain and therefore predicts a particular referential access to its elements ('red' compared to 'not-red,' 'focused' compared to 'not-focused').

This unified framework allows to confront the various contextual information, and to model the implicit whatever its origin between perception, speech and gesture. Considering reference domains, a referring action can:

1. **'new-ref' mode:** introduce a new referent in a linguistic manner;
2. **'ext-any-ref' mode:** extract any element from an activated reference domain;
3. **'ext-par-ref' mode:** extract a particular element from an activated reference domain;
4. **'ind-par-ref' mode:** indicate a particular referent that is focused elsewhere;
5. **'ind-par-dom' mode:** indicate a particular reference domain whose one element is focused;
6. **'gen-ref' mode:** refer to a generic entity, that is not a set of particular referents nor a reference domain.

In this list, the introduction of a new referent by a multimodal referring action is seen as the linguistic mention of a referent that is focused by the coreferent ostensive gesture ('ind-par-ref' mode). One important point is that we consider that referring directly to a particular object is impossible without an activated domain. Consequently, the direct reference mode of Reboul corresponds here to 'ext-par-ref' mode. Mentional expressions (see (Corblin, 1999)) with "first," "second" or "last"

Mode	Mechanism details		Examples with singular, plural, numeral adjective
ext-par-ref	The activated domain can be:	delimited by a coreferent pointing gesture	"**The triangle**" with a gesture delimiting a group of geometrical forms including one triangle (all possibilities).
		delimited by a previous referential term	"Select the blue triangle and the green square" followed by "delete **the triangle**," "select the triangles" followed by "**the two red triangles**," "the red triangle, the green one and the blue one" followed by "**the first**," "the group" followed by "**the triangle**" (all possibilities).
		delimited by a previous focussing on a visual space	After some references to objects at the left of the visual scene, "**the triangle**" can refer to "the triangle on the left" (all possibilities).
		delimited by a precision in the referential term	"**The triangle on the left of the scene**" (all possibilities).
		unprecised (in which case we consider a salient focus space)	"**The triangle**" interpreted as "the salient triangle" (all possibilities).
ind-par-ref	The referent can be:	given by a coreferent pointing gesture	"**The triangles**" with a gesture pointing a group of triangles (all possibilities)
		given by a previous referential term	"Select a red triangle" followed by "**the triangle**," "the triangle and the square" followed by "**the two forms**" (all possibilities).
ind-par-dom	The focused element can be:	given by a coreferent pointing gesture	"**The triangles**" with a gesture pointing out one triangle (in this particular example, a pointed object is extended to a group of similar objects, so only the plural is relevant).
		given by a previous referential term	"The square with circles around" followed by "**the group**" (this is also a particular case).
gen-ref	No coreferent pointing gesture.		"**The triangle** is a simple geometrical form", "**the triangles** have three sides" (numeral adjectives are impossible).

Figure 3: Definite noun phrases.

also correspond to 'ext-par-ref' mode, the differentiation criterion for the referents identification being the rank. Words like "other" and "next" have particular mechanisms. They refer to not-focused elements of a domain that has just been used, and are then included in 'ext-par-ref' mode.

3 The Modes Linked to a Referential Term

The possible modes considering the type of determiner are grouped in the following tables: Figure 2 for indefinite noun phrases (including headless ones), Figure 3 for definite noun phrases, Figure 4 for demonstrative noun phrases, Figure 5 for personal pronouns, and Figure 6 for demonstrative pronouns. With all these possibilities, we show how complex the relation between terms and modes is. A system that has to interpret spontaneous multimodal expressions must know all this information.

4 Reference Resolution in Theory

In this section we present the foundations for a model of reference resolution using reference domains. From the components of the verbal utterance and from the possible ostensive gesture, we deduce a list of clues that the system may exploit to identify the correct referring mode, the correct reference domain and the correct referent. We start with the determiner and then we detail the role of the predicate and of the other linguistic components. Following (Corblin, 1987), we make the hypothesis that the propositional context (and not only the determiner in the referential term) will favour the specific or the generic interpretation. Indeed, we consider that generic references are not usual in human-computer interaction. Thus, we ignore here the 'gen-ref' mode.

For an indefinite noun phrase, the system may choose between 'new-ref,' 'ext-any-ref,' and 'ind-par-ref' referring modes. The presence of a point-

Mode	Mechanism details		Examples with singular, plural, numeral adjective
ext-par-ref	A coreferent pointing gesture is impossible. Focussing is necessarily due to a previous referential term.		"Select the blue triangle and the green square" followed by "delete **this square**" (all possibilities).
ind-par-ref	The referent can be:	given by a coreferent pointing gesture	"**This triangle**" with a gesture pointing out one triangle. This is the most common multimodal referring expression (all possibilities).
		given by a previous referential term	"Select the blue triangle" followed by "**this triangle**," "the triangle" followed by "**this form**" (all possibilities).
ind-par-dom	The focused element can be:	given by a coreferent pointing gesture	"**These triangles**" with a gesture pointing out one triangle with a particular aspect (extension to a group of similar objects, so only with a plural).
		given by a previous referential term	"The square with circles around" followed by "**this group**" (the same particular example than in definites).
gen-ref	Three referring modes can be distinguished:	transition from a gestural antecedent to a generic interpretation	"**These forms**" with a gesture pointing out one triangle (numeral adjectives are impossible).
		transition from a linguistic antecedent to a generic interpretation	"This strange form" followed by "**these forms**" (only plural with no numeral adjective).
		direct multimodal generic interpretation	"**This form**" with a gesture pointing out one triangle, that can be interpreted as "this type of form," and is by consequence ambiguous with a specific interpretation (only singular).

Figure 4: Demonstrative noun phrases.

Mode	Mechanism details	Examples
ind-par-ref	A coreferent pointing gesture is impossible. The referent can be given by an obvious intention (it depends on the situation) or by a previous referential term.	"Sélectionne le triangle bleu"/ "select the blue triangle" followed by "supprime-**le**"/ "delete **it**". One other case is when the pronoun refers to another specimen of the referent linked to the antecedent: "j'ai supprimé le triangle mais **il** est revenu"/ "I deleted the triangle but **it** appears again" (singular and plural are possible).
ind-par-dom	A coreferent pointing gesture is impossible. The focused element is given by a previous referential term.	"Ajoute un triangle vert"/ "add a green triangle" followed by "supprime-**les**"/ "delete **them**" (the plural form is necessary to build on the domain).
gen-ref	A coreferent pointing gesture is impossible. This case corresponds to the transition from a linguistic antecedent to a generic interpretation.	"J'ai ajouté un triangle rouge parce qu'**ils** attirent le regard"/ "I added a red triangle because **they** are eye-catching" (the plural form is necessary).

Figure 5: Personal pronouns.

ing gesture may help the system: if the gesture delimits a set of objects not reduced to the referent(s), the only possible mode is 'ext-any-ref;' if the gesture is pointing out the referent(s), the only possible mode is 'ind-par-ref.' If no coreferent gesture is produced, there is an ambiguity between 'new-ref' and 'ext-any-ref' modes. The presence of an activated linguistic domain will favour the 'ext-any-ref' mode. In the other case, the predicate will disambiguate: a verb that denotes the introduction of new referent(s) like "add" and "create" will force the 'new-ref' mode. The 'ext-any-ref' mode will be chosen otherwise.

In the 'new-ref' interpretation, the system has to add the new object(s) in the visual domain corresponding to the scene. In this domain, a new partition is created, with a differentiation criterion linked to the predicate. The chosen referent(s) are focused in this partition. In the 'ext-any-ref' interpretation, the considered domain is the activated

Mode	Mechanism details	Examples
ext-par-ref	A coreferent pointing gesture is impossible. The focussing is necessarily due to a previous referential term.	"Le triangle, le carré et le rond"/"the triangle, the square and the circle" followed by the mentional reference "**celui-ci**"/"**this one**" (singular and plural are possible).
ind-par-ref	Demonstrative pronouns combine a demonstrative reference and an anaphor. They are associated to a pointing gesture to refer to a new object with the characteristics of a previous referent. The focussing is then necessarily due to a coreferent gesture.	In "sélectionne ce triangle bleu"/"select this blue triangle" followed by "supprime **celui-ci**"/"delete **this one**," "celui-ci"/"this one" together with a coreferent gesture refers to another blue triangle (singular and plural are possible).
gen-ref	See gen-ref mode for personal pronouns.	"J'ai ajouté un rond vert et un triangle rouge. **Ceux-ci** attirent le regard"/"I added a green circle and a red triangle. **These ones** are eye-catching" (the plural form is necessary).

Figure 6: Demonstrative pronouns.

one and the process is the same. In the 'ind-par-ref' interpretation, the process is the same, except that the choice of referents is not free but constrained by the gestural interpretation.

For a definite noun phrase, the system may choose between 'ext-par-ref,' 'ind-par-ref,' and 'ind-par-dom' modes. A fine analysis of the referential term and the possible pointing gesture is not sufficient to disambiguate. All hypotheses have then to be kept. In the 'ext-par-ref' interpretation, the system has to extract and to focus the referent from the activated domain. This referent has to be isolated with the category and its modifiers. For the 'ind-par-ref' and 'ind-par-dom' modes, the system has to build a new domain around the focused referent. The differentiation criterion of the new partition in this domain is the referent category.

For a demonstrative, the process is nearly the same than for definites, except that for 'ind-par-ref' and 'ind-par-dom' modes, the differentiation criterion of the new partition is given by the predicate or by the intervention of a pointing gesture. That shows the main difference between definites and demonstratives: the contrast between the referent and the other elements of the reference domain is due to category (and modifiers) for definites, and to focussing for demonstratives.

For a personal pronoun, the system may choose between 'ind-par-ref' and 'ind-par-dom' modes. The clue to disambiguate is a change in the use of singular or plural forms: if a transition occurs from a singular to a plural form, then the 'ind-par-dom' is identified. In this case, the system has to build a new domain around the focused element, the differentiation criterion of the new partition being the category. In the other case, the focussing nature does not change and then no new domain has to be built.

For a demonstrative pronoun, the presence of a pointing gesture forces the 'ind-par-ref' interpretation ('ext-par-ref' interpretation otherwise). In this case, the system has to extract and to focus the referent from the activated domain, the differentiation criterion being the order of mention. For the 'ext-par-ref' interpretation, the system has to build a new domain around the focused element, the new differentiation criterion being the gestural intervention.

Now that we have these interpretation rules, we can precise what a complete reference resolution model may do. The main point is the creation of a new reference domain, especially for definites and demonstratives. The linguistic and contextual clues are sometimes not sufficient for the delimitation of such a domain. For this reason, we propose to manage underdetermined reference domains, as it is done in (Salmon-Alt, 2001) and then in (Landragin et al., 2002). The linguistic and gestural information allow to build an underdetermined domain that groups all constraints. Then, the reference resolution process consists in the unification of this underdetermined domain with the domains that appear in the context. The one with the best unification result is kept for the referent identification.

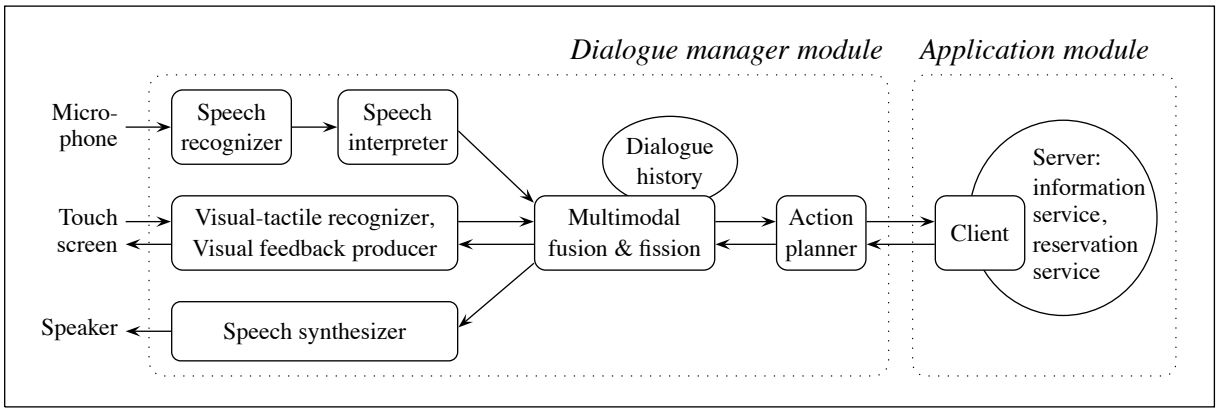


Figure 7: Architecture of the INRIA demonstrator for the OZONE European project.

5 Reference Resolution in Practice

One of the purposes of the OZONE European project is to design a dialogue system with two clearly separated modules, the first devoted to the application and the second (the dialogue manager) to the communicative abilities. Thus, several applications like finding a theater or a cinema, or like reserving a a train ticket or a room in a hotel, can be plugged to the same dialogue manager. A demonstrator, which is currently under construction, will implement that, including a multimodal interaction implying the microphone and the touch screen of a Tablet PC.

Considering a transport information and reservation application, the visual context is a map displayed on the screen. Gestures pointing out streets or train stations can be done. Here is a sample of a dialogue between the user and the system:

User: "How can I get to Paris?"

System: "You can either take a bus, a taxi or a train" (displaying some ways on the map).

User: "How long does it take to go from here to there?" (with an imprecise gesture pointing out a way between two train stations).

System: "Forty minutes..." This answer shows that the system is able to use the dialogue history for managing reference domains linked to the possible transport means (bus, taxi and train), and to exploit the gesture to build on a reference domain that groups the visible train stations.

The input and output are both multimodal, so fusion and fission algorithms are required, as it is showed in Figure 7. The resultant 'multimodal fusion & fission' module has to resolve multimodal referring expressions. Reference domains are managed in this module. The dialogue history is needed to resolve anaphorical expressions. This module also has to translate the whole utterance in a logical form that will be treated by the 'action planner', in order to choose an answering strategy.

Even if the objects are not triangles nor circles but streets and train stations, all that we have explored above has an importance in this demonstrator. With these objects, with spontaneous speech and gesture on a touch screen, all the described referring phenomena and ambiguities are possible. For example: "what are the two stations on the left of the map?," "show me a way to go to Paris," "I want to go to Versailles-Chantiers" followed by "this station is far away from Paris," etc. Thus, the resolution of multimodal referring expressions is an important part of the dialogue management. Moreover, sub-contexts exist and have to be taken into account using reference domains. Our approach and our model of reference seem to be relevant for an implementation like the one of the OZONE project.

6 Conclusion and Future Work

Reference to objects can take several forms which are not linked to particular mechanisms of identification. The choice of a determiner, of the singular or plural form, of a coreferent pointing gesture, lead to clues that specify some aspects of the interpretation process. In this paper we investigate multimodal human-computer interaction in-

volving simple objects. We explore the multiple possibilies of referring modes. We show that even in such a limited context many ambiguities can occur. We propose a list of disambiguation principles based on the notion of reference domain and of the concrete examples we found in the corpus of (Wolff et al., 1998) and in linguistic classical works like (Moeschler and Reboul, 1994), (Reboul and Moeschler, 1998) or (Sperber and Wilson, 1995). The examples we investigate illustrate a number of reference possibilities in terms of anaphor, transition from specific to generic interpretation, associations of referential terms and pointing gestures, etc.

Our proposition is focused on the identification of the referring mode and has to be complemented by an algorithm of reference resolution. The global method presented in (Landragin et al., 2002) fits well this concern, and the implementation of the OZONE demonstrator follows this method. As future research, we plan to make more precise the details of the interpretation process under the light of the classification we present here. One problem, given our focussing on complex phenomena (for example when the pointed objects are not the referents), is the lack of multimodal corpora suitable for evaluating such a classification. Nevertheless, the phenomena can easily be found in human-human communication; we need algorithms for a system to understand these phenomena, even if their evaluation is difficult for the moment.

Acknowledgements

This work was supported by the IST 2000-30026 OZONE EC project (http://www.extra.research.philips.com/euprojects/ozone/).

References

Robbert-Jan Beun and Anita Cremers. 1998. Object Reference in a Shared Domain of Conversation. *Pragmatics and Cognition*, 6(1/2):121–152.

Xavier Briffault. 1991. Cognitive, Semantic and Linguistic Aspects of Space. *Proceedings of the 9th IASTED International Symposium on Applied Informatics*, Innsbruck.

Francis Corblin. 1987. *Indéfini, défini et démonstratif*. Droz, Genève.

Francis Corblin. 1999. Mentional References and Familiarity Break. *Hommages à Liliane Tasmowski-De Ryck*, Unipress, Padoue.

Annette Herskovitz. 1986. *Language and Spatial Cognition*. Cambridge University Press, Cambridge.

Annette Karmiloff-Smith. 1979. *A Functional Approach to Child Language. A Study of Determiners and Reference*. Cambridge University Press, Cambridge.

Andrew Kehler. 2000. Cognitive Status and Form of Reference in Multimodal Human-Computer Interaction. *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000)*, Austin.

John Kelleher. 2003. A Perceptually Based Computational Framework for the Interpretation of Spatial Language. Ph.D. Thesis, Dublin City University.

Frédéric Landragin. 2002. The Role of Gesture in Multimodal Referring Actions. *Proceedings of the fourth IEEE International Conference on Multimodal Interfaces*, Pittsburgh.

Frédéric Landragin, Susanne Salmon-Alt, and Laurent Romary. 2002. Ancrage référentiel en situation de dialogue. *Traitement Automatique des Langues*, 43(2):99–129.

Jacques Moeschler and Anne Reboul. 1994. *Dictionnaire encyclopédique de pragmatique*. Seuil, Paris.

Anne Reboul and Jacques Moeschler. 1998. *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*. Armand Colin, Paris.

François Récanati. 1993. *Direct Reference: From Language to Thought*. Blackwell, Oxford.

Susanne Salmon-Alt. 2001. Reference Resolution within the Framework of Cognitive Grammar. *Proceedings of the Seventh International Colloquium on Cognitive Science*, San Sebastian, Spain.

Dan Sperber and Deirdre Wilson. 1995. *Relevance. Communication and Cognition (2nd edition)*. Blackwell, Oxford UK and Cambridge USA.

Frédéric Wolff, Antonella De Angeli, and Laurent Romary. 1998. Acting on a Visual World: The Role of Perception in Multimodal HCI. *Proceedings of AAAI Workshop on Multimodal Representation*, Madison.

Interactive Communication Management in an Issue-based Dialogue System

Staffan Larsson
Department of Linguistics
Göteborg University
sl@ling.gu.se

Abstract

We first give an overview of Interactive Communication Management (ICM), and especially feedback, in human-human communication. We then select a limited range of ICM behaviour useful for dialogue systems, and show how these utterances and responses to them are managed in a dialogue system implementing a theory of issue-based dialogue management.

1 Introduction

In this paper[1], we will give an account of the generation of feedback and sequencing in a dialogue system. Allwood (1995) uses the concept of "Interactive Communication Management" (ICM) to designate all communication dealing with the management of dialogue interaction. This includes feedback but also *sequencing* and *turn management*. Sequencing "concerns the mechanisms, whereby a dialogue is structured into sequences, subactivities, topics etc.".

Here, we will use the term ICM as a general term for coordination of the common ground, which in an information state update approach (Traum and Larsson, forth) comes to mean explicit signals (formalised as dialogue moves) enabling coordination of updates to the common ground. While feedback moves are associated with specific utterances (usually the utterance preceding the feedback move), ICM in general does not need to concern any specific utterance.

We start by briefly discussing ICM in current dialogue systems. We then proceed to give an overview of ICM, and especially feedback, in human-human communication. Finally, we select a limited range of ICM behaviour which we consider useful for dialogue systems, and show a brief example of how these utterances and responses to them are managed (selected, generated, and/or integrated) in GoDiS , a dialogue system implementing a theory of issue-based dialogue management (Larsson, 2002).

2 Verification

In the literature concerning practical dialogue systems, e.g. San-Segundo et al. (2001), grounding is often reduced to verification of the system's recognition of user utterances. Two common ways of handling verification are described as "explicit" and "implicit" verification, exemplified below (example from San-Segundo et al. (2001)).

- I understood you want to depart from Madrid. Is that correct? [explicit]
- You leave from Madrid. Where are you arriving at? [implicit]

Actually, both "explicit" and "implicit" feedback contain a verbatim repetition or a reformulation of the original utterance, and in this sense they are

[1] This paper is based on chapter 4 of (Larsson, 2002).

both explicit. The actual base for the distinction is that the first utterance (1) signals understanding but also a lack of confidence in the interpretation and (2) is trying to elicit a response regarding the correctness of the interpretation from the hearer, whereas the second utterance (1) signals confident understanding and (2) does not try to elicit a response.

In human-human dialogue, explicit confirmations are not very common. Presumably, they occur primarily in noisy environments and in situations where understanding is critical (e.g. when arranging a meeting in a busy airport). However, in dialogue systems, verification of user utterances are of central importance given the quality of current speaker-independent speech recognition. This explains to some extent why verification is often the only aspect of feedback handled by current systems - it is simply necessary. There is no reason, however, not to explore the possible uses of a wider range of feedback behaviour in dialogue systems. The obvious starting point for such exploration is human feedback behaviour.

3 Feedback and related behaviour in human-human dialogue

Clark and Schaefer (1989) describe *grounding* as the process of adding to the common ground. By *feedback* we mean behaviour whose *primary* function is to deal with grounding of utterances in dialogue[2]. This distinguishes feedback from behaviour whose primary function is related to the domain-level task at hand, e.g., getting price information. Non-feedback behaviour in this sense includes asking and answering task-level questions, giving instructions, etc. (cf. the "Core Speech Acts" of Poesio and Traum (1998)).

Answering a domain-level question (e.g., saying "Paris" in response to "What city do you want to go to?") certainly involves aspects of grounding and acceptance, since it shows that the question was understood and accepted. However, the primary function of a domain-level answer is to resolve the question, not to show that it was understood and accepted.

To get an overview of the range of explicit feedback behaviour that exists in human-human dialogue, we will classify feedback according to two basic criteria. (Below, we will assume a situation where a DP (Dialogue Participant) S has just uttered or is uttering u to a DP H, when the feedback utterance f, uttered by H to S, occurs.)

- level of action (basic communicative function): contact, perception, understanding or acceptance
- polarity: positive, negative or checking

These two criteria are loosely based on (Allwood et al., 1992) and (Allwood, 1995). Additional possible criteria, which will not be discussed further here (but see Larsson (2002)), are eliciting/non-eliciting, object-level/meta-level content, and surface form.

3.1 Levels of action in dialogue

Both Allwood (1995) and Clark (1996) distinguish four levels of action involved in communication[3].

- Acceptance (reaction, consideration): whether H has accepted and integrated (the content of) u
- Understanding (recognition): whether H understands u
- Perception (identification): whether H perceives u
- Contact (attention): whether H and S have contact, i.e., if they have established a channel of communication

These levels of action are involved in all dialogue, and to the extent that contact, perception, understanding and acceptance can be said to be negotiated, all human-human dialogue has a minimal element of negotiation built in.

[2] Since this paper is not concerned with multimodal dialogue, we will only discuss verbal feedback.

[3] Allwood and Clark use slightly different terminologies; here we use Allwood's terminology (and, for the acceptance level, also Ginzburgs) and add Clark's corresponding terms in parenthesis. The definitions are mainly derived from Allwood.

Given that grounding is concerned with all levels, it follows that feedback may concern any of these levels (and possibly several levels simultaneously).

3.2 Levels of understanding

We can make further distinctions between different levels of understanding, corresponding to three levels of meaning. These sublevels give a finer grading to the level of understanding[4].

- discourse-independent (but possibly domain-dependent) meaning (roughly corresponding to "meaning" in the terminology of Barwise and Perry (1983) and Kaplan (1979)), e.g., static word meanings
- domain-dependent and discourse-dependent meaning (roughly, "content" in the terminology of Barwise and Perry (1983) and Kaplan (1979))
  - referential meaning, e.g., referents of pronouns, temporal expressions
  - pragmatic: the relevance of u in the current context

By "discourse-independent" we mean "independent of the dynamic dialogue context." However, discourse-independent meaning may still be dependent on static aspects of the activity/domain. It is obvious that these levels of meaning are intertwined and do not have perfectly clear boundaries. Nevertheless, we believe they are useful as analytical approximations.

Since dialogue systems usually operate in limited domains, we will assume that we do not have to deal with ambiguities which are resolved by static knowledge related to the domain. For example, a dialogue system for accessing bank accounts does not have to know that "bank" may also refer to the bank of a river; it is simply very unlikely (though of course not impossible) that the word will be used with this meaning in the activity. It can be questioned whether this is always a good strategy, but for now we accept this as a reasonable simplification.

[4] A similar distinction is also used by Ginzburg (forth).

In this paper, we will not be dealing with referential meaning (simply because referent identification is currently not implemented in GoDiS). We will use the term *semantic meaning* to refer to discourse-independent meaning, and contrast this to pragmatic meaning.

3.3 Positive, negative, and checking feedback

Positive feedback indicates one or several of contact, perception, understanding, and acceptance, while negative feedback indicates lack thereof.

Negative feedback can be caused by failure to integrate U on any of the levels of action in dialogue:

- lack of contact - H and S have not established a channel of communication, or the channel has broken down
- lack of perception - H did not hear what S said
- lack of understanding on a semantic/pragmatic level - H recognized all the words, but could not extract a content
  - semantic meaning, e.g., word meanings
  - pragmatic meaning, i.e., the relevance of S's utterance in relation to the context
- rejection (lack of acceptance) of content

While there are clear cases of positive ("uhuh", "ok") and negative ("pardon?", "I don't understand") feedback, there are also some cases which are not so clear. For example, are check-questions (e.g., "To Paris?" in response to "I want to go to Paris") positive or negative? If positive feedback shows understanding, and negative feedback lack of understanding, then check-questions are somewhere in between; they indicate understanding but also that the lack of confidence in that understanding.

We assume a third category of *checking* feedback for check-questions and similar feedback types. If negative feedback indicates a lack of understanding, checking feedback indicates lack of confidence in one's understanding.

4 Feedback in GoDiS

In this section, we first show how feedback dialogue moves in GoDiS are represented. We then review the full range of feedback moves, starting with system-generated feedback and then moving on to user feedback.

The general notation for feedback ICM moves used in GoDiS is the following:

- icm:$L * P\{:Arg\}$

where L is an action level, P is a polarity, and Arg is an argument.

For user utterances, GoDiS will be able to produce the following kinds of feedback utterances (for the examples, assume that the user just said "I want to go to Paris"):

- L: action level
  - con: contact ("Are you there?")
  - per: perception ("I didn't hear anything from you", "I heard you say 'to Paris"')
  - sem: semantic understanding ("I don't understand", "To Paris.")
  - und: pragmatic understanding ("I don't quite understand", "You want to know about price.")
  - acc: acceptance/reaction ("Sorry, I can't answer questions about connecting flights", "Okay.")
- P: polarity
  - neg: negative
  - chk: checking
  - pos: positive
- Arg: argument

The arguments are different aspects of the utterance or move which is being grounded, depending action level:

- for per-level: $String$, the recognized string
- for sem-level: $Move$, a move interpreted from the utterance, including (possibly underspecified) semantic content
- for und-level: C, where C : Proposition is the propositional (fully specified) content of the utterance
- for acc-level: C : Proposition, the content of the utterance

For example, the ICM move icm:und$*$pos:dest-city(paris) provides positive feedback regarding an utterance that has been understood as meaning that the destination city is Paris.

For user utterances, GoDiS will be able to produce e.g., the following kinds of feedback utterances (for the examples, assume that the user just said "I want to go to Paris"):

- contact
  - negative; icm:con$*$neg ("I didn't hear anything from you")
- perception
  - negative; icm:per$*$neg realized as fb-phrase ("Pardon?", "I didn't hear what you said.")
  - positive; icm:per$*$pos:$String$ realized as metalevel verbatim repetition ("I heard 'to paris' ")
- understanding (semantic)
  - negative; icm:sem$*$neg realized as fb-phrase ("I don't understand.")
  - positive; icm:sem$*$pos:$Content$ realized as repetition/reformulation of content ("To Paris."[5])
- understanding (pragmatic)
  - negative; icm:und$*$neg realized as fb-phrase ("I don't quite understand.")
  - positive; icm:und$*$pos:$Content$ realized as repetition/reformulation of content (object-level) ("To Paris.")
  - checking; icm:und$*$chk:$Content$ realized as ask about interpretation ("To Paris, is that correct?")

[5] For a more informative example, see utterance 9 in the dialogue on page 5.

- acceptance
  - negative: icm:acc∗neg:$Content$ realized as explanation ("Sorry, Paris is not a valid destination city")
  - positive; icm:acc∗pos realized as fb-word ("Okay")

As a subcase of negative acceptance (i.e., rejection), GoDiS is able to reject user questions using the icm:acc∗neg:issue(Q) move, where Q is a question, as illustrated below (U is the user, S is the system)[6]:

U> `What about connecting flights?`
S> `Sorry, I cannot answer questions about connecting flights.`

We are not claiming that humans always make these distinctions explicitly or even consciously, nor that the link between surface form and feedback type is a simple one-to-one correspondence; for example, "mm" may be used as positive feedback on the perception, understanding, and acceptance levels. Feedback is, simply, often ambiguous. However, since GoDiS is making all these distinctions internally we might as well try to produce feedback which is not so ambiguous.

Of course, there is also a tradeoff in relation to brevity; extremely explicit feedback (e.g., "I understood that you referred to Paris, but I don't see how that is relevant right now.") could be irritating and might decrease the efficiency of the dialogue. We feel that the current choices of surface forms are fairly reasonable, but testing and evaluation on real users would be needed to find the best ways to formulate feedback on different levels. This is an area for future research.

A general strategy used by GoDiS in ICM selection is that if negative or checking feedback on some level is provided, the system should also provide positive feedback on the level below. For example, if the system produces negative feedback on the pragmatic understanding level, it should also produce positive feedback on the semantic understanding level:

S> `To Paris. I don't quite understand.`

For system utterances, GoDiS will react appropriately to the following types of user-initiated feedback:

- perception level
  - negative; fb-phrase ("Pardon?", "Excuse me?", "Sorry, I didn't hear you") interpreted as icm:per∗neg
- acceptance level
  - positive; fb-phrase ("Okay.") interpreted as icm:acc∗pos
  - negative; issue rejection fb-phrase ("I don't know", "Never mind", "It doesn't matter") interpreted as icm:acc∗neg:issue

In addition, irrelevant followups to system ask-moves are regarded as implicit issue-rejections. The coverage of user feedback behaviour is thus more limited than the coverage for system behaviour. The main motivation for this is that system utterances are less likely to be problematic for the user to interpret than vice versa.

5 Issue-based grounding in GoDiS

Different kinds of feedback from the system carry with them different expectations on and options for the user. This is especially important in the case where the system is unsure whether it has correctly interpreted the user's utterance, or in cases where the system misinterprets the user. In the issue-based approach to feedback management, this is accounted for by bringing under discussion questions related to grounding.

On our analysis, checking feedback on the understanding level with content C introduces a meaning-question ?und(C), roughly paraphrasable as "Is C a correct interpretation the latest utterance?". This question is available for resolution of elliptical answers (for y/n-questions, "yes" and "no"), and must be resolved before the dialogue can proceed. If the system produces this kind of feedback, the user may respond "yes" or

[6] Note that this requires the system to recognize and understand common questions that it cannot answer; in current systems this is usually not (if ever) the case.

"no" depending on whether the interpretation was correct or not; only when a positive answer to the understanding-question is received does the system assume that C is a part of the shared dialogue state. If no answer is received, the system cautiously assumes that the interpretation was erroneous.

U> `To Lund please`

S> `To London?`
(understanding-question now raised and unresolved, and available for resolution of short answers)

U> `(silence)`
(since the understanding-question is still not resolved, neither is the issue of destination)

S> `So, what city do you want to go to?`

When producing positive feedback on the understanding level, the understanding-question is not raised explicitly as was the case with checking feedback. The interpretation is tentatively assumed to be correct and is added to the shared dialogue state. The understanding-question is thus (tentatively) assumed to be resolved. However, on our analysis, positive feedback still makes the understanding-issue available for elliptical responses, so that the user may respond to the positive feedback using "yes" or "no".

This is possible in GoDiS because of the distinction between a local ("QUD", see Ginzburg (1996)) and a global ("ISSUES") stack of questions under discussion, where the former represents questions available for ellipsis resolution and the latter contains questions which have been raised but not yet resolved. If a question q is on QUD but not on ISSUES, an answer to q will trigger an accommodation rule copying q from QUD to ISSUES, thus reraising q and retracting any previously stored answers to q. In the case of grounding, where q is an understanding-question, this amounts to retracting the assumption that the previous user move was correctly understood.

Thus, if a negative response to positive feedback on the understanding level is received, the system retracts its assumption about the correctness of its interpretation:

U> `To Lund please`

S> `OK, to London.`
(understanding-question now tentatively resolved but available for resolution of short answers)
[formally, ?und(dest-city(london)) $\in$ QUD[7]]

U> `No!`
(understanding-question addressed and accommodated to ISSUES)
[formally, ?und(dest-city(london)) $\in$ ISSUES]
(negative answer to understanding-question integrated; proposition dest-city(london) retracted)

S> `Not to London. So, what city do you want to go to?`
(the question of destination city is reraised)

If no response to the understanding-question is received, the system continues to assume that the interpretation was correct and eventually proceeds to the next item in its dialogue plan:

U> `To Lund please`

S> `OK, to London.`
(understanding-question now tentatively resolved but available for resolution of short answers)

U> `(silence)`
(understanding-question no longer available for resolution of short answers)

S> `What city do you want to go from?`
(system proceeds to next item on the plan)

A more detailed account of issue-based grounding in GoDiS can be found in Larsson (2002).

6 Sequencing ICM in GoDiS

We will now briefly discuss sequencing ICM in GoDiS. We believe it is good practice to try to keep the user informed about what's going on inside the system, at least to a degree that facilitates a natural dialogue where system utterances "feel

[7] Although it lies outside the scope of this paper to present the formalism used in GoDiS for representing semantic content, it has been included here for those interested. We hope the formalism is sufficiently straightforward to be understood by the reader; if not, the formal details can be ignored.

natural". One way of doing this is to produce sequencing ICM indicating significant updates to the information state which are not directly related to specific user utterances but instead concerns subactivities, topics etc.[8].

For GoDiS, we have implemented three types of sequencing ICM. First, GoDiS will indicate when it is loading a dialogue plan[9]. Second, GoDiS will produce ICM to indicate when an issue is being reraised (in contrast to being raised for the first time). Third, GoDiS generates ICM moves for indicating accommodation of questions (which may involve plan recognition). This is meant to indicate that some inference has been performed, and that it has been successful.

The choice of actual wordings for these moves are currently only based on intuitions; both corpus studies and usability tests can serve to give a firmer ground to the exact formulation of sequencing moves.

- icm:reraise: indicate reraising implicitly ("So, ...")
- icm:reraise:Q: reraising an issue Q explicitly ("Returning to the issue of price.")
- icm:loadplan ("Let's see.")
- icm:accommodate ("Alright.")

7 A sample dialogue with GoDiS

The dialogue below gives an impression of what feedback and sequencing capabilities our system handles. The basic principle for move selection in GoDiS is that utterances may not raise more than one question. Surface form generation in GoDiS is template-based. The dialogue has been annotated with ICM dialogue moves.

1 S> Welcome to the travel agency!

2 U> price information please

3 S> Okay. [icm:acc∗pos]
You asked about price.
[icm:und∗pos:issue(?x.price(x))]
Let's see. [icm:loadplan]
How do you want to travel?

4 U> pardon [icm:per∗neg]

5 S> How do you want to travel?

6 U> by plane, london

7 S> By plane, is that correct?
[icm:und∗chk:transport(plane)]

8 U> yes

9 U> London. [icm:sem*pos:london]
I don't quite understand. [icm:und∗neg]
Do you mean to London or from London?

10 U> to london

11 S> Okay, [icm:acc∗pos]
to London. [icm:und*pos:dest-city(london)]
What city do you want to go from?

12 U> lyons

13 S> Sorry, Lyons is not a valid destination city. [icm:acc∗neg:dept-city(lyons)]
So, [icm:reraise]
What city do you want to go from?

14 U> paris

15 S> Okay, [icm:acc∗pos]
to Paris. [icm:und∗pos:dept-city(paris)]
When do you want to travel?

16 U> Do I need a Visa?

17 S> Okay, [icm:acc∗pos]
you want to know about visa.
[icm:und∗pos:issue(?need-visa)]
What country are you from?

18 U> Sweden

19 S> Okay, from Sweden. No, you do not need a visa. Returning to the issue of price. [icm:reraise:?x.price(x)]
So, [icm:reraise]
when do you want to travel?

...

20 U> What about connecting flights?

21 S> Sorry, I cannot answer questions about connecting flights.
[icm:acc*neg:issue(?x.connecting-flights(x))]

8 Conclusions and future research

Starting from a fairly broad categorization of various types of feedback and sequencing moves in human-human dialogue, we have implemented a subset of these moves in a dialogue system. We believe that a view of the full range of ICM behaviour can be useful for suggesting useful additions to the range of moves handled by dialogue

[8] While able to produce sequencing ICM, GoDiS is currently not able to understand sequencing ICM from the user.

[9] Due to space limitations we will not here go into the internal workings which trigger sequencing ICM. See Larsson (2002) for detailed accounts of these.

systems, both by making systems more flexible in terms of dialogue management and by making dialogues more natural.

Our account of grounding and ICM is so far only partial in coverage; phenomena that remain to be accounted for and/or implemented include clarification ellipsis (Cooper and Ginzburg, 2001), semantic ambiguity resolution, collaborative completions and repair (Traum, 1994), and turntaking ICM. While we have included some rudimentary sequencing ICM, further investigations of the appropriateness and usefulness of these utterances are needed; here, research on discourse markers (Schiffrin, 1987) and cue phrases (Grosz and Sidner, 1986; Polanyi and Scha, 1983; Reichman-Adar, 1984) can be of use. We also want to explore turntaking in asynchronous dialogue management, and how this relates to turntaking ICM.

We are currently implementing several applications for GoDiS to enable evaluation and further development of the system. A VCR application will be reachable by phone and allow users to program and control a (computerized) VCR using spoken dialogue. We are also implementing a robot control application allowing users to interactively control and program a robot's movements using spoken dialogue.

References

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsen. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26.

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

J. Barwise and J. Perry. 1983. *Situations and Attitudes*. The MIT Press.

H. H. Clark and E. F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259 – 94.

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

Robin Cooper and Jonathan Ginzburg. 2001. Resolving ellipsis in clarification. In *Proceedings of the 39th meeting of the Assocation for Computational Linguistics, Toulouse*, pages 236–243.

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

J. Ginzburg. forth. Questions and the semantics of dialogue. Forthcoming book, partly available from `http://www.dcs.kcl.ac.uk/staff/ginzburg/papers.html`.

B. J. Grosz and C. L. Sidner. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

D. Kaplan. 1979. Dthat. In P. Cole, editor, *Syntax and Semantics v. 9, Pragmatics*, pages 221–243. Academic Press, New York.

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

Massimo Poesio and David R. Traum. 1998. Towards an axiomatization of dialogue acts. In *Proceedings of Twendial'98, 13th Twente Workshop on Language Technology: Formal Semantics and Pragmatics of Dialogue*, pages 207–222.

L. Polanyi and R. Scha. 1983. On the recursive structure of discourse. In K. Ehlich and H. van Riemsdijk, editors, *Connectedness in Sentence, Discourse and Text*, pages 141–178. Tilburg University.

R. Reichman-Adar. 1984. Extended man-machine interface. *Artificial Intelligence*, 22(2):157–218.

Ruben San-Segundo, Juan M. Montero, Juana M. Guitierrez, Ascension Gallardo, Jose D. Romeral, and Jose M. Pardo. 2001. A telephone-based railway information system for spanish: Development of a methodology for spoken dialogue design. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, pages 140–148.

D. Schiffrin. 1987. *Discourse Markers*. Cambridge University Press, Cambridge.

David Traum and Staffan Larsson. forth. The information state approach to dialogue management. In Ronnie Smith and Jan Kuppevelt, editors, *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers.

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

A Dialogue Game to Agree to Disagree about Inconsistent Information

Henk-Jan Lebbink
Institute of Information and Computing Sciences
Utrecht University
henkjan@cs.uu.nl

Cilia Witteman
Diagnostic Decision Making
Faculty of Social Sciences
University of Nijmegen
c.witteman@ped.kun.nl

John-Jules Meyer
Institute of Information and Computing Sciences
Utrecht University
jj@cs.uu.nl

Abstract

This paper proposes a dialogue game in which coherent conversational sequences at the speech act level are described of agents that become aware they have a disagreement and settle the dispute by agreeing to disagree when they believe insufficient propositions to resolve the situation. A dialogue game is formulated in which agents can offer information possibly resulting in non-reconcilable, mutually inconsistent belief states. These states are handled by a dialogue rule that defines when to 'agree to disagree'.

1 Introduction

If we are to understand conversations we may need to carefully model the underlying principles that drive them, but we would probably be just as satisfied if we could build computational models that generate useful conversations. In general conversations, participants may have autonomy over their cognitive states but they may also have intentions to change those of others. This autonomy and intention may result in discussions about non-reconcilable beliefs. How to cope with these disputes and how to devise a computational model that identifies them?

A dialogue game by Beun (2001) describes speech acts between agents and identifies three structures with accompanying properties that form a dialogue game that agents need to play to communicate in a sensible way. Agents need to have a cognitive state, dialogue rules to generate speech acts to convey information to other agents and update rules to process incoming information. In Lebbink et al. (2003), a multi-valued logic is introduced to describe inconsistent and biased information in dialogue games without forcing agents to perform belief revision. In the same vein as the FIPA work on agent communication languages (Labrou, 2001), we formulate semantics of the speech acts to offer information and to agree to disagree.

What is lacking in Beun (2001) and Lebbink et al. (2003), is the possibility for agents to recognize irresolvable disagreements and based on this recognition utter an agreement to disagree making the disagreement common belief. This common belief can motivate dialogues on the definition of the terms used in the disagreement or dialogues to weaken arguments in order to retract propositions to resolve the disagreement.

Agents may be motivated to persuade others to accept to believe certain propositions. Consequently, agents may accept to believe proposed propositions, but they may also deny to accept to believe propositions when accepting them would result in inconsistent beliefs. The objective is to present a dialogue game in which cognitive agents become aware of non-reconcilable beliefs and manifest this awareness to others while preserving their private beliefs. We use a formal method to represent classical and inconsistent information in the agents' cognitive states, enabling us to devise dialogue rules that can cope with these types of information. The resulting formalism allows for embedded dialogues and verification of existing dialogues, but notably it is computational which makes it possible to generate all valid dialogues in a dialogue game.

In the following (fictitious) dialogue, Tv (from Sesame Street) tries to insure his new car; he rings

an insurance company and explains his desire to an insurance agent (Ia). The Ia wants to sell Tv an expensive insurance policy because Sesame Street puppets are prone to fast and dangerous driving. The Ia wants Tv to accept that its car is not safe justifying an expensive policy. For the sake of argument, agents do not accept information that renders their beliefs inconsistent. This dialogue is mathematically checked as a valid sequence of utterances of our dialogue game. We omit proofs and conclude that this dialogue is one of the 177 different valid dialogues.

Dialogue 1 (Car insurance in Sesame Street)

1. *Tv to Ia* *'My car is a Ferrari.'*
2. *Ia to Tv* *'Ok.'*
3. *Tv to Ia* *'(and) My car is safe.'*
4. *Ia to Tv* *'I don't believe that.'*
5. *Ia to Tv* *'(actually) I think your car is not safe.'*
6. *Tv to Ia* *'I don't accept that my car is not safe.'*
7. *Ia to Tv* *'Do you accept that if a car is a Ferrari then it is not a safe car?'*
8. *Tv to Ia* *'(no) I don't want to accept that.'*
9. *Ia to Tv* *'Lets agree to disagree whether your car is safe or not.'*
10. *Tv to Ia* *'Ok.'*

2 A Dialogue Game to Offer Information

In accordance with Wittgenstein's dictum "meaning is use", we propose a dialogue game that gives meaning to the conception of offering information. To understand a word or sentence is to know how to use it. And to be able to use a word or a sentence is to be capable of recognizing the contexts in which it is appropriate to utter it (Ellenbogen, 2003). This activity of speaking is described in a normative way, governed by dialogue rules that dictate correct and incorrect use of communicative acts. We could say that an agent understands a word when it can distinguish between correct and incorrect uses.

Following the approach taken in Beun (2001) and Lebbink et al. (2003), a dialogue game is a set of dialogue rules that describe which communicative acts an agent may utter given its current cognitive state. A dialogue game also has a set of update rules that describe the changes of the cognitive state given an uttered communicative act.

We assume that information can only accumulate in the participants' cognitive states and cannot be retracted. In this information-monotonic approach additions may introduce inconsistent beliefs. Although the dialogue rules we are to present will prohibit inconsistencies, agents use the possibility of inconsistencies in a look-ahead fashion when deciding to believe propositions. Agents are considered omnipotent and use equal consequence relations. In addition, agents can only speak to one agent at a time via an ideal half-duplex communication channel which means that no information is lost and that information can only flow in one direction at a time. No restrictions are made on the number of participants.

2.1 Ordering of Information: Bilattices

Whereas in classical logic terms are assigned a truth-value true or false, in multi-valued logic (MVL) new truth-values are introduced to represent uncertain, non-determined or other epistemic attitudes (Rescher, 1969). In previous work (Lebbink et al., 2003), truth-values from a bilattice structure (Ginsberg, 1988; Fitting, 1991) are used to define a MVL, and theories of this MVL are used to represent the agent's cognitive state. Theories of MVL are sets of multi-valued propositions which are terms taken from some ontology with an assigned truth-value from a bilattice structure. Next to these propositions, an implication operation for four truth-values imp_4 is added with a reading similar to the one from classical logic: if the antecedent of the implication is part of a theory, then the consequent is also present.

Two terms are used to denote the information of the example dialogue: $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{ferrari})$ and $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{safe})$ stating that it is true that this car denoted by $\mathsf{this\_car}$ is a Ferrari and a safe one respectively. The multi-valued proposition $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{ferrari}){:}\mathsf{f}$ is read as 'term $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{ferrari})$ has truth-value f'. If this proposition is part of a theory T that represents the beliefs of some agent, we say that the agent believes that 'this car is not a Ferrari'. An implication between the fact that if some car is a Ferrari then that car is not safe, is represented by the proposition $\mathsf{imp}_4(\mathsf{is\_a}(\mathsf{X}, \mathsf{ferrari}){:}\mathsf{t}, \mathsf{is\_a}(\mathsf{X}, \mathsf{safe}){:}\mathsf{f}){:}\mathsf{t}$.

A bilattice is an algebraic structure that formal-

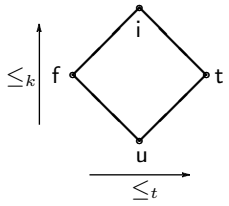


Figure 1: Bilattice representing partial and inconsistent information.

izes an intuitive space of generalized truth-values with two lattice orderings. In Figure 1 the bilattice for a four-valued logic (Belnap Jr., 1977) is graphically depicted; t and f stand for the classical truth-values true and false respectively; non-orthodox truth-value u and i represent a complete lack of information (unknown) and the inconsistent information state. Truth-values are ordered by the amount of truth $\leq_t$ and the amount of information $\leq_k$; only the latter is of interest to us. For instance, u has less information than t and f, denoted by u $\leq_k$ t and u $\leq_k$ f, but t and f are unrelated to one another in the k-order, that is, t $\not\leq_k$ f and f $\not\leq_k$ t. Bilattices with more truth-values and even a continuum of truth-values can be used to represent biased information or probabilities (Ginsberg, 1988); we use only the four from Figure 1.

2.2 Communicative Acts

The communicative act of offering information and its two corresponding answers to accept and reject information are defined next. The act of an *offer* $[x, y, p]^!$ is uttered by a speaker (x) directed to a listener (y) and is read as *'Are you (y) willing to accept to believe proposition p?'* In the first line of the example dialogue, Tv states that its car is a Ferrari, which we consider equal to the phrase *'Are you, Ia, willing to accept to believe that it is true that my car is a Ferrari?'*. In answer to this, in the second line, the Ia grants Tv's offer. The act of *granting an offer* $[x, y, p]^{!+}$ is read as *'I (x) am willing to accept to believe p.'* An agent may also deny an offer, which is done in line four when the Ia denies to believe that Tv's car is safe. The act of *denying an offer* $[x, y, p]^{!-}$ is read as *'I (x) am not willing to accept to believe p.'* The act of agreeing to disagree $[x, y, (p, q)]^\Delta$ is read as *'Are you (y) willing to agree that we are in disagreement over proposition p and q.'* Precise contexts for correct use of these acts are defined in Section 2.6.

A rendition of the dialogue from Section 1 is used in the remainder of this paper.

Dialogue 2 (Car insurance in Sesame Street)

1. [tv, ia, is_a(this_car, ferrari):t]$^!$
2. [ia, tv, is_a(this_car, ferrari):t]$^{!+}$
3. [tv, ia, is_a(this_car, safe):t]$^!$
4. [ia, tv, is_a(this_car, safe):t]$^{!-}$
5. [ia, tv, is_a(this_car, safe):f]$^!$
6. [tv, ia, is_a(this_car, safe):f]$^{!-}$
7. [ia, tv, imp(is_a(X, ferrari):t, is_a(X, safe):f):t]$^!$
8. [tv, ia, imp(is_a(X, ferrari):t, is_a(X, safe):f):t]$^{!-}$
9. [ia, tv, (is_a(this_car, safe):f, is_a(this_car, safe):t)]$^\Delta$
10. [tv, ia, (is_a(this_car, safe):f, is_a(this_car, safe):t)]$^\Delta$

2.3 The Agent's Cognitive State

An agent's cognitive state consists of a number of mental constructs which are theories of MVL. An agent's belief state is probably the most important construct next to its desire state. A proposition p is said to be believed by an agent x if p is part of mental construct B_x, that is, $p \in B_x$. Analogously, an agent x desires that an agent y believes a proposition p if p is part of mental construct $D_x^{B_y}$, that is, $p \in D_x^{B_y}$. For example, Tv desires that the Ia believes that it is true that this car is safe, that is, is_a(this_car, safe):t$\in D_{\mathsf{tv}}^{B_{\mathsf{ia}}}$.

Agents keep record of all other agents' explicitly communicated beliefs, desires and accompanying consequences. The mental construct for manifested beliefs $M_x B_y$ represents agent y's beliefs that agent x is aware of. For instance, ψ:t $\in M_x B_y$ states that x is aware that y believes that ψ has at least truth-value t. An agent also records other agents' communicated desires, this is done in $M_x D_y^{B_z}$. For instance, Tv is aware that the Ia desires that Tv believes that this car is not safe, that is, is_a(this_car, safe):f$\in M_{\mathsf{tv}} D_{\mathsf{ia}}^{B_{\mathsf{tv}}}$. Also, agents need to keep record of explicitly stated ignorance of other agents; the third type of mental construct is the manifested ignorance state. $p \in B_x I_y$ states that agent x is aware that agent y is ignorant towards p.

In addition, higher-order manifested mental constructs are needed for agents to remember to whom they stated their desires and beliefs. This information is needed to prevent them from uttering offers more than once; this is addressed by dialogue rules (Section 2.6). Mind-bending mental

constructs are needed to encode this information; the construct $M_x M_y B_x$ states that x is aware of y's awareness that x believes p. From a usage perspective, if an agent has answered an offer regarding some proposition then this proposition is part of this construct; this is addressed by update rules (Section 2.7). Likewise, if an agent has proposed an agreement to disagree it is not allowed to utter the same agreement again. Therefore, a record needs to be kept of these agreements. For instance, $M_{\mathsf{tv}} M_{\mathsf{ia}} \Delta(\mathsf{tv}, \mathsf{ia})$ states that Tv is aware of the Ia's awareness of their disagreement, this situation is described in Section 3.2.

A dialogue game defines a space of different dialogues that unfold by applying dialogue rules and update rules. Given a (initial) cognitive state of all agents participating in the dialogue (hereafter collective state), all sequences of communicative acts can be generated that are considered valid in the game. For our example dialogue the following initial collective state is used. The Ia has the desire that Tv believes its car is not safe, and Tv has the desire that the Ia believes that its car is a Ferrari and safe. The Ia believes that if a car is a Ferrari then the car is not safe, and Tv believes that its car is a Ferrari and a safe one. Formally,
$\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{safe}){:}\mathsf{f} \in D_{\mathsf{ia}}^{B_{\mathsf{tv}}}$,
$\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{ferrari}){:}\mathsf{t} \in D_{\mathsf{tv}}^{B_{\mathsf{ia}}} \cap B_{\mathsf{tv}}$,
$\mathsf{imp}_4(\mathsf{is\_a}(\mathsf{X}, \mathsf{ferrari}){:}\mathsf{t}, \mathsf{is\_a}(\mathsf{X}, \mathsf{safe}){:}\mathsf{f}){:}\mathsf{t} \in B_{\mathsf{ia}}$, and
$\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{safe}){:}\mathsf{t} \in D_{\mathsf{tv}}^{B_{\mathsf{ia}}} \cap B_{\mathsf{tv}}$.

2.4 Cognitive Processes

In our formalism, agents can perform two cognitive processes: deciding to believe offered information and deducing consequences of newly accepted beliefs. Other central concepts, such as choice between permissible communicative acts, or which strategy to use to persuade others, will not be included in the descriptions presented here.

From a mathematical perspective, an agent is persuaded to believe a proposition when its cognitive state changes from not believing the proposition to believing it (Walton and Krabbe, 1995); the proposition is set-theoretically added to the belief state. Agents can have different criteria for accepting to believe something. For simplicity, we assume agents to be very credulous: they accept to believe offered propositions that are consistent with their current beliefs.

Reasoning is restricted to the agents' capacity to draw conclusions based on believed implication rules and antecedents. If an implication rule is believed by the Ia, $\mathsf{imp}_4(\mathsf{is\_a}(\mathsf{X}, \mathsf{ferrari}){:}\mathsf{t}, \mathsf{is\_a}(\mathsf{X}, \mathsf{safe}){:}\mathsf{f}){:}\mathsf{t} \in B_x$ and the Ia is persuaded to believe an antecedent that the car is a Ferrari, $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{ferrari}){:}\mathsf{t} \in B_{\mathsf{ia}}$, then it also concludes to believe the consequent that the car is not safe, $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{safe}){:}\mathsf{f} \in B_{\mathsf{ia}}$. However, if the Ia already believes $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{safe}){:}\mathsf{t}$, then its belief state becomes inconsistent, that is, $\mathsf{is\_a}(\mathsf{this\_car}, \mathsf{safe}){:}\mathsf{i} \in B_{\mathsf{ia}}$. The closure $cl(B_{\mathsf{ia}} \cup \{p\})$ corresponds with the set of propositions including Ia's beliefs plus p with consequents.

2.5 Motivations to Communicate

In Beun (2001) an agent's motivation to utter a question is to balance its belief and desire state. Stated in our terminology, an agent may pose a question regarding some proposition if it has the desire to be in a belief state in which it believes the proposition and it currently does not (yet) believe the proposition. We give a similar motivation to offer information. An agent x may offer information to an agent y regarding some proposition p if x has the desire that y is to believe the proposition, and x is not aware that y already believes the proposition. Stated differently, an agent's motivation to utter an offer is to balance its desire and manifested belief state; an agent x is motivated to offer p to y if $p \in D_x^{B_y}$ and $p \notin M_x B_y$.

According to the Gricean maxims of cooperation, speakers are forbidden to ask anything they already believe (Grice, 1975). Analogously, speakers are forbidden to put forward information if they are aware the listener already believes the information. Next to giving restrictions, these maxims also provide motivations for granting and denying questions and offers: both should always be answered. In the next paragraph, the motivations and restrictions are combined to form the preconditions for 'correct' usage of our communicative acts; these preconditions provide the semantics of the acts and give communicative acts meaning in the context of a dialogue game.

2.6 Dialogue Rules

In group decisions, different experts make decisions as a group by agreeing on the assumptions they need to make to come to one common decision. This means that the assumptions (which are beliefs) should be non-conflicting. A motivation for someone who facilitates group decision making is to introduce dialogue rules that enable experts to offer information with the objective that experts become aware of other agent's attitudes towards their beliefs. One way to check whether assumptions are conflicting is to offer these to others and conclude from their responses whether they agree to believe these. In this section, the dialogue rules are defined allowing an agent to utter communicative acts given its cognitive state.

An offer $[x, y, p]^!$ is defined applicable when the speaker (x) is motivated to utter an offer. As stated before, the speaker desires the listener (y) to believe proposition p and the speaker is not aware that the listener already believes p, that is, $p \in D_x^{B_y}$ and $p \notin M_x B_y$. Of course, a dialogue game can be conceived in which meaning is given to offering information even when the speaker is aware the listener already believes this. Such a different game is played when the speaker wants to convey that it was *not* aware that the listener believed the proposition, although it was. Furthermore, it is assumed that in this dialogue game, agents are not allowed to propose information that they do not believe themselves. Due to the ideal communication channel, agents are also not allowed to pose a communicative act more than once: after uttering an offer, they are aware that the listener is aware of the speaker's desire to induce a belief change. This is represented by the mental construct that the speaker is aware that the listener is aware of the speaker's desire to induce a belief change associated with p in the listener. Also, proposition p should not be part of this mental construct, that is, $p \notin M_x M_y D_x^{B_y}$ which can only be true if speaker has already offered p.

Definition 1 (offer) *If $p \in D_x^{B_y}$, $p \notin M_x B_y$, $p \in B_x$ and $p \notin M_x M_y D_x^{B_y}$ then an offer $[x, y, p]^!$ is applicable.*

The communicative act of granting an offer $[x, y, p]^{!+}$ is applicable when the (granting) speaker (x) believes the proposition p. Whether agent x is persuaded to believe the proposition as a result of the offer x is responding to, does not matter. Obviously, a speaker may only grant an offer if it is aware the listener has the desire to make the speaker believe p, that is, $p \in M_x D_y^{B_x}$. In this dialogue game, an agent x can only be aware of this if the other agent y has uttered an offer. To ensure that the information represented by granting an offer is not superfluous, e.g. stated more than once, the speaker may not be aware that the listener is aware that the speaker believes the proposition it granted, that is, $p \notin M_x M_y B_x$.

Definition 2 (granting an offer) *If $p \in B_x$, $p \in M_x D_y^{B_x}$ and $p \notin M_x M_y B_x$ then granting an offer $[x, y, p]^{!+}$ is applicable.*

An offer of information can be answered by granting or denying to accept to believe the proposition. The act of denying an offer $[x, y, p]^{!-}$ is similar to the act of granting with the difference that the speaker had not been persuaded to believe p, that is, $p \notin B_x$. Equal to the act of granting an offer, the speaker must be aware that the listener has the desire to induce a cognitive state change in the speaker, that is, $p \in M_x D_y^{B_x}$. To prevent that the denial is not superfluous, the speaker may not be aware that the listener is already aware that it has explicitly stated that it does not believe the proposition, that is, $p \notin M_x M_y I_x$.

Definition 3 (denying an offer) *If $p \notin B_x$, $p \in M_x D_y^{B_x}$ and $p \notin M_x M_y I_x$ then denying an offer $[x, y, p]^{!-}$ is applicable.*

A follow-up offer is an offer that substantiates some claim to believe another proposition. This offer is syntactically indistinguishable from the offer defined in Definition 1. However, the follow-up offer is a different speech act from a semantic perspective: it has the following preconditions. The speaker has the desire that the listener believes some proposition p, but the listener does (not yet) believe p and the speaker has already offered him p. The speaker may utter a follow-up offer regarding proposition q if q added to the listeners belief state would make him accept to believe p. Formally, if q is added set theoretically to $M_x B_y$, then p becomes part of the manifested belief state, that is, $p \in cl(M_x B_y \cup \{q\})$. The proposition q may in

this case be offered when the listener does not believe it and the speaker has not proposed it before.

Definition 4 (follow-up offer) *If* $p \in D_x^{B_y}$, $p \notin M_x B_y$, $p \in M_x M_y D_x^{B_y}$, $p \in cl(M_x B_y \cup \{q\})$, $q \notin M_x B_y$ *and* $q \notin M_x M_y D_x^{B_y}$ *then offer* $[x, y, q]^!$ *is applicable.*

2.7 Update Rules

Update rules define the agent's change of cognitive state given a communicative act directed at the agent.

After a speaker (x) has offered proposition p to a listener (y), that is, after $[x, y, p]^!$, the following properties for the cognitive states hold. The listener is aware that the speaker desires the listener to believe proposition p, that is, $p \in M_y D_x^{B_y}$, and the speaker is aware that the listener is aware of this, that is, $p \in M_x M_y D_x^{B_y}$. In addition, after an offer the listener is aware that the speaker believes the proposition, that is, $p \in M_y B_x$, and the speaker is aware that the listener is aware of this, $p \in M_x M_y B_x$. Offering a proposition may have the effect that the listener is persuaded to believe the proposition ($p \in B_y$). Note that being persuaded is not encoded in update rules but in the agent's cognitive processes (Section 2.4).

Definition 5 (offer) *after the update of an offer* $[x, y, p]^!$ *holds that* $p \in M_y D_x^{B_y}$, $p \in M_x M_y D_x^{B_y}$, $p \in M_y B_x$ *and* $p \in M_x M_y B_x$.

After a speaker (x) has granted an offer regarding proposition p to a listener (y), that is $[x, y, p]^{!+}$, the following properties for the cognitive state hold. The listener is aware the speaker believes proposition p, that is, $p \in M_y B_x$, and the speaker is aware that the listener is aware of this, that is, $p \in M_x M_y B_x$. Remember, this mental construct is used to represent that the speaker has granted the offer.

Definition 6 (granting an offer) *after the update of granting an offer* $[x, y, p]^{!+}$ *holds that* $p \in M_y B_x$ *and* $p \in M_x M_y B_x$.

After a speaker (x) has denied an offer regarding proposition p to a listener (y), that is $[x, y, p]^{!-}$, the following properties for the cognitive state hold. Similar to granting an offer, after denying an offer, the listener is aware that the speaker does not believe the proposition, that is, it is ignorant towards it, $p \in M_y I_x$. Also, the speaker is aware of this, which is represented by $p \in M_x M_y I_x$.

Definition 7 (denying an offer) *after the update of denying an offer* $[x, y, p]^{!-}$ *holds that* $p \in M_y I_x$ *and* $p \in M_x M_y I_x$.

3 Agree to Disagree

If a group of experts are unable to agree on a decision when for example two experts disagree on some propositions that are needed to agree, a persuasion dialogue may resolve the disagreement by adding information to the expert's belief state (Walton and Krabbe, 1995). If all methods to persuade have become exhausted, the experts can conclude that they disagree on a specific subject and that they both agree on this. This agreement may trigger the meta dialogue in which for example a coin flipping method is proposed to resolve the problem. In this section, a dialogue rule for agreeing to disagree is proposed, making the disagreement common belief.

3.1 Disagreement

Two pieces of information are in disagreement when they are not subsumed under each other in the information order. Stated differently, two pieces of information disagree when the truth-values representing the information are not related with respect to the information order $\leq_k$. For example, t disagrees with f because t $\not\leq_k$ f and f $\not\leq_k$ t, but u agrees with t because u $\leq_k$ t. An agent believing a proposition ψ : u and another agent believing ψ : t do not disagree about ψ, the latter agent is just more informed than the former naive agent. Equally, t agrees with i because t $\leq_k$ i, see also Figure 1.

A *disagreement* between two agents x and y about the truth-value of term ψ exists if (and only if) x believes a proposition $\psi : \theta_1$ and y believes proposition $\psi : \theta_2$, and the truth-values disagree. In line 1 of the example dialogue, Tv states that the car is a Ferrari. After the update of cognitive states, the Ia believes this and he concludes that this car is not a safe car, but he is not yet aware that Tv believes that this car is safe.

3.2 Awareness of Disagreements

An agent (x) is aware of a disagreement with another agent (y) if and only if x believes a proposition ψ:θ_1 and x is aware that y believes proposition ψ:θ_2 and θ_1 disagrees with θ_2. In line 3 of the example dialogue Tv states that its car is safe. After the cognitive state update of this act the Ia is aware that Tv believes that the car is safe; the Ia is now aware of a disagreement because it believes that the car is safe. After line 4, Tv is also aware of the disagreement.

Under the assumption that the dialogue only results in additions of the agents' cognitive states, it can be proven that disagreement awareness always implies the existence of a disagreement and that it is not possible that agents are incorrectly aware of a disagreement. Note that if agents have a disagreement, they need not be aware of this.

A second-order disagreement awareness exists when an agent (x) is aware that another agent (y) believes a proposition ψ:θ_1 and x is aware that y is aware that x believes proposition $\psi:\theta_2$ and θ_1 disagrees with θ_2. In line 5 the Ia states that it believes that the car is not safe, after the update, the Ia is aware of a second order disagreement. After line 6, Tv is also aware of this disagreement.

3.3 Resolving Disagreements

The minimal piece of information that is needed to resolve a disagreement between two agents about a term ψ is represented by proposition $\psi:\xi$ that if added to one of the agent's belief state resolves the disagreement. Remember that in the current dialogue game only additions of information are possible and consequently, resolving disagreement can only take place by adding sufficient information to one of the two agent's, rendering it possibly inconsistent.

3.4 Dialogue Rule to Agree to Disagree

The situation in which participants may utter an agreement to disagree can now be equated in a dialogue rule. The speaker (x) may propose to agree to disagree about term ψ to the listener (y) if:

1. The speaker is aware that it has a disagreement about term ψ with the listener, that is, ψ:$\theta_1 \in B_x$, ψ:$\theta_2 \in M_x B_y$ and θ_1 disagrees with θ_2.

2. The speaker is aware that the listener is also aware of the disagreement, that is, $\psi : \theta_3 \in M_x M_y B_x$ and θ_3 disagrees with θ_2.

3. The speaker is *not* aware of a set of propositions that it has not offered to the listener before that could have resolved the disagreement if the listener had accepted to believe them. Suppose the proposition $\psi:\xi_1$ represents the minimal amount of information that if added to the listener's belief state resolves the disagreement. If for a set of propositions Φ that is believed by the speaker ($\Phi \subseteq B_x$) holds that if Φ were added to the listener's belief state then the disagreement would have been resolved, that is, ψ:$\xi_1 \in cl(M_x B_y \cup \Phi)$. Furthermore, if a set of proposition $\Phi \subseteq B_x$ has been offered to the listener then holds that $\Phi \subseteq M_x M_y D_x^{B_y}$. We can now state that the speaker has no methods (sets of propositions Φ) left to persuade the listener by $(\forall \Phi \subseteq B_x)(\psi$:$\xi_1 \in cl(M_x B_y \cup \Phi) \Rightarrow \Phi \subseteq M_x M_y D_x^{B_y})$.

4. According to the speaker, the listener is *not* aware of a set of proposition not offered to the speaker before that can resolve the disagreement if the speaker accepts to believe them. That is, for all sets of propositions Ψ believed by the listener according to the speaker ($\Psi \subseteq M_x B_y$) holds that if these propositions Ψ were believed by the speaker then the speaker would have believed ψ:ξ_2 representing the minimal information that is needed to resolve the disagreement (according to the listeners that the speaker is aware of), that is, $\psi:\xi_2 \in cl(M_x M_y B_x \cup \Psi)$. If all these propositions Ψ have been offered and apparently this did not resolve the disagreement, that is $\Psi \subseteq M_x M_y I_x$, the speaker is aware that the listener has no methods left to resolve the situation. Formally, $(\forall \Psi \subseteq M_x B_y)(\psi : \xi_2 \in cl(M_x M_y B_x \cup \Psi) \Rightarrow \Psi \subseteq M_x M_y I_x)$.

5. The speaker is not aware that it proposed to agree to disagree regarding the disagreement before. Agreements to disagree are kept record of similar to manifested beliefs and desires. $M_x M_y \Delta(x, y)$ states that x is aware of the agreement to disagree by x between the agents.

If these five preconditions hold, the speaker may propose an agreement to disagree regarding ψ, denoted $[x, y, (\psi$:θ_3, ψ:$\theta_2)]^{\Delta}$.

3.5 Generation of the Example Dialogue

The proposed dialogue game gives preconditions to offer information. Combined with the formal rule of the speech acts, the update rules and the (not presented) axioms of theories of MVL, the example dialogue can be proven to follow from the initial collective state. Note that the agreement to disagree is based on information regarding the two agents involved in the disagreement. This disagreement may need to be retracted when new information is gained on how to resolve the situation. This is because the agreement is based on the absence of beliefs of only the two agents that actually have the disagreement. If another agent offers a proposition that resolves the disagreement, the agreement to disagree needs to be retracted.

With the help of software tools, the complete state space of the dialogue game with its initial collective state is generated in one tenth of a second, resulting in a graph with 37 nodes representing the collective states and with 66 edges representing speech act utterances with associated cognitive state updates. This network comprises 177 different dialogues with three different final collective states. One has to remember that an agent accepts to believe a proposition if it is consistent with its current belief state. This makes offering information of crucial importance, resulting in the three different endings.

4 Conclusions

We have given a formal semantics to the act of uttering a proposal to agree to disagree; these semantics are defined by formulating the rules of usage in the context of a computational dialogue game for offering information. We have shown that semantics of communicative acts can be given with a dialogue game in an intuitive manner, and that given the dialogue game a formal system emerges in which sequences of communicative acts can be checked valid dialogues. Also dialogues can be generated from the dialogues and update rules providing the possibility to analyse dialogue games on useful properties like termination or whether the unbalanced desire/belief states are resolved in the terminating collective states.

If two agents utter their agreement to disagree they are mutually aware of the disagreement they have about some proposition, this awareness is vital in group decisions because these decisions should not be based on information that agents disagree upon. Once agents agree to disagree about a proposition it cannot be used in future reasoning, even previous decisions in which this proposition played a role may be compromised.

Future research addresses agents that strategically select which speech acts to utter with the intention to arrive at a collective state in which desirable properties hold. Other research centres around speech acts for retracting information. Retracting information is an act of offering *not* to believe a proposition, that is, an offering to forget. Retractions of information enhance the current dialogue game by to retracting agreements to disagree when new information comes to light.

References

[Belnap Jr.1977] N.D. Belnap Jr. 1977. a useful four-valued logic. In J.M. Dunn and G. Epstein, editors, *Modern Uses of Multiple-Valued Logic*, pages 8–37.

[Beun2001] R.J. Beun. 2001. On the Generation of Coherent Dialogue: A Computational Approach. *Pragmatics & Cognition*, 9(1):37–68.

[Ellenbogen2003] S. Ellenbogen. 2003. *Wittgenstein's Account of Truth*. SUNY series in philosophy. State University of New York Press.

[Fitting1991] M. Fitting. 1991. Bilattices and the semantics of logic programming. *J. of Logic Programming*, 11:91–116.

[Ginsberg1988] M.L. Ginsberg. 1988. Multivalued Logics: A Uniform Approach to Reasoning in Artificial Intelligence. *Comput. Intelligence*, 4:265–316.

[Grice1975] H.P. Grice. 1975. Logic and conversation. In P. Cole and J.L. Morgan, editors, *Speech Acts*, volume 11 of *Syntax and Semantics*, pages 41–58.

[Labrou2001] Y. Labrou. 2001. Standardizing Agent Communication. *LNCS*, 2086:74–98.

[Lebbink et al.2003] H.-J. Lebbink, C.L.M. Witteman, and J.-J.Ch. Meyer. 2003. Dialogue Games for Inconsistent and Biased Information. *presented at LCMAS03 (www.win.tue.nl/ẽvink/lcmas03.html)*.

[Rescher1969] N. Rescher. 1969. *Many-valued Logic*. McGraw-Hill.

[Walton and Krabbe1995] D.N. Walton and E.C.W. Krabbe. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press. Albany, NY, USA.

Denial and correction in Layered DRT

Emar Maier
Department of Philosophy
University of Nijmegen
e.maier@phil.kun.nl

Rob van der Sandt
Department of Philosophy
University of Nijmegen
rob@phil.kun.nl

Abstract

The central characteristic of denials is that they perform a non-monotonic correction operation on discourse structure. A second characteristic is that they may be used to object to various kinds of information including presuppositions and implicatures. In this paper we first use standard DRT to capture these features, implement an earlier proposal of van der Sandt (1991) in DRT and point out a shortcoming of that approach. We then adopt *Layered DRT*. LDRT is an extension of standard DRT designed to represent and interpret different types of information conveyed in a conversation by distributing them over separate layers of the same LDRS. We will then show how LDRT allows us to solve the problems of the classic monostratal system. The resulting system makes use of a *directed reverse anaphora* mechanism to locate, remove and negate the material objected to.

1 Introduction

In the literature on denial and correction we can roughly distinguish two views. The dominant view has it that these phenomena are best captured within a general view of the semantics and pragmatics of negation. In terms of speech act theory this means that negatory force is located in the negative morpheme. A particular influential statement of this view is found in Horn (1985; 1989), who distinguishes between an ordinary 'descriptive' negation operator and a so-called metalinguistic negation. The former is the standard logical connective contributing to truth-conditional content, the latter is a non-truth-functional device that is meant to account for rejection of non-propositional material. This comprises objections to implicatural and presuppositional information, and to infelicities arising from style and register. Horn characterizes this operator as a metalinguistic device which can be used to signal an objection to an utterance on whatever grounds. This view has a number of drawbacks. It implies that some instances of of natural language negation are not interpretable by semantic means. It also implies an ambiguity in natural language negation which is not realized in any known language.

We defend a second view according to which the semantic notion of negation and the speech act notion of denial are fully independent. Denial patterns with assertion in that its nature and function should be accounted for in terms of the discourse effects it gives rise to. Just as the primary function of assertion is to convey new information, the primary function of a denial is to object to information which has been entered before and to remove it from the discourse record. The central characteristic of denial then is that it performs a non-monotonic correction operation on contextual information. This approach has several advantages. It yields a uniform account of denial, which comprises the standard 'propositional' and

the marked cases, it does not force us to postulate an ambiguity in the natural language negation, and by incorporating ideas from Levinson (2000), van Leusen (1994; to appear) and van der Sandt (1991; 2003) it can be implemented in a natural way in a dynamic theory of discourse interpretation.

2 Denials in DRT

2.1 Dialogue

Van der Sandt (1991; 2003) gives a mechanism to locate and remove the information conveyed by a previous utterance in a discourse in order to account for the non-incremental behavior of denials in discourse. An implementation of this account in DRT requires some minor extensions to standard DRT; we need to keep track of who said what and when in a dialogue in order to locate and remove a previous speaker's contribution. We let a discourse be a sequence of sentences, $\sigma_1, \ldots, \sigma_n$.[1] The first addition to the syntax of the DRS language is harmless: we index every DRS condition and reference marker with a natural number specifying the sentence σ_i it originated from.[2]

Our goal is now to incrementally build DRSs, $\varphi_1, \ldots, \varphi_n$, representing the various stages of the ongoing discourse. We assume a given background representation φ_0 as starting point for the incrementation process.

2.2 Star conditions and reverse anaphora

We will assume that whenever an utterance is analyzed as a denial of a previous utterance, this denial is not further parsed, but leaves a simple negated $\star$-condition ($\neg[\ \star : \star\ ]$) in the DRS. DRS-construction now proceeds as follows. We first construct a preliminary sentence DRS from the parse of the sentence and merge it with the background DRS. The standard presuppositional resolution mechanism will then bind or accommodate the unresolved presuppositions and other anaphoric expressions (van der Sandt, 1992; Geurts, 1999). In the case of an assertoric utterance the output of this process is a new DRS which monotonically accumulates the new information conveyed by the utterance.

However, if the utterance under consideration gives rise to a $\star$-condition, we resolve it by a mechanism we call *reverse anaphora*. This mechanism collects all material originating from the previous utterance and *moves* it to the position of the $\star$. Reverse anaphora thus has a non-monotonic effect on the discourse structure established: the contribution of the previous utterance is removed from the main DRS and entered under the scope of the negation introduced by the denial.[3]

2.3 Examples I

The following example of a presupposition denial illustrates the procedure in some detail:

(1) σ_1 The King of France walks in the park.
σ_2 No, he doesn't,
σ_3 France doesn't have a king.

Let the background φ_0 be empty. The presuppositional expression of σ_1 (represented here by underlining) thus has to be accommodated. We denote the algorithms responsible for building preliminary sentence DRSs, resolution, and reverse anaphora, as **Prel**, **Res**, and **RA** respectively. $\oplus$ is the merge operator.

(2) a. $\mathbf{Prel}(\sigma_1) =$
$[\ \underline{x_1} : \underline{\text{King of France}_1(x)},$
$\text{walk in park}_1(x)\ ]$

b. $\varphi_1 = \mathbf{Res}(\varphi_0 \oplus \mathbf{Prel}(\sigma_1)) =$
$[\ x_1 : \text{King of France}_1(x),$
$\text{walk in park}_1(x)\ ]$

c. $\varphi_1 \oplus \mathbf{Prel}(\sigma_2) =$
$\big[\ x_1 : \text{King of France}_1(x),$
$\text{walk in park}_1(x), \neg_2[\ \star : \star\ ]\ \big]$

d. $\varphi_2 = \mathbf{RA}\big(\mathbf{Res}(\varphi_1 \oplus \mathbf{Prel}(\sigma_2))\big) =$
$\big[\ \sout{x_1} : \sout{\text{King of France}_1(x)},$
$\sout{\text{walk in park}_1(x)},$
$\neg_2[\ x_1 : \text{King of France}_1(x),$
$\text{walk in park}_1(x)\ ]\ \big]$

[1] We will sometimes use the terms 'utterance' and 'sentence' rather loosely. The idea behind our terminology is that an utterance can be conceived of as a sentence combined with a representation of the *context*, where our notion of context comprises both a DRS representing the common ground and a Kaplanian context specifying the actual world, the speaker and the time of the sentence under discussion.

[2] Who said what and when can now be encoded in DRS conditions like *speaker(x,37)* and *time(t,37)*, meaning that x uttered σ_{37} at time t.

[3] Cf. the account of Levinson (2000) who calls this process *quasi-anaphora*

e. $\varphi_3 = \mathbf{Res}(\varphi_2 \oplus \mathbf{Prel}(\sigma_3)) =$
$\big[$: $\neg_2$[x_1 : King of France$_1$(x),
walk in park$_1$(x)],
$\neg_3$[x_3 : King of France$_3$(x)] $\big]$

The upshot is that the final interpretation, φ_3, is equivalent to the predicate logical formula $\neg\exists x\big[\text{KF}(x) \wedge \text{walk}(x)\big] \wedge \neg\exists x\big[\text{KF}(x)\big]$ which is easily seen to be equivalent to the intuitively correct $\neg\exists x[\text{KF}(x)]$.

2.4 Problems

Strawson's famous (3) illustrates that we cannot always just remove whatever is conveyed by the offensive utterance:

(3) a. A man jumped off the bridge.
b. He didn't jump, he was pushed.
Strawson (1952)

when processing (4b) we need to retain the discourse referent introduced for *the man* in (4a) in order to bind the anaphoric pronoun.

Geurts (1998) argued against the strategy of removing the *full* contribution of the utterance objected to, pointing out that the problem is a general one. Consider an example where several presuppositions are involved:

(4) a. The King of France knows I quit smoking.
b. No he doesn't, France doesn't have a king.

The presupposition objected to is that France has a king and should be removed from the discourse record. The presupposition that I quit smoking seems to pass unharmed. We thus need to allow for the possibility that the final representation retains the latter information.

Another argument against removing the full contribution of a sentence is the fact that speakers sometimes overtly acknowledge parts of the last utterance's contribution while rejecting other parts. A case in point is (5) where the second speaker confirms that the person referred to is nice, but denies the implicature evoked by the use of the indefinite term 'a lady'.[4]

(5) a. Now, THAT's a nice lady
b. Yes, she is, but she's not a LAdy, she's my WIfe

In order to account for examples like (5) we need to represent and *interpret* different types of information conveyed by the very same utterance separately, thus enabling us to remove specific types information while retaining others.

3 Using layers

We will use layered representations in order to *direct* the reverse anaphora module at the *offensive* layers only. This requires a further extension of DRT which enables us to both encode and interpret the various kinds of information. We adopt Layered DRT (or LDRT for short).[5]

3.1 LDRT's syntax

The syntax of LDRT is like that of standard DRT, except that every discourse referent and DRS condition is paired with a layer label, specifying which kind of information it encodes. For the purpose of this paper we index every label with a number corresponding to the number of the sentence from which it originated (and 0 for background information). We will here restrict ourselves to the set of labels Λ_0 given in (6):

(6) $\Lambda_0 = \{0,\ fr_1, \ldots, fr_n,\ acc_1, \ldots, acc_n,$
$imp_1, \ldots, imp_n,\ p_1, \ldots, p_n\}$
- $fr_i \approx$ the Frege layer, what is "said" in the contribution of σ_i
- $acc_i \approx$ accommodated material coming from p_i-layer
- $imp_i \approx$ implicature invoked by σ_i
- $p_i \approx$ presupposition triggered by σ_i

The primitive symbols of our LDRT language fragment are:

- a set $\mathcal{X}$ of reference markers

[4] Adapted from the classic:
A: That was a nice lady I saw you with last night.
B: That wasn't a LAdy, that was my WIfe
Note that by his utterance B certainly does not imply that his wife is not a lady as a purely truth-functional analysis would predict. Grice (1989: 37) remarks: 'Anyone who uses a sentence of the form *X is meeting a woman this evening* would normally implicate that the person to be met was someone other than X's wife, mother, sister, or perhaps even close platonic friend.'

[5] See Geurts and Maier (ms).

- (for some n) a set $\mathcal{P}red^{(n)}$ of n-place predicates
- the set Λ_0 of layer labels

The syntactic rules are as follows:

(7)
a. if $x \in \mathcal{X}$, $L \subseteq \Lambda_0$, then $x_L = \langle x, L\rangle \in \mathcal{X} \times \wp(\Lambda)$ is a labeled reference marker
b. if $P \in \mathcal{P}red^n$, $L \subseteq \Lambda_0$, then P_L is a labeled predicate
c. if $x, y \in \mathcal{X}$, $L \subseteq \Lambda_0$, then $x =_L y$ is a labeled condition
d. if $x_1, \ldots, x_n \in \mathcal{X}$, P_L a labeled n-place predicate, then $P_L(x_1, \ldots, x_n)$ is a labeled condition
e. if φ and ψ are labeled conditions, $L \subseteq \Lambda_0$, then $\neg_L\varphi$, $\varphi \vee_L \psi$, and $\varphi \Rightarrow_L \psi$ are labeled conditions
f. if U is a set of labeled reference markers and Con a set of labeled conditions, then $\langle U, Con\rangle$ is an LDRS (notation: $\varphi = \langle U(\varphi), Con(\varphi)\rangle$)

The following is an example of a sentence and its preliminary LDRS representation:[6]

(8)
a. σ_1 = It is possible the Pope is right.
b. $\big[x_{p_1}$: pope$_{p_1}(x)$,
$\Diamond_{fr_1}$[: right$_{fr_1}(x)$],
$\neg\Box_{imp_1}$[: right$_{imp_1}(x)$]$\big]$

This representation encodes the fact that the definite description 'the Pope' has a presuppositional status. Given an individual that satisfies this predicate the classical truthconditional (Fregean) contribution is that *he* is possibly right, and with a typical utterance of σ_1 the speaker furthermore conveys the implicature that *he* is not necessarily right. The crucial feature of the LDRT representation is that all conditions inhabit their own layer but nevertheless employ one and the same reference marker; the presuppositional, assertoric and implicature expressions are all linked to the same discourse referent, thus capturing the fact that they all attribute some property to the same individual.

[6] We will use $\neg\Box_{imp_1}$[: right$_{imp_1}$(x)] as a notational shorthand for $\neg_{imp_1}\big[$: $\Box_{imp_1}$[: right$_{imp_1}$(x)] $\big]$

3.2 LDRT's semantics

The idea of LDRT's semantics is that the truth definition is relativized to a set of labels, which means that we ignore all material labeled otherwise. The truth definition is thus only a minor extension of the standard definition of truth for DRSs in terms of verifying or truthful embeddings, requiring only two small adaptations: First, since ignoring certain layers often causes the remaining layers to form only an incomplete representation, missing perhaps some essential reference markers, the truth definition should be partial, i.e. for some choices of layer sets a truthvalue is undefined. In the following we will use $\uparrow$ and $\downarrow$ to abbreviate undefinedness and definedness respectively.

Secondly, since we are going to talk about propositions, we need at least an intensional version, i.e. we need possible worlds. An *intensional model* for our LDRT language fragment is a triple $\langle D, W, R\rangle$, where D is a set of individuals, W a set of 'possible worlds', which we take to be interpretation functions mapping basic predicates onto their extensions, and an an accessibility relation R ($\subseteq W^2$). With respect to such models the clauses below (9) give a partial definition of truth:

(9) For $L \subseteq \Lambda_0$, an LDRS φ ,and embedding functions f, g, define:
a. $U_L(\varphi) = \{x \in \mathcal{X} \mid \exists K[K \cap L \neq \emptyset \wedge x_K \in U(\varphi)]\}$
b. $Con_L(\varphi) = \{\psi \mid \psi$ is an LDRS condition $\wedge\ \exists K[K \cap L \neq \emptyset \wedge \psi_K \in Con(\varphi)]\}$
c. $f[X]g = f \subseteq g \wedge Dom(g) = Dom(f) \cup X$

Let $\mathcal{M} = \langle D, W\rangle$ be an intensional model. Let f be a partial embedding of the set of reference markers into D, $L \subseteq \Lambda_0$, $w \in W$, and φ an LDRS.

- $\|\varphi\|^f_{L,w} =$

$$\begin{cases} \{g \mid f[U_L(\varphi)]g \wedge \wedge\forall\psi \in Con(\varphi)\big[\|\psi\|^g_{L,w} = 1\big]\} & \text{if } \exists g\big[f[U_L(\varphi)]g \wedge \wedge\forall\psi \in Con(\varphi)\big[\|\psi\|^g_{L,w} \downarrow\big]\big] \\ \uparrow & \text{otherwise} \end{cases}$$

- $\|x =_K y\|^f_{L,w} =$

$$\begin{cases} 1 & \text{if } K \cap L = \emptyset \text{ or} \\ & \quad x, y \in Dom(f) \wedge f(x) = f(y) \\ 0 & \text{if } K \cap L \neq \emptyset \text{ and} \\ & \quad x, y \in Dom(f) \wedge f(x) \neq f(y) \\ \uparrow & \text{otherwise} \end{cases}$$

- $\|P_K(x_1, \ldots, x_n)\|^f_{L,w} =$

$$\begin{cases} 1 & \text{if } K \cap L = \emptyset \text{ or} \\ & \quad \big(x_1, \ldots, x_n \in Dom(f) \wedge \\ & \quad \wedge \langle f(x_1), \ldots, f(x_n)\rangle \in w(P)\big) \\ 0 & \text{if } K \cap L \neq \emptyset \text{ and} \\ & \quad x_1, \ldots, x_n \in Dom(f) \wedge \\ & \quad \wedge \langle f(x_1), \ldots, f(x_n)\rangle \notin w(P) \\ \uparrow & \text{otherwise} \end{cases}$$

- $\|\neg_K \psi\|^f_{L,w} =$

$$\begin{cases} 1 & \text{if } K \cap L = \emptyset \text{ or } \|\psi\|^f_{K\cap L,w} = \emptyset \\ 0 & \text{if } K \cap L \neq \emptyset \text{ and } \|\psi\|^f_{K\cap L,w} \downarrow \text{ and} \\ & \quad \|\psi\|^f_{K\cap L,w} \neq \emptyset \\ \uparrow & \text{otherwise} \end{cases}$$

- $\|\psi \vee_K \chi\|^f_{L,w} =$

$$\begin{cases} 1 & \text{if } \|\psi\|^f_{K\cap L,w} \downarrow \wedge \|\chi\|^f_{K\cap L,w} \downarrow \text{ and} \\ & \quad \|\psi\|^f_{K\cap L,w} \cup \|\chi\|^f_{K\cap L,w} \neq \emptyset \\ 0 & \text{if } \|\psi\|^f_{K\cap L,w} = \|\chi\|^f_{K\cap L,w} = \emptyset \\ \uparrow & \text{otherwise} \end{cases}$$

- $\|\psi \Rightarrow_K \chi\|^f_{L,w} =$

$$\begin{cases} 1 & \text{if } \|\psi\|^f_{K\cap L,w} \downarrow \wedge \forall g \in \|\psi\|^f_{K\cap L,w} : \\ & \quad \|\chi\|^g_{K\cap L,w} \downarrow \wedge \|\chi\|^g_{K\cap L,w} \neq \emptyset \\ 0 & \text{if } \exists g \in \|\psi\|^f_{K\cap L,w} : \ \|\chi\|^g_{K\cap L,w} = \emptyset \\ \uparrow & \text{otherwise} \end{cases}$$

- $\|\Box_K \psi\|^f_{L,w} =$

$$\begin{cases} 1 & \text{if } \forall w' R w : \ \|\psi\|^f_{K\cap L,w'} \downarrow \wedge \neq \emptyset \\ 0 & \text{if } \exists w' R w : \ \|\psi\|^f_{K\cap L,w'} = \emptyset \\ \uparrow & \text{otherwise} \end{cases}$$

Layered propositions or contents expressed by LDRSs are defined as follows:

(10) $\mathcal{C}^f_L(\varphi) =$

$$\begin{cases} \{w | \exists g \in \|\varphi\|^f_{L,w}\} & \text{if } \exists w[\|\varphi\|^f_{L,w} \downarrow] \\ \uparrow & \text{otherwise} \end{cases}$$

$\mathcal{C}_L(\varphi) = \mathcal{C}^{\emptyset}_L(\varphi)$

This defines a wide variety of content notions at once. For example $\|(8b)\|_{\{p_1, fr_1\}}$ is the proposition that the Pope is possibly right, and[7] $\|(8b)\|_{p_1, imp_1}$ the proposition that the Pope is not necessarily right.

In order to determine e.g. $\|(8b)\|_{imp_1}$, the proposition *implicated* by a sentence (a type of content we will often need), we might use a more sophisticated two-dimensional semantics, allowing us to determine such *open* propositions with respect to a backgroundable set of layers in addition to the layer under evaluation. See Geurts and Maier (ms) for such a two-dimensional account. Here we will follow a simpler strategy to determine a workable truthconditional substitute for *prima facie* undefined open propositions like $\|(8b)\|_{imp_1}$. The idea is to enlarge the set of evaluated layers until a proposition is expressed. Assuming that the layers we add, also express a (more general) proposition we get a substitute for the original undefined content by subtracting these two from each other.

(11) a. $\mathcal{C}^{[K]}_L(\varphi) =$
$W - (\mathcal{C}_K(\varphi) - \mathcal{C}_{K\cup L}(\varphi)) =$
$\big\{w \in W \,\big|\, w \in \mathcal{C}_K(\varphi) \rightarrow w \in \mathcal{C}_{K\cup L}(\varphi)\big\}$

b. $\mathbf{Cl}(\varphi, L) =$ the smallest $K \subseteq \Lambda_0$ s.t. $\mathcal{C}_{K\cup L}(\varphi) \downarrow$ ('**Cl**osure set')

c. $\mathcal{C}^*_L(\varphi) = \mathcal{C}^{[\mathbf{Cl}(\varphi, L)]}_L(\varphi)$

It is easily verified that $\mathcal{C}^*_L(\varphi) = \mathcal{C}_L(\varphi)$ if both are defined. And if the plain layered proposition $\mathcal{C}_L(\varphi)$ is not defined, we can think of $\mathcal{C}^*_L(\varphi)$ as modeling the (weakest possible) truthconditional contribution of φ's L layers.

As an example, consider again the possibly right pope of (8): we saw that $\|(8b)\|_{imp_1}$ is undefined because the implicature layer lacks a reference marker. The *closure* of imp_1 is p_1, so:

(12) $\mathcal{C}^*_{imp_1}((8b)) =$
$W - (\mathcal{C}_{p_1}((8b)) - \mathcal{C}_{p_1 imp_1}((8b))) =$
the proposition that if there is a Pope, then there is a not necessarily right Pope

[7] leaving out subscripted and other obvious braces for convenience here and in the following

3.3 Directed reverse anaphora

The reverse anaphora algorithm depends on **Off**, a mechanism for singling out offensive material from an utterance that has been denied by a subsequent utterance. **Off** is defined as in (13), where L and K are sets of labels.

(13) $\mathbf{Off}(\varphi, K)$ = the smallest $L \subseteq \Lambda_0$ s.t. $\mathcal{C}^*_L(\varphi) \cap \mathcal{C}^*_K(\varphi) = \emptyset$

So $\mathbf{Off}(\varphi, K)$ gives us a set of labels that, in φ, clashes with a given set K, in other words, the smallest set of layers disjoint from K that cause φ as a whole to be contradictory.

To direct the reverse anaphora mechanism in an assertion-denial-correction sequence, σ_1-σ_2-σ_3 (as in (1)), we again represent the contribution of σ_2 by a condition of the form $\neg_{fr_2}[\ \star : \star\ ]$, as in 2.2. This however does not suffice for **Off** to determine which layer is offensive, since there is no contradiction yet. We thus first add the layered contribution of σ_3, causing a contradiction somewhere in our growing LDRS (say ψ). We then determine the cause of the contradiction on the basis of the fr-content of the correction: $\mathbf{Off}(\psi, \{fr_3\})$. The conflicting material in the offensive layer is now moved, by the modified reverse anaphora mechanism, to the place of the $\star$s thus ending up under the negation in layer fr_2.

Formally this comes down to the following redefinition of reverse anaphora, $\mathbf{RA}^*$

(14) $\mathbf{RA}^*_i(\varphi)$ = move $U_{\mathbf{Off}(\varphi, fr_i)}$ and $Con_{\mathbf{Off}(\varphi, fr_i)}$ from their original position in φ to where the $\star$'s are in φ (then erase the $\star$'s)[8,9]

[8] $\varphi \ominus \varphi_{\mathbf{Off}(\varphi, \{fr_{i+2}\})}\Big[[\ \star : \star\ ] \mapsto \varphi_{\mathbf{Off}(\varphi, \{fr_{i+2}\})}\Big]$

[9] Instead of moving the whole offensive layer determined by **Off** we may need an even more finegrained Reverse Anaphora algorithm moving only the smallest inconsistent subDRS within that layer. This is essentially what Van Leusen's (to appear) 'nonmonotonic updating' does in order to account for the discourse effects of denials and corrections in a compositional, monostratal variant of DRT. Something like this finer grained targetting of offensive stuff is needed to solve the problems posed by multiple accommodation (4b) and the Strawson example (3b). A definition of 'offensive material, conflicting with correction σ_i', depending on the notion of syntactic DRS inclusion, $\sqsubseteq$, would have to be added:

(i) $\mathbf{OFF}(\varphi, i)$ = the $\sqsubseteq$-smallest subDRS ψ, $\psi \sqsubseteq \langle U_{\mathbf{Off}(\varphi, fr_i)}, Con_{\mathbf{Off}(\varphi, fr_i)}\rangle$, s.t. $\mathcal{C}^*_{\mathbf{Off}(\varphi, fr_i)}(\psi) \cap \mathcal{C}^*_{fr_i}(\varphi) = \emptyset$

and $\mathbf{RA}^*$ would be modified accordingly:

(ii) $\mathbf{RA}^{*\prime}$ = move $\mathbf{OFF}(\varphi, i)$ to the position of the $\star$'s.

Interpretation now proceeds as follows:

- If $\langle \sigma_i, \sigma_{i+1}, \sigma_{i+2}\rangle$ is an assertion-denial-correction sequence, then $\varphi_{i+1} = \mathbf{Res}(\varphi_i \oplus \mathbf{Prel}(\sigma_{i+1})) = \varphi_i \oplus \big[\ : \neg_{\{fr_{i+1}\}}[\ \star : \star\ ]\ \big]$ and $\varphi_{i+2} = \mathbf{RA}^*_{i+2}\big(\mathbf{Res}(\varphi_{i+1} \oplus \mathbf{Prel}(\sigma_{i+2}))\big)$
- Otherwise, $\varphi_{i+1} = \mathbf{Res}(\varphi_i \oplus \mathbf{Prel}(\sigma_{i+1}))$ and similarly for φ_{i+2}

In the next section we will illustrate the workings of the definitions given above by applying them to some selected examples.

3.4 Examples II

Consider first an example of an implicature denial, cf. (8):

(15) σ_1 It is possible the Pope is right.
σ_2 No, it is not POssible,
σ_3 it is NEcessary that he is right.

We assume that the denial σ_2 pertains only to part of the content of σ_1 and that the correction σ_3 indicates that only the implicature is at stake. The step-by-step interpretation process of this dialogue according to the Directed Reverse Anaphora algorithm is shown below:

- $\varphi_0 = [x_0 : \text{pope}_0(x)]$
- $\mathbf{Prel}(\sigma_1) =$ $\big[x_{p_1} : \text{pope}_{p_1}(x),\ \Diamond_{fr_1}[: \text{right}_{fr_1}(x)],\ \neg\Box_{imp_1}[: \text{right}_{imp_1}(x)]\big]$
- $\varphi_1 = \mathbf{Res}(\varphi_0 \oplus \mathbf{Prel}(\sigma_1)) =$ $\big[x_0 : \text{pope}_0(x),\ \Diamond_{fr_1}[: \text{right}_{fr_1}(x)],\ \neg\Box_{imp_1}[: \text{right}_{imp_1}(x)]\big]$
- $\mathbf{Prel}(\sigma_2) = \big[\ : \neg_{fr_2}[\star : \star]\big]$
- $\varphi_2 = \mathbf{Res}(\varphi_1 \oplus \mathbf{Prel}(\sigma_2)) =$ $\big[x_0 : \text{pope}_0(x),\ \Diamond_{fr_1}[: \text{right}_{fr_1}(x)],\ \neg\Box_{imp_1}[: \text{right}_{imp_1}(x)],\ \neg_{fr_2}[\star : \star]\big]$
- $\mathbf{Prel}(\sigma_3) =$ $[z_{p_3} : \text{masc}_{p_3}(z),\ \Box_{fr_3}[: \text{right}_{fr_3}(z)]]$

- $\psi = \mathbf{Res}(\varphi_2 \oplus \mathbf{Prel}(\sigma_3)) =$
$\big[x_0 : \text{pope}_0(x),\ \Diamond_{fr_1}[: \text{right}_{fr_1}(x)],$
$\neg\Box_{imp_1}[: \text{right}_{imp_1}(x)],\ \neg_{fr_2}[\star : \star],$
$\Box_{fr_3}[: \text{right}_{fr_3}(x)]\big]$

- $\mathbf{Cl}(\psi, fr_3) = \{0\}$ so $\mathcal{C}^*_{fr_3}(\psi)$ is the proposition that if there is a pope, there is a necessarily right one (see (12)).

- $\mathbf{Off}(\psi, fr_3) = \{imp_1\}$, because $\mathcal{C}^*_{imp_1}(\psi)$ is the proposition that if there is a pope, there is a not necessarily right one, which contradicts $\mathcal{C}^*_{fr_3}(\psi)$, on the assumption that it is common ground between the participants that there exists some unique pope.

- $\varphi_3 = \mathbf{RA}^*(\psi) =$
$\big[x_0 : \text{pope}_0(x),\ \Diamond_{fr_1}[: \text{right}_{fr_1}(x)],$
$\cancel{\neg\Box_{imp_1}[: \text{right}_{imp_1}(x)]},$
$\neg_{fr_2}[: \neg\Box_{imp_1}[: \text{right}_{imp_1}(x)]],$
$\Box_{fr_3}[: \text{right}_{fr_3}(x)]\big]$

Note that van der Sandt (1991) and our implementation thereof was able to handle this example, too, giving an equivalent representation as the final output. However, in the course of the resolution process it temporarily threw out the whole contribution of the assertion objected to.

The example in (5), repeated below as (16), shows that such an omnilayered removal is problematic:

(16) σ_1 Now, THAT's a nice lady
σ_2 Yes, she is,
σ_3 but she's not a LAdy,
σ_4 she's my WIfe

The earlier monostratal proposal would in effect first have to somehow acknowledge that the person referred to is both nice and female with σ_2, but then by reverse anaphora remove that information from the discourse representation along with everything else conveyed by σ_1. The information that this person is female is then restored with the correction σ_4, but the suggestion that she is nice does not survive. To see how the layered version handles this case, note first that this example is a little more involved than the previous ones since it features an *affirmation* or acknowledgement, σ_2, of (part of the content of) σ_1, before the familiar denial (σ_3) and correction (σ_4).[10]

- $\varphi_0 = [x_0 : \text{woman}_0(x), \text{pointed\_at}_0(x)]$
(we start with a representation of someone's pointing at a woman)

- $\mathbf{Prel}(\sigma_1) = [y_{p_1} : \text{pointed\_at}_{p_1}(y),$
$\text{lady}_{fr_1}(y), \text{nice}_{fr_1}(y), \text{stranger}_{imp_1}(y)]$
(assuming the use of 'a lady' invokes the implicature that the person referred to is not a close relative of the addressee)

- $\varphi_1 = [x_0 : \text{woman}_0(x), \text{pointed\_at}_0(x),$
$\text{lady}_{fr_1}(x), \text{nice}_{fr_1}(x), \text{stranger}_{imp_1}(x)]$

- $\mathbf{Prel}(\sigma_2) = [z_{p_2} : \text{fem}_{p_2}(z), \text{nice}_{fr_2}(z)]$
(affirmation is treated just like asserted content)

- $\varphi_2 = \mathbf{Res}(\varphi_1 \oplus \mathbf{Prel}(\sigma_2)) =$
$[x_0 : \text{woman}_0(x), \text{pointed\_at}_0(x),$
$\text{lady}_{fr_1}(x), \text{nice}_{fr_1 fr_2}(x), \text{stranger}_{imp_1}(x)]$

- $\mathbf{Prel}(\sigma_3) = \big[\ : \neg_{fr_3}[\star : \star]\big]$
(the negative sentence σ_3 is marked as a denial)

- $\varphi_3 = \big[x_0 : \text{woman}_0(x), \text{pointed\_at}_0(x),$
$\text{lady}_{fr_1}(x), \text{nice}_{fr_1 fr_2}(x), \text{stranger}_{imp_1}(x),$
$\neg_{fr_3}[\star : \star]\big]$

- $\mathbf{Prel}(\sigma_4) = [w_{p_4} : \text{fem}_{p_4}(w), \text{wife}_{fr_4}(w)]$

- $\psi = \mathbf{Res}(\varphi_3 \oplus \mathbf{Prel}(\sigma_4)) =$
$\big[x_0 : \text{woman}_0(x), \text{pointed\_at}_0(x),$
$\text{lady}_{fr_1}(x), \text{nice}_{fr_1 fr_2}(x), \text{stranger}_{imp_1}(x),$
$\neg_{fr_3}[\star : \star], \text{wife}_{fr_4}(x)\big]$

- $\mathbf{Off}(\psi, fr_4) = imp_1$, since $\mathcal{C}^*_{fr_4} \cap \mathcal{C}^*_{imp_1} = \emptyset$
('if there is a woman pointed at, a woman pointed at is a stranger' and 'if there is a woman pointed at, a woman pointed at is my wife' do not strictly speaking contradict each other, but they do once we add the by

[10] The formal handling of affirmations like this in the current framework necessitates among other things some minor revision of the assumed consecutive numberings in the formal definitions of section 3.3. We will not spell out the details.

all means reasonable assumption that the dialogue partners assume the object of the pointing to be uniquely determined)[11]

- $\varphi_4 = \mathbf{RA}^*(\psi) =$
$\big[x_0 : \text{woman}_0(x), \text{pointed\_at}_0(x),$
$\text{lady}_{fr_1}(x), \text{nice}_{fr_1 fr_2}(x), \sout{\text{stranger}_{imp_1}(x)},$
$\neg_{fr_3}[: \text{stranger}_{imp_1}(x)], \text{wife}_{fr_4}(x)\big]$

4 Conclusion

In this paper we presented a representational account of the discourse function of denial in an extension of DRT and discussed two variants of what we called reverse anaphora. The first version follows van der Sandt (1991) by removing the full contribution of a previous utterance from the discourse record. We then presented *directed reverse anaphora* which removes the main objection against the undirected version by allowing us to direct and limit the retraction procedure to the offensive part of a previous utterance. The mechanism is able to do so by relying on a DRT-extension that enables us to encode and interpret information that historically has been labeled pragmatic and non-truth-conditional in nature. The account has moreover the advantage of being more general than Horn (1989) and Geurts (1998) in retaining a unified account of the propositional, implicature and presuppositional denials, providing a general semantics and accounting for the function of denials in terms of their non-monotonic discourse effects.[12]

[11] The two-dimensional treatment of open propositions alluded to on page 5 and discussed in (Geurts and Maier, ms) will give an immediate contradiction here.

[12] One might think that the style and register cases, discussed in Horn (1989) would escape our treatment since the information objected too does not seem propositional in an intuitive sense.

(i) They didn't call the POlice—they called the poLIce.

We should remark that LDRT does allow us to account of denials like (i), by positing a layer for intonational and other surface features. However, the semantics of formal layers turns out to require heavier semantic apparatus than we have space here to develop, so we leave out examples like (i) in our treatment of denial, and refer the interested reader again to (Geurts and Maier, ms).

References

Bart Geurts and Emar Maier. ms. Layered DRT. (2003).

Bart Geurts. 1998. The mechanisms of denial. *Language*, 74:274–307.

Bart Geurts. 1999. *Presuppositions and Pronouns*. Elsevier, Amsterdam.

H. Paul Grice. 1989. *Studies in the Way of Words*. Harvard University Press, Cambridge, Massachusetts.

Laurence R. Horn. 1985. Metalinguistic negation and pragmatic ambiguity. *Language*, 61(1):121–174.

Laurence R. Horn. 1989. *A Natural History of Negation*. Chicago University Press, Chicago.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic: Introduction to Modeltheoretic Semantics in Natural Language, Formal Logic and Discourse Representation Theory, Vol. 1*. Kluwer Academic Publishers, Dordrecht.

Stephen C. Levinson. 2000. *Presumptive Meanings*. The MIT Press, Cambridge, Massachusetts.

Peter F. Strawson. 1952. *Introduction to Logical Theory*. Methuen, London.

Rob A. van der Sandt. 1991. Denial. In *Papers from CLS 27(2): the parasession on negation*, pages 331–344. CLS.

Rob A. van der Sandt. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377.

Rob A. van der Sandt. 2003. Denial and presupposition. In Hannes Rieser Peter Kühnlein and Henk Zeevat, editors, *Perspectives on Dialogue in the New Millennium*. John Benjamins, Amsterdam.

Noor van Leusen. 1994. The interpretation of corrections. In Peter Bosch and Rob van der Sandt, editors, *Focus and Natural Language Processing*, pages 523–532, Heidelberg. IBM Deutschland, Institute for Logic and Linguistics.

Noor van Leusen. to appear. *Nonmonotonic update in description grammar*. Ph.D. thesis, University of Amsterdam, Amsterdam.

An empirical study of acknowledgment structures

Philippe Muller and Laurent Prévot
Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier
{muller,prevot}@irit.fr
http://www.irit.fr/~Philippe.Muller/
http://www.irit.fr/~Laurent.Prevot/

Abstract

The subject of our study is one type of "response" in dialogue, usually called acknowledgment or positive feedback. We show here how distinguishing between different acknowledgments is central to the establishment of information. The study is based on a french corpus of direction-giving dialogues which we have gathered. The factors that we investigate about the acknowledgments are their producer, their target and their scope. We focus on the relations between those features and linguistic discourse markers.

1 Introduction

The subject of our study is one type of "response" in dialogue, usually called acknowledgment, and how it relates to the dynamics of settled information in a conversation. Dialogue acts are commonly divided between initiations (assertions, questions, commands,...) and responses to initiations; something also called "forward communicative" and "backward communicative" functions in the work of (Core and Allen, 1997).

There have been several studies detailing the many roles that assertions, questions and answers can have in a conversation. Coding schemes for dialogue (Core and Allen, 1997; Carletta et al., 1997) take great care in distinguishing these functions. A lot of attention has been given to the question/answer pair and answers and how it interact semantically and pragmatically within a conversation (Asher and Lascarides, 1998; Ginzburg, 1994). Less emphasized is the role of all speech turns ensuring that information exchanged is properly interpreted (*feedback*). The important work of (Traum, 1994) has studied in some detail how these utterances play a role in deciding the status of information exchanged during a dialogue (mutually accepted or under discussion). He emphasizes that different levels of acknowledgment exist as proposed by (Clark, 1996; Allwood et al., 1992). It has often been noted that some utterances signal something has been heard and are marking expectations, while (Clark and Schaefer, 1989) for instance, mention different kinds of evidence that a speaker understands what has been previously said. We want to show here how distinguishing between such turns is central to the establishment of information, along with question/answer pairs; and how they can be accounted for in a structural theory for representing dialogue. We have thus studied the role and influence of several discourse markers on acknowledgments, in a french corpus of direction-giving dialogues which we have gathered.

Since we want to explicit the relational nature of such dialog acts, we have also studied the scope of such acts within the structure of a dialogue. We tried to integrate it in SDRT (Asher and Lascarides, 2003) a theory in which dialogue structure is defined as relations between utterances.

We present here a preliminary quantitative study of factors taking a part in various types of

acknowledgments. We based our study on empirical data, complementing more qualitative studies such as (Allwood et al., 1992; Novick and Sutton, 1994).

2 Types of acknowledgments and underlying processes

The feedback effects range from rejection to acceptance. Here, we will focus on positive feedback or acknowledgment. The speaker uttering the response might have heard, understood or agreed on the target of the backchannel. However, it is not often clear what factors take part in defining this level of acceptance, nor how they interact with the other functions of feedback. As example, an acknowledgment can support the current *initiative* state (*continuers, assessment* (Schegloff, 1982) or try to modify it (*incipient speakership* (Jurafsky et al., 1998)).

We do not make a difference *a priori* between acknowledgment of assertions, question/answer pair or complete sub-dialogues[1]. We see the question/answer pair as defining a kind of assertion about the topic given by the question. We believe the nature of the corresponding acknowledgment is the same, even though the conditions they impose on the actual form of the acknowledgment can vary.

In a first analysis, we have listed the following functions for the different kinds of feedback, in accordance with other works (Allwood et al., 1992; Clark, 1996; Carletta et al., 1997; Jurafsky et al., 1998) considering this topic (terminology may vary). We do not claim that it is very original or more relevant than those cited above, but gives a picture of the different concepts that seem to be at play.

weak acknowledgment (*continuer*, *support acknowledgment*) signals that what has been said and heard without necessarily accepting it. We will see that the most common markers for this phenomenon are *(oui, ouais, mhmm)* (in English *yes, mh*). In (Traum, 1994), this acknowledgment is a "grounding" act and belongs to the " utterance" level.

strong acknowledgment (*agreement, acceptance*) accepts an utterance either as true or as committing the receiver. It is mainly uttered with *oui, ok, d'accord* (in English *yes, okay*)[2]. In Traum's taxonomy, acceptance is "core speech act" and belongs to the "discourse" level. We put also *confirmation* in this category often associated with (*c'est ça, exactement*) (in English *that's it, exactly*). Indeed their originality comes from the status of the speaker (informant or not). Nevertheless, we will see that the form of such confirmations allow us to recognize them most of the time.

This classification is less fine grained (Clark, 1996) or (Allwood et al., 1992) who distinguish four levels of communication (*(i) contact-execution/attention, (ii) perception-presentation/identification, (iii) understanding-meaning/understanding, (iv) attitudinal reactions-proposition/consideration)*[3]. What we called *weak acknowledgment* covers the levels (i) and (ii); but, the determination of any difference between marking attention and perception seems hard to include in an annotation scheme without accurate prosodic analysis. We prefer to *infer* such difference from our basic annotated data. For the same kind of reasons our annotation will not integrate the fourth level. These remarks lead us to the simple weak/strong division. But, during the annotation task we will have only one kind of acknowledgment. The weak/strong division will come from succeeding inferences.

Feedback is associated with the establishment of different kind of objects:(i) propositions and/or their truth on the one hand (grounding), (ii) referents and their identity on the other hand (anchoring). (Clark and Schaefer, 1989) and (et

[1] Here we consider a question/answer compound strictly as one question and one answer. Any other kind of question/answer structures will be regarded as a subdialogue.

[2] Since *oui* and *ouais* are already markers for weak acknowledgment, the markers will be often ambiguous. We still think that is possible to go a bit further in the analysis of these turns by taking into account other information sources.

[3] In our list the first item belongs to Allwood's terminology and the second to the Clark's one. There is two terms in the second item because in Clark's grounding levels the actions of the speaker and of addressee are separated.

D. Wilkes-Gibbs, 1986) focus respectively on these different aspects.

grounding (understanding, settling proposition) is not related to the truth of utterances or the acceptance of an order. It is just a coordination between the speakers on what has been said. The participants agree on the content of an utterance but not necessarily on the truth of this constituent nor the acceptance of this constituent for the current purpose (Clark and Schaefer, 1989; Traum, 1994).
For instance: *y a un café [...] qui s'appelle le Matin.- le Matin* (there is a café called le matin - le matin) possibly followed by . *I don't know where it is*..

accepting (agreement), opposite to grounding, leads to the acceptance of the truth about the information agreed or at least it leads to the acceptance of the information regarding to the current purpose. In Clark's terminology, *grounding* describes the whole process of information establishment, thus accepting is just one level of this global process.
For instance: *tu prends la rue des Filatiers à gauche de là ou tu es - la rue des filatiers? - oui - ok* (you take the Filatiers street on the left of where you are - the Filatiers street? - yes - ok.).

anchoring (establishing referent) is finding an internal anchor in one's beliefs for what has been said by the other speaker (e.g. a common referent has been found, something crucial in our examples). This explains the difference between the following example and the one mentioned on *grounding* above: *c'est à dire après t'arrives à Esquirol quand tu continues.- ouais ouais je vois où c'est Esquirol* (then you arrive at Esquirol [Plaza] when you go on - yeah yeah i see where Esquirol is).

closing is terminating a span of discourse as a sub-dialogue (i.e a transaction in the discourse analysis literature – or an exchange) either successfully *voilà, c'est fini!* ("here, we're done"), or unsuccessfully (*en fait c'est pas grave*...) ("actually it doesn't matter"). Here we'll study only the positive (or successful) closure. Generally only the speaker with the initiative is allowed to perform a closure.

However it is not easy to find systematically the function of a given feedback utterance. Moreover interactions between those processes are complex. There is not direct entailment between them except maybe that *accepting* requires *grounding*. Accept an utterance requires also most of the time to anchor (establish all the referents within) it before. Finally a speaker could be wrong about what he understood without realizing it right away, allowing for later corrections.

To analyze how these notions are at play in conversations and how they interact with each other, we have studied our corpus with a special attention to the following factors:

- the linguistic cues of agreement (discourse markers, redundancies);
- the kind of response acknowledgments take part in;
- the kind of target acknowledgments have (mood and function);
- the kind of structure is agreed on (the contexts of acknowledgments).
- the role of the speaker (Is he the informant or the informee? Does he have the initiative at this very moment?)

3 A corpus of acknowledgments

To support our study of conversation structures, we have recorded a set of dialogues between French speakers located at two different places. Speakers talked to each other on a phone. Speaker A (*the giver*) had the task of explaining to speaker B (*the receiver*) how to get from where B was to the place where A was. A and B didn't know each other in advance. The corpus is made of 21 dialogues (about 9000 words) involving 23 speakers[4]. The conditions of the experiment gave little

[4]Most of the participants only recorded one dialogue, but some of them were involved in more than one (either in the same or in different roles).

indications to the *receiver* outside of the conversation itself (no signs, no *a priori* common knowledge), so we expected a lot of speech turns explicitly devoted to the settlement of information. This was all the more important as the task itself was discursive (speakers have to agree on the basis of a linguistic description of a route)[5]. We focus here on the conditions of acknowledgments during such dialogs and on their occurrence in acknowledgment structures. We use the *acknowledgment structure* term to emphasize our special attention on acknowledgment in discourse structures where such phenomenon occurs.

3.1 Lexical cues for acknowledgments and roles of the speaker

We have isolated a set of markers indicating positive feedback of the other speaker utterances, listed in table (1), along with approximate English equivalents. The determination of the set was not an easy task. The question was to determine (i) what can count as a discourse marker (DM) and (ii) what can be associated to positive feedback.

With respect to the former issue, the literature offers long (but non-exhaustive) lists of DM in English (Schiffrin, 1987; Aijmer, 2002; Stenström, 1994), in French (Auchlin, 1981; Colineau, 1997; Reboul and Moeschler, 1998) and in other languages. These studies conclude that any word: small group of words, or sound can be considered as a DM when it becomes grammaticalized.

With respect to positive feedback, existing works about DM in french, even when they consider interactional DM, are not focused on their feedback aspect (Auchlin, 1981). There is a lot of studies about linguistic clues signaling feedback in English (mainly developed to enrich dialogue act taxonomies) but it seems difficult to use them for french markers (See in Table 1). In fact, we decided to devote more attention to feedback because we observed a significant number of speech turns without propositional content (at least in a strict sense) Our first set of DM was created from the elements of these turns signaling a positive feedback obviously enough (e.g. 1). Within this set we only kept markers who can form a speech turn by themselves. We hope this selection filters markers of other phenomena like hesitating (*euh*) or attitudinal changes toward information (*ah, en fait*) (*ah, actually*)[6].

(1)
F_{41c}. c'est au 27 rue des Polinaires
R_{42a}. ouais je vois,
F_{42b}. en fait c'est rue des Polinaires,
R_{42c}. d'accord.
F_{43}. voilà
(it's the 27, Polinaires street – ok i see – in fact, its on Polinaires street – ok – that's it)

We add to the DM analysis the observation of *informationally redundant utterances* who help participants to infer acceptance as described in (Walker, 1996). Such redundancies are also illustrated in example (1:F_{42b}).

We found 337 various acknowledgments in our corpus, only 34 without either of these markers[7]. We indicate how many times each marker appears in an acknowledgment by one participant (and in parentheses, the number of times where it is the only marker in the utterance).

About speaker's variability, since there was many participants, none of them will have a too big influence on data. Local behaviour don't modify global picture.

The asymmetry between the two participants allowed us to investigate which markers were preferably used by someone with or without the initiative. In our corpus examples, the *giver* had globally the initiative of the explanation. The two "strong" acknowledgment markers *voilà* and *d'accord* seem respectively typical of the *giver* with the initiative and of the *receiver* without it. These are noisy data obviously, since our dialogues are mixed-initiative ones (Walker and Whittaker, 1990) (i.e in our context initiative can be locally taken by the receiver).

[5] In the end, we isolated 337 acknowledgment acts, in a total of 746 speech turns.

[6] See the very accurate studies of *ah, oh* and *actually* proposed by (Schiffrin, 1987; Aijmer, 2002).

[7] Half of which are turns that repeat part of the previous utterance, and the rest are marginal synonyms of cases listed. Note also that Table (1) makes up more than 337 since several markers can appear in the same utterance.

[8] For technical reasons, this category covers only utterances containing exactly a string from the target. It will be interesting to extend the treatment in order to detect cases where strings do not exactly match but are still redundant.

Table 1: Count and English equivalents of reported french acknowledgment markers

Count	French markers	Produced by Giver(alone)	by Receiver(alone)	English equivalents
141	*oui, ouais*	34(28)	107(85)	yes/yeah
67	*d'accord*	18(13)	49(33)	ok, I see
47	*voilà*	38(31)	9(9)	exactly, that's it
37	*ok*	9(6)	28(18)	ok
29	*mhmm*	8(8)	21(20)	mmmh
18	*bon*	10(5)	8(2)	now, ok, well
14	*je vois*	1(1)	13(2)	I see
12	repeat[8]	8	4	-
22	other	-	-	-

The study of which utterances contain multiple markers is also an indication of the strength of the acknowledgment in the rough scale mentioned before, with the extreme example 2.

As it was pointed out by one reviewer, this kind of acknowledgment could be a clue about attitudinal changes in the speaker's mind. We agree with this conclusion; but, we still believe that even when the producer of such feedback has changing opinions about the target information, at the end of the turn the information is strongly grounded or rejected. The reason is just that by producing such turns the speaker emphasizes her attention to this piece of information, and thus has to signal acceptance or rejection.

(2) F_{13a}. et voilà c'est là
R_{13b}. en face la Poste.
R_{14}. ah okay okay okay bon ben ouais d'accord.
(and there it is – facing the Post Office – ah ok ok ok well yes i see)

Conversely, the mumbling *mhmm* is practically always alone. It seems to confirm the intuition that it is only a weak form of backchannel (the only other case is *mhmm ... ouais*, *ouais* being a weak form also). We prefer to consider the combination of several acknowledgments markers as only one act instead of considering that each marker produces an act. We made this choice because in many cases DM were uttered together and very quickly. Another frequent phenomenon is the repetition of the same DM many times in a row. But we agree the case could be made for considering strictly one act for one marker. In the end we think that these two working methods should lead to the same conclusions. On one hand, the first one will be considering the complex properties of the combination of markers within an utterance. On the other hand, the second will study the combination of the acts.

3.2 Scope of the acknowledgments

Now we will turn to the difficult question of the scope of the feedback. Backward acts scope is a notoriously difficult issue. Here, the acknowledgment structure is partly based on the target(s) of the acknowledgments (see Table 2). If the target was a single segment or a set of segments within the same turn, we consider it as a *narrow scope* acknowledgment. If the target is one utterance performing a question/answer with a previous question, we say that it is a *QAP-scope* acknowledgment. Finally, if the target is another acknowledgment whose target is not a simple initiation, the segment under consideration will be set as a *wide-scope* acknowledgment[9].

We are aware of some flaws in this classification. For example, it is quite possible for the number of segments concerned by a QAP-scope acknowledgment to be the same as the a wide-scope one (in case of *assertive-ack-ack* and *interrogative-answer-ack* sequences). Another problem is when an answer is elaborated on several segments or turns, their acknowledgment will still be marked as a QAP-scope even when the scope is very wide. Further work is needed in order to make really accurate propositions; but, we think this work depends very much on the interpretation of question/answer structures.

[9] Agreement between annotators was good regarding the labelling of acknowledgments ($\kappa = 0.82$). Determining the targets of such acts was less convincing ($\kappa \approx 0.6$).

Table 2: Acknowledgments scope by cue words

	ouais	oui	ok	d'accord	voilà	mhmm	bon	Total
Narrow Scope	65	21	17	31	17	24	4	179
QAP Scope	12	1	5	8	7	1	7	41
Wide Scope	1	0	2	3	10	3	0	19
Total	78	22	24	42	34	28	11	239

3.3 Function of the acknowledgment target

As another preliminary step in the study of acknowledgment structures, we have looked at the function of the the previous utterance with the context (her relational function).

The data presented in table (3) distinguishes between task-related assertions (describing an itinerary: introduction of landmarks (e.g. 3:F_{1b}) or description of landmarks (e.g. 3:F_{1c}), instructions (e.g. 3:F_{1a}) and comments) and interactional segments which are not related to the task (mainly feedback turns).

We think these distinctions are important with respect to the difference between acceptance and anchoring, since anchoring is mainly about landmark management. The segment concerned by the management of landmarks mainly aims to anchor the referents they include. On the other side, the instruction needs to be grounded/accepted. To sum up, anchoring underlies the establishment of "managing referent segments" and grounding underlies the establishment of "instruction segments".

(3) F_{1a}. euh tu remontes
F_{1b}. il y a une pizzeria.
F_{1c}. elle est à peu près au milieu de la rue.
(er you go up / there is a pizzeria / it is about the middle of the street)

It is to be noted that these markers seem to have very different functions since they appear in significantly different contexts. This is an indication of different kinds of agreements at play.

3.4 Closure

Closure has not be annotated. It's a quite risky task to determine at a given point of a discourse which segments are closed and which are not. We do not consider necessarily that a segment is closed definitely. Participants can still go back on it but it will require an explicit signaling of this re-opening.

The tables 2 and 3 in conjunction give some information about the preferred closure scope of markers. We can notice that *voilà* is used to close a lot of sub-dialogs (actually, it closes 30% of all our dialogues). We see in table (3) that the previous utterance is often an acknowledgment itself, thus another indication that something larger than just one speech turn has been closed or is in the process of being closed. In comparison, the seemingly close marker "d'accord" is mainly used to confirm recent, task-related pairs; it is rarely after another acknowledgment.

It could also seem that *voilà* is ambiguous since we noticed that it occurs a lot as an acknowledgment of only one speech turn. However in that case this marker is almost always produced by the informant. It is used after a request for confirmation ("alignment"), not an acknowledgment, so it is easy to separate the two uses of the cue.

On the basis of this analysis, for each of these marker, we gave a default feedback function corresponding to those introduced in section 2 (table 4). Lack of space prevents a detailed analysis here of every marker but we hope to have shown what is to be gained by a multiple factor analysis of this corpus to determine the forms of acknowledgments.

4 Representation of acknowledgment structures

We place ourselves within the framework of Segmented Discourse Representation Theory applied to dialogue (Asher and Lascarides, 2003). In this perspective dialogue acts realized by utterances are linked by "rhetorical" relations expressing their respective functions (semantic, intentional or conventional functions). The SDRT hypothesis about the relational nature of speech act (Asher and Lascarides, 2003) fits pretty well our representation of acknowledgments. In fact, One

Table 3: Acknowledgment targets by cue word

	ouais	oui	ok	d'accord	voilà	mhmm	bon	Total
Landmarks	46	14	13	24	9	12	4	122
Instructions	26	8	5	13	2	13	0	67
Comments	2	0	1	0	1	0	1	5
Feedback	4	0	5	5	22	3	6	45
Total	78	22	24	42	34	28	11	239

Table 4: Preliminary default properties of positive feedback french markers

Marker	Grounding	Accepting	Confirming	Closing	Anchoring
oui,ouais,yes	+	neutral	neutral	neutral	neutral
mhmm	+	-	-	-	-
ok, d'accord	+	+	-	neutral	neutral
voilà	+	-	+	+	-
bon	+	-	-	+	-
je vois	+	+	-	neutral	+
c'est ça, exactement	+	-	+	neutral	-
repeat[10]	+	neutral	neutral	neutral	neutral

of the most striking features of acknowledgments is precisely their relational nature. The determination of a turn as an acknowledgment highly depends on the nature of the target. As already signaled in (Allwood et al., 1992) a turn consisting of "yes" is interpreted as an answer if it targets a yes/no question but as an acknowledgment if it is related to an affirmative sentence.

About SDRT, some relations are considered hierarchical ("subordinating") so they induce a partial order and a tree that defines the structure of the dialog. By attaching left-to-right the most recent dialogue act only to the rightmost nodes of the tree, SDRT imposes constraints to possible continuations of a dialogue situation (achieving in a similar but arguably more flexible way what is realized in other frameworks with a dialogue stack, e.g. the QUD of (Ginzburg, 1994). Moreover parts of the dialogue can be combined to make complex nodes in the structure, open for further attachments. It is thus easy to define acknowledgment scopes as attachment at various levels of the rhetorical structure of the conversation. Thus a question/answer pair defines a superseding topic node which is a possible site for a closure.

The actual treatment of positive feedback in SDRT introduce only an acknowledgment relation which corresponds to an acceptance. We propose here to refine this point of view (i) by taking a more cautious position on the default nature of positive feedback (i.e. considering it as a grounding act and not as an accepting act) and (ii) by adding a *closure* relation. A weak acknowledgment must leave all segments available as possible attachments, whereas stronger forms of acknowledgment seem to settle the topic under discussion. So we have two relations: *acknowledgment* which is a *subordinating relation*, and *closure* which is a *coordinating* one. We do not have a relation for strong acknowledgment because strength scale in communication is not directly usable, at least without taking an *a posteriori* position about interpretation.

We still have to define precisely the semantics (Inference Rules and semantic/structural effects) of these relations in SDRT terms. And we have also to bring in the picture the other phenomena presented here (anchoring, accepting). We are not planning to represent these processes by new relations but rather by a combination of *weak acknowledge*, *closure* and more information about semantic content. Thus our model will take the form of predicates taking as arguments speech acts (SDRT's label) for *accepting*[11] and discourse

[10] Repetitions are not often by themselves. Thus they are neutral regarding to their function. For example, *repeat* + *mhmm* is totally different from *repeat* + *voilà* or *repeat* + *c'est ça*.

[11] This is actually already evoked by the *settled* predicate

referents for *anchoring*. Reasoning about feedback will also include inferences of implicit closure when it is not signaled by an explicit marker as what we studied here. Finally SDRT will be also useful when we will put in a same picture acknowledgment and question/answers structures.

5 Conclusion

Our goal here was twofold: (i) refine analysis of linguistic positive feedback (specially in french language) by showing what factors can be isolated (ii) focus on the form and on the targets of acknowledgment acts (in a broad sense). The study of markers of positive feedback is an invaluable help, even though it has to be continued to fully validate the choices made here. Moreover we also still have to precise the representations proposed within a formal theory of dialogue (SDRT), by fully formalizing the conditions under which they arise. This implies more complex interaction between semantics and the discourse structure.

References

Karin Aijmer. 2002. *English Discourse Particles, Evidence from a Corpus*, volume 10 of *Studies in Corpus Linguistics*. Benjamins.

J. Allwood, J. Nivre, and E. Ahlsn. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9.

N. Asher and A. Lascarides. 1998. Questions in dialogue. *Linguistics and Philosophy*, 21:237–309.

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

A. Auchlin. 1981. Mais, heu, pis bon, ben alors voilà, quoi! marqueurs de la structuration et complétude. *Cahiers de Linguistique Française*, 2:141–159.

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*.

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

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

of (Asher and Lascarides, 2003)

N. Colineau. 1997. *Etude des marqueurs discursifs dans le dialogue finalisé*. Ph.D. thesis, Université Joseph Fourier, Grenoble I.

M. Core and J. Allen. 1997. Coding dialogs with the damsl annotation scheme. In *Working Notes of the AAAI Fall Symposium on Communicative actions in Humans and Machines*, pages 28–35, Cambridge, MA.

H. H. Clark et D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22.

Jonathan Ginzburg. 1994. An update semantics for dialogue. In *Proceedings of the workshop on computational semantics*, Tilburg.

D. Jurafsky, E. Shriberg, B. Fox, and T. Curl. 1998. Lexical, prosodic and syntactic cues fo dialogs acts. In *ACL/COLING-98 Workshop on Discourse Relation and Discourse Markers*.

D. G. Novick and S. Sutton. 1994. An empirical model of acknowledgment for spoken language systems. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 96–101.

A. Reboul and J. Moeschler. 1998. *Pragmatique du discours : de l'interprétation de l'énoncé à l'interprétation du discours*. Paris, Armand Colin.

E. A. Schegloff. 1982. Discourse as an interactional achievement: somes uses of 'uh huh' and other things that come between sentences. In D. Tannen, editor, *Analysing Discourse : Text and Talk*, pages 71–93. Georgetown University Press.

D. Schiffrin. 1987. *Discourse Markers*. Cambridge University Press.

A. B. Stenström. 1994. *An introduction to spoken interaction*. Longman, London.

David Traum. 1994. *A computational theory of grounding in natural language conversation*. Ph.D. thesis, University of Rochester.

M. Walker and S. Whittaker. 1990. Mixed initiatives in dialogue: an investigation into discourse segmentation. In *Proceedings of the 28th Annual meeting of the ACL*, pages 70–78.

M. A. Walker. 1996. Inferring acceptance and rejection in dialogue by default rules of inference. *Language and Speech*, 2(39).

Robust Multimodal Discourse Processing

Norbert Pfleger and **Ralf Engel** and **Jan Alexandersson** *
DFKI GmbH
D-66123 Saarbrücken, Germany
{pfleger,rengel,janal}@dfki.de

Abstract

Providing a generic and robust foundation for the correct processing of short utterances is vital for the success of a multimodal dialogue system. We argue that our approach based on a three–tiered discourse structure in combination with partitions provide a good basis for meeting the requirements in such a system. We present a detailed description of the underlying representation together with some show cases.

1 Introduction

Continuous recognition of speech and gesture puts extra requirements on the part of a dialogue system concerned with semantic processing. The successful processing of tasks, like selecting the correct or best analysis out of competing ones belong to the basic processing abilities, but also resolving ambiguities are of great importance.

Utilizing some kind of discourse context is vital for disambiguation, especially for vague, reduced, or partial expressions, e. g., elliptical or referential expressions. A typical example is depicted in figure 1, where the correct interpretation of U2 and U3 relies on an elaborated discourse model.

In this paper we argue that our approach to discourse modelling which has borrowed its main inspiration from (Luperfoy, 1992; Salmon-Alt, 2000) and (Wahlster, 2000) provides a generic, simple and yet powerful basis for multimodal discourse modelling and processing. We construe an unified representation of user and system contributions for mono- as well as multimodal systems supporting the resolution of elliptical and cross-modal referential expressions in a simple and generic manner. Our work described here (see also (Pfleger, 2002)) contributes to the DFKI core dialogue backbone for multimodal dialogue systems.

U1: *I'd like to see a film tonight.*
S1: [Displays a list of films] *Here [↗] you see a list of the films running in Heidelberg.*
U2: *Hmm, none of these films seems to be interesting... Please show me the TV program.*
S2: [Displays a list of broadcasts] *Here [↗] you see a list of broadcasts on TV tonight.*
U3: *Then tape the first one for me!*

Figure 1: Dialogue excerpt 1

2 The Dialogue Backbone

Our long term effort is to develop a reusable dialogue backbone. Though the present status has been heavily coloured by the SMARTKOM project (www.smartkom.org), parts of the backbone are and have been used in several mono- as well as multimodal dialogue systems including monomodal and multimodal as well as typed and spoken input/output (including gesture and facial expressions), e. g., Miamm (www.miamm.org),

*The research presented here is funded by the German Ministry of Research and Technology under grant 01 IL 905. The responsibility for the content is with the authors. We would like to thank Stephan Lesch and Massimo Romanelli for their help with implementation and evaluation.

Comic (www.hcrc.ed.ac.uk/comic) and NaRaTo, an industrial project aiming at a typed NL interface for the ARIS tool-set (see www.ids-scheer.com).

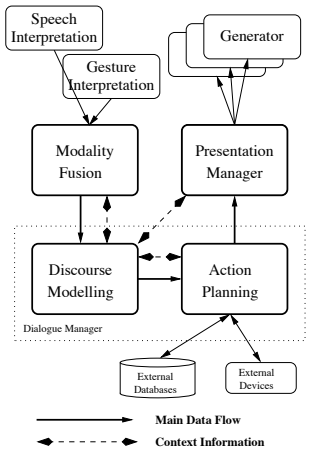


Figure 2: Architecture of the backbone

We use a pipe-line approach enhanced with several request-response interfaces (forward and backward) between different components. (Löckelt et al., 2002) contains a quite detailed description of our architecture. In our current system, some additional interfaces were introduced, e.g., the output from the system is completely processed by the discourse modeller (henceforth DIM) supporting for instance the processing of cross-modal referring expressions (see also (Pfleger et al., 2003)).

The main communication representation within the backbone is the so-called *intention lattice* containing instances of our domain model, syntactic information and scoring information etc. Our domain model (Gurevych et al., 2003) is a hand crafted ontology encoded in OIL-RDFS (Fensel et al., 2001). Basis for its development are the ideas of (Russel and Norvig, 1995; Baker et al., 1998). The current version comprises more than 700 concepts and about 200 relations. We apply *closed world reasoning* - everything that can be communicated is encoded explicitly in our ontology.

Instances of the top-level types in the ontology are called *application objects*. These are complete descriptions of actions, such as accessing a database or zooming a map. Parts of application objects are called *subobjects* which are more often atomic objects, e.g., channels and cities, although they might be structured for example as time expressions and seat collections (which in our ontology have cardinality and a set of seats). Some subobjects are meaningful for the action planner (henceforth AP), these subobjects are called *slots*. A slot is a pair consisting of a name and a path. Slots are unique, so given a slot name we can uniquely find its corresponding path (and vice versa). The effect of the presence of a slot in an user intention is described in (Löckelt et al., 2002).

Central to this paper is the notion of *path*. Some paths are defined by the *action plans* which, in a sense, connect to the ontology by pointing into it. Action plans define *states* and *slots*. Some states are called *goals* and in each plan there is at least one goal. For each goal there is a path called *goal path*. The goal path is usually not a slot. A slot is a pair consisting of a symbol and a path. Although the plans are mainly used by AP they are additionally utilized by several components of the system.

We give a very short recapitulation of the processing within the backbone (see also (Löckelt et al., 2002)): user actions - speech and gesture - are analyzed and interpreted and finally brought together in MF. DIM enriches the hypotheses with contextual information and - based on the different scores - finally selects the most probable hypothesis and sends it to AP. Depending on input and the dialogue state, AP may choose to access some external devices before the modality fission is requested to generate and present the system reaction.

The analysis components of our backbone score each hypothesis in different ways; a task we call *validation*. Scoring is based on the knowledge in the respective modules and on different views of the user intention. Whereas, e.g., the language analysis computes a score based on how *linguistic* a certain path in the word lattice is, DIM computes its score based on how well the hypothesis fits the discourse context (see (Pfleger et al., 2003)).

2.1 Language Understanding

The task of the language understanding component is to analyze the different hypotheses of the speech recognizer and to assign to them a semantic

meaning in terms of the domain model. We use a template based semantic parser (henceforth SPIN (Engel, 2002)). The basic idea of the used approach is to apply so-called *templates* to a working memory (WM) in a depth-first search fashion. Initially, the WM is filled with the recognized words. In a first phase, templates capable of transforming the initial words to simple objects (typically subobjects) are applied. Then, these objects are combined to more complex objects (typically application objects or nested subobjects). In case some objects (or words) contribute nothing to the final interpretation, a lower score is assigned to that particular interpretation.

Referring expressions are internally represented as subobjects together with a feature called `reference`. It is filled with information about the characteristics of referring expression, e.g., definiteness and/or position in a list.

Referring expressions containing no type information of the referenced object, like *this one*, are initially given the most common type - PHYSICALOBJECT. However, it might, during template application, be refined due to the intrasentential context. For instance, if a template responsible for creating an application object of type INFORMATIONSEARCH expects a domain object of type BROADCAST and the WM contains a domain object of type PHYSICALOBJECT, then the type of the object in the WM is (destructively) refined to the type BROADCAST.

2.2 Modality Fusion

The task of the modality fusion component is to combine and integrate multiple hypotheses produced by the analyzers for the different modalities. Pointing gestures are integrated into a speech recognition hypothesis containing deictic expressions by replacing the referring expression with the object associated with the gesture. A different strategy is applied if a gesture is recognized accompanying a spoken utterance without a deictic expression. In that case, the ontology is utilized in order to find possible insertion places.

In case a spoken utterance contains referring expressions that cannot be resolved with gestures, MF requests DIM and replaces the referring expression with the possible discourse objects.

Following (Nigay and Coutaz, 1993), we currently process *synergistic* input, i.e., a combination of coherent information from gesture and speech which can be mapped onto a single domain object and *exclusive* input, i.e., either the input is speech- or gesture-only.

3 Discourse Modelling

Context Representation: Our approach to discourse modelling is based on a generalization of (Luperfoy, 1992) together with some ideas from (Salmon-Alt, 2000) and (Wahlster, 2000). Following the ideas of (Luperfoy, 1992), we use a three-tiered context representation where we have extended her linguistic layer to a *modality* layer (see Figure 3). Additionally, we have adopted some ideas from (Salmon-Alt, 2000) by incorporating directly perceived objects and compositional information of collections. Basic discourse operations used in (Wahlster, 2000) has been further developed (see (Alexandersson and Becker, 2003)). For more details please see (Pfleger, 2002; Pfleger et al., 2003)

The advantage of our approach to discourse representation lies in the *unified* representation of discourse objects introduced by the different modalities. As we show below, this not only supports the resolution of elliptical expressions but allows for, i.e., cross-modal reference resolution. The context representation of the discourse modeller consists of three levels:

• **Modality Layer:** The objects at the modality layer (MOs) encapsulate information about the concrete realization of referential objects. We employ three types of objects: (i) *Linguistic Objects* (LOs) providing information about linguistic features, e.g., number and gender, (ii) *Visual Objects* (VOs) providing information about the position on the screen, and (iii) *Gesture Objects* (GOs) providing no realization information but which are used to group objects together. Each modality object is linked to a corresponding discourse object and shares its information about its concrete realization with the discourse object (see figure 4). Important for this paper is that a MO provides information about its original position within the event structure - its path in the corresponding application object it is embedded in.

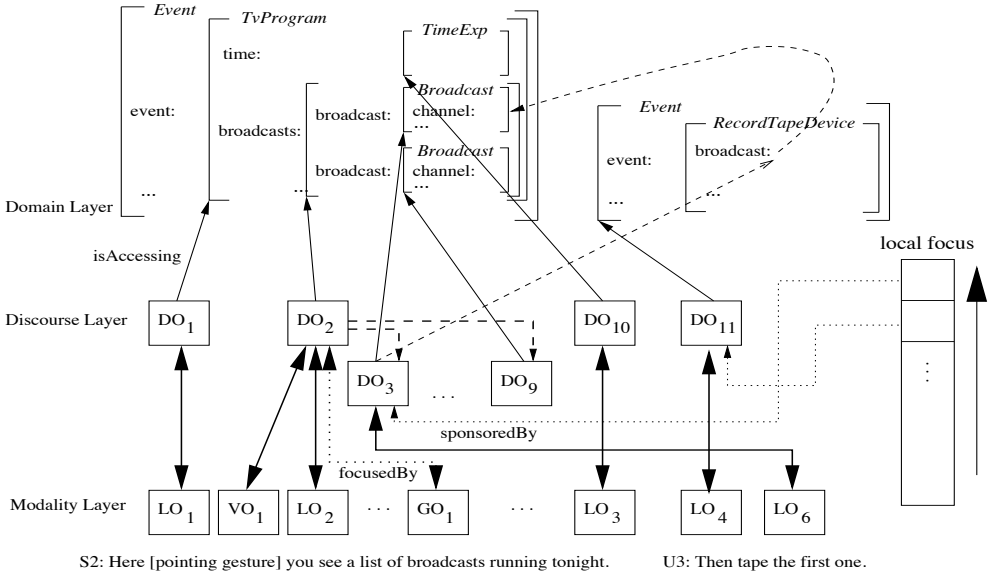


Figure 3: The Multimodal Context Representation. The dashed arrow(s) indicates that the value of the broadcast in the (new) structure to the right is shared with that of the old one (to the left).

• **Discourse Object Layer:** This layer contains discourse objects (DOs) which serve as referents for referring expressions. A DO is created every time a concept is newly introduced into the discourse by speech and for directly perceived concepts, e. g., graphical presentations (Salmon-Alt, 2000).

Two classes of information are used by a DO (i) modality specific information, and (ii) domain information. For each concept introduced during discourse there exists only one DO independent of how many MOs mention this concept. Each DO is hence unique.

The compositional information of DOs representing collections of objects is provided by *partitions* (Salmon-Alt, 2000).

Partitions represent collections of objects and are based either on perceptive information, e. g., the list of broadcasts visible on the screen, or discourse information stemming from grouping discourse objects. The elements of a partition are distinguishable from one another by at least one *differentiation criterion*. Yet, one element alone of a partition may be in focus, according to gestural or linguistic salience. Figure 4 depicts a sample configuration of a discourse object (DO2) with a partition.

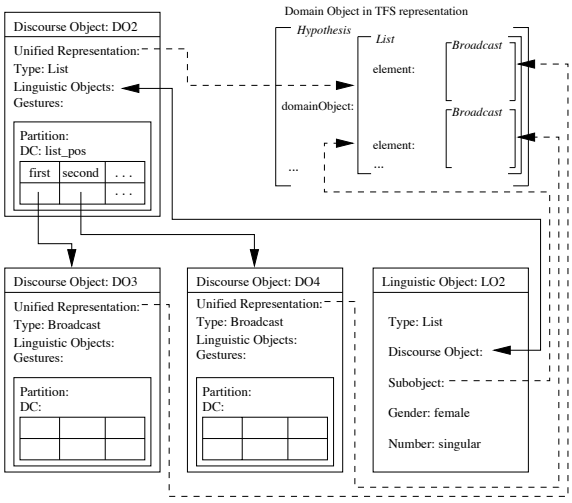


Figure 4: Discourse Objects

• **Domain Object Layer:** The domain object layer encapsulates the instances of the domain model and provides access to the semantic information of objects, processes, and actions. Initially, the semantic information of a DO is defined by a subobject (possibly embedded in an application object) representing the object, process, or action it corresponds to. This information is only accessed by the DO. However, the semantic information of a DO might be extended as soon as the

object is accessed again. Consider for example DO42 initially representing a movie with the title "The Matrix" (created by the user request *"When will the movie The Matrix be shown on television?"*). The system will respond with presenting a list of broadcasts of the movie "The Matrix" (accompanied with information specific to the different broadcasts, like time, database key, channel etc). Now, if the user selects one of the movies - *"tape this [↗] one"* - the initial information of DO42 will be extended with this additional information.

3.1 Modelling Attentional State

We differentiate between two focus structures restricting access to objects stored in the discourse model: (i) a *global* focus structure and (ii) a *local* focus structure. The former represents the topical structure of discourse and resembles a list of focused items - *focus spaces* - ordered by salience. In SMARTKOM, the global focus is imposed by the action planner providing a flat structure of discourse in terms of discourse topics. A focus space covers all turns belonging to the same topic and enables access to a corresponding local focus structure (see also (Carter, 2000)).

A local focus structure provides and restricts access to all discourse objects that are antecedent candidates for later reference. Also on this level, the content of the structure, i.e., discourse objects, are ordered by salience. For each user or system turn, the local focus structure for this topic is extended with all presented concepts (see also figure 3).

3.2 Initiative–Response Units

To further restrict and provide access to referents we use simplified, flat *initiative–response units* or IR-units (Ahrenberg et al., 1991), mirroring who is having the initiative. This information has effects on the interpretation of partials (Löckelt et al., 2002). Our flat treatment is robust and despite its simpleness capable of processing one-level subdialogues, i.e., cases where, if the user has the initiative, the system imposes a sub-dialogue by stealing the initiative by requesting additional required information.

4 Discourse Processing

We now turn to the task of processing, e.g., referring expressions. In this section we describe how the structures presented in the last section are utilized.

4.1 Context Dependent Interpretation

Our main operations for the manipulation on instances of our domain model is unification and a default unification operation we call OVERLAY (Alexandersson and Becker, 2003). The starting point for the development of the latter operation was twofold: First we view our domain model as typed feature structures and employ closed world reasoning on their instances. Second, we saw that adding information to the discourse state can be done using unification as long as the new information is consistent with the context. However, as the user changes her mind and specifies competing information, unification will fail. Instead, we saw the need for a non-monotonic operation capable of overwriting parts of the old structure with the new information *and* at the same time keep the information still consistent with the new information. The solution is default unification, e.g., (Carpenter, 1993; Grover et al., 1994) which has proven to be a powerful and elegant tool for doing exactly this: Overwriting old, contextual information - *background* - with new information - *covering* - thereby keeping as much consistent information as possible.

U4: *What is on TV tonight?*
S4: [Displays a list of broadcasts] *Here* [↗] *you see a list of the broadcasts running tonight.*
U5: *What is running on CBS?*
S5: [Displays a list of broadcasts for CBS tonight] *Here* [↗] *you see a list of the broadcasts running tonight on CBS.*
U6: *and CNN?*
S6: ...

Figure 5: Dialogue excerpt 2

We distinguish between full or partial utterances. Example of the former is a complete description of an user action, e.g., U4 in Figure 5, whereas partial utterances often but not necessarily are elliptical responses to a system request. We

handle these cases differently as described below.

4.2 Full Utterances

For, e. g., task-oriented dialogues, there are many situations where information can and should be inherited from the discourse history as shown in the dialog excerpt in figure 5. Due to spatial restrictions on the screen it may be impossible for the system to display every broadcast for all channels, e. g., in S4. The system therefore chooses some broadcasts of some channels. Clearly, the intention in U5 is to ask for the program on CBS *tonight* thus requiring the system to inherit the time expression from U4. Default unification provides an elegant mechanism for inheriting information from the background for these cases. Full utterances are processed by traversing the global focus structure and pick the focused application object in each focus space (if any). In figure 5, default unifying U5 (covering) with U4 (background) results in *what is running on TV on channel CBS tonight*.

4.3 Partial Utterances

For the interpretation of partial utterances (henceforth partials) we gave a detailed description in (Löckelt et al., 2002). The general idea is to convert the partial to an application object - referred to as bridging - and then use this application object as covering and the focused application object as background. There is, however one more challenge we have to face: resolving referring expressions. Given resolved referring expressions *and* the correct bridging we can use the basic processing technique as described above in section 4.2. Next, we concentrate of the latter whereas the former task is described in section 4.5.

4.4 Resolving Referring Expressions

There have been many proposals in the literature for finding antecedents for referring expressions, e. g., (Grosz et al., 1995). These approaches typically advocate a search over lists containing the potential antecedents where information like number, gender agreement etc. are utilized to narrow down the possible candidates. Our approach to reference resolution is a bit different. Additionally, in a multimodal scenario, the modality fusion first has to check for accompanying pointing gestures before accessing the discourse memory. In case of a missing gesture, DIM receives a request from modality fusion containing a subobject as specific as possible inferred by the intra-sentential context. This goes together, if possible, with the linguistic features and partition information. DIM searches the local focus structure and returns the first object that complies with the linguistic constraints in the respective LO and unifies with the object of the corresponding DO.

We resolve three different referring expressions:

- **Total Referring Expressions:** A *Total Referring Expression* is the condition where a referring expression co-refers with the object denoting its referent. Those referring expressions are resolved through the currently active local focus. The first discourse object that satisfies the type restriction and that was mentioned by a linguistic object with the same linguistic features is taken to be the intended referent. In this case, there is a linguistic sponsorship relation established between the referring expression and its referent.

  If no linguistic sponsorship relation can be established, DIM tries to establish a discourse sponsorship relation. This condition is characterized by a mismatch between the linguistic features but the objects themselves are compatible (unifiable).

  However, if both conditions are not fulfilled, the focused discourse objects (but only the ones most focused on) of the other global focus spaces are tested thereby searching for a discourse object that allows for a linguistic sponsorship relation.
- **Partial Referring Expressions:** In the case of a partial referring expression, the focus structures are searched for a discourse object that shows compositionality and satisfies (i) the differentiation criterion specified in the `partition` feature of the request, and (ii) has a discourse object - DO_i - in its value feature of the partition that satisfies the value feature of the partition (see figure 4). The first such DO_i is returned.
- **Discourse Deictic Expressions:** If the type of the referring expression cannot be identified, the focused discourse object of the currently active local focus is tested as to whether it shares the linguistic features with the request. If it does, that discourse object is taken to be the intended referent, otherwise the referring expression is inter-

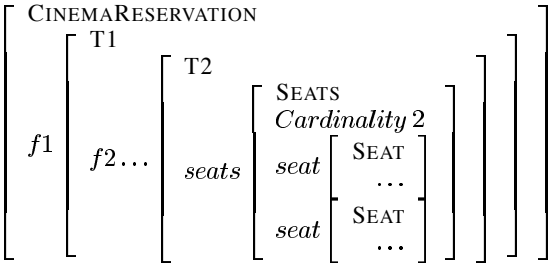


Figure 6: An application object representing the reservation of two seats.

preted as being a discourse deictic one in which case the discourse object representing the last system turn is returned.

4.5 Partial Utterances Revisited

We return to the interpretation of partials again. After being processed by MF, an intention is now guaranteed to be either an application object or, as we will focus on now, a subobject representing a partial. Processing partials consists of two steps (see also (Löckelt et al., 2002)):

1: Find an anchor for the partial. If present, the anchor is searched for and possibly found under (i) the list of expected slots, (ii) in the local focus stack, or, finally, (iii) in the list of possible slots.

2: Compute the bridge by using the *path* in either expected or possible slots, or in the MOs in the local focus stack.

There are, however, some exception to this general scheme of which we provide two examples:

• **User provides too much information:** If the system has the initiative thus requesting for information, the user might provide more, still compatible information than asked for. A good example is the case where, during seat reservation for a performance, the system asks the user to specify where she would like to sit. The expected slot is in this case pointing to a seat. In our domain model the seat part of a set construction contains not just one seat but, e. g., a set of seats and cardinality. If the user contribution specifies something like "*Two seats here [↖]*[1]" then the user contribution will not fit the expectation.

[1]...where [↖] stands for an encircling gesture.

A representation of the reservation is schematically depicted in figure 6. The system asks for a specification of a seat, i. e., an information at the end of the path *f1:f2:...:seats:seat* which is a partial of type SEAT. The answer *contains* the expectation which, however, is embedded in an object consistent with the expectation. The correct processing of such an answer is to walk along the expectation until the answer is found. If this happens before the end of the expected path, the rest of the path (*seat*) has to be part of the subobject.

• **Manipulations of Sets:** For the correct processing of the example above, we had to extend OVERLAY with operations on sets, like union. Using almost the same example, we have the case where the user is not satisfied with the reservation and replies to the system request "*Is the reservation OK?*" with a modification containing manipulations of, in this case, a set of seats by uttering "*two additional seats here [↖]*". The processing consists of two steps: (i) SPIN marks the seats in the intention hypothesis with a set modification flag. (ii) MF requests the focused set of seats from DIM and computes the allowed set of, in this case additional seats. (iii) The intention is now containing two seats (pointed at with the set modification flag) which are then processed by DIM in the same way as described above: The path from the root of the focused application object corresponding to the set of seats in the discourse memory. These are found in the local focus stack. After computing the covering, OVERLAY is performed where the sets of seats in the background and covering are unified with union.

5 Evaluation

The evaluation of DIM is part of a bigger adventure where we are seeking the answer to the following questions:

Ellipses and Anaphora: including cross-modal anaphora. In how many cases do we find the right antecedent for spoken referential expressions, e. g., "Tape it!", "Tape this one", "Tape the first". In addition to speech recognition, the performance of MF and SPIN plays a central role.

Enrichment: Using default unification as basic operation for discourse has the drawback that sometimes too much - still consistent - contextual

information is inherited. Consequently we are currently rather bothering about what *not* to inherit than what to inherit.

Score: Did the scoring from all components contributed to the selection of the correct hypothesis?

The evaluation is at the moment of writing not completed, but we hope for some indication of the system performance at the workshop.

6 Conclusions

We presented a generic robust discourse module which has been developed for and used in several mono- as well as multi-modal dialogue systems. Our "largest" system is a multimodal system for which about 50 different functionalities have been implemented, e. g., (Reithinger et al., 2003). During the development, we have tested the system on far more than 300 test dialogs. Our next, obligatory step is evaluation. Another topic for future development is more support for modality fission.

References

Lars Ahrenberg, Arne Jönsson, and Nils Dahlbäck. 1991. Discourse Representation and Discourse Management for a Natural Language Dialogue System. Research Report LiTH–IDA–R–91–21, Institutionen för Datavetenskap, Universitetet och Tekniska Högskolan Linköping, August.

Jan Alexandersson and Tilman Becker. 2003. The Formal Foundations Underlying Overlay. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, Tilburg, The Netherlands, February.

Collin F. Baker, Charles J. Fillmore, and John Lowe. 1998. The berkeley framenet project. In *Proceedings of COLING-ACL*, Montreal, Canada.

Bob Carpenter. 1993. Skeptical and credulous default unification with application to templates and inheritance. In A. Copestake E. J. Briscoe and V. de Paiva, editors, *Inheritance, Defaults and the Lexicon*, pages 13–37. Cambridge University Press, Cambridge, England.

David Carter. 2000. Discourse focus tracking. In Harry Bunt and William Black, editors, *Abduction, Belief and Context in Dialogue*, volume 1 of *Studies in Computational Pragmatics*, pages 241–289. John Benjamins, Amsterdam.

Ralf Engel. 2002. SPIN: Language understanding for spoken dialogue systems using a production system approach. In *Proceedings of 7th International Conference on Spoken Language Processing (ICSLP-2002)*, pages 2717–2720, Denver, Colorado, USA.

D. Fensel, F. van Harmelen, I. Horrocks, D. McGuinness, and P. F. Patel-Schneider. 2001. OIL: An ontology infrastructure for the semantic web. *IEEE Intelligent Systems*, 16(2):38–45.

B.J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. Technical Report IRCS Report 95-01, The Institute For Research In Cognitive Science, Pennsylvania.

Claire Grover, Chris Brew, Suresh Manandhar, and Marc Moens. 1994. Priority union and generalization in discourse grammars. In *32nd. Annual Meeting of the Association for Computational Linguistics*, pages 17–24, Las Cruces, NM. Association for Computational Linguistics.

Iryna Gurevych, Robert Porzel, Hans-Peter Zorn, and Rainer Malaka. 2003. Semantic coherence scoring using an ontology. In *Proceedings of the Human Language Technology Conference - HLT-NAACL 2003*, Edmonton, CA, May, 27–June, 1.

Markus Löckelt, Tilman Becker, Norbert Pfleger, and Jan Alexandersson. 2002. Making sense of partial. In *Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002)*, pages 101–107, Edinburgh, UK, September.

Susan Luperfoy. 1992. The Representation of Multimodal User Interface Dialogues Using Discourse Pegs. In *Proceedings of ACL-92*, pages 22–31.

L. Nigay and J. Coutaz. 1993. A design space for multimodal systems: Concurrent processing and data fusion. In *Proceedings of INTERCHI-93*, pages 172–178, Amsterdam, The Netherlands.

Norbert Pfleger, Jan Alexandersson, and Tilman Becker. 2003. A robust and generic discourse model for multimodal dialogue. In *Workshop Notes of the IJCAI-03 Workshop on "Knowledge and Reasoning in Practical Dialogue Systems"*, Acapulco, Mexico, August.

Norbert Pfleger. 2002. Discourse processing for multimodal dialogues and its application in smartkom. Master's thesis, Universität des Saarlandes.

Norbert Reithinger, Jan Alexandersson, Tilman Becker, Anselm Blocher, Ralf Engel, Markus Löeckelt, Jochen Müeller, Norbert Pfleger, Peter Poller, Michael Streit, and Valentin Tschernomas. 2003. Smartkom - adaptive and flexible multimodal access to multiple applications. In *Proceedings of ICMI 2003*, Vancouver, B.C.

Stuart Russel and Peter Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Englewood Cliffs, NJ.

Susanne Salmon-Alt. 2000. Interpreting referring expressions by restructuring context. In *Proceedings of ESSLLI 2000*, Birmingham, UK. Student Session.

Wolfgang Wahlster, editor. 2000. *VERBMOBIL: Foundations of Speech-to-Speech Translation*. Springer.

Utterances as Update Instructions

Matthew Purver and **Raquel Fernández**
Department of Computer Science
King's College, London
Strand, London WC2R 2LS, UK
matthew.purver@kcl.ac.uk, raquel@dcs.kcl.ac.uk

Abstract

We present an approach to utterance representation which views utterances and their sub-constituents as instructions for contextual update: programs in a dynamic logic defined with respect to the dialogue gameboard of (Ginzburg, 1996; Fernández, 2003b). This approach allows utterance processing protocols to be represented within the grammar rather than postulated as separate dialogue processes: it also allows a view of incremental grounding and clarification which reflects the insights of dynamic semantics and allows us to treat salience and information structure in a coherent manner.

1 Introduction

The Information State (IS) approach to dialogue modelling has received much attention in recent years (see e.g. (Cooper et al., 1999; Larsson et al., 2000; Lemon et al., 2001)). This approach allows modelling of the information available to each participant at each stage of the dialogue, with updates to this information being defined in terms of update protocols postulated as part of one's general dialogue capability.

In this paper we discuss an alternative view of IS update processes: that updates can be defined as part of the grammatically conveyed content of utterances. We present an approach to utterance representation which views utterances and their sub-constituents as instructions for contextual update: programs in a dynamic logic defined with respect to the dialogue gameboard of (Ginzburg, 1996; Fernández, 2003b). We argue that this approach has several advantages: transferring the burden of definition from general protocols to utterance content allows us to simplify the protocols, transparently express the update rules as part of our linguistic/grammatical competence, and reflect the insight of Gricean pragmatics that utterance content includes the speaker's intended effect on the hearer. It also provides us with a framework within which each sub-utterance can give its own effect on the context, allowing us to reflect the basic insights of dynamic semantics and define a simple approach to grounding and clarification.

The rest of the paper is structured as follows: section 2 describes some theoretical background underlying our approach. Our proposal is then presented in section 3, and applied to concrete phenomena such as grounding and the given/new distinction. In section 4, we sketch the HPSG-based grammatical framework we assume. Section 5 and section 6 show how our approach can be extended to integrate update rules and clarification questions, respectively. We present our conclusions and directions for further work in section 7.

2 Background

2.1 The Dialogue Gameboard

We adopt the theory of dialogue context developed in the KOS framework (Ginzburg, 1996; Ginzburg, ms). In KOS, conversational interaction involves updates by each dialogue participant of her own

dialogue gamebord (DGB), a data structure characterised by the following components: a set of FACTS, which the dialogue participants take as common ground; a partially ordered set of questions under discussion QUD; and the LATEST-MOVE made in the dialogue.

2.2 Grounding and Clarification

Following Ginzburg (1996)'s work, (Ginzburg and Cooper, 2001; Ginzburg and Cooper, forthcoming) present an analysis of clarification questions which motivates a highly contextually dependent representation of utterances. Utterance types are viewed as lambda-abstracts over a set of contextual parameters, i.e. as functions from context to content. In HPSG terms, this is achieved by introducing a new C-PARAMS feature, which encodes the set of contextual parameters of an utterance, and is amalgamated over syntactic daughters by a C-PARAMS Amalgamation Principle. The grounding process for an utterance then involves finding values for these contextual parameters. Failure to do this results in the formation of a clarification question.

2.3 The Utterance Processing Protocol

To account for how utterances get integrated into a dialogue participant's IS, Ginzburg and Cooper (2001) formulate a set of instructions – the Utterance Processing Protocol (UPP) – for a dialogue participant to update her IS, leading either to grounding or clarification.

```
1. Add U to PENDING
2. Attempt to ground U by instantiating
   each P in C-PARAMS
3. If successful:
      3(a). remove U from PENDING;
      3(b). add content(U) to LATEST-
          MOVE;
      3(c). If content(U) = assert(P):
            push whether(P) onto QUD
      3(d). If content(U) = ask(Q):
            push Q onto QUD
4. Else: form clarification question C(P
   ) and add ask(C(P)) to AGENDA
```

Listing 1: Utterance Processing Protocol

Listing 1 shows a simple version: utterances are first added to PENDING for the grounding process, which consists of attempting to identify the referents of the elements of C-PARAMS in context. Failure leads to the formation of a clarification question relevant to that parameter. Success leads to removal from PENDING and addition to LATEST-MOVE. In the case of assertions, a question relevant to the asserted proposition is also added to QUD; in the case of ask moves, the asked question is added to QUD; other moves such as orders may have their own specific actions specified (although in the case of seemingly "empty" moves such as greetings, they may not).

2.4 Formalising the DGB with Dynamic Logic

In (Fernández, 2003a; Fernández, 2003b) the DGB is formalised using first-order Dynamic Logic (DL) as it is introduced in (Harel et al., 2000) and (Goldblatt, 1992). Thus, Fernández (2003b) uses the paradigm of DL familiar from AI approaches to communication modelling[1] to formalise not motivational attitudes but information states and update processes on information states.

In short, DL is a multi-modal logic with a possible worlds semantics, which distinguishes between expressions of two sorts: *formulae* and *programs*. The language of DL is that of first-order logic together with a set of modal operators: for each program α there is a box $[\alpha]$ and a diamond $<\alpha>$ operator. The set of possible worlds (or states) in the model is the set of all possible assignments to the variables in the language. Programs are interpreted as relations between states. Atomic programs change the values assigned to particular variables. They can be combined to form complex programs by means of a repertoire of program constructs, such as *sequence* **;** , *choice* $\cup$, *iteration* * and *test* ?.

Given that in DL transitions between states are changes in variable assignment, the components of the DGB are modelled as variables ranging over different domains, while update operations are brought about by program executions that involve changes in variable assignments. See (Fernández,

[1] For some DL-inspired formalisations within the Beliefs, Desires and Intentions (BDI) tradition, see e.g. (Cohen and Levesque, 1990; Sadek, 1991; Moore, 1995).

2003a; Fernández, 2003b) for the details of the formalisation, as well as for a basic introduction to first-order DL.

3 Utterances as Programs

Our approach in this paper is to view utterances as DL programs: as such, an utterance U denotes a transition between states $s \xrightarrow{U} s'$. As long the program contains all relevant instructions for updating the IS, the UPP no longer needs to be specified separately – instead, we merely specify that an IS s can integrate an utterance U iff $\mathcal{M} \models_s < U > \top$ (i.e. U suceeds for the current IS).

3.1 About the formalism

Following (Fernández, 2003a; Fernández, 2003b), we use the variable names `FACTS`, `QUD` and `LM` to represent the three different components of the DGB. To distinguish between the information that has been commonly agreed on during the dialogue (stored in the initially empty `FACTS`) and the more general context available to a dialogue participant, we also include an additional variable `BG` (background).

We distinguish between individual variables ranging over terms, and *stack* variables ranging over strings of terms. The atomic programs we use to manipulate these variables are simple assignments ($x := t$), where x is an individual variable and t is a term; and $\texttt{X}.\textbf{push}(x)$ and $\texttt{X}.\textbf{pop}$ programs, where `X` is a stack variable (i.e. a string of elements) and x stands for the element to be pushed onto `X`.

`LM` is an individual variable ranging over moves; `QUD` is a stack variable ranging over strings of questions. Although we think of `BG` and `FACTS` as sets, we also model them as stack variables ranging over strings of terms. This allows us to use the **pop** program to check whether some term t belongs to `FACTS/BG` or not: if t is in `FACTS/BG` and we pop the stack repeatedly, t will show up at some point as the head of the stack. Thus, we will use the notation $t \in \texttt{FACTS/BG}$ as an abbreviation for $< \texttt{FACTS/BG}.\textbf{pop}^* > \textbf{head}(\texttt{FACTS/BG}) = t$.

3.2 Replacing the UPP

The most basic form an utterance program can take is:

$$\texttt{LM} := m$$

thus assigning a conversational move m to `LM`, as per part 3(b) of the original UPP (see listing 1 above). The effects of parts 3(c,d) of the UPP are achieved by more complex programs for questions and assertions, which are sequences of atomic programs, as follows:

$$\texttt{LM} := ask(q); \texttt{QUD}.\textbf{push}(q)$$

$$\texttt{LM} := assert(p); \texttt{QUD}.\textbf{push}(whether(p))$$

We assume that these programs are assigned to utterances by the sentence grammar (we currently use a HPSG grammar which relates program to sentence type – see section 4). We can visualise the effect of the above as a transition between ISs:

$$\begin{bmatrix} \text{LM} & m_0 \\ \text{QUD} & qud_0 \end{bmatrix} \xrightarrow{\quad U \quad} \begin{bmatrix} \text{LM} & ask(q) \\ \text{QUD} & q \circ qud_0 \end{bmatrix}$$

Figure 1: Utterance as IS transition

In this way, the meaning associated with a particular move encapsulates the speaker's intention (in the case of an *ask* move, to introduce the asked question to `QUD`).

3.3 Grounding

This approach also allows us to formulate programs which achieve the same effect as Ginzburg and Cooper (2001)'s application of the utterance abstract to the context – namely the identification of referents in the contextual background. The program associated with an expression which requires such a referent to exist must be a program which finds that referent in context (i.e. that succeeds only when the referent is present). Given our formalism, this will be a program $(t \in \texttt{BG})?$ (where $\phi?$ is a program which checks that ϕ holds in the current state).

We therefore propose that referential subutterances (e.g. certain NPs) can contribute such

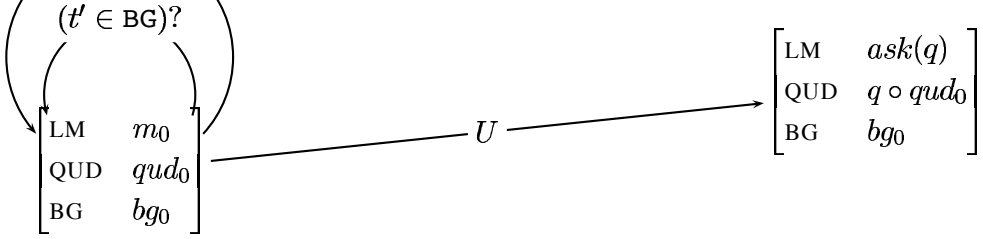


Figure 2: Grounding existing referents

programs to the utterance (again, the details of this must depend on the grammar – our HPSG approach uses a general amalgamation principle over syntactic daughters). A suitable representation for the sentence *"john snores"*, in which a referent for *john* must be found, might be the following complex program:

$$(name(x, john) \in \mathtt{BG})?;$$
$$\mathtt{LM} := assert(snore(x));$$
$$\mathtt{QUD}.push(whether(snore(x)))$$

Here we take x to be a variable ranging over a finite set of people. In this case the test program $(name(x, john) \in \mathtt{BG})?$ would succeed iff there is a *john* in BG and x happens to be assigned to *john* in the current world.[2] We are thus narrowing down the set of possible assignment functions to those that have this property.

In the same way, common nouns & verbs (which, following (Purver and Ginzburg, 2003) we take to refer to a predicate which must be identified) will require a predicate referent to be identified in context.

Thus, we take an utterance to be a program made up of a sequence of tests (that check whether the reference of particular expressions can be found in the context) followed by several atomic programs which update the relevant DGB components, as shown in figure 2. An utterance U can therefore be grounded iff $< U > \top$ holds.

3.4 The Given/New Distinction

This approach allows us to distinguish between *given* referents (e.g. definites) which must be *found* in context as above, and *new* referents (e.g. indefinites) which should be *added* to the context, by associating indefinites with a program which introduces a new referent (thus following the dynamic semantic tradition of (Heim, 1982; Kamp and Reyle, 1993; Groenendijk and Stokhof, 1991), and subsequent DRT-based dialogue theory such as (Poesio and Traum, 1998)).

"the dog": $(dog(x) \in \mathtt{BG/FACTS})?$

"a dog": $\mathtt{FACTS}.\mathbf{push}(dog(x))$

Thus the program associated with a sub-utterance with a given referent tests for existence in the current state, while that for a new referent involves a state change introducing that referent.

This need not be restricted to definites and indefinites, but can be extended to the given/new distinction in general, including the information-structural focus/ground distinction. Following (Engdahl et al., 1999; Ginzburg, ms), we express this distinction by a condition on membership of QUD: that a particular focus/ground partition presupposes a corresponding question on QUD:

"JOHN snores": $(\lambda x.snore(x) \in \mathtt{QUD})?$

"john SNORES": $(\lambda R.R(john) \in \mathtt{QUD})?$

4 Grammar

We use an HPSG grammar similar to that of (Ginzburg and Sag, 2000), with the utterance programs assigned to a new feature C(ONTEXTUAL)-PROG(RAM); this replaces the C-INDICES feature of (Ginzburg and Sag, 2000), or the C-PARAMS feature of (Ginzburg and Cooper, 2001).

By default, this C-PROG feature is built up by phrases, by linear combination of the programs associated with its syntactic daughters, using the sequence operator as shown in AVM [1].

$$[1] \quad \begin{bmatrix} \text{C-PROG} & A;\ldots;B \\ \text{DTRS} & \left\langle \left[\text{C-PROG} \quad A\right], \ldots, \left[\text{C-PROG} \quad B\right] \right\rangle \end{bmatrix}$$

The default program associated with a phrase is therefore a purely sequenced combination of

[2] In fact, for proper names and definites, it will not be enough to require that there is *a* known referent: we need there to be a *unique/most salient* referent. The statement of a suitable test program will depend on one's theory of definiteness, but we do not see this as affecting our general approach.

the programs contributed by its daughters, but this default is overwritten for particular phrase types which by their nature make their own contributions to the overall program. For example, the clause type *root-clause* is specified to add the sub-program which updates LM:

$$[2]\quad \begin{bmatrix} \textit{root-clause} & \\ \text{CONT} & \boxed{1}[\textit{illoc-rel}] \\ \text{C-PROG} & A; \texttt{LM} := \boxed{1} \\ \text{HEAD-DTR} \mid \text{C-PROG} & A \end{bmatrix}$$

while clauses of type *declarative* (which have propositions as their semantic content) and *interrogative* (questions) add the sub-program which updates QUD:

$$[3]\quad \begin{bmatrix} \textit{declarative} & \\ \text{CONT} & \boxed{1}[\textit{proposition}] \\ \text{C-PROG} & A; \texttt{QUD}.\mathbf{push}(\textit{whether}(\boxed{1})) \\ \text{HEAD-DTR} \mid \text{C-PROG} & A \end{bmatrix}$$

$$[4]\quad \begin{bmatrix} \textit{interrogative} & \\ \text{CONT} & \boxed{1}[\textit{question}] \\ \text{C-PROG} & A; \texttt{QUD}.\mathbf{push}(\boxed{1}) \\ \text{HEAD-DTR} \mid \text{C-PROG} & A \end{bmatrix}$$

Phrases which contribute given or new referents are specified in an entirely parallel way, e.g. for a definite NP:

$$[5]\quad \begin{bmatrix} \textit{definite} & \\ \text{CONT} & \boxed{1}[\textit{parameter}] \\ \text{C-PROG} & A; B; (\boxed{1} \in \texttt{BG/FACTS})? \\ \text{DTRS} & \left\langle [\text{C-PROG} \quad A], [\text{C-PROG} \quad B] \right\rangle \end{bmatrix}$$

The interaction with information structure is expressed at the top *root-clause* level:

$$[6]\quad \begin{bmatrix} \textit{root-clause} & \\ \text{CONT} & \boxed{1}[\textit{illoc-rel}] \\ \text{CTXT} \mid \text{IS} & \begin{bmatrix} \text{FOCUS} & \boxed{2} \\ \text{GROUND} & \boxed{3} \end{bmatrix} \\ \text{C-PROG} & A; (\lambda\boxed{2}.\boxed{3} \in \texttt{QUD})?; \texttt{LM} := \boxed{1} \\ \text{HEAD-DTR} \mid \text{C-PROG} & A \end{bmatrix}$$

In figure 3 we show an example derivation of a sentence: the resulting C-PROG program contains instructions for grounding a referent, for establishing the required information structure, and applying the UPP.

5 Integrating Update Rules

Typical IS-based dialogue system behaviour is defined by means of *update rules*, either stated in terms of plans and agendas (Larsson et al., 2000; Larsson, 2002) or in terms of obligations (Poesio and Traum, 1997; Poesio and Traum, 1998). The rules can be quite general but must state the expected behaviour for particular types of move: e.g. one might specify that a move which asks a question causes an obligation (or plan) to respond to the question to arise, that a greeting leads to an obligation (or plan) for a reciprocal greeting, and so on.

In section 3.2 we have seen that part of the speaker's intentions associated with particular move types (e.g. in the case of an *ask* move, to make the question "under discussion") can be specified within the grammar. We could actually go one step further and allow moves to introduce other update effects usually brought about by update rules, allowing us to replicate the rules of (Poesio and Traum, 1998) or pragmatic interpretations of (Stone, 2003).

Assuming that dialogue move types are integrated into the grammatical analysis of utterances (Ginzburg et al., 2001), the need for domain independent rules can be removed by specifying suitable programs as being directly associated with a particular move type. For example, an update rule such as the following (extracted from (Matheson et al., 2000)), would be replaced by the complex DL program below:

act	ID:2, **info_request**(DP,Q)
effect	push(OBL, **address**(**o**(DP),ID))

$$\texttt{LM} := \mathit{info\_request}(DP, Q);$$
$$\texttt{OBL}.\mathbf{push}(\mathit{address}(\mathbf{o}(DP), Q)$$

where OBL is a stack of obligations and $\mathbf{o}(DP)$ refers to *the other dialogue participant*.

A clear motivation for integrating certain contextual updates as part of the grammatical analysis of utterances is the existence of moves whose meaning can only be represented as an update of the dialogue context. An example of such moves are acknowledgements. Acknowledgements as 'okay' or 'uh-huh' are grounding acts with no further descriptive content associated. Thus, given

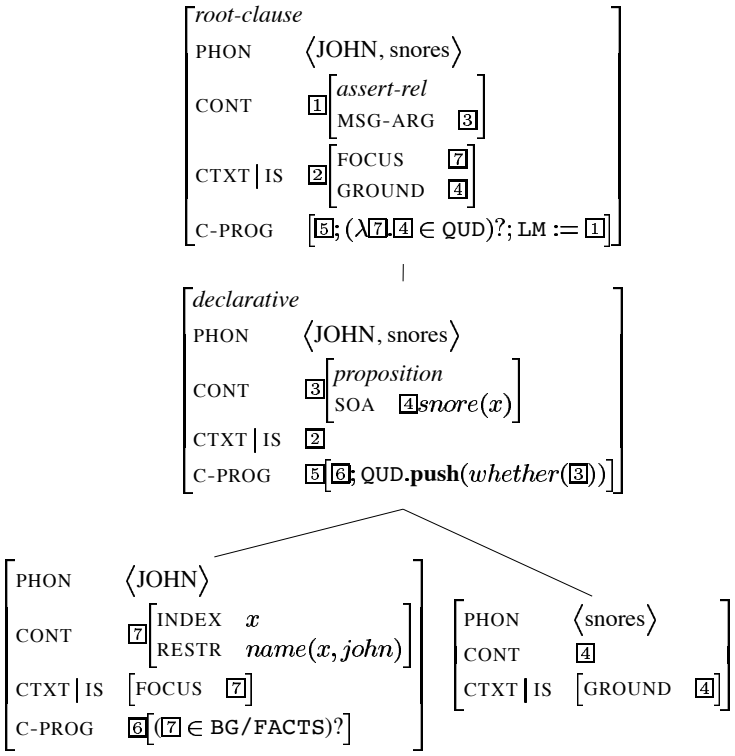


Figure 3: Example derivation: *"JOHN snores"*

our formalism, acknowledgements will be assigned the following program by the grammar:

$$\texttt{LM} := ack; \texttt{FACTS}.\mathbf{push}(\mathbf{head}(\texttt{QUD})); \texttt{QUD}.\mathbf{pop}$$

Although one may argue that this approach contributes to a more unmodular theory of dialogue, we think that one of the advantages of the framework we present here is that it offers a means of specifying the contextual effect of utterances in a general fashion, from sub-utterances to dialogue acts.

6 Clarification

The approach to grounding in section 3.3 is simplistic in that it does not consider the possibility of clarification of a referent that cannot be identified: the program U will simply fail in such cases. We can take this possibility into account by modifying the programs associated with sub-utterance referents. Such programs, rather than being simple instructions to find a referent, now become instructions to find a referent *or* determine it via clarification if one cannot be found. The program associated with *"john"* will therefore take the form:

$$\mathbf{while}(\neg(name(x, john) \in \texttt{BG/FACTS}), \mathbf{clarify}(S))$$

where S is the *sign*[3] corresponding to the sub-utterance. Here, $\mathbf{while}(\phi, \alpha)$ abbreviates $(\phi?; \alpha)^*; \neg\phi?$, and $\mathbf{clarify}(t)$ is a program which governs generation of a clarification question relevant to a term t, together with interpretation of any answer given. Its form therefore depends on one's overall system philosophy; for a plan-based approach, it might take the form:

$$\texttt{AGENDA}.\mathbf{push}(ask(CQ(t))); \alpha_T$$

where $CQ(t)$ is a clarification question relevant to the sign t, and α_T is the program which governs the generation & production of the next turn together with the interpretation of its response.

Note that this sub-program will only succeed once the required information is present in

[3] As (Ginzburg and Cooper, 2001) point out, clarification questions must take into account many properties (including phonology, syntax etc.): we therefore assume all sign information is required.

FACTS, allowing the overall utterance program to continue (eventually e.g. updating QUD). Note also that this may result in LM now being set to the initial move (rather than the latest move in any clarification sequence) – but this is the behaviour we desire, as it is this move that is now being responded to.[4]

We can further modify this to take account of the possibility of *accommodation* of the referent (adding it to the common ground without clarification):

$$\textbf{while}(\neg(name(x, john) \in \texttt{BG/FACTS}),$$
$$\quad \textbf{clarify}(S) \cup \texttt{FACTS}.\textbf{push}(name(x, john)))$$

Note that this treatment need not be confined to programs associated with sub-utterances that require given referents to be identified (e.g. names and definites), but can be used in general to account for clarification caused by the failure of any program (including those for indefinites, focus/-ground partition, and the overall move made) in a uniform way.

7 Summary

7.1 Conclusions

- This approach reduces the amount of our dialogue competence which is specified in processing protocols / update rules, and instead specifies it as part of the grammar. Utterance processing now merely consists of applying the state changes specified by the utterance itself.
- The resulting representation expresses the Gricean intuition that interpretation depends upon recognising the speaker's intention: utterance meaning includes the intended effect on the hearer (e.g. that they add the move to their IS, and make the desired question "under discussion").
- The approach also allows sub-utterances to specify their effect on the context, allowing us to express the given/new distinction neatly, and define a process for grounding & clarification.
- The representation therefore allows all immediate contextual effects of an utterance to be specified on the same level: from the update effect of NPs familiar from dynamic semantics, to the update effect of utterances familiar from IS-based theories of dialogue.

[4] If one has a different view of the role of LM, a program can be constructed that avoids this.

7.2 Further Work

The work presented in this paper is an initial proposal for the representation of utterances within a dialogue grammar. In order to investigate it further, we plan:

- A thorough study of dialogue update rules across systems/approaches to determine the level of domain-dependence and thus the extent to which specification in the grammar is desirable;
- Extension to other dialogue phenomena such as revision;
- Extension of our HPSG grammar fragment and implementation within a prototype dialogue system.

7.3 Acknowledgements

The authors would like to thank Jonathan Ginzburg, Ulle Endriss and the anonymous DiaBruck reviewers for several helpful comments. They are supported by EPSRC grant GR/R04942/01 and ESRC grant RES-000-23-0065, respectively.

References

Philip Cohen and Hector Levesque. 1990. Rational interaction as the basis for communication. In P. Cohen, J. Morgana, and M. Pollack, editors, *Intentions in Communication*. MIT Press.

Robin Cooper, Staffan Larsson, Massimo Poesio, David Traum, and Colin Matheson. 1999. Coding instructional dialogue for information states. In *Task Oriented Instructional Dialogue (TRINDI): Deliverable 1.1*. University of Gothenburg.

Elisabet Engdahl, Staffan Larsson, and Stina Ericsson. 1999. Focus-ground articulation and parallelism in a dynamic model of dialogue. In *Task Oriented Instructional Dialogue (TRINDI): Deliverable 4.1*. University of Gothenburg.

Raquel Fern´andez. 2003a. A dynamic logic formalisation of inquiry-oriented dialogues. In *Proceedings of the 6th CLUK Colloquium*, pages 17–24, Edinburgh. CLUK.

Raquel Fern´andez. 2003b. A dynamic logic formalisation of the dialogue gameboard. In *Proceedings of the Student Research Workshop, EACL 2003*, pages 17–24, Budapest. Association for Computational Linguistics.

Jonathan Ginzburg and Robin Cooper. 2001. Resolving ellipsis in clarification. In *Proceedings of the 39th Meeting of the ACL*, pages 236–243. Association for Computational Linguistics, July.

Jonathan Ginzburg and Robin Cooper. forthcoming. Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*.

Jonathan Ginzburg and Ivan Sag. 2000. *Interrogative Investigations: the Form, Meaning and Use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications.

Jonathan Ginzburg, Ivan A. Sag, and Matthew Purver. 2001. Integrating conversational move types in the grammar of conversation. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue (BI-DIALOG 2001)*, pages 45–56.

Jonathan Ginzburg. 1996. Interrogatives: Questions, facts and dialogue. In S. Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 385–422. Blackwell.

Jonathan Ginzburg. ms. A semantics for interaction in dialogue. Forthcoming for CSLI Publications. Draft chapters available from: http://www.dcs.kcl.ac.uk/staff/ginzburg.

Robert Goldblatt. 1992. *Logics of Time and Computation*. Number 7 in Lecture Notes. CSLI Publications.

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

David Harel, Dexter Kozen, and Jerzy Tiuryn. 2000. *Dynamic Logic*. Foundations of Computing Series. The MIT Press.

Irene Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts at Amherst.

Hans Kamp and Uwe Reyle. 1993. *From Discourse To Logic*. Kluwer Academic Publishers.

Staffan Larsson, Peter Ljunglöf, Robin Cooper, Elisabet Engdahl, and Stina Ericsson. 2000. GoDiS - an accommodating dialogue system. In *Proceedings of ANLP/NAACL-2000 Workshop on Conversational Systems*.

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

Oliver Lemon, Anne Bracy, Alexander Gruenstein, and Stanley Peters. 2001. Information states in a multimodal dialogue system for human-robot conversation. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Proceedings of the Fifth Workshop on Formal Semantics and Pragmatics of Dialogue*, pages 57–67. BI-DIALOG.

Colin Matheson, Massimo Poesio, and David Traum. 2000. Modeling grounding and discourse obligations using update rules. In *Proceedings of the First Annual Meeting of the North American Chapter of the ACL*, Seattle, April.

Robert C. Moore. 1995. *Logic and Representation*. Lecture Notes. CSLI Publications.

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

Massimo Poesio and David Traum. 1998. Towards an axiomatization of dialogue acts. In J. Hulstijn and A. Nijholt, editors, *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*, pages 207–222, Enschede, May.

Matthew Purver and Jonathan Ginzburg. 2003. Clarifying nominal semantics in HPSG. In *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar (HPSG-03)*, East Lansing, Michigan, July. Michigan State University.

M. D. Sadek. 1991. Dialogue acts as rational plans. In *Proceedings of the ESCA/ETR Workshop on Multimodal Dialogue*.

Matthew Stone. 2003. Linguistic representation and gricean inference. In *IWCS-5: Proceedings of the Fifth International Workshop on Computational Semantics*, pages 5–21, University of Tilburg, January.

Engagement By Looking: Behaviors for Robots When Collaborating with People

Candace L. Sidner
Mitsubishi Electric Research Labs
201 Broadway
Cambridge, MA 02139
sidner@merl.com

Christopher Lee
Mitsubishi Electric Research Labs
201 Broadway
Cambridge, MA 02139
lee@merl.com

Neal Lesh
Mitsubishi Electric Research Labs
201 Broadway
Cambridge, MA 02139
lesh@merl.com

Abstract

This paper reports on research on developing the ability for robots to engage with humans in a collaborative conversation for hosting activities. It defines the engagement process in collaborative conversation, and reports on our progress in creating a robot to perform hosting activities. The paper then presents the analysis of a study that tracks the looks between collaborators and discusses rules that will allow a robot to track humans so that engagement is maintained.

1 Introduction

This paper reports on our research toward developing the ability for robots to participate with humans in a collaborative interaction for hosting activities. Engagement is the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake. Engagement is supported by conversation (that is, spoken linguistic behavior), ability to collaborate on a task (that is, collaborative behavior), and gestural behavior that conveys connection between the participants. While it might seem that conversational utterances alone are enough to convey connectedness (as is the case on the telephone), gestural behavior in face-to-face conversation conveys much about the connection between the participants.

Engagement is a process to further collaborations. It accounts for how to undertake a collaboration, and how to maintain it once it begins. Engagement is the means by which one collaborative partner tells the other that he or she intends to continue the interaction. Engagement is conveyed not only by the collaborator with the speaking turn in the interaction, but is also by the non -speaking collaborator. Since the non-speaker cannot use linguistic devices, gestures are a means to indicate the desire to further or discontinue the collaboration. Grounding (Clark, 1996) is a device that is part of engagement. It is a backward functioning device to indicate that what has just been said has been understood. Successful grounding is evidence that the collaboration will continue, but it is only partial evidence. Grounding failures offer evidence, inclusive at best, that one of the partners may wish to disengage. One challenge for the research on engagement is to understand which gestures serve grounding purposes, which serve conversational devices such as turn taking, and which perform other engagement roles.

Collaborative interactions cover a vast range of activities from call centers to auto repair to physician-patient dialogues. In order to narrow our research efforts, we have focused on hosting activities. Hosting activities are a class of collaborative activity in which an agent provides guidance in the form of information, entertainment, education or other services in the user's environment and may also request that the user undertake actions to support the fulfillment of those services. Hosting activities are situated or embedded activities, because they depend on the surrounding environment as well as the participants involved. They are social activities because, when undertaken by humans, they depend upon the social roles that people play to determine the choice of the next actions, timing of

those actions, and negotiation about the choice of actions. In applying our research, physical robots, serving as guides, replace human hosts in the environment. To do so, our goals include understanding the nature of human-to-human engagement, especially the role of gestures. We then apply our findings to robots interacting with people.

The gestures discussed in this paper generally concern looks at/away from the conversational partner, pointing behaviors, (bodily) addressing the conversational participant and other persons/objects in the environment, all in appropriate synchronization with the conversational, collaborative behavior. Other gestures, especially with the hands and face play a role in human interactions as some researchers (Cassell et al 2000, Pelachaud et al, 1996, among others) are discovering. The paper limits its focus to the types of gesture mentioned above. These engagement gestures are culturally determined, but every culture has some set of behaviors to accomplish the engagement task. These arise from two tasks that participants undertake: the need to pay attention to the environment around them, and the need to convey some of the intentions of the participants via the head, heads and body. Other intentions are conveyed by linguistic means. When collaborators are not face-to-face, they have only linguistic devices, and cultural conventions expressed linguistically, to tell their partner that they wish to dis/continue the interaction, and if proceeding, how to further their joint goals. In face-to-face interaction, gestures can take some of the load of collaborative information. Some gestures serve to convey ongoing engagement, while others, such as pointing, fill in details about collaborative actions and beliefs. The engagement rules explored here include both purposes because in robotic behavior the two purposes are often intertwined.

Not only must the robot produce these gestures, but also it must interpret similar behaviors from its collaborative partner (hereafter CP). Proper gestures by the robot and correct interpretation of human gestures dramatically enhance the success of conversation and collaboration. Inappropriate behaviors can cause humans and robots to misinterpret each other's intentions. For example, a robot might look away for an extended period of time from the human, a signal to the human that it wishes to disengage from the conversation and could thereby terminate the collaboration unnecessarily. Incorrect recognition of the human's behaviors can lead the robot to press on with a conversation in which the human no longer wants to participate.

While other researchers in robotics are exploring aspects of gesture (for example, Breazeal, 2001, Kanda et al, 2002), none of them have attempted to model human-robot interaction to the degree that involves the numerous aspects of engagement and collaborative conversation that we have set out above. Robotics researchers interested in collaboration and dialogue (Fong et al, 2001) have not based their work on extensive theoretical research on collaboration and conversation, as we will detail later. Our work is also not focused on emotive interactions, in contrast to Breazeal among others. For 2D conversational agents, researchers (notably, Cassell et al, 2000 and Johnson et al, 2000) have begun to explore agents that produce gestures in conversation. This work complements that research while also focusing on the special demands of 3D physical devices.

In this paper we discuss our research agenda for creating a robot with collaborative conversational abilities, including gestural capabilities in the area of hosting activities. We will also discuss the results of a study of human-human hosting and how we are using the results of that study to determine rules and associated algorithms for the engagement process in hosting activities. We will also critique our current engagement rules, and discuss how our study results might improve our robot's future behavior.

2 Communicative capabilities for collaborative robots

To create a robot that can converse, collaborate, and engage with a human interactor, a number of different communicative capabilities must be included in the robot's repertoire. Most of these capabilities are linguistic, but some make use of physical gestures as well. These capabilities are:

(1) Engagement behaviors: initiate, maintain or disengage in interaction;

(2) Conversation management: turn taking (Duncan, 1974) interpreting the intentions of the

conversational participants, establishing the relations between intentions and goals of the participants and relating utterances to the attentional state (Grosz and Sidner, 1996) of the conversation.

(3) Collaboration behavior: choosing what to say or do next in the collaboration, to foster the shared collaborative goals of the human and robot, as well as how to interpret the human's contribution (either spoken acts or physical ones) to the collaboration.

Turn taking gestures serve to indicate engagement because the overall choice to take the turn is indicative of continuing the interaction. CPs produce and observe in their partners other types of gestures during the conversation (such as beat gestures, which are used to indicate old and new information (Halliday, 1973, Cassell, 2000). These types of gestures are significant to robotic participation in conversation because they allow the robot to communicate using the same strategies and techniques that are normal for humans, so that humans can quickly perceive the robot's communication.

We assume that humans do not necessarily turn off their own engagement and conversational capabilities when interacting with robots. While this assumption is a strong one and can be tested with operational robots in collaboration with people, we start with this assumption because many human capabilities are not always consciously under human control. If humans do use their normal engagement and conversational capabilities, then robots must recognize these capabilities. At the same time, robots can themselves use equivalent capabilities to successfully communicate with humans. We hypothesize that such use will make interactions with robots easier and more predictable. Obtaining operational robots that recognize human engagement behaviors and perform them requires that the robot must fuse data gathered from its visual and auditory sensors to determine the human gestures and infer the human intentions conveyed with these gestures. It must also make decisions as it takes part in the collaboration about which intentions it will convey by gesture and which by linguistic means through conversation.

Our engagement model describes an engagement process in three parts, (1) initiating a collaborative interaction with another, (2) maintaining the interaction through conversation, gestures, and, sometimes, physical activities, and (3) disengaging, either by abruptly ending the interaction or by more gradual activities upon completion of the goals of the interaction. The rules of engagement, which operate within this model, provide choices to a decision-making algorithm for our robot about what gestures and utterances to produce.

Our robot, which looks like a penguin, as shown in Figure 1, uses its head, wings and beak for gestures that help manage the conversation and also express engagement with its human interlocutor (3 DOF in head/beak, 2 in wings). The robot can only converse with one person at a time because the collaboration models we use for conversation only posit one partner. However, the robot performs gestures that acknowledge the onlookers to the conversation without their being able to converse. Gaze for our robot is determined by the position of its head, since its eyes do not move. Since our robot cannot turn its whole body, it does not make use of rules we have already created concerning addressing with the body. Because bodily addressing (in US culture) is a strong signal for whom a CP considers the main other CP, change of body position is a significant engagement signal. However, we will be mobilizing our robot in the near future and expect to test these rules following that addition.

To create an architecture for collaborative interactions, we use several different systems, largely developed at MERL. The conversational and collaborative capabilities of our robot are provided by the Collagen™ middleware for collaborative agents (Rich et al, 2001, Rich and Sidner, 1998, Lochbaum, 1998) and commercially available speech recognition software (IBM ViaVoice). We use a face detection algorithm (Viola and Jones, 2001), a sound location algorithm, and an object recognition algorithm (Beardsley, 2003) and fuse the sensory data before passing results to the Collagen system. The robot's motor control algorithms use the engagement rule decisions and the conversational state to decide what gestures to perform. Further details about the architecture and current implementation can be found in (Sidner and Lee, 2003).

Figure 1: Mel, the penguin robot

3 Current engagement capabilities

The greatest challenge in our work on engagement is determining rules governing the maintenance of engagement. Our first set of rules, which we have tested in scenarios as described in (Sidner and Lee, 2003), are a small and relatively simple set. The test scenarios do not involve pointing to or manipulating objects. Rather they are focused on engagement in simpler conversation. These rules direct the robot to initiate engagement with gaze and conversational greetings. For maintaining engagement, gaze at the speaking CP signals engagement, when speaking, gaze at both the human interlocutor and onlookers maintains engagement while gaze away for the purpose of taking a turn does not signal disengagement. Disengagement from the interaction is understood as occurring when a CP fails to take an expected turn together with loss of the face of the human. When the human stays engaged until the robot has run out of things to say, the robot closes the conversation using known rules of conversational closing (Schegeloff and Sacks, 1973, Luger, 1983).

Though the above list is a fairly small repertoire of engagement behaviors, it was sufficient to test the robot's behavior in a number of scenarios involving a single CP and robot, with and without onlookers. Much of the robot's behavior is quite natural. However, we have observed oddities in its gaze at the end of its turn (for example, it will incorrectly look at an onlooker instead of its CP when ending its turn, which signals that the onlooker is expected to speak) as well as confusion about where to look when the CP leaves the scene.

These conversations have one drawback: they are of the how-are-you-and-welcome-to-our-lab format. However, our goal is hosting conversations, which involve many more activities. While other researchers have made considerable progress on the navigation involved for a robot to host visitors [e.g. (Burgard et al, 1998)] and gestures needed to begin conversation [e.g. (Bruce et al, 2002)], many aspects of interaction are open for investigation. These include producing extended explanations, pointing at objects, manipulating them (on the part of the humans or robots), moving around in a physical environment to access objects and interacting with them. This extended repertoire of tasks requires many more gestures than our initial set. In addition, some of these gestures needed in hosting would be understood as disengagement cues by our first repertoire (looking away from the human speaker for an extended time is indicative of disengagement). So engagement gestures are sensitive to the conversational and collaborative context of use.

To explore hosting collaborations, we have provided our robot with some additional gestural rules and new recipes for action (in the Collagen framework), so that our penguin robot now undertakes hosting through a demonstration of an invention created at MERL. This hosting activity includes engagement behaviors as well as utterances and physical actions to jointly perform the demo. Mel greets a visitor (other visitors can also be present), convinces the visitor to participate in a demo, and proceeds to show the visitor the invention. Mel points to demo objects (a kind of electronic cup sitting on a table), talks (with speech) the visitor through the use of the cup, asks the visitor questions, and interprets the spoken answers, and includes the onlookers in its comments. The robot also expects the visitor to say and do certain activities, as well as look at objects, and will await or insist on such gestures if they are not performed. The entire interaction lasts about five minutes. Not all of Mel's behaviors in this interaction appear acceptable to us. For example, Mel often looks away for too long (at the cup and table) when explaining them, it (Mel is "it" since it is not human) fails to make sure it's looking at the

visitor when it calls the visitor by name, and it sometimes fails to look for a long enough when it turns to look at objects. To make Mel perform more effectively, as well as to understand how people perform in their interactions, we are investigating gesture in human-human interactions.

4 Evidence for engagement in human behavior

Much of the available literature on gestures in conversation (e.g. Duncan, 1974, Kendon, 1967) provides a basis for determining what gestures to consider, but does not provide enough detail about how gestures are used to maintain conversational engagement, that is, to signal that the participants are interested in what the other has to say and in keeping the interaction going.

The significance of gestures for human-robot interaction can be understood by considering the choices that the robot has at every point in the conversation for its head movement, its gaze, and its use of pointing. The robot must also determine whether the CP has changed its head position or gaze and what objects the CP points to or manipulates. Head position and gaze are indicators of engagement. Looking at the speaking CP is evidence of engagement, while looking around that room, for more than very brief moments, is evidence of disinterest in the interaction and possibly the intention to disengage. However, looking at objects relevant to the conversation is not evidence of disengagement. Furthermore, the robot needs to know that the visitor has or has not paid attention to what it points at or looks at. If visitor fails to look at what the robot looks at, the visitor might miss something crucial to the interaction.

A simple hypothesis for maintaining engagement (for each listening CP) is: Do what the speaking CP does: look wherever the CP looks, look at him if he looks at you, and look at whatever objects are relevant to the discussion when he does. This simple hypothesis is effective because it assures that the listening CP will have the most information from the speaking CP about the interaction. It will allow the listening CP to be prepared to ground the conversation whenever needed as well. When the robot is the speaking CP, this hypothesis means it will expect perfect tracking of its looking by the human interlocutor. The hypothesis does not constrain the speaking CP's decision choices for what to look and point at.

Note that there is evidence that the type of utterances that occur in conversation affect the gaze of the non-speaking CP. Nikano (Nikano et al, 2003) provides evidence that in direction giving tasks, the non-speaking CP will gaze more often at the speaking CP when the speaking CP's utterance pairs are assertion followed by elaboration, and more often at a map when the utterance pairs are assertion followed by the next map direction.

To evaluate the simple hypothesis for engagement, we have been analyzing interactions in videotapes of human-human hosting activities, which were recorded in our laboratory. In these interactions, a member of our lab hosted a visitor who was shown various new inventions and computer software systems. The host and visitor were followed by video camera as they experienced a typical tour of our lab and its demos. The host and visitor were not given instructions to do anything except to give/take the lab tour. We have obtained about 3.5 hours of video, with three visitors, each on separate occasions being given a tour by the same host. We have transcribed portions of the video for all the utterances made and the gestures (head, hands, body position, body addressing) that occur during portions of the video. We have not transcribed facial gestures. We report here on our results in observing gestural data (and its corresponding linguistic data) for just over five minutes of one of the host-visitor pairs.

The purpose of the investigated portion of the video is a demonstration of an "Iglassware" cup which P (a male) demos and explains to C (a female). P produces a gesture, with his hands, face, and body, gazes at C and other objects. He also point to the cup, holds it and interacts with a table to which the cup transfers data. He uses his hands to produce iconic and metaphorical gestures (Cassell, 2000), and he uses the cup as a presentation device as well. We do not discuss iconic, metaphorical or presentation gestures, in large part because our robot does not have hands with which to perform similar actions.

We report here on C's tracking of where P looks (since P speaks the overwhelming majority of the utterances in their interaction). Gaze in this analysis is expressed in terms of head movements (looking). We did not code eye movements due to video quality.

There are 82 occasions on which P changes his gaze by moving his head with respect to C. Seven additional gaze changes occurred that are not counted in this analysis because it is unclear to where P changed his gaze. Of the counted look changes, C tracks 45 of them (55%). The remaining failures to track looks (37, or 45% of all looks) can be subclassed into 3 groups: quick looks, nods (glances followed by gestural or verbal feedback), and uncategorized failures (see Table 1).

These tracking failures indicate that our simple hypothesis for maintaining engagement is incorrect. Of these tracking failures, the *nod failures* can be explained because they occur when P looks at C even though C is looking at something else (usually the cup or the table). In all these instances, P offers an intonation phase, either at his looks or a few words after, to which C nods and often articulates with "Mm-hm," "Wow" or other phrases to indicate that she is following her conversational partner. In grounding terms [23], P is attempting to ascertain by looking at C that she is following his utterances and actions. When C cannot look, she provides feedback by nods and comments. She is able to do this because of linguistic information from P indicating that her contribution is called for. She grounds P's comments and thereby indicates that she is still engaged. In the nod cases, she also is not just looking around the room, but paying attention to an object that is highly relevant to the demo. In two instances of nods, P looks away from C to something else. In both cases, C is attending to P, and at his intonation pause, C nods.

	Count	% of tracking failures	% of total failures
Quick looks	11	30	13
Nods	14	38	17
Uncategorized	12	32	15

Table 1: Failures to Track Changes in Looking

The quick looks and uncategorized failures represent 62% of the failures and 28% of the look changes in total. Closer study of these cases reveals significant information for our robot vision detection algorithms.

In the *quick look* cases, P looks quickly (including moving his head) at something without C tracking his behavior. In eight instances, P looks to something besides C; in three instances, he looks up to C from the cup or table without her awareness. Is there a reason to think that C is not paying attention or has lost interest? P never stops to check for her feedback in these instances. Has C lost interest or is she merely not required to follow? In all of these instances, the length of the quick look is brief (under 1 second, in most cases under .6 seconds). During these, C is either occupied with something else (looking at something, laughing or nodding) and thus misses the look, or the look occurs without an intonation pause to signal that acknowledgement is expected. In only one instance, does P pause intonationally and look at C. One would expect an acknowledgement here even without tracking P's looks. However, in that instance, C is distracted by the glass and simply fails to notice the look, which is very brief, only .10 seconds.

Of the *uncategorized failures*, the majority (8 instances) occur when C has other actions or goals to undertake. In addition, all of the uncategorized failures are longer in duration (2 seconds or more). For example, C may be finishing a nod and not be able to track P while she's nodding. Of the remaining three tracking failures, each occurs for seemingly good reasons to us as observers, but may not be known at the time of occurrence. For example, one failure occurs at the start of the demo when C is looking at the new (to her) object that P displays and does not track P when he looks up at her.

This data clearly indicates that our rules for maintaining engagement must be more complex than the "do whatever the speaking CP does" hypothesis. In fact, tracking is critical to the interaction because it allows the listening CP to observe the speaking CP's behavior. It is not completely necessary at every moment because quick glances away can be disregarded. Furthermore, the speaking CP can be relied upon to pause for verbal feedback when needed. In

those instances where arguably one should track the speaking CP, failure to do so may lead the speaking CP to pause to wait for return of visual attention and perhaps even to restate an utterance, or alternatively to just go on because the lost information is not critical to the interaction. The data in fact suggest that for our robot engagement rules, tracking the speaking CP in general is the best move, but when another goal interferes, providing verbal feedback, when required, will maintain engagement. Furthermore, the robot as tracker can ignore head movements of brief duration, as these quick looks do not need to be tracked.

We can ask whether the rule of "track whenever possible, but allow other goals to interfere" makes sense, in general terms. Since humans often find themselves collaborating in environments that are not peaceful or may not be benign, lookaways to check up on the world around one are sensible and perhaps necessary. When speaking, lookaways to objects of interest may serve the cognitive function of reminders of how to continue the current utterance. So while tracking serves the very useful function of keeping on top of the other CP's current behavior, it cannot be performed all of the time. Furthermore as long as the other CP (when speaking) intends to maintain engagement, that individual can be relied upon to provide feedback about what is happening when a CP fails to track.

When the robot is the speaking CP, our data suggest that it would be most natural for the robot to seek acknowledgements from the human, especially when the human is looking at something besides the robot. Of course just when the robot should linguistically seek an acknowledgement remains to be accounted for by a theory of grounding in conversation.

5 Future directions

An expanded set of rules for engagement requires a means to evaluate them. We want to use the standard training and testing set paradigm common in computational linguistics and speech processing. However, training and test sets are hard to come by for interaction with robots because one must first have a robot that can perform sufficiently complex interactions to create the sets. Our solution has been to approach the process with a mix of techniques. We have developed a graphical display of simulated, animated robots running our engagement rules. We plan to observe the interactions in the graphic display for a number of scenarios to check and tune our engagement rules. In addition we are now undertaking an evaluation of the robot's demonstration of the Iglassware cup with subjects who interact when the robot uses a varied set of gestural tracking rules with each subject group.

6 Summary

This paper has discussed the nature of engagement in human-robot interaction, and outlined our methods for investigating rules for engagement for the robot. We report on analysis of human-human look tracking where the humans do not always track the changes in looks by their conversational interlocutors. We conclude that such tracking failures indicate both the default behavior for a robot and when it can fail to track without its human conversational partner inferring that it wishes to disengage from the interaction.

Acknowledgements

The authors wish to acknowledge the work of Charles Rich on aspects of Collagen critical to this effort.

References

Beardsley, P.A. 2003. *Piecode Detection*, Mitsubishi Electric Research Labs TR2003-11, Cambridge, MA, February.

C. Breazeal. 2001. "Affective interaction between humans and robots," *Proceedings of the 2001 European Conference on Artificial Life (ECAL2001)*. Prague, Czech Republic.

A. Bruce, I. Nourbakhsh, R. Simmons. 2002. "The Role of Expressiveness and Attention in Human Robot Interaction," In *Proceedings of the IEEE International Conference on Robotics and Automation,* Washington DC, May.

Burgard, W., Cremes, A. B., Fox, D., Haehnel, D., Lakemeyer, G., Schulz, D., Steiner, W. & Thrun, S. 1998. "The Interactive Museum Tour Guide Robot," *Proceedings of AAAI-98*, 11-18, AAAI Press, Menlo Park, CA.

J. Cassell. 2000. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (eds.), Cambridge, MA: MIT Press.

J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. 2000. "Human Conversation as a System Framework: Designing Embodied Conversational Agents," in Embodied Conversational Agents, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (eds.), MIT Press, Cambridge, MA.

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

S. Duncan. 1974. Some signals and rules for taking speaking turns in conversation. in *Nonverbal Communication*, S. Weitz (ed.), New York: Oxford University Press.

T. Fong, C. Thorpe, and C. Baur. 2001. Collaboration, Dialogue and Human-Robot Interaction, *10th International Symposium of Robotics Research*, Lorne, Victoria, Australia, November.

B. J. Grosz and C.L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3): 175—204.

M.A.K. Halliday. 1973. *Explorations in the Functions of Language*, London: Edward Arnold

W.L. Johnson, J.W. Rickel, and J.C. Lester. 2000. "Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments," *International Journal of Artificial Intelligence in Education*, 11: 47-78.

T. Kanda, H. Ishiguro, M. Imai, T. Ono, and K. Mase. 2002. "A constructive approach for developing interactive humanoid robots. *Proceedings of IROS 2002*, IEEE Press, NY.

A. Kendon. 1967. Some functions of gaze direction in social interaction. *Acta Psychologica*, 26: 22-63.

K.E. Lochbaum. 1998. A Collaborative Planning Model of Intentional Structure. *Computational Linguistics*, 24(4): 525-572.

H.H. Luger. 1983. "Some Aspects of Ritual Communication," *Journal of Pragmatics.* Vol. 7: 695-711.

Y. Nikano, G. Reinstein, T. Stocky, J. Cassell. 2003. "Towards a Model of Face-to-Face Grounding," *Proceedings of the 41st ACL meeting*, Sapporo, Japan.

C. Pelachaud, N. Badler, M. Steedman. 1996. "Generating facial expressions for speech," *Cognitive Science ,* 20(1): 1-46.

C. Rich, C.L. Sidner, and N. Lesh. 2001. "COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction," *AI Magazine, Special Issue on Intelligent User Interfaces*, AAAI Press, Menlo Park, CA, Vol. 22: 4: 15-25.

C. Rich and C.L. Sidner. 1998. "COLLAGEN: A Collaboration Manager for Software Interface Agents," *User Modeling and User-Adapted Interaction*, Vol. 8, No. ¾, 1998, pp. 315-350.

E. Schegeloff, & H. Sacks. 1973. Opening up closings. *Semiotica*, 7(4): 289-327.

C.L. Sidner and C. Lee. 2003. *An Architecture for Engagement in Collaborative Conversations between a Robot and Humans,* Mitsubishi Electric Research Labs TR2003-13, Cambridge, MA.

P. Viola and M. Jones. 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features, IEEE *Conference on Computer Vision and Pattern Recognition*, Hawaii, pp. 905-910.

Coordinating Understanding and Generation in an Abductive Approach to Interpretation

Matthew Stone
Department of Computer Science
Rutgers University
mdstone@cs.rutgers.edu

Richmond H. Thomason
Philosophy Department
University of Michigan
rich@thomason.org

Abstract

We use a dynamic, context-sensitive approach to abductive interpretation to describe coordinated processes of understanding, generation and accommodation in dialogue. The agent updates the dialogue uniformly for its own and its interlocutors' utterances, by accommodating a new context, inferred abductively, in which utterance content is both true and prominent. The generator plans natural and comprehensible utterances by exploiting the same abductive preferences used in understanding. We illustrate our approach by formalizing and implementing some interactions between information structure and the form of referring expressions.

1 Introduction

The idea of interpretation as abduction is explained in (Hobbs et al., 1993) in terms of the following recipe for the interpretation of a sentence:

> Prove the logical form of this sentence, together with the constraints that predicates impose on their arguments, allowing for coercions, merging redundancies where possible, making assumptions where necessary.

In (Stone and Thomason, 2002), we modify and extend this idea. We use a modal logic of context to represent the logical form of utterances in terms of their potential to change the context. Accordingly, our abductive interpretations are proofs that describe a new context created by an utterance, in which its content is both true and prominent. Our earlier paper shows that the extension is both essential and effective in deriving correct abductive interpretations for sequences of utterances.

Here we go further to show how our approach supports the reasoning of a full conversational agent, capable of generation as well as understanding. Indeed, the approach reveals systematic commonalities in the reasoning required for understanding, generation and dialogue management—commonalities that spring from their use of shared representations and of abduction as the core reasoning process.

We illustrate the approach with an example in which the generation of appropriate information depends on attentional information, as well as on linguistic information and real-world knowledge.

2 Framework for Dialogue

We base our discussion on an information-state framework for dialogue management of the sort described, for instance, in (Larsson and Traum, 2000). Specifically, we assume that the information state takes the form of a comprehensive representation cgn of the common ground at state n, including not only the dynamic aspects of the dialogue but also the grammar and ontology that are consulted in proving interpretations. Since the information state will be used as a resource by inductive processes of understanding, the common

ground may consist not only of propositional information, but of attentional information that induces preferences between interpretations.

Participation in an ongoing dialogue is maintained by an update operation that revises the common ground in response to an utterance interpretation i, and a selection operation that draws on the agent's private knowledge k as well as the common ground cgn to select a goal for generation. Understanding and generation mediate between utterances and communicative goals, and thereby construct the representations of interpretation over which dialogue update is defined.

Overall, the dialogue agent acts as a turn-taking manager, conforming to the following schematic perception-deliberation-action loop, where u, u' are utterances and i, i' are dynamic interpretations.

loop { input(u);
understand(u, cgn, i);
$cgn \leftarrow$ update(i, cgn);
generate(select(k, cgn), cgn, u', i');
output(u');
$cgn \leftarrow$ update(i', cgn) }

This schema clearly shows the representational constraints that are imposed on interpretation and generation by the need for a single common ground update operation. The interpreter must produce the same update-supporting representations that are produced by the generator.

Constraints on the reasoning processes themselves that are used in understanding and generation follow from the further assumption that the achievement of mutuality in dialogue depends on similarities between the two processes. Suppose that a dialogue agent performs the step

generate(select(k, cgn), cgn, u', i'),

associating utterance u' with interpretation i'. It is natural to suppose that the same agent would also produce

understand(u', cgn, i');

that is, its interpretive component should derive the same effect that the generator intends. On this assumption, we can use similarities between dialogue agents to explain the maintenance of coordination in the course of a dialogue.[1]

[1]Since mistakes can be made about the discourse context, this model needs to be supplemented with an account of how miscoordinations can be prevented, identified and repaired.

3 Representing interpretation

To conform to this framework for dialogue, an abductive approach must provide parallel accounts of generation and understanding, in which the processes construct the same abductive interpretations despite the different goals and premises they use in reasoning. We now outline such an account, building on (Stone and Thomason, 2002), and illustrate it with the example utterance 'He left'.

We continue to work with meaning representations formulated in a modal extension of Prolog. Interpretive goals for abductive proof are modalized atomic formulas and the clauses are modalized Prolog clauses whose component formulas may themselves be modalized. Modal operators represent contexts, which incorporate attentional as well as informational components.

We continue to assume that asserting a proposition has two effects; (i) the purely assertional effect of adding the proposition to the common ground suppositions of the conversation,[2] and (ii) the side effect of changing attentional features of the context. Developing this theme further, we use the formula *add-to-cg*(P, c_1, c_2) to say that adding the proposition P to the common ground and making the required attentional changes to c_1 will produce the new context c_2.

Both understanding and generation seek to link an utterance with an intended change to the context. But in generation the content of the change is given as the goal of the plan, and what needs to be assumed is the utterance, as the means of achieving the goal. In understanding, what is given is the means of achieving a change, and what needs to be assumed is the intended content of this change.

Consider the utterance 'He left'. The generator wishes to assert a proposition P, represented, say, by '*leave*(P, X)', a combination of a predicate, '*leave*', with a variable 'X' specifying a particular domain referent.[3]

The generation process in our example begins with the goal of deriving c_1 : *add-to-cg*(P, c_1, c_2) from c_1 : *leave*(P, X)—in the context c_1 the generator postulates a language-neutral description of

[2]See, for instance, (Stalnaker, 1981).

[3]In order to keep this example as simple as possible, we are ignoring all considerations having to do with the past tense.

(3.1)

c_1 : *he*(X)	Assumed, with low cost if X is masculine and in focus.
c_1 : *utter*('he left', E, P)	Proved using the grammar, which selects the utterance as a way of expressing P.
c_1 : *do*(E)	An intention is formed by hypothesizing an action: this is a low-cost assumption.
c_1 : *leave*(P, X)	Postulated in formulating the goal.
---	Inference about the effects of communication.
c_1 : *add-to-cg*(P, c_1, c_2)	

(3.2)

c_1 : *he*(X)	Assumed, with low cost if X is masculine and in focus.
c_1 : *utter*('he left', E, P)	Postulated based on observation.
c_1 : *do*(E)	Postulated based on observation.
c_1 : *leave*(P, X)	Proved from utterance-type using grammar.
---	Inference about the effects of communication.
c_1 : *add-to-cg*(P, c_1, c_2)	

a proposition, and attempts to use grammatical and contextual resources to find least-cost assumptions that allow the conversational goal to be proved. The generator is free to make assumptions that hypothesize the occurrence of a new utterance and describe its intended interpretation. As a side effect of our example proof, an intention is formed to utter 'He left'.

Schematically, the proof has the structure of the abductive derivation shown in Proof (3.1). This proof uses shared information to explain how a proposed action or series of actions can update the context to assert P. Such a proof constitutes the generator's discourse plan.

Understanding derives this same schematic proof, but by a different strategy. Understanding is given an eventuality E (the utterance), and the words that are uttered. This utterance needs to be classified as communicating a certain content. In this example, the process begins by postulating c_1 : *utter*('he left', E, P) and adding it to the database. The goals of understanding and generation are exactly the same: proving that asserting the proposition P in c_1 will yield a new context c_2. In this case, however, the new assumption that is added by the proof explains what P is.

When generation and understanding act reciprocally, generation's intended P matches understanding's inferred P and Proofs (3.1) and (3.2) give a common interpretation to both interlocutors. Looking ahead to Section 6, we can adopt more fine-grained representations of context-dependence and context change in these interpretive proofs, so that an interlocutor can update the dialogue context simply by executing the transition specified there. In general, contexts incorporate different knowledge resources for conversational reasoning—we have already mentioned our division into informational resources (represented as axioms and rules) and attentional resources (represented as abductive preferences). We divide informational resources into old information and new information, which here has to do with new utterance events. Where c_n is a context, let i_n be the component of c_n representing old information, e_n be the component representing new information about events, and a_n be the attentional component. We can now revise Proofs (3.1) and (3.2) to make explicit the dependence of the utterance on these components of context, and the potential of the utterance to change them:

(3.3)

a_1 : *he*(X)
e_1 : *utter*('he left', E, P)
e_1 : *do*(E)
i_1 : *leave*(P, X)

---

c_1 : *add-info*(P, i_1, i_2)
c_1 : *put-in-focus*(X, a_1, a_2)

The interpretation now provides two instructions

for the dialogue manager: it should update the informational context from state i_1 to state i_2 by assuming the proposition P that X left; and it should update the attentional state from a_1 to a_2 so that X remains a prominent potential referent for a pronoun.[4] We will refer to this proof in our discussion of the implementation in Section 5.

4 Coordination

We now turn to the problem of coordination. Under what conditions can we in fact expect understanding to construct the interpretation that is intended by generation? And how easy will it be for understanding to do so? Concretely, consider examples (3.1) and (3.2), where generation produces an utterance U by deriving *add-to-cg*(P) from *leave*(P, X), and understanding derives *add-to-cg*(P) assuming the occurrence of U. Under what conditions will we expect understanding's proof to precipitate the assumption *leave*(P, X) intended by the generator? And what inference will be involved?

4.1 Mutuality of information

We believe that coordination depends crucially on separating mutual and private information, as many dialogue architectures do, including the information-state architecture. In fact, this provides an important motivation to extend (Hobbs et al., 1993) along the lines we propose in (Stone and Thomason, 2002). Otherwise, the correct resolution of almost every utterance, including those discussed in (Hobbs et al., 1993), would depend on *ad hoc* assumptions about what is left out of the database, and/or *ad hoc* assignments of costs to axioms. In 'He left', for example, what if understanding knows about many people who could serve as the referent for 'He'? Then its knowledge base will license an assumption *he*(Y) for many individuals Y. Abductive understanding will not infer that 'He' is X unless its preferences for this interpretation outweigh all the alternatives. These preferences must vary with circumstances.

In (Stone and Thomason, 2002), we propose to handle such effects in a general way by a straightforward modification of the scheme for abductive weights of (Hobbs et al., 1993). Discourse contexts specify the abductive weights that attach to assumptions. After an utterance where X has been set up as the most prominent referent (as a sentence subject, for example), the weights induce a low cost for assuming *he*(X) in the next utterance. Such specifications provide a very general approach to focus of attention; assumptions with relatively high costs become invisible, and different priorities can be assigned to the assumptions that are visible.

The reasoning needs of generation reveal another side to this requirement of mutuality. Generation starts from a specific communicative goal, but in deriving its discourse plan, it must respect the fact that this goal is a privileged, private resource. The generator must use the goal to guide the planning process, but not to constrain its reasoning about abductive interpretation. Otherwise, consider what would happen in contexts with many people who could serve as the referent for 'He'. In building candidate interpretations, the generator would already know from its communicative goal that it wanted to produce an utterance whose subject referred to X. So to the generator—but not to another agent—it would look as if there were no alternative interpretations.

Our symmetric representations make it easy to maintain the dual perspective required in generation. The generator maintains multiple copies of interpretive proofs, which share the same structure. A distinguished proof records the instantiation required to establish the intended goal in generation; other proofs record the alternative instantiations derived only from shared information. The generator succeeds when only the distinguished proof remains, ensuring that the natural shared interpretation will achieve the goal. This dual representation extends the insights of the SPUD generator to abductive interpretation; see (Stone, 1998; Stone et al., 2003).

4.2 Coordination and preferences

Coordination in generation involves being clear, as well as being comprehensible. In our example, there will be many ways to identify X. Suppose it's grounded that X is not only 'He', but also

[4] Thus we have c_1 : *put-in-focus*(X, a_1, a_2) in place of $a_1[X]a_2$ from (Stone and Thomason, 2002), and a_1 : *he*(X) in place of a_1 : *in-focus*$^*(X)$ and i_1 : *man*(X).

'The conference's second presenter'. The generator hardly seems cooperative if it asks the understander to infer who 'The conference's second presenter' is, when 'He' would do.

Again, the symmetric representations we adopt make it easy for the generator to take into account the interpretive effort required in understanding. In understanding, the cost of abductive proofs controls the search for possible interpretations. To keep this search space small, the generator should formulate an interpretation with a low abductive cost. Since the generator maintains its intended interpretation in the same representation it expects understanding to reconstruct, the generator can use the cost of this proof to guide its search. When the generator must make a choice between alternative expressions that could achieve its communicative goal, it can choose the one with the lowest cost. More generally, the generator can factor the cost of interpretation into its heuristics for searching the space of possible utterances.[5] Concretely in our scenario, as long as the reference to X as 'He' has lower cost than a reference to X as 'The conference's second presenter' (which it certainly will), the generator will prefer it.

The heuristics that underlie choice in generation architectures are often weakly motivated. That's certainly true of SPUD. Our use of interpretive preferences draws on a collaborative view of dialogue to suggest a more principled alternative. And the approach offers a new perspective on the abductive weights themselves. They represent preferences over pairings of forms and meanings, which are used declaratively and reversibly throughout the dialogue architecture.

5 Implementation

We have implemented our model through a series of simple dialogue agents in Prolog. The rather limited ability of these agents to deliberate about appropriate conversational goals could be elaborated in a specific conversational domain. In addition, our current agents abstract away from surface realization and parsing; the utterances they exchange are represented as syntactic derivation trees rather than strings of words. It would be feasible to relax this simplification too. Nevertheless, these agents suffice as a testbed for exploring the predictions and opportunities of our architecture. For example, they realize the context-dependent patterns of reference generation and reference resolution described in Section 6.

Our implementation incorporates the abductive theorem prover described in (Stone and Thomason, 2002). It implements a modal extension of Prolog, and records its assumptions by manipulating *marked queries* consisting of a query $c : A$ and a label classifying the query as resolved, assumed or unsolved. An abductive proof is completed when all of its steps are either resolved or assumed. The reasoner prefers minimal-cost proofs, where cost is computed in terms of *abductive weights* that are attached to possible assumptions. The incorporation of abductive weights in contexts, formalized as modal operators, permits localized cost distributions to be used in proofs. A query $c : A$ is proved using the assumption costs incorporated in c.

The dialogue manager implements the loop described in Section 2. It includes several types of context updates, corresponding roughly to "speech acts", which affect representations of the discourse state incorporating the notion of a *question under discussion* (see (Ginzburg, 1996)). At the same time, the current context is routinely revised in the course of a dialogue. Revision involves a mechanism for updating abductive costs using abstract, qualitative representations of preference, as in (Stone and Thomason, 2002).

Understanding starts from the syntactic structure of the utterance and the current context. It accesses the grammar to compute possible logical forms, and proves them abductively. Its interpretation is the overall least-cost proof. Conversely, generation implements a best-first search over grammatical derivations and assesses its progress based on the same model of utterance interpretation. As described in Section 4, the generator con-

[5] We don't pursue here the question of whether the generator must model the interpreter's abductive cost, or can coordinate directly based on its own costs. The simplest and most direct way to achieve coordination would be for the generator to assume an interpreter that is like it in important respects. But this assumption will work only if dialogue agents can attune themselves to one another in ways that crucially affect the processes of generation and interpretation, and if the reasoning architecture treats these processes symmetrically.

structs the lowest-cost interpretation that will be understood as intended.

6 An Example

We will illustrate our approach with correspondences between informational state and the form of English referring expressions that are described in (Gundel et al., 1993). (For a related discussion of identifiability and activation, see (Lambrecht, 1994)[Chapter 3].) This paper postulates a "givenness hierarchy" with six different levels; here we will consider only the following part of the hierarchy: *IF* (*in focus*), *Act* (*activated*), UI (*uniquely identifiable*), and *TI* (*type identifiable*).

The levels of this hierarchy, in the order in which they were presented, correspond to the progressively weaker prominence of a reference. A referent that is in focus is not only known and actively under consideration, but because of the recent discourse or the mutual environment it is maximally prominent. A referent that is activated is in the current "short-term memory" of the conversation; it is readily available for retrieval. A referent is uniquely identifiable if there is an easy way of constructing a predicate that applies to the referent, and that distinguishes it for the hearer from all other referents. Finally, we treat type identifiability as a residual category.[6] Note that we depart from (Gundel et al., 1993) in that their formulations are characterized in terms of the hearer's knowledge. In view of the importance of mutuality, we substitute 'grounded' for 'known by the hearer'.

Gundel et al. do not provide a detailed model of the cognitive state of a conversational agent. To incorporate their hierarchy in a discourse agent, we need to be more explicit. For current purposes, we can assume that the cognitive state consists of the following components:

(i) A partially ordered set $\langle F, \preceq \rangle$ of items that are in focus (in conversational short-term memory and at the center of attention). The ordering $\preceq$ represents relative prominence.

(ii) A set $STM \supseteq F$ of items in short-term memory.

iii) A set $D \supseteq STM$ of referential distractors, and

(iv) A set CG of common-ground predications, where a common-ground predication is a pair $\langle P, d \rangle$ consisting of a predicate P and a distractor d in D.

(v) A set $DR \supseteq STM$ of discourse referents.

Our four givenness categories can then be defined in terms of cognitive state as follows:

(1) $IF = F$.

(2) $Act = STM$.

(3) $UI =$ $\{i : i \in D$ and there is a conjunction of propositions $P_1(d_1) \wedge \ldots \wedge P_n(d_n)$, where each proposition is in CG, such that i is the only member of D satisfying this conjunction$\}$.

(4) $TI = DR$.

For simplicity, we assume that *IF* is a unit set.

Say that an item *is classified* by givenness category X if it belongs to X and to no more restrictive definiteness category. According to Gundel et. al, the pronouns 'he', 'she', and 'it' are appropriate for items classified by IF; definite NP's with determiner 'that' or 'this' are appropriate for items classified by Act; definite NP's with determiner 'the' are appropriate for items classified by DR; and indefinite NP's with determiner 'a' are appropriate for items classified by TI. The following examples illustrate the correspondences.

(6.1) The neighbor's dog is a neighborhood nuisance.
It kept me awake last night.

(6.2) I'm going to talk to the neighbor with the terrier.
That dog kept me awake last night.

(6.3) [A dog is heard, barking outside the window.]
That dog kept me awake last night.

[6] An item is type identifiable according to Gundel et al. if the hearer "is able to access a representation of the type of object described by the expression." This last characterization strikes us as problematic. For one thing, we believe that the hierarchy is best thought of as applying to (discourse) referents, prior to referring expression generation. But in this case, it is not clear what would be meant by "the expression." If you amend the characterization to require that the hearer should be able to access some predicate that in fact (whether or not the hearer knows it) applies to the item, then the classification applies to everything if 'thing' is a predicate.

(6.4) I'm beginning to regret moving into this place.
The neighbor's dog kept me awake last night.

(6.5) I'm short on sleep.
A dog kept me awake last night.

We accept these correspondences. We now show how our conversational agent realizes them.

Our agent uses inference about interpretation to maintain the elements (i–v) of its cognitive state. The grammar of referring expressions includes rules such as these that update cognitive status:

(6.6) A sentence whose subject is discourse referent i creates a cognitive state in which $IF = \{i\}$.

(6.7) A discourse referent is activated when it is mentioned, and this activation persists until the end of the conversational episode.

These rules determine the specifications for context change in utterance interpretations, as specified in Proof (3.3). The discourse manager executes these specifications as part of context update. Nonlinguistic events induce similar updates:

(6.8) A discourse referent is assigned to a new entity that is perceptually salient in the mutual perceptual environment, and is assigned an activated status.

Now, our agent exploits these connections, by once more reasoning about interpretation in context. Our grammar specifies preferred associations between cognitive state and linguistic patterns within a specific context. For instance, we link pronouns to referents by (6.9).

(6.9) A context in which $IF = \{i\}$ is associated with linguistic preferences for a pronominal NP when i is the reference of that NP,

In combination with (6.6), (6.9) will cause the interpreter to prefer X as the referent to the utterance of "He" in (6.1). The very same mechanism will cause the generator to prefer "He" here: among the recognizable references to X, this utterance has the lowest cost.

7 Discussion

The original theory of interpretation as abduction contributed uniform analyses for a wide range of pragmatic phenomena in isolated sentences. We wish to claim that our new approach offers a similar benefit for dialogue; it formalizes analyses in common terms that can be combined or reconciled with one another. However, readers familiar with the formalization of interpretation in (Hobbs et al., 1993) will note that proofs such as (3.3) are quite different from the interpretation format of that paper. These differences are natural consequences of rethinking interpretation so that it applies to utterances rather than merely to sentences, and to ensure that it takes only into account information that is shared. The format of Proof (3.3) also differs slightly from the account of interpretive proofs given in (Stone and Thomason, 2002); these differences have mainly to do with the need to provide a common form for interpretation and generation. The differences between Proof (3.1) and the account of the same example in (Stone, 2003) have mainly to do with the introduction of explicit speech acts, represented as operators on the common ground.

Individually, the principles of our account of English referring expressions are familiar. Interpretation involves inference that combines linguistic meaning and an extensive common ground: see e.g., (Clark and Marshall, 1981). Grounding a new utterance involves coordinated updates to the attentional state of the dialogue; see e.g., (Brennan, 1998). Interpretation means recovering the most preferred interpretation; see e.g., (Hobbs et al., 1993). Conversely, generation means using the most preferred form that can be understood in context; see e.g., (Buchwald et al., 2002). We regard it as a strength of our formalism that its key principles are not controversial.

Nevertheless, formalizing (6.1–6.5) as we have done crucially requires a framework in which all three principles are simultaneously available. No previous formalism does this. For example, treatments of reference in generation such as (Stone et al., 2003) model the information that interlocutors use to disambiguate expressions, but require additional mechanisms if they are to select expres-

sions with a suitable form. Conversely, treatments of pronouns in discourse such as (Buchwald et al., 2002) can determine whether a pronoun naturally evokes a target referent, but do not generalize to the construction of referring expressions when a pronoun would be ambiguous. Our proposal reconciles the insights of both approaches in a single computational formalism.

8 Conclusion

In this paper, we have presented an approach to the coordination of generation and interpretation in dialogue that adds a principled treatment of discourse context to the interpretation as abduction framework. Coordination depends on shared background that can inform interpretation and on shared preferences that say which interpretations are natural. In our architecture, generation and understanding both rely on these preferences to derive good interpretations. The architecture explains how speakers can use knowledge of language to produce and understand the concise and comprehensible utterances they seem naturally to use.

The flexibility of the approach gives new impetus to efforts to analyze extended natural dialogues in formal terms. We are optimistic that such projects will confirm the elegance and power not only of the formalism itself, but of the intuitive principles of coordination in dialogue that it embraces.

References

Susan E. Brennan. 1998. Centering as a psychological resource for achieving joint reference in spontaneous discourse. In Marilyin A. Walker, Aravind K. Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*, pages 227–249. Oxford University Press.

Adam Buchwald, Oren Scwartz, Amanda Seidl, and Paul Smolensky. 2002. Recoverability optimality theory: Discourse anaphora in a bidirectional framework. In Johan Bos, Mary Ellen Foster, and Colin Matheson, editors, *EDILOG 2002: Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialogue*, pages 37–44. Cognitive Science Centre, University of Edinburgh, Edinburgh.

Herbert H. Clark and Catherine R. Marshall. 1981. Definite reference and mutual knowledge. In Arivind Joshi, Bonnie Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, Cambridge, England.

Jonathan Ginzburg. 1996. Interrogatives: Questions, facts, and dialogue. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 385–422. Blackwell Publishers, Oxford.

Jeanette K. Gundel, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142.

Knud L. Lambrecht. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press, Cambridge, England.

Staffan Larsson and David Traum. 2000. Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering*, 6:323–340.

Paul Smolensky. 1996. Generalizing optimization in OT: A competence theory of grammar 'use'. Paper presented at the Stanford Workshop on Optimality Theory.

Robert C. Stalnaker. 1981. Assertion. In Peter Cole, editor, *Radical Pragmatics*. Academic Press, New York.

Matthew Stone and Richmond H. Thomason. 2002. Context in abductive interpretation. In Johan Bos, Mary Ellen Foster, and Colin Matheson, editors, *EDILOG 2002: Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialogue*, pages 169–176. Cognitive Science Centre, University of Edinburgh, Edinburgh.

Matthew Stone, Christine Doran, Bonnie Webber, Tonia Bleam, and Martha Palmer. 2003. Microplanning with communicative intentions: The SPUD system. *Computational Intelligence*, 19(4). To appear.

Matthew Stone. 1998. *Modality in Dialogue: Planning, Pragmatics and Computation*. Ph.D. dissertation, Computer Science Department, University of Pennsylvania, Philadelphia, Pennsylvania.

Matthew Stone. 2003. Communicative intentions and conversational processes in human-human and human-computer dialogue. In John Trueswell and Michael Tanenhaus, editors, *World Situated Language Use*. The MIT Press, Cambridge, Massachusetts. To appear.

Modelling Concession Across Speakers in Task-Oriented Dialogue

Kavita E. Thomas and Colin Matheson
School of Informatics
University of Edinburgh
{kavitat, colin}@cogsci.ed.ac.uk

Abstract

We determine criteria for modelling concession signalled via *plan-based* "but" in task-oriented dialogue (TOD) (following (Thomas, 2003)) based on some examples from the corpora, analysing where Speech Act (SA) and planning information fails to predict concession. In our approach we focus on cases involving cross-speaker concession, where the speaker accepts part of the previous speaker's turn but signals contrast via "but", and we argue that this contrast can (in the examples shown) be modelled as concessive. Then given a representation of task-plan history in the Information State (IS) model of the dialogue, we present an initial framework for an algorithm that predicts concessive interpretation across speakers in two situations encountered in the corpora. We motivate this work by showing how it updates beliefs in the PTT (Poesio and Traum, 1998) model of dialogue and can be used to facilitate recognition of planning mismatches and potentially avoid misunderstandings, and more generally improve discourse understanding and natural language generation (NLG).

1 Introduction

In this paper we determine criteria for modelling concession across speakers in task-oriented dialogue (TOD), extending the treatment of cross-speaker concession presented in (Thomas, 2003) to propose an algorithm that models two cases of cross-speaker concession signalled by the contrastive cue "but". We focus our analysis on what is communicated at the planning level in examples from the Maptask and TRAINS TOD corpora[1] and present an algorithm that interprets concession and generates an interpretation based on the presence of certain salient (predominantly plan-related) features. Due to the sparseness of cross-speaker concession data involving "but" (Thomas, 2003), we model two examples from the TRAINS and Maptask TODs rather than model a statistically frequent type of cross-speaker concession. We argue that semantic and speech act (SA) representations don't provide enough information to distinguish what's really being communicated at a task level, and motivate our work by claiming that understanding task-plan-related communication and inferences is essential for full interpretation and generation of utterances in TOD.

Our approach is based in the Information State (IS) framework, assuming the PTT (Poesio and Traum, 1998) model of dialogue. We motivate our work by showing how it updates beliefs in this model of dialogue and facilitates interpretation and generation by predicting the in-

[1] MAPTASK dialogues can be browsed interactively at www.ltg.ed.ac.uk/ amyi/maptask/demo.html and TRAINS dialogues can be found at www.cs.rochester/research/trains/.

ferences involved, which aids the interpretation of speakers' intentions and resolves potential planning mismatches. Our treatment of concession follows work by (Lagerwerf, 1998), extending the approach presented in (Thomas and Matheson, 2003) and (Thomas, 2003) which argued that Lagerwerf's treatment of concession can be extended to dialogue, and connects concessive cases in TOD to the planning information necessary to interpret them.

2 Determining Salient Characteristics of Cross-speaker Concession

Given the cue of contrast signalled by cross-speaker "but"[2], we start by determining what criteria are salient for distinguishing concession in a constructed example illustrating the sort of situation we're initially considering, and then analyse examples from the Maptask and TRAINS TOD corpora. We then determine how concession might be generated in the IS framework and focus on how the utterances connect with planning operations in the task, employing plan-recognition models like those described in (Litman and Allen, 1987) and the COLLAGEN project (Lesh et al., 1999). Our approach follows the concessive analysis presented in (Lagerwerf, 1998) and extended in (**?**) to TOD. Lagerwerf claims that concession requires a contextually available claim (or *tertium comparationis*, TC), for which both a positive and a negative argument are provided. Extending this approach to TOD, (Thomas, 2003) argues that the positive and negative arguments for and against the TC are presupposed as expectations launched by the relevant assertions in the adjacent speaker turns in these cross-speaker "but" cases with respect to the contextually relevant claim (i.e., the TC). Here we determine criteria for predicting concession in order to generate the TC and stance of the speakers with respect to the TC in cases where interpreting concession is reasonable. We motivate this work by demonstrating that recognising concession, the TC, and speakers' expectations regarding this TC, facilitates interpretation and generation of subsequent utterances in the dialogue.

[2]We consider that an occurrence of "but" involves contrast across speakers if the constituent it modifies (i.e., usually the main clause of the turn it appears in) is contrasted with something in a previous speaker turn. Cases where "but" is turn-initial or preceding the first clause of the turn are almost always cross-speaker cases. Here we look for contrast with something in the immediately preceding turns for simplicity.

2.1 A Motivating Example

Let's consider how concession might be interpreted in actual cases by tracing how we'd analyse a simple example first. The concessive interpretation algorithm proposed in section 4 describes two particular situations. In the first case, there is an opening question about which both speakers debate, as in the example below:

(1) B1: Should we add the mushrooms to the sauce now?

A2: That should be fine.

B3: But I thought we need to add the beans first.

Here the TC is the question asked in B1, and A argues for, while B argues against. Presumably B raises the question because of her own doubts about the validity of adding mushrooms at this point in the task, and wants to debate the issue with A in order to resolve what to do. Her doubt is emphasised by the objection she raises in B3. A would then need to respond to this objection by either denying B's concern that beans need to be added first, or by agreeing to B's implicit proposal to either add the beans first, or not add the mushrooms to the sauce. The other case that the algorithm addresses is the case in which A proposes a plan or goal (as in the example below), and B objects to this goal:

(2) A1: Let's add the mushrooms to the sauce.

B2: But don't we need to add the beans first?

B2': But the sauce is already too thick.

Notice that here the TC simply asks whether A's proposal is a valid one, with A implicitly supporting it and B objecting explicitly in B2 and indirectly in B2'. Sometimes (as B2' illustrates) the plan is proposed indirectly, as we shall later see in the TRAINS example. Similarly, interpreting B2' as an objection to the proposed plan requires recognising that B is pointing out a negative effect which will be exacerbated by A's proposal of adding something more to the sauce.

2.2 Analysing an Example from Maptask

Maptask dialogues involve two participants with slightly different versions of a map with various landmarks on it, and the goal is for the route giver (who has a route drawn on her map) to direct the follower along this route from start to finish, so a large part of the task involves establishing that the other participant has the same salient landmarks along the route in order to communicate clear directions. This is the sort of situation assessment we see going on in Maptask dialogue q3nc4 below (with disfluencies removed):

(3) F1: right so i'm going underneath this mountain and along to the abandoned truck ... not underneath the mountain ... above the mountain?

G2: above the mountain yeah.

F3: so it's to the right of the mountain that i'm going?

G4: the abandoned truck's to the left of the old temple

F5: yes, BUT the mountain's between the abandoned truck and the old temple

The DAMSL annotation scheme (Allen and Core, 1996), which annotates utterances by coding their multi-level direct SAs, would predict that G4 has a Forward-Looking Function (FLF) of a statement assertion, and F5 has a Backward-Looking Function (BLF) of partial-acceptance (under the agreement category) and also has a FLF of statement assertion. In terms of planning operator type, both G4 and F5 are facts. Neither semantic information nor direct SAs nor planning operator-type seem to shed much light on the issue of how concession might be interpreted here. Clearly we need to delve deeper into the actual problem the agents are trying to solve, namely their plans, in order to get at the underlying meaning of what's being conveyed by F5's "but". The follower (F) has the mountain on her map which the giver (G) doesn't have, and G only realises this discrepancy in maps after hearing F5.[3] This portion of the dialogue establishes the environment (and therefore path) that must be followed in this part of the map. I.e., the overall goal is to establish the route that must be traversed, which involves the subgoal of establishing salient landmarks around the given route that must be circumnavigated.

Concessive analysis F5 can be seen as expressing concession, since F agrees with G4's assertion but asserts previously unmentioned information which is contrastive because it indicates that despite G4's assertion being acceptable, G4's attempt to respond to her question (F3) hasn't answered it. Notice also that F3 indirectly proposes a route around the mountain, and that G4's indirect answer (i.e., assertion of a fact) can only be interpreted with respect to F3, since G must believe that the orientation of the truck with respect to the temple helps F to recognise the route she is trying to navigate. F5 accepts this information, but implies that it fails to answer F3 via the introduction of a contrastively cued (i.e., "but") assertion. This assertion then needs to be checked against the speaker's plan in order to determine where the contrast lies. In this case, the plan (which includes relative locations of objects on the map as constraints and preconditions of navigational actions) indicates that the path needs to be defined with respect to the mountain (i.e., she needs to know whether to go above or below the mountain).

So the concession expresses acceptance of G4's assertion but indicates that this assertion fails to answer the question, which is a Speech Act (SA) level relation rather than something at the planning level, but planning information is necessary for determining this. Determining where the contrast lies with this planning information also enables detection of discrepancies and deficiencies in their maps and facilitates alignment of their perceived environments. After hearing F5, G understands F's difficulty (they've aligned their maps), and communicates the route to F taking into account the mountain that's on F's map. The TC in this case can be framed as the proposal raised in F3 to go to the right of the mountain, and G4 can be assumed to be against this TC since it doesn't answer F3's question satisfactorily (according to F5), leaving F5 to support the TC.

Is this really argumentative? However, thinking of the two turns as supporting and denying the TC doesn't add much useful information here, and

[3] G2 thinks F1 refers to the white mountain (a different one), which is further along on the route, and responds accordingly.

may not even be accurate, since the speakers are not disputing a claim but resolving a confusion; it might be more useful instead to think of the two turns as expressing a different perspective on the environment. That is, in G's map, because there's no mountain nearby, he expresses the salient features he thinks F must circumnavigate, while in F's map, the route must be specified with respect to the mountain that's between the two features (i.e., the truck and the temple) that G mentions. So while argumentative stance *per se* is not an issue here, determining where the discrepancy lies is crucial, and viewing G4 and F5 as conveying different perceived environments with respect to the salient feature specified in the TC (i.e., the mountain) resolves this confusion. In other words, determining the TC and the speakers' perspective with respect to this TC facilitates this realignment by isolating the discrepancy in their plans and establishing how the preconditions raised in G4 and F5 both need to be accounted for in their joint plan.

2.3 Analysing the TRAINS example

In the TRAINS TODs[4], the agents' plans are considerably more complicated than in many other TODs (e.g., Maptask) and involve planning future actions rather than interleaving task and speech actions. Here the agents plan how to achieve a transportation task together where one agent knows additional constraints and guides the user. In the dialogue subsection below (from d.93-15.2, with disfluencies removed), one needs to understand where in the plan the speakers are in order to recognise that the user (U) is actually proposing a new subplan with her assertion of an effect, and that the system (S) agrees that this effect holds but rejects the validity of the indirectly proposed subplan (on the basis that the desired task goal, in this case getting the maximum amount of oranges to Bath by 7 a.m., wouldn't be met).

(4) U: well actually I'll already have an engine in Bath after I unload the boxcars right

S: right but you wouldn't have it in time

Where semantic and SA analyses fail Semantic representations of the utterances clearly can't capture the inferences that must be made in order to determine that U is actually indirectly proposing a plan. Considering the speech act (SA) level, the DAMSL annotation scheme would predict for U the FLF of information-request and statement assertion and the BLF of reject-part (cued by the "actually"). (Indirect SAs are not marked, though if they were, then the utterance would be annotated as an action-directive.) DAMSL would classify S' utterance as involving the BLFs of both partial-acceptance and partial-rejection (for the "right" and the "but you wouldn't have it in time" parts of the utterance respectively), along with the FLF of statement-assertion. Simply mapping the assertional content of the turns to planning operators maps both turns to effects in this case, which doesn't give enough information to infer that U is proposing a new plan. Clearly we need to account for where in the joint plan these operators occur and how they are related, i.e., we need to account for prior planning actions when annotating cases like this, which we argue can be facilitated by accessing a task-plan history (TPH) for the dialogue so far that can be maintained in a dynamic semantic representation like the IS framework.

Where plans can help This particular example involves an implicit plan proposed by U to take the engine already in Bath back up to Corning with the boxcars to collect more oranges and return to Bath. This implicit proposal is recognised by S, who then criticises the effect of this plan with respect to achieving the agents' main goal of getting to Bath by 7 a.m. with the maximum amount of oranges by indicating that they can't get the oranges back to Bath by 7. In our approach, we assume that a planner that can recognise the implicit proposal in U's assertion and the criticism of this proposal's effect in S's turn is available[5] and communicates with the dialogue manager in order to keep the IS updated with what's happening in terms of the agents' salient planning actions. We also assume that an interpreter is available to map utterances into planning operators and SAs, which is feasible in restricted TOD domains and has been implemented in other systems (e.g., the

[4] See www.cs.rochester.edu/research/trains/

[5] Via scripts we write that communicate with the planner and request forward-chaining or plan-recognition with various inputs.

BEE tutorial dialogue project[6] and the COLLAGEN project (Lesh et al., 1999).

Concessive analysis In this example the TC asks whether U's proposed plan is valid, and U argues for it (inferred since U proposes it) while S argues against. Notice that in order to determine this TC we need to infer U's proposed plan and record it in the IS, requiring a dynamic representation of plan history for the dialogue so far. In fact, given an evolving record of the plan history in the IS, we can interpret this example as involving Denial of Expectation (DofE, following (Thomas and Matheson, 2003)) with the expectation $having\_engine\_at\_Bath\_at\_current\_stage > plan\_of\_getting\_more\_oranges\_to\_Bath-in\_time\_succeeds$, which is more informative than the DofE interpretation proposed using their algorithm, which would predict $having\_engine\_at\_Bath > having\_it\_there\_in\_time$, since their approach doesn't have the benefit of prior planning history. The bonus of being able to predict useful expectations for DofE in these indirect TOD cases provides an additional argument in favour of maintaining planning history in the IS. In the next two sections, we will describe how to derive concessive interpretations in plan-based cases like this.

3 Incorporating Task-Plan History in the IS

In (Thomas, 2003) we saw that annotation based on shallow analysis didn't give conclusive results or indicate trends that help predict rhetorical relations, so we bypass the problems of sparse and diverse TOD data by resorting to a "deeper" analysis that requires a dynamic representation of task-plan history in the IS. Consider the IS shown in fig. 5 which contains salient task-plan information that helps us interpret the TRAINS example above.

We omit irrelevant fields and acts here for brevity, and just show what the relevant part of the IS[7] (following the IS structure given in (Matheson et al., 2000))would look like for S (after hearing U utter CA1[8]) in Ex.4. Notice that TPH contains Task Actions (TA), mirroring Conversational Acts (CAs) in the Dialogue History (DH). The recognition of CAs as signalling specific TAs (e.g., TA4 and TA5) would occur within a CA interpretation process which is called by the dialogue manager, and would involve plan-recognition (e.g., (Litman and Allen, 1987) and (Lesh et al., 1999)). This process would require communication with the planner, since many CAs communicate information about planning operators (e.g., the assertions of effects in the TRAINS example) rather than directly making plan proposals. In these cases the interpreter would need to call the planner with the specific planning operator communicated and have the planner forward-chain from this operator to see if the TC is achievable given this input. These processes would be called by update rules in the dialogue manager that call the interpreter and planner in order to determine what intentions are being communicated in the utterances.

How the TPH helps resolve anaphoric actions For example, before update CA3 would contain $assert[S, not\_enough\_time(do(X))]$; in the IS above, the argument X of the anaphoric action communicated by S in CA3 has already been resolved to TA4 in the update process by recognising that U is making a proposal. U's indirect plan proposal can be inferred via the interpreter, planner and the TPH; i.e., TA4 is inferred by the planner upon receiving CA3, thereby enabling recognition (in the planner) that an alternative to TA3 is being proposed, and resulting in TA5. This resolution should occur before the update module determines whether or not to generate concession, and the algorithm proposed in the next section takes as input an IS that's been updated with this information, which simply requires the dialogue manager to call the algorithm after applying the update rules which interpret the input and updates the IS with the TAs recognised by the planner. So to resolve anaphoric actions like $do(X)$ in CA3, we need to incorporate into the update module a

[6] See www.cogsci.ed.ac.uk/ jmoore/tutoring/index.html.

[7] The abbreviated fields are: Previous and Current Dialogue Unit (PDU and CDU), Ground (GND) and Conditions (COND). TPH is omitted in CDU for brevity, since it's empty.

[8] The field values for the $take$ action in TPH are: *with (where E1-3 are engines), from, to and via*.

(5)
$$\left[\begin{array}{ll} \text{PDU} & \left[\begin{array}{ll} \text{GND} & [\text{TA3: } take(E3, Elmira, Bath, Corning)] \\ \text{DH} & [\text{CA1: } assert(U, have[engine, in(Bath), after(unload(boxcars))])] \\ \text{TPH} & \left[\begin{array}{ll} \text{TA4:} & propose\_plan(take([E2, 2boxcars], Bath, Bath, Corning)) \\ \text{TA5:} & propose\_alternative(TA4, TA3) \end{array}\right] \end{array}\right] \\ \text{CDU} & \left[\begin{array}{ll} \text{DH} & \left[\begin{array}{l} \text{CA2: } acknowledge(S, CA1) \\ \text{CA3: } assert(S, not\_enough\_time(TA4)) \\ \text{CA4: } concession(S, TC[willTA4succeed?], for(TC, U), \\ against(TC, S)) \end{array}\right] \\ \text{COND} & \left[\begin{array}{l} accept(U, CA4) \rightarrow scp(U, concession(S, TC[willTA4succeed], \\ for(TC, U), against(TC, S))) \end{array}\right] \end{array}\right] \end{array}\right]$$

rule of the form:[9]

1. If PDU.TOGND.TPH contains a TA that matches $propose\_alternative(TA_i, TA_j)$ then
   (a) If PDU.TOGND.DH contains a CA_n of the form $assert(Y)$ where Y contains $do(X)$, then replace $do(X)$ with TA_i in Y.

3.1 The Benefits of Modelling Cross-speaker Concession

If U accepts the concession communicated by S, e.g., by saying "Yeah, I guess you're right, going back up to Corning for more oranges wouldn't work", then she is socially committed (Matheson et al., 2000) to the concession relation being communicated with the given TC and S's expectation that her proposed plan wouldn't work. In fact, she needs to recognise that S is responding to her implicit plan in order to know that S recognised it. On the other hand, if she doesn't accept S's concession, she can indicate this by saying "But it only take an hour to travel between Bath and Corning, so I should get it there in time", which requires that she understands that S is signalling a rejection of the implicit plan she proposed, and counters S's argument that her proposed plan wouldn't work. So both recognising these implicit plans and also the concession relation S communicated is essential for her to generate the appropriate response in cases like these.

4 An Initial Framework for an Algorithm for Concession

We propose an update algorithm to model the cases of cross-speaker concession discussed above. (We will take into account other cases after more examples from the corpora are analysed.) This algorithm would be called by an update rule in the dialogue manager triggered by "but". If the required circumstances (outlined in the algorithm) apply in the given case, concession is predicted. How do we distinguish the simplicity of the Maptask example above, where the TC for concessive interpretation arose from the immediately preceding question in F(3) from the plan-dependendent interpretations in the TRAINS example? As a first step we could check for questions to which alternative answers are being given in the turns containing and preceding the "but", and if so, predict that the question or proposal is the TC and generate concessive interpretations if the planner verifies that the relevant turns argue for and against (or favour and disfavour–given the planner) the TC (case 1 in the algorithm below).

If there's no preceding question then one should check whether the PDU proposes a plan, goal or action[10], and if so, whether the "but" turn counters this proposal also via communication with the planner (case 2). Here we only address cases where the planning operator communicated in CDU (the "but" turn) is an effect or precondition (referred to as *Operator(TA$_i$)* below, where TA_i is the task action proposed in PDU) that can't be met given the proposal made in PDU, and we'll focus on other possible planning operator combinations in future work. So the initial distinctions an algorithm modelling these cases should make are as follows, given a dialogue with B uttering "but" in CDU (i.e., turn i):

1. If turn B($i-2$) contains a question in CA_k in

[9] This rule is an example of many similar rules for other situations involving anaphoric reference to actions that need to be resolved in order to interpret the dialogue.

[10] Here we only deal with assertions and not indirect proposals, though this will be addressed in later work. We will also address preconditions and effects (both future ones and failed past ones) in PDU in future work, since speakers can object to actions they think are forthcoming or point out problems they think have occurred.

UDU.TOGND.DH,[11]

(a) Call interpreter with (CA_l, CA_j, CA_k) and get back (TA_l, TA_j, TA_k), where CA_l comes from the assertion in turn A$(i-1)$ found in PDU.TOGND.DH and CA_j comes from the assertion in B(i) found in CDU.TOGND.DH where both are uttered by different speakers. TA_k contains the task action queried by the question contained in CA_k; if CA_k queries an effect or precondition, then replace TA_k by CA_k.[12]

(b) Pass $alternatives(TA_l, TA_j, TA_k)$ to the planner, which returns "yes" if CA_l and CA_j provide alternate answers, arguments or comments w.r.t. the question in CA_k and also returns $for[TC[CA_k], speaker(CA_l)]$ and $against[TC[CA_k], speaker(CA_j)]$ if the arguments in TA_l and TA_j support and counter[13] the question in CA_k and vice versa otherwise.

i. If "yes" then

A. Update the IS by adding to CDU.TOGND.DH a new CA of the form $concession(TC[CA_k], for[TC[CA_k], speaker(CA_l)], against[TC[CA_k], speaker(CA_j)])$ or vice versa the arguments of for and against depending on which turn supports and counters the TC and resolving the speakers by determining where CA_k and CA_l occur in the IS.

ii. Else do nothing.

2. Else if [[PDU.TOGND.DH contains a CA_n of the form $assert(Y)$[14] (where Y corresponds to TA_j, which was returned by the planner after the interpreter passed it Y)] and [PDU.TOGND.TPH contains $TA_j = propose\_plan(Y, Plan\_Y)$ (where the planner instantiates $Plan\_Y$ with the plan Y proposes directly or implicitly)] and [CDU.TOGND.TPH contains $CA_k = Operator(TA_i)$, where $Effect$ can be either an effect or precondition and is instantiated in CA_k (e.g., here it's not_enough_time).]] then

(a) Update the IS by adding to CDU.TOGND.DH a new CA of the form $concession(TC["IsPlan\_Yvalid?"], for[TC, speaker(CA_n)], against[TC, speaker(TA_k)])$ where speaker gets resolved appropriately by determining whose turn it is; the speaker of PDU is assumed to favour the proposal they put forth in CA_n while the speaker of the "but" utterance in CDU presumably argues against this proposal.

3. Else do nothing. (This case will be elaborated after analysing more examples.)

Our approach addresses cases like the Maptask and TRAINS examples via update rules in the dialogue manager which call the interpreter (which then sends the planner the planning operators communicated to get back TAs). The interpreter then passes the TAs it receives back to the update rules which update the IS accordingly with the given TAs. An update rule in the dialogue manager triggered by "but" then calls the concessive interpretation algorithm, which tests whether the given situation can be interpreted as concession. The algorithm orders the concessive cases so that we search for the simpler situation of a leading question first (the Maptask case), before exploring whether an alternative plan is being proposed (as in the TRAINS example). The algorithm then spells out how to derive concessive interpretations in either of these cases. We leave for future work (via the "else do nothing" statements) examples that convey concession in different ways. Future work will also address possible social obligations (following (Matheson et al., 2000)) launched by proposals raised (e.g., the TC).

5 Conclusions and Future Directions

In this paper we present a novel treatment of cross-speaker concession when signalled by "but" in two specific cases. We motivate our work by analysing how the examples modelled communicate concession, and show that recognising concession across speakers gives useful information about speakers' expectations and goals and helps avoid potential planning mismatches by enabling the speakers to recognise each others' intentions and respond appropriately and meaningfully. While the sorts of cases we argued were concessive here are far-removed from examples displaying the sort of argumentative force one usually associates with concession (*epistemic* relations, following (Sweetser, 1990)), we call these TOD cross-speaker "but"

[11] UDU was left out of the IS shown in (5) for brevity, and contains ungrounded dialogue units prior to PDU. See (Matheson et al., 2000; Poesio and Traum, 1998) for more information on the IS model and fields involved.

[12] For now we will only allow the question to lie in the turn before the turn preceding the "but", i.e., B$(i-2)$, though a more developed treatment should allow for (possibly via search) salient unanswered questions which have been possibly interrupted by (e.g.) clarificatory or subgoal-resolving subdialogue.

[13] We determine whether TA_l and TA_j support or counter the TC (CA_k) by performing forward-chaining from both TAs in turn to see whether the TC is achievable given that input.

[14] We don't require Y to be an effect, as occurred in the TRAINS example to allow for some generalisability.

examples concessive because they display different stances with respect to some proposal, and identifying these stances gives useful information about what the speakers believe, how they perceive their environment, and what their plans are, and as shown in the Maptask example, they facilitate alignment and grounding and help expedite clarification of misunderstandings.

In future work we will extend this analysis to more cases of concession across speakers found in task-oriented corpora to expand our coverage, and we'll use these examples to also enlarge upon our current conception of concession across speakers in dialogue. Exploring how the combinations of planning operator-type pairs in concessive cases convey potential or previous problems and how these can be resolved will be a practical benefit of this work.

References

J. Allen and M. Core. 1996. DAMSL: Dialog Act Markup in Several Layers. University of Rochester.

L. Lagerwerf. 1998. Causal Connectives Have Presuppositions. *Catholic Univ. of Brabant, Holland Academic Graphics, The Hague, The Netherlands*.

N. Lesh, C. Rich, and C. Sidner. 1999. Using Plan Recognition in Human-Computer Collaboration. *International Conference on User Modelling*.

D. Litman and J. Allen. 1987. A Plan Recognition Model for Subdialogues in Conversation. *Cognitive Science*, volume 11.

C. Matheson, M. Poesio, and D. Traum. 2000. Modelling Grounding and Discourse Obligations Using Update Rules. *Proceedings of the North American Association for Computational Linguistics*.

M. Poesio. and D. Traum. 1998. Towards an Axiomatization of Dialogue Acts. *Proceedings of Twente Workshop*.

E. Sweetser. 1990. From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure. *Cambridge University Press, Cambridge*.

K. Thomas. 2003. Modelling 'but' in Task-Oriented Dialogue. *4th International and Interdisciplinary Conference on Modeling and Using Context*.

K. Thomas and C. Matheson. 2003. Modelling Denial of Expectation in Dialogue. *5th International Workshop on Computational Semantics*.

K. Thomas. 2003. Criteria for Modelling "But" in Task-Oriented Dialogue. *To appear in Proceedings of ESSLLI Workshop on Discourse Particles*.

Topicality in the Semantics and Pragmatics of Questions and Answers: Evidence for a File-Like Structure of Information States

Katsuhiko Yabushita
Department of English
Naruto University of Education
yabuchan@naruto-u.ac.jp

Abstract

In the literature of the semantics and pragmatics of questions and answers, little attention has been paid to topicality in stark contrast to focus. In this work, we will examine a phenomenon from Japanese that strongly suggests the relevance of topicality, and propose a formal analysis in a dynamic-semantic framework, "Extended File Change Semantics" developed by Portner and Yabushita (1998) for the treatment of topic phrases. Then, we will consider its empirical and theoretical implications. Specifically, we will argue that the success of the proposed analysis gives evidence for the thesis that information structures have a file-like structure being segmented for each discourse referent and the update by an utterance in general is a local operation with its effects being restricted primarily to the segment, or file-card for an discourse referent which the discourse is "about". Furthermore, we will discuss the issue whether the alleged relevance of topicality attested by Japanese, which has an explicit topic marker, is motivated also for languages that do not have an explicit topic marker, using data from English.

1 Relevance of Topicality for Semantics and Pragmatics of Questions and Answers: Data from Japanese

First, consider the following question-answer dialogue in Japanese, which has an explicit marker for a topic phrase, i.e. *-wa*:

(1)

Q. Dare ga hashitte
who Nom running

imasu ka.
be Q
'Who is running?'

A1. Jon ga hashitte imasu.
John Nom running be
'John is the one who is running./Only John is running.'

A2. #Jon wa hashitte imasu.
John Top running be
'As for John, he is running.'

What is to be noted here is the difference between (1A1) and (1A2) in interpretation as answers to (1Q). First, (1A1) and (1A2) are minimally different from each other in that the subject, *John* is nominative-marked in (1A1), while it is topic-marked in (1A2). Interpreted out of context, there are no truth-conditional differences between (1A1) and (1A2); both of them are true if and only John is running. Interpreted as

answers to (1Q), however, there are truth-conditional differences between them. As the glosses indicate, (1A1) implies that only John is running, which is considered an instance of 'exhaustiveness' (Groenendijk and Stokhof 1982, 1984, 1990). On the other hand, (1A2) commits itself only to the truth of John's running, staying away from the issue whether the other relevant people are running or not. A word is in order about (A2) being #-marked. The sharp sign indicates that (A2) is not as felicitous or "congruent" an answer to (1Q) as (A1); (A2) does not strike one as straight an answer as (A1), giving the impression that you are withholding some information relevant to the question under consideration, or sidestepping the issue.

We take the above interpretational phenomena to imply among others, the following facts about topicality in relation to the semantics and pragmatics of questions and answers:

(2) a. Topicality has some semantic or truth-conditional significance.[1]

b. Topicality is relevant in the semantics and pragmatics of questions and answers, at least with respect to exhaustiveness.

2 Formal Analysis

In the following, we will propose a formal analysis of the phenomenon that does justice to the observations in (2), and discuss its implications for the interpretation of dialogue, especially with respect to the structure of information states and how they get updated.

2.1 Extended File Change Semantics

As the semantic framework in which the analysis to be proposed will be couched, we adopt the one that was developed by Portner and Yabushita (1998) for a semantic treatment of topicality, called "Extended File Change Semantics". As the name suggests, the framework was an extension of File Change Semantics proposed by Heim (1982: Chapter 3) for an analysis of the semantics of (in)definite noun phrases and nominal anaphora. Portner and Yabushita extended it for a semantic treatment of topic phrases in that information states (Portner and Yabushita call them "common grounds") are constructed to have a file-like structure in having a file card, or segment for each discourse referent. In the setting, the topic phrase of a sentence was analyzed as a pointer of a discourse referent whose file card is to be selectively updated by the propositional content of the sentence, along with the so-called "aboutness condition" approach to topicality (cf. Vallduví 1990).

Here, let us review Extended File Change Semantics to the extent that it is essential to understand the following discussion. Therein, an information state ('file' in Heim 1982, 'common ground' in Portner and Yabushita 1998) is defined to be a set of infinite sequences of pairs, where each pair consists of an entity and a set of possible worlds. Suppose A is an element of an information state, i.e. an infinite sequence of pairs of entities and sets of possible worlds. Intuitively, the *i*th pair denoted $<e_{i, A}, I_{i, A}>$ represents a discourse referent numbered *i*; $e_{i, A}$ is a possible value for the discourse referent *i*, and $I_{i, A}$ is propositional information having been attributed to the discourse referent so far in the discourse. With an information state having a segment for each discourse referent, the update function for a formula can be specified to selectively update the segment designated for a particular discourse referent, as in (3). There the updating function is denoted $+_k$, with index *k* indicating that the file card for discourse referent *k* is to be updated, an information state is denoted CG (common ground) following Portner & Yabushita, and for an expression α, Int(α) is the intension of α.

(3) For an information state CG, an *n*-place predicate R ($n \geq 1$), and variables $x_i, \ldots, x_j$,
$CG +_k R(x_i, \ldots, x_j) =$
$\{A \in CG :$ for every $w \in I_{k, A}$,
$< e_{i, A}, \ldots, e_{j, A} > \in Int(R)(w)\}$.

Using the (indexed) update function, the topic phrase of a sentence can be analyzed in terms of information-state update as directing which discourse referent's segment is to be updated with the propositional content of the sentence, as in (4).

[1] For more evidence for the semantic or truth-conditional significance of topic phrases other than the data in (1), see (Portner and Yabushita, 1998).

(4) For any information state CG, and any sentence of the logical form $[T_i, \phi]$, where T_i is a topical discourse referent *i* and ϕ is a formula for the sentence,
$CG + [T_i, \phi] = CG +_i \phi$.

2.2 Extension of Extended File Change Semantics for the Treatment of Questions and Answers

Now that the basic features of Extended File Change Semantics have been reviewed, we will present an analysis of questions and answers as we extend the framework by adding necessary features specific to questions and answers. First, logical forms will be enriched to incorporate focus structure along with the structured-meaning approach to focus (von Stechow 1989, Krifka 1991), according to which the logical form of a sentence in general has a binary structure <B, F>, where B is the background part and F is the focus part. Given a sentence S with a focused constituent A with their "ordinary" logical forms ϕ and α, respectively, the structured-meaning logical form of the sentence will be $<\lambda X.\phi[\alpha/X], \alpha>$, where $\phi[\alpha/X]$ is the result of replacing α in ϕ with an appropriate variable X. On the assumption that a WH-phrase is inherently focused (Rooth 1985 among others), a WH-sentence will have a logical form of the background-part form. The logical form actually coincides with the 'relational' meaning of a question on the "categorial" approach to questions and what Groenendijk and Stokhof (1982, 1984, 1990) called the "(*n*-place) abstract". We adopt the background-focus form logical form augmented with the topic structure abstracting over the focus part as the meaning of a WH-question because of fact (2b). That is, a WH-question will now have a logical form of the form of $\lambda Y[T_i, <B, Y>]$, with the focus part being abstracted, called "focus abstract". For example, question (5a) will be considered to have the logical form as in (5b), and a possible, congruent answer to (5a), e.g. (6a) will be considered to have the logical form as in (6b).

(5)

a. Jon_1 wa dare o
John Top who(m) Acc
aishite-imasu ka.
love Q
'Who(m) does John love?'

b. $\lambda Y[John_1, <\lambda X.LOVE(John_1, X), Y>]$

(6)

a. Jon_1 wa $Meari_2$ o
John Top Mary Acc
aishite-imasu.
love
'John loves Mary.'

b. $[John_1, <\lambda X.LOVE(John_1, X), Mary_2>]$

With formulas of the background-focus structure now introduced, it is necessary to define the (index) updating of an information state with so structured a formula. It is defined as follows:

(7) $CG +_i <B, F> = CG +_i B(F)$, where B is an expression of the form $\lambda X.\phi$ and B(F) is the result of substituting F for every occurrence of X in ϕ.

What about the logical forms of (1Q), (1A1), and (1A2), which prompted the current discussion in the first place? The obvious problem with (1Q) and (1A1) is that they do not have a topic phrase present. Here, we assume that some contextually determined situation is the topic for (1Q) and (1A1); furthermore, the situation occurs as an argument of the predicate in question (Davidson 1967, Kratzer 1988/1995). That is, the logical forms of (1Q), (1A1), and (1A2) are considered something as in (1Q)′, (1A1)′, and (1A2)′, respectively.

(1Q)′ $\lambda Y[s_3, <\lambda X.RUNNING(s_3, X), Y>]$

(1A1)′ $[s_3, <\lambda X.RUNNING(s_3, X), John_1>]$

(1A2)′ $[John_1, < John_1, \lambda X.RUNNING(s_3, X)>]$

With the logical forms of the relevant sentences determined, let us proceed to the interpretation of questions and answers and an analysis of exhaustiveness. We take answering to be an illocutionary act of assertion in the environment

of a question, which is interpreted to be a (partial) function from information states to information states. In this setting, exhaustiveness will be analyzed as a conversational implicature arising from an (optional) operation on the information state resulted from updating by answering. Schematically, the current analysis of answering a question and exhaustiveness can be represented by the following diagram:

(8)
<common ground CG, question Q, answer A>
|
|←illocutionary act of answering (ANSWER)
|
<common ground CG′, question Q >
|
|←exhaustivization (EXHAUST) (optional)
|
common ground CG″

Diagram 1: Interpretation Schema of Answering Optionally Followed by Exhaustivization

The operation of ANSWER is defined as follows:

(9) ANSWER(CG, Q: $\lambda Y[T^q, <B^q, Y>]$, A: $[T^1, <B^1, F^1>], \ldots, [T^n, <B^n, F^n>]$)
maps a common ground CG to a common ground CG′ such that
CG′ = CG + A: $[T^1, <B^1, F^1>], \ldots, [T^n, <B^n, F^n>]$ on the following felicity condition among others: The question and the answer sentence(s) have the same topic and the background parts, i.e. $T^q = T^1 = \ldots = T^n$ and $B^q = B^1 = \ldots = B^n$.

What ANSWER does is basically to update a common ground CG in such a way that for every $A \in CG$, the information segment for the topical discourse referent, say *k*, i.e. $I_{k, A}$ should entail the propositional content of the (the conjunction of) the answer sentence(s).

Next, we go on to the definition of the exhaustivization operation, EXHAUST. It is an operation on an information state resulted from the operation of ANSWER, as is indicated in Diagram 1. Before we go into the formal definition of EXHAUST, it will be expositional to outline what EXHAUST does first so that we will not be bogged down in the details involved.

Suppose CG′ is an information state resulted from the application of ANSWER to some information state CG in the environment of a question Q and some answer sentence(s). The operation of EXHAUST will modify CG′ modulo Q: $\lambda Y[T^q, <B^q, Y>]$ into an information state CG″, which is minimally different from CG′ in that for every $A \in CG''$, for the topic discourse referent, say *k* (= T^q), every possible world $w \in I_{k, A}$ is such that the extension of $\lambda Y[T^q, <B^q, Y>]$ in w contains only the individuals 'minimally' warranted by the answer sentence(s); in other words, w is "minimal" in $I_{k, A'}$ for some $A' \in CG'$ with respect to the extension of B^q.

Here, let us illustrate what we mean by "the individuals 'minimally' warranted by the answer sentence(s)". Consider the following examples of question-answer dialogue.

(10) Q. Who came to see Mary?

A1. John did.
A2. A man did.
A3. Siskel or Ebert did.

As answers to (10Q), (10A1), (10A2), and (10A3) are normally interpreted exhaustively that only John came to see Mary, that one man and only one man did, and that either Siskel or Ebert, but not both did, respectively. Those interpretations all can be characterized as those in which the extension of the predicate in question 'came to see Mary' is specified to be a minimal set containing the individual(s) the answer asserts to have come to see Mary.

In terms of the current view of interpretation as a process of information-state updating, exhaustivization will be analyzed as pragmatically induced, optional updating to be applied post to the initial updating by the literal content of the answer sentence(s). (See Diagram 1.)

For the formal definition of EXHAUST, we need a couple of auxiliary notions to be defined. First, we need to determine the extension of a focus abstract as the meaning of an interrogative.

(11) Definition (Extension of a Focus Abstract)

Given a focus abstract Ψ of the form $\lambda Y[T_i, <B, Y>]$, a model M, a possible world w, a common ground CG, and an index $j \notin Dom(CG)$, the extension of Ψ with respect to M, w, and CG, denoted $|\Psi|^{M, w, CG}$ is defined as follows:

$|\Psi|^{M, w, CG}$

$= \{e_{A, j}: A \in CG + \Psi(j) \ \& \ w \in I_{A, i}\}$.

Based on the extension of a focus abstract Ψ of the form $\lambda Y[T_i, <B, Y>]$ with respect to a model M, a possible world w, and a common ground CG, we will define a partial order on $I_{i, A}$ for $A \in CG$, denoted $\leq_{\Psi, CG}$.

(12) Definition ($w \leq_{\Psi, CG} w'$)

Given a Topic-Background abstract Ψ of the form $\lambda Y[T_i, <B, Y>]$, a common ground CG, a sequence $A \in CG$, for any two possible worlds, w and $w' \in I_{i, A}$,

$w \leq_{\Psi, CG} w'$

if and only if $|\Psi|^{M, w, CG} \subseteq |\Psi|^{M, w', CG}$.

Given a set of possible worlds ordered with respect to the partial order just defined, the "minimal" elements of the set can be defined as follows:

(13) Definition ("Minimal" Possible Worlds)

Given a Topic-Background abstract Ψ of the form $\lambda Y[T_i, <B, Y>]$, a common ground CG and a sequence $A \in CG$, let $<I_{i, A}, \leq_{\Psi, CG}>$ be $I_{i, A}$ with the partial order $\leq_{\Psi, CG}$. Then, $w \in I_{i, A}$ is a minimal element of $I_{i, A}$ if and only if for any $w' \in I_{i, A}$, $w' \leq_{\Psi, CG} w$ implies $w' = w$.

With all the necessary auxiliary notions having been defined, we can finally go on to the formal definition of EXHAUST, which is as follows:

(14) Definition (EXHAUST)

Let CG′ be an information state resulted from answering to a question whose focus abstract is Ψ with an answer sentence(s) on an input information state CG, i.e. CG′ = ANSWER(CG, Ψ, A), which is CG + $[T^1, <B^1, F^1>]$, $[T^2, <B, F>]$, …, $[T^n, <B, F>]$. When the answer A is uttered with some linguistic signal indicating that the utterance is complete such as the falling tone in English, and there is no expression to explicitly defy an exhaustive reading, CG′ is subjected to the following operation EXHAUST.

EXHAUST(CG′, Ψ)

$= \{A': A \in CG' \ \& \ A'$ is exactly like A except that $I_{i, A'} = \{w: w$ is minimal in $< I_{i, A}, \leq_{\Psi, CG'}>\}\}$.

2.3 The Current Analysis' Account of the Data

In the above we have presented a semantic and pragmatic analysis of questions and answers incorporating topicality, which has been shown to be relevant to the semantics and pragmatics of questions and answers. It is time to see how the current analysis fares well with the data introduced at the outset, specifically the interpretational facts surrounding the question-answer dialogues: (1Q)-(1A1) and (1Q)-(1A2). The sentences, (1Q), (1A1), and (1A2) will be reproduced here along with their information-structural logical forms, (1Q)′, (1A1)′, and (1A2)′, respectively.

(1Q) Dare ga hashitte
who Nom running

imasu ka.
be Q
'Who is running?'

(1Q)′ $\lambda Y[s_3, <\lambda X.RUNNING(s_3, X), Y>]$

(1A1) Jon ga hashitte imasu.
John Nom running be
'John is the one who is running./Only John is running.'

(1A1)′ $[s_3, <\lambda X.RUNNING(s_3, X), John_1>]$

(1A2) #Jon wa hashitte imasu.
John Top running be
'As for John, he is running.'

(1A2)′ [$John_1$, < $John_1$, λX.RUNNING(s_3, X)>]

Now let us see how the interpretation of (1A1) as an answer to (1Q) is analyzed. The felicity condition for ANSWER in (9) is satisfied as their topic parts and background parts coincide with each other. Then the answering operation proceeds to update a given CG to CG_{A1}′ as in (15).

(15) ANSWER(CG, (1Q)′, (1A1)′)
= CG + (1A1)′ = CG + [$John_1$, < $John_1$, λX.RUNNING(s_3, X)>] = CG_{A1}′
= {A ∈ CG : for every w ∈ $I_{3, A}$, John is running in the situation 3 in w}.

CG_{A1}′, then, will undergo the exhaustivization operation, i.e. EXHAUST(CG_{A1}′, (1Q)′), and will be modified into CG_{A1}″ as in (16).

(16) EXHAUST(CG_{A1}′, (1Q)′) = CG_{A1}″ =
{A′: A ∈ CG_{A1}′ & A′ is exactly like A except that $I_{3, A'}$ = {w: w is minimal with in < $I_{3\ A}$, $\leq_{(1Q)', CG_{A1}'}$>}}.

As for every A ∈ CG_{A1}′, for every w ∈ $I_{3\ A}$, w is such that the extension of (1Q)′ contains at least John, a possible world w is minimal in < $I_{3\ A}$, $\leq_{(1Q)', CG_{A1}'}$> when the extension of (1Q)′ at w contains only John. Therefore, the resulted information state, CG_{A1}″ can be expressed in more plain language as in (17).

(17) EXHAUST(CG_{A1}′, (1Q)′) = CG_{A1}″ =
{A ∈ CG_{A1}′, for every w ∈ $I_{3, A,}$ the extension of running in the situation 3 in w contains only John, or equivalently, only John is running in the situation 3 in w}.

We have seen that answering (1Q) with (1A1) felicitously updates an information state and the resulted information state is further modified to one such that (the segment "about" the situation 3 of) it entails that only John is running (exhaustiveness), which coincides with the empirical data. Next, let us see the case of answering (1Q) with (1A2)? The felicity condition for the answering operation is not satisfied as neither the topic parts nor the background parts of (1Q)′ and (1A2)′ coincide with each other. In terms of topic, (1Q) is "about" the situation 3, while (1A2) is "about" John, and in terms of focus, (1Q) focuses on who is running in the situation 3, while (1A2) focuses on what John is doing in the situation 3. We claim that these discrepancies between (1Q) and (1A2) are responsible for the infelicity of (1A2) as an answer to (1Q). Coerced to be interpreted despite the violation of the felicity condition, (1A2) would update a given information state CG to CG_{A2}′ as in (18).

(18) ANSWER(CG, (1Q)′, (1A2)′)
= CG + (1A2)′ = CG + [$John_1$, < $John_1$, λX.RUNNING(s_3, X)>] = CG_{A2}′
= {A ∈ CG : for every w ∈ $I_{1, A}$, $John_1$ is running in the situation 3 in w}.

As is shown in (18), (1A2) would update the segment "about" John with respect to what he is doing when question (1Q) is "about" the situation 3 and solicits information about who is running in the situation. This, we claim, coincides with the fact that (1A2) as an answer to (1Q) gives the impression that it is "sidestepping the issue". As for exhaustiveness, the exhaustiveness operation would not affect the segment "about" John, which has been updated by (1A2) because the operation was defined to affect the segment about the topic of the question, in this case, that of the situation 3. Consequently, the segment "about" John remains to entail that John is running, no more or no less, which accounts for the absence of exhaustiveness from (1A2).

3 Cross-Linguistic Considerations: Is Topicality Universally Relevant to Semantics and Pragmatics of Questions and Answers?

The current analysis of the semantics and pragmatics of questions and answers in relation to topicality is essentially based on the data from Japanese, which has an explicit morphological marker for topic phrase. The questions that might occur to the reader naturally at this point include among others the following. Is the current analysis relevant to the cases of other languages, especially those that have no explicit

topic marker? Isn't the background-focus structure sufficient for an analysis of the data in question? In other words, is it necessary to bring in topicality into the picture?

For those questions, let us consider the following examples of English question-answer dialogue, where there is a phonological prominence, or sentential stress on the capitalized phrases.

(19) Q. Who is running?

A1. JOHN is running.
A2. #John is RUNNING.

This set of examples is a perfect reflex of that in (1) in that (19A1) is interpreted to mean that only John is running as (1A1) is, and that (19A2) is infelicitous as an answer to (19Q) giving the impression that your are "sidestepping the issue" as (1A2) is to (1Q) and if coerced to be interpreted, (19A2) could only be interpreted non-exhaustively, i.e., that John is running with no implications as to whether the other people are running or not.

On the assumption that a WH-phrase is inherently focused and the phonological prominence is a focus marker in English; in (19A1) and (19A2), the subject *John* and the predicate *is running* are considered to be focused, respectively, the background-focus structures of (19Q), (19A1), and (19A2) will be as in (19Q)′, (19A1)′, and (19A2)′, respectively.

(19Q)′ $\lambda Y[<\lambda X.\mathrm{RUNNING}(s_3, X), Y>$

(19A1)′ $<\lambda X.\mathrm{RUNNING}(s_3, X), \mathrm{John}_1>$

(19A2)′ $< \mathrm{John}_1, \lambda X.\mathrm{RUNNING}(s_3, X)>$

It is observed that in the case of an felicitous question-answer dialogue, i.e. (19Q)-(19A1), their background parts are identical, while in an infelicitous case, i.e. (19Q)-(19A2), the background parts are distinct. In light of the observation, we could propose the following felicity condition for questions and answers, which is weaker than that in (9), not requiring the identity in the topic part.[2]

(20) Given a question and an answer with their background-focus structures, Q: $\lambda Y<B^q, Y>$, A: $<B^1, F^1>, \ldots, <B^n, F^n>$, Q and A are a felicitous question-answer pair only if the background of the question and that of the answer are identical, i.e., $B^q = B^1 = \ldots = B^n$.

With the condition above, which is free from reference to topicality, we can explain the felicity facts of (19). As for the exhaustive vs. non-exhaustive difference in reading between (19A1) and (19A2). One might propose an account like the following. As an infelicitous answer, (19A2) does not quality for the exhaustivization operation, which is applicable only to felicitous answers; therefore, the reading available to (19A2) on a coerced interpretation would be at most the one of the literal meaning, in this case, the non-exhaustive reading.

With the putative success of the above topicality-free account of the semantic and pragmatic facts of (19), one might argue that the same account should be applicable to the Japanese data in (1) as well; therefore, the relevance of topicality will be in doubt.

For such a contention, we would like to present the following question-answer example.

(21) Q. Who(m) did John introduce Mary to?

A1. John introduced Mary to BILL.
A2. #Mary was introduced to BILL by John.

What is to be noted here is that as the sharp sign indicates, (21A2) is not felicitous as an answer to (21Q) unlike (21A1).

If the background-focus structure were a sufficient articulation for logical forms of questions and answers, (21A1) and (21A2) should not show any differences in felicity as answers to (21Q), for (21A1) and (21A2) are considered to have the same background-focus structure. Actually, however, they do as we have noted above. On the other hand, the felicity facts in question are not a problem to our current analysis, which adopts a more articulated structure for logical forms. On the widely accepted assumption that the subject is a default position for a topic phrase in English, the logical forms of (21Q), (21A1),

[2] Conditions of the same effects have been proposed in e.g. (Yabushita, 1991, 1992), (Rooth, 1992), and (Krifka, 1999).

and (21A2) will be something as in (21Q)′, (21A1)′, and (21A2)′, respectively.

(21Q)′ $\lambda Y[John_1, <\lambda X.INTRODUCED(s_3, John_1, Mary_2, X), Y>]$
(21A1)′ $[John_1, <\lambda X.INTRODUCED(s_3, John_1, Mary_2, X), Bill_4>]$
(21A2)′ $[Mary_1, <\lambda X.INTRODUCED(s_3, John_1, Mary_2, X), Bill_4>]$

Then, the current analysis, specifically, the specification of the ANSWER operation correctly predicts that (21A2) will not be felicitous as an answer to (21Q) as they do not have an identical topic-background structure, while (21A1) and (21Q) do. We take what has been seen above to be evidence that topicality is crucially involved in the semantics and pragmatics of questions and semantics even in languages that have no explicit topic-marker like English.

4 Conclusion

We have presented a phenomenon from Japanese, i.e. data in (1) as evidence for the relevance of topicality to the semantics and pragmatics of questions and answers, and have proposed a formal analysis couched in Extended File Change Semantics, which was developed by Portner and Yabushita (1998) for the treatment of topic phrases. We take the success of the analysis to imply the legitimacy of the framework with regards to the structure of information states and the mode of updating information states by utterances.

Acknowledgement

The author is grateful to the two anonymous reviewers for carefully reading the submitted extended abstract carefully and giving detailed comments and suggestions. They indeed have turned out to be very helpful in preparing the current version of the paper although not all of them have been able to be incorporated. Of course, the author would be solely responsible for any remaining inadequacies.

References

Donald Davidson. 1967. The Logical Form of Action Sentences. In N. Rescher (Ed.). The Logic of Decision and Action, 81-95. University of Pittsburgh Press.

Jeroen Groenendijk and Martin Stokhof. 1982. Semantic Analysis of Wh-Complements. *Linguistics and Philosophy*, 5: 175-233.

Jeroen Groenendijk and Martin Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph. D. Dissertation, University of Amsterdam.

Jeroen Groenendijk and Martin Stokhof. 1990. *Partitioning Logical Space*. Annotated Handout for the Second European Summer School on Logic, Language and Information.

Irene Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph. D. Dissertation. University of Massachusetts, Amherst.

Angelika Kratzer. 1995. Stage-Level and Individual-Level Predicates. In Gregory N. Carlson and Francis Jeffry Pelletier (Eds.) *The Generic Book*, 125-175. University of Chicago Press.

Manfred Krifka. 1991. A Compositional Semantics for Multiple Focus Constructions. *Proceedings of the First Semantics and Linguistic Theory Conference* (SALT I).

Manfred Krifka. 1999. For a Structured Account of Questions and Answers. Ms. University of Texas at Austin.

Paul Portner and Katsuhiko Yabushita. 1998. The Semantics and Pragmatics of Topic Phrases. *Linguistics and Philosophy*, 21:117-157.

Mats Rooth. 1985. *Association with Focus*. Ph. D. Dissertation. University of Massachusetts, Amherst.

Mats Rooth. 1992. A Theory of Focus Interpretation. *Natural Language Semantics* 1: 75-116.

Arnim von Stechow. 1989. Focusing and Backgrounding Operators. Arbeitspapier 6 der Fachgruppe Sprachwissenschaft, Universität Konstanz.

Enric Vallduví. 1990. *The Informational Component*. Ph. D. Dissertation. University of Pennsylvania.

Katsuhiko Yabushita. 1991. Topic, Focus, and Information. Talk Presented at the Eighth Amsterdam Colloquium.

Katsuhiko Yabushita. 1992. On the Exhaustive-Listing Use of Japanese Nominative Marker *Ga*. Talk Presented at LSA annual meeting.

The Syntax Semantics Interface of Speech Act Markers

Henk Zeevat

Abstract

The paper proposes a semantics for speech acts as a generalisation of update semantics and an account of the meaning of speech act markers where they are default resetting devices. The speech act marker works on specific parameters of a given speech act and defines new valus for them.

The update semantics of (Veltman, 1996) and (Heim, 1983) runs into its limits if it tries to handle non-monotonic phenomena and non-assertions like requests, exclamations or —it can be argued contra (Jäger, 1996), (Zeevat, 1994) or (Groenendijk, 1999)— questions.

This would be a minor problem only if it would be the case that these phenomena can be isolated from the syntactic/semantic system of natural language. But that does not appear to be the case in the least. Discourse particles, intonation and aspectual markers can have the kind of pragmatic contribution to the utterance in which or on which they appear that is best understood as a modification of the speech act performed by the utterance. This is not normally derivable from the semantic content, from the anaphoric properties of the marker, or from the kinds of operators that can be handled in update semantics.

This paper makes the case in point using a number of such particles and some recent proposals for a meaning of intonation.

The generalisation of update semantics that appears to be required is to extend the coverage from just plain updates (the case of standard assertions) to the full range of possible speech acts and the effects they have on the context. This includes next to updates downdates. But it is also necessary to be able to recover topics, pass over the authority to the other speaker, to make links to previous utterances, to have a range of evidential and modal operators and to recover various forms of bias. The proposal I am making here is no doubt not sufficiently general.

I assume that speech acts have at least three dimensions. The first dimension is the set of preconditions for the speech act: what must be the case with the context of the utterance that makes it possible to carry out the speech act. This includes the presence of a pivot for the discourse relations or of a topic that is to be addressed (often the relation to the pivot is given by the topic, so arguably these are not separate cases). The second dimension is the aim that the speaker wants to achieve with her speech act. This aim may be of various kinds and it does not depend on the speaker alone whether she is going to reach it. It is therefore necessary to distinguish carefully between her goal with the utterance on the one hand to be achieved together with the hearer and the effect that the utterance will achieve without any further action from the hearer, if the utterance as such is successful (i.e. it is perceived and recognised as such). I call this last aspect the minimal effect.

The first dimension can be called context marking. It is concerned with the preconditions of the speech act and changes to the defaults assumed there. Too and $instead$ are both additive markers in the sense that they presuppose an old topic that has been addressed before. But their contribution differs in the second and third dimension. With too, we intend to bind an old topic to the new value that is obtained by adding the value specified in the sentence to the old value. With $instead$, we intend to replace the old binding by the new value specified in the utterance. This also affects the third dimension: in the case of *too* the speaker endorses

the old value of the topic in addition to the value specified in the sentence, whereas, in the case of *instead*, she disagrees with the old value and expresses her belief to know that the value expressed in the utterance is the value that should have been given in the first place.

The second dimension, the intention, is a proposal for bringing about a new common ground. The intention can only be reached if the hearer in some sense cooperates and takes over the goal of the speaker. This cooperation can be minimal, in the sense of not speaking out against it, as with proper assertions and promises (the speaker is the authority in such speech acts, even if her authority can be challenged). But in a question (a proposal to create a common ground which contains the answer to the question), it is the hearer who has to create (part of) the content of the proposal.

Updates belong to the context change effects. They are most clearly exemplified by assertions but also play a role in requests, promises and questions. These other speech acts give updates with respect to the knowledge of the speaker and with respect to the courses of action that the speaker wants to be pursued. They do this even if the goal of the speaker is not achieved. The full semantics for speech acts is composed of conditions on the common ground, the proposed change to the common ground and the updates about the speakers beliefs and desires. It allows inferences about the common ground before the speech act, about the choices facing the hearer and about the information the hearer and third parties have gained if the speech act was successful. When the hearer acts as the speaker had planned and the speaker does not challenge her authority, the intended effect is reached. This characterises what is expected of the hearer in response to a speech act.

Assertions have the following preconditions, intentions and minimal effects. Affirmatives are the least marked form in natural language and all other speech acts require more marked forms. I pursue here the view that all other marking devices (intonation, particles, special syntactic forms, etc.) are there to override some of the parameter settings of the assertion. A speech act marker must in this view be seen not as something that marks a speech act by itself, but as something that together with other devices constructs the speech act from the default speech act.

Formalising our speech act theory requires some choices. The simplest representation of a speech act is as a set of parameters with default settings. Let's give them some names and give the settings for the speech of an assertion of $\varphi(c)$ (these are the default settings).

POSITIVE BIAS	none
NEGATIVE BIAS	none
TOPIC BIAS	none
TOPIC	$?x\varphi(x)$
CONTENT	$\varphi(c)$
OPERATOR	nil
INTENTION	OPERATOR(CONTENT)
AUTHORITY	speaker
MINIMAL EFFECT	$B_s K_s \varphi(c)$

POSITIVE BIAS is the presence in the common ground of $\varphi(c)$ or a commitment to $\varphi(c)$ or evidence for $\varphi(c)$. The value *none* for the assertion indicates that it is assumed that there is no positive bias. Assertions fail if there is positive bias or rather they change into confirmations of different kinds: of information earlier incorporated in the common ground (it is a reminder then), of an opinion of the hearer or a third party or of the evidence for $\varphi(c)$. NEGATIVE BIAS is the opposite: it is constituted by the presence in the common ground of $\neg\varphi(c)$ or a commitment to it or evidence for it. An assertion comes with the implicature that there is no NEGATIVE BIAS and is incorrect if there is such. With negative bias, it gives way to corrections of various kinds: corrections of the common grounds, of parties involved in the conversation and others, and of the evidence (roughly, the concessive relationship).

TOPIC BIAS is the assumption that the topic has been addressed before, i.e. that there is information about the topic already in the common ground, possibly again only as commitment of the conversation partners or others. In assertions, there is no TOPIC BIAS: the topic is assumed to be unaddressed, though it normally has been raised.

The TOPIC parameter is a suitable topic that the current assertion asserts. I assume that it is given

as a question. The assertion supplies an answer for this topic in its focus.

CONTENT contains the propositional content of the speech act and is one of the parameters that together with AUTHORITY and OPERATOR determines the proposal: to make OPERATOR(CONTENT) common ground on AUTHORITY.

AUTHORITY contains the person who is assigned the role of the decision maker and the OPERATOR is the operator under which the assertion is to be entered into the common ground. Evidentiality is the main target of this parameter, distinction between hearsay, direct evidence, belief and possibilities should be made here.

The last parameter, the MINIMAL EFFECT records the propositional attitude of the speaker towards CONTENT that she makes manifest with her utterance. The minimal effect is what makes the proposal rational.

The parameters making up the speech act define it: under the circumstances indicated by the bias-parameters, the speech act indicates a proposal for a new entry in the common ground, possibly correcting earlier information, the proposal to be decided by the speaker or the hearer.

Chains of speech acts form conversations and there is one aspect that I am neglecting here: the fact that speech acts license other speech acts both by the same agent and by the other agent. Licensing can be described to some extent by restrictions on the TOPIC parameter. The TOPIC must then be set to one of the available topics, with a mechanism describing which ones are available. Ginzburg e.g. describes a mechanism of this kind.

The use of speech acts however brings with it the idea of expected responses of other people. Asking a question is expecting the other to do one of range of things: answering the question, possibly overanswering it or underanswering it, refusing to answer the question, asking a counterquestion or an elucidation question. The range of these reactions is part of our understanding of the question speech act. We can only say that expected reactions are what is necessary to effect the proposal (or a related one) or ways of declining it.

Speech acts can be modelled by a system of changes common grounds. The best choice is here to model common grounds by their basis, a set of propositions σ. φ is common ground iff it is a logical consequence of σ and the common ground inference rules (Zeevat, 1997). The downdates that we need to assume can all be given as removal of specifically targeted propositions in σ. Some special predicates and operators. W for a shared desire, entailing W_s (the speaker wants) and W_h (the hearer wants), common ground belief B entailing B_s and B_h (speaker and hearer belief) and possibly a similar operator for knowledge. Further ways of recovering raised topics, addressed topics and an operator for proposals. A speech act is only defined on σ if its preconditions are common ground. The change is that its minimal effect and the speaker's proposal become common ground.

Speech acts change a number of things in the common ground: the facts, the goals, the beliefs and goals of the speaker and the hearer. It is thereby not purely about information, and consequently, speech act semantics is not a semantics for a logical formalism. It contains however the semantics for a number of logical formalisms, like e.g. update semantics for the content, and for operators like may or the definedness operator for presupposition.

A full speech act semantics would be able to take a common ground and predict for a given speech act whether it is allow given the preceding discourse, what changes it brings to the common ground before the hearer's intervention, what choices it gives to the hearer and what the effects are of the hearer's intervention.

1 Speech Act Markers

Speech act markers come in many shapes and flavours. Syntactic form is perhaps the most basic. Many languages have special forms for imperatives, interrogatives and affirmatives. E.g. in German, fronted WH-XPs mark WH-questions, V2 affirmatives, V1 without a subject imperatives and V1 interrogatives.

Performative verbs and modal verbs form a second category of markers. Intonation is important in speech act marking, in dimensions like bias, topic and authority. And a final group is particles, possibly appearing as clitics or morphs.

The reason for calling them speech act markers is

that they can change the values of one or more of the parameters defining the speech act. The other parameter values are inherited from the speech act to which it applies. There is no evidence here for full operators: the parameters can be affected only once. There can be no conflict. It also does not seem necessary to assume an order in which the markers apply. The markers seem to provide a set of instructions to reset the default parameter values in a particular way. What is possible though is that some markers behave differently given other markers that apply at the same time. This is similar to normal contextual disambiguisation and does not change the concept of a speech act marker.

In principle and in practice it is possible that a marker both contributes to the content and acts as a speech act marker. We can use content operators (e.g. embedding verbs) to limit the interpretation of the utterance as a speech act. (1) is in principle ambiguous between a description of what I am doing and carrying out the promise to be good,

(1) I promise to be good.

in a way another way (2) of carrying out the same speech act is not.

(2) I will be good.

"I promise" is thereby not a pure speech act marker. In (3), for example, it does not carry out a promise at all.

(3) If I promise to be good, will you not be angry?

Mixed speech act markers give an interesting way of thinking about the grammaticalisation processes underlying the pure speech act markers which —by the principle that all markers derive from lexical words— must have come into being by such processes operating on mixed speech act markers. Assuming this process allows some insight into the fact why the less central speech act markers —and the speech acts they mark— are so different even in closely related languages. The explanation is that new markers grammaticalise not just in response to general communicative needs but also in response to the available inventory of speech acts markers and the frequency with which they are used and that aspects of its earlier history of the speech act marker may still be encoded in its distribution. Much more work is necessary to gain further insight in the reasons for the diversity of speech act marking as well as the universal patterns.

Speech act markers —like other markers— come in two flavours: obligatory markers and optional markers. Optional markers typically compete with a range of similar markers and the unmarked form. It is typical that what they mark can also be assumed without the marker in question. Our last example can be interpreted as a promise (but also as a prediction about the speaker's future behaviour) and there is competition for the explicit marker with other markers.

Obligatory markers are markers that cannot fail to be there if the speech act has the property that is marked. Obligatory markers can be replaced by another equivalent marker perhaps, but it is not possible to use the unmarked form. This seems to arise from the default interpretation of the unmarked form. If the form is unmarked, the interpretation includes that the marked interpretation does not obtain. This gives the reason why the marker is necessary and may the most important reason for the genesis of grammaticalised speech act markers. The lexical markers are turned into functional ones and lose their additional meaning by semantic epenthesis (Fong, 2001).

The most important claims here are the following two. Speech act markers can be described as mapping speech acts into other speech acts by resetting the value of one or more speech act parameters. And: if the parameter is not reset by an obligatory marker, the speech act retains the default value.

2 Some Markers

Syntactic patterns seem to play only a minor role in marking speech acts. Affirmative sentence can have many functions, as can interrogatives. Imperatives and constituent questions —by the restrictions inherent in their form— seem more on target. But affirmatives and interrogatives conspire with auxiliaries, particles and intonation to define a wide range of speech acts. Typical of the interrogatives is the transfer of authority from speaker to hearer. But there are exceptions to that. In a suggestion like (4) or in rhetorical questions, the transfer is basically a fiction.

(4) Isn't there a pub near here?

Proper questions can be defined by syntactic form, intonation, particle, morphs and tags. They vary considerably in properties though all transfer authority to the hearer.

(5)
a. John is in Berlin,isn't he?
b. Ni hao ma? (Are you doing well?)
c. John is in Berlin?
d. Is John in Berlin?
e. Isn't John in Berlin?
f. John is doch in Berlin? (I thought John was in Berlin?)

(5b,d) are unbiased questions that are inditinguishable form assertions except in the authority parameter. (5a) expresses (in the minimal effect parameter) the speaker's opinion that John is in Berlin and allows positive bias. (5c) requires $B_h K_h \varphi$ as positive bias, while (5d) has $B_h K_h$*John not in Berlin* as positive bias and as part of the minimal effect $B_s K_s$*John is in Berlin* and as possible value of negative bias. (5f) inverts these values, the speaker indicates that she believes to know John to be in Berlin (minimal effect which can be possible value of positive bias), while the hearer believes to know that John is elsewhere as part of negative bias.

Auxiliary-pronoun combinations like $can\ you$, $you\ can$, $you\ may$, $may\ I$, $will\ you$, $you\ will$ and $I\ will$ and embedding verbs + pronouns like $I\ promise$ or $I\ request$ can be added to affirmatives and interrogatives to form requests, permissions, requests for permission, offers and promises in the obvious way.

I am taking the naive view here that we can believe to know on the authority of one of the speakers that one of them will do something, is permitted to do something or can do something. That seems to be the fiction underlying promises, request and orders.

Let us go through a promise in our representation scheme for speech acts.

POSITIVE BIAS	$W_h \varphi$
NEGATIVE BIAS	none
TOPIC BIAS	none
TOPIC	$?\varphi$
CONTENT	φ
OPERATOR	nil
INTENTION	OPERATOR(CONTENT)
AUTHORITY	speaker
MINIMAL EFFECT	$B_s K_s \varphi$

Threats are made by inverting the bias, offers to do φ by shifting the authority to the hearer and changing the minimal effect to $W_h \varphi \rightarrow \varphi$. Requests shift the authority and have as minimal effect that $W_s \varphi$.

It is generally assumed that intonation can mark transfer of authority by itself, in the so-called sentence final rising. It also serves to indicate surprise and can thereby be used to mark both reminders that contradict what the other just said or implied and confirmation questions where the surprise indicates that the speaker assumes that there is a conflict with what assumed before. Within our interpretational scheme we could also accommodate the case where there is an intonational marking of the fact that a weaker topic is addressed than the one given in the context (the so-called "twiddle"). It seems a good deal easier to address the semantic/pragmatic contribution of intonation as setting parameters in speech act concepts than in a framework like modal logic as in (Steedman, 2003).

My personal motivation for looking at speech act marking is to make some progress with the semantic/pragmatic properties of particles. Following the presupposition literature, (Kripke, ms)'s observation on the particle too and the ideas of (Blutner and Jäger, 2000) on optimality theoretic versions of presupposition resolution and accommodation, it seemed that progress could be made (Zeevat, 2002). But as I indicated, already too and its cousin $instead$ suffice for making it quit dubious that this will work. Both particles have the same presupposition (or mark the same contextual property) that the current topic has been addressed before. But that does not clear up their difference. We need to talk about the different context change that they define: one adds to the old information on the topic, the other replaces it. This cannot be coded in the informational content of the utterance or in its contextual marking properties, but needs the difference between an assertion and a correction.

In the current framework, too and $instead$ both have a topic bias: they assume that the topic has

been addressed before. But they make a different proposal: *too* proposes the content of the clause that it marks, *instead* also proposes to eliminate the earlier information (and has it as minimal effect that the speaker believes to know that the earlier information was not correct.

Another old problem is the Dutch particle *wel* having both an emphatic use, which seems to mark negative bias in the common ground and correction, and various hard to describe nonemphatic uses like:

(6) Het komt wel goed.
Don't worry.

Like its English translation, *wel* seems to refer to the worry, i.e. the thought that things will not be allright and corrects it. But it is not a proper correction: the speaker has no proper evidence and the main effect of the particle is to tone down BK to B or even weaker. Negative bias is what the two uses share. It can perhaps be described as a way of refusing to make the hearer's belief that it will go wrong common ground. The intonational difference seems due to to the fact that the emphatic *wel* requires an overt negative antecedent with which it stands in a contrast relationship. Where the antecedent is merely supposed, contrastive intonation is unnecessary. There are however uses of *wel* which seem to lack any relation to a negative opinion of the hearer like (7).

(7) We gaan wel wat eten in de stad.
(??)We can go and eat something in town

The same picture arises with Dutch/German *toch*/*doch*. Most share negative bias, but within that they do different things.

(8) 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 most curious thing about *toch* is that the nonemphatic uses decay into markers of reminders (like (a) and (e)). Again the emphatic uses require proper antecedents. The development of a use as a reminder particle probably derives from its use as the marker of a reconfirmation question for an element of the common ground in the face of conflicting evidence (b).

Reconfirmation is also the typical function of *indeed*. Here we have positive bias like in reminders, but there is a crucial difference. Where unemphatic *doch* calls on the hearer's memory for the reconfirmation, *indeed* brings the idea of new and proper evidence. We can capture this in our representation by instantiating the positive bias parameter with a weaker antecedent.

3 Conclusion

I have given the outline of a theory of speech act marking. Syntactic form, content, intonation and particles combine in setting a number of parameters that together determine the speech act. The parameters are set by overriding a default and the default itself has a strong content that gives rise to a range of unexpressed implicatures. inferences. The semantics of the representation is given by possible developments of the common ground under the speech act. This needs a generalisation of update semantics that incorporates corrections and the notion of authority, as well as a series of quasi-

modal operators.

Speech act semantics is the proper place to study particles and marked syntactic forms. In a pure update semantics, characterisations remain partial only and the connections between the different uses of particles remains unclearer than they need to be.

In a theory of speech acts in which there are default settings for various parameters, it is possible to understand better why the use of particles is governed by marking principles. If the particle is not there to override the default setting, the default is assumed. This means that the speaker will be misunderstood if she omits the particle when she intends a speech act with a property expressed by the particle. Obligatory marking can therefore arise from the zero situation with an ambiguous speech act form. If marking has become possible, if there is a statistically based preference for one of the readings (possibly resulting from optional marking) and if there are sufficiently many misunderstandings, the optionality of the marking will decrease and the bias for misunderstanding the unmarked form as expressing the standard case will increase. This promotes further marking and stronger bias towards misinterpretation of the standard form. Under the appropriate statistical conditions this leads to obligatory marking.

References

Reinhard Blutner and Gerhard Jäger. 2000. Against lexical decomposition in syntax. In A. Z. Wyner, editor, *The Proceedings of the Fifteenth Annual Conference*, pages 113–137. The Israeli Association for Theoretical Linguistics.

Vivienne Fong. 2001. "already" in singapore english. Abstract, Aspect Conference, Utrecht, December 12-14.

Jeroen Groenendijk. 1999. The logic of interrogations. In T. Matthews and D.L. Strolovitch, editors, *The Proceedings of the Ninth Conference on Semantics and Linguistic Theory*, CLC.

Irene Heim. 1983. On the projection problem for presuppositions. In Michael Barlow, Daniel Flickinger, and Michael Westcoat, editors, *Second Annual Westcoast Conference on Formal Linguistics*, pages 114–126. Stanford University.

Gerhard Jäger. 1996. Only updates. on the dynamics of the focus particle only. In P. Dekker and M. Stokhof, editors, *Proceedings of the 10th Amsterdam Colloquium*, pages 387–406, ILLC, University of Amsterdam.

Saul Kripke. ms. Presupposition.

Mark Steedman. 2003. Information-structural semantics of english intonation. ftp://ftp.cogsci.ed.ac.uk/pub/steedman/prosody/santabarbara.pdf.

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

Henk Zeevat. 1994. Applying an exhaustivity operator in update semantics. In H. Kamp, editor, *Ellipsis, Tense and Questions*, DYANA-2 Deliverables. ILLC, University of Amsterdam.

Henk Zeevat. 1997. The common ground as a dialogue parameter. In Gerhard Jaeger and Anton Benz, editors, *Proceedings Mundial 97*. CIS, Universitaet Muenchen.

Henk Zeevat. 2002. Explaining presupposition triggers. In Kees van Deemter and Rodger Kibble, editors, *Information Sharing*, pages 61–87. CSLI Publications.

Project I1-OntoSpace: Ontologies for Spatial Communication[1]

John A. Bateman bateman@uni-bremen.de
Kerstin Fischer kerstinf@uni-bremen.de
Reinhard Moratz moratz@informatik.uni-bremen.de
FB10: Faculty of Linguistics and Literary Sciences
Center for Computer Studies
University of Bremen, Germany

Scott Farrar farrar@uni-bremen.de
Thora Tenbrink tenbrink@sfbtr8.uni-bremen.de
FB10: Faculty of Linguistics and Literary Sciences
University of Bremen, Germany

Abstract

The poster describes the essential ingredients of project I1-OntoSpace of the "SFB/ TR8 on Spatial Cognition: Reasoning, Action, Interaction" (Bremen/ Freiburg). The SFB/TR8 is an interdisciplinary collaborative research center that has been financed by the DFG since January 2003. The poster highlights the empirical studies in linguistic human-robot interaction being carried out, including first results.

1 Project Outline

Natural language is an essential mode of interaction between users and sophisticated spatially-aware systems such as mobile assistance robots. Providing suitably sophisticated natural language capabilities for ever more complex interaction scenarios is a major problem. An important contributing factor is the lack of appropriate modularizations of the technical components involved in complete systems that combine spatial and linguistic capabilities. Such interaction needs to be as natural and non-intrusive as possible in order to support the widest possible range of potential users, and so sophisticated accounts are required. Most of the features that make interaction 'natural' still present substantial challenges for dialog systems and solutions are not to be found within single research areas.

One particularly effective strategy for modularization is the adoption of linguistically motivated ontologies that mediate between domain/application knowledge and Human Language Technology (HLT) components. But current strategies for ontology mediation exhibit a rigidity that is inappropriate for the relationships observed between domain and linguistic knowledge in real interactions. Here, a negotiation of mediation within 'conversational' interaction appears crucial both for ontology design and for achieving interoperability.

This project therefore has the primary goal of developing a toolbox of ontology-based methods suitable for supporting natural language interaction and technology re-use within a range of interaction scenarios involving spatially-aware systems. The methods developed will be motivated by detailed empirical investigations of the actual communicative strategies of negotiation employed by users of robotic systems. The results of these experiments will feed into a situation/scenario-parameterized formalization of inter-ontology mappings.

[1] The authors listed are the members of Project I1-OntoSpace. The poster is presented by Thora Tenbrink.

The project will contribute to the further development of emerging standards in ontological engineering, their application in the management of natural language, and the development of ontological modules involving spatial representations for mobile robots. It will also use the ontology mediation strategy to overcome a persistent lack of interaction between spatial system design, robotic interaction, etc. and HLT. This has limited both the wider re-usability of important research results concerning the natural language analysis and generation of spatial relations, including fine-grained semantics for spatial expressions, route description planning and understanding, resource-adaptive generation, etc.

2 Empirical Studies

This project is concerned with two primary ontology levels: the linguistic and the spatial; these are defined by importing existing research results and then refined by targeted empirical investigation of linguistic behavior in specially tuned spatial settings. For maximal re-usability of generic components, any language components employed must refer only to the linguistic ontology level. Contact with the spatial knowledge is achieved by inter-level mediation.

The particular mappings to be defined between the spatial and linguistic ontologies are derived from the communicative strategies revealed in the empirical investigations. The empirical problems we address are located in two problem areas: first, the complexity of the interpretation of spatial expressions based on the considerable variability of implicitly underlying reference systems and the associated negotiation processes between the interactants; second, the peculiarities pertaining to the choice of linguistic expressions in an unfamiliar interaction situation involving an artificial interlocutor. Both problem areas combine in linguistic interaction scenarios in which users are required to communicate with an unfamiliar robot about spatial surroundings.

We address the parameterization of ontology mappings related to spatial configurations by making the relationship between situational variables and linguistic properties transparent. For instance, the users' choices of spatial reference systems and of strategies for referring to landmarks are central parameters of linguistic variability in human-robot interaction. Users may (justifiably!) be uncertain about what the robot can perceive, and so the lack of mutual and reflexive common ground for the interactants regarding the spatial situation may lead to insecurity about which objects may serve as landmarks and how they can be referred to. This variability can be experimentally controlled. Closely related to this factor are the participants' linguistic and spatial choices concerning group-based reference – a kind of reference system often neglected in the literature – using further similar objects instead of a different object as a relatum to specify the target object's position. Such issues are addressed by confronting users with tasks involving different configurations of diverse (similar and differing) objects, the robot, and the user.

Results achieved so far suggest that in human-robot interaction, irrespective of the particular spatial configuration, human instructors reliably take the robot's point of view. This behavior differs from that found in human-human interaction and so needs to be considered in designing dialog interpretation strategies for such systems. Furthermore, the dialogue history demonstrably influences users' decisions about their choice of reference systems: With previous success using group-based reference, users tend to employ this format even in a scenario in which the robot only perceives *one* box, a situation where group-based reference does not standardly apply. Users attend to the robot's (linguistic and motional) reactions, relying on previous successful instruction formats as well as aligning to the robot's output on all linguistic levels: morphological, syntactic, semantic and conceptual. A further major area of results concerns the degree to which the spatial arrangement – including the robot's view direction – influences the users' instructions with regard to choice and applicability range of reference system, choice and direction of perspective, complexity of reference, vagueness or redundancy, etc. Functional aspects such as perceived distance and accessibility play a major role, for example in interpreting expressions like *front* and *back*.

Such findings can – and will in the project's future – be used for the parameterization of inter-ontology mappings in order to link abstract spatial representations with their possible linguistic expressions in a flexible manner.

Language Phenomena in Tutorial Dialogs on Mathematical Proofs

Christoph Benzmüller[1], Armin Fiedler[1], Malte Gabsdil[2], Helmut Horacek[1], Ivana Kruijff-Korbayová[2], Dimitra Tsovaltzi[2], Bao Quoc Vo[1], Magdalena Wolska[2]
[1]Fachrichtung Informatik [2]Fachrichtung Computerlinguistik,
Universität des Saarlandes, Postfach 15 11 50, D-66041 Saarbrücken, Germany
{chris,afiedler,horacek,bao}@ags.uni-sb.de, {gabsdil,korbay,dimitra,magda}@coli.uni-sb.de

1 Introduction

Empirical findings (Moore, 1993) show that flexible natural language interaction promotes active learning. In the DIALOG project [1] (Pinkal et al., 2001), we are investigating and modelling semantic and pragmatic phenomena in tutorial dialogs focused on problem solving skills in mathematics. The language used in this domain is particularly challenging because of the mixture of natural language and formal symbolic expressions. We have collected a corpus of dialogs with a simulated tutoring system for proofs in naive sets theory. Corpus investigation into the use of natural language in mathematics enables us to identify the linguistic phenomena that will impose specific requirements on dialog management. The goal of the project is to develop a prototype system gradually embodying the results of our analyses.

2 Empirical Study

To collect a corpus of tutorial dialogs, we have conducted a *Wizard-of-Oz (WOz)* experiment (Benzmüller et al., 2003a). 24 subjects with varying educational backgrounds, and little to fair prior mathematical knowledge participated in the experiment which consisted of 3 phases: (1) *preparation and pre-test* (on paper), (2) *tutoring session* (mediated by the *WOz* tool (Fiedler and Gabsdil, 2002)), (3) *post-test and evaluation questionnaire* (on paper). Subjects were asked to prove 3 theorems[2]: (a) $\overline{(A \cup B) \cap (C \cup D)} = (\overline{A} \cap \overline{B}) \cup (\overline{C} \cap \overline{D})$; (b) $A \cap B \in \wp((A \cup C) \cap (B \cup C))$; and (c) When $A \subseteq \overline{B}$, then $B \subseteq \overline{A}$. The subjects were instructed to enter proof steps, rather than complete proofs at once, to encourage dialog with the system. The wizard's task was to respond to the student's utterances following a given algorithm (Fiedler and Tsovaltzi, 2003): first evaluate their completeness, accuracy, and relevance with respect to a valid proof, then select the next dialog move and verbalize it. The subjects and the wizard were free to mix natural and formal language. The dialogs were carried out in German, and had between 1 and 15 student turns; 5 on average.

3 Language phenomena in utterances

Figure 1 shows example responses; only English translations are included for brevity. We have observed that mathematical and natural language expressions are tightly interleaved, and the language is telegraphic: sentences are fragmented, symbols are used to name relations. We have identified recurring structural and semantic phenomena, examples of which we present below.

Structural phenomena, firstly, concern parsing the interleaved symbolic and natural language ("A also $\subseteq B$") and identifying the status of mathematical expressions (considering scoping issues, as in "B contains no $x \in A$"). Secondly, they concern recognizing typical patterns expressing derivation steps ("If..., then...", "So..."), and possible structural ambiguities in cases such as coordination ("[$x \in B$] and so [$x \subseteq \overline{B}$ and $x \subseteq \overline{A}$] given

[1]DIALOG is part of the Saarland University Collaborative Research Center on *Resource-Adaptive Cognitive Processes* (SFB 378) (http://www.coli.uni-sb.de/sfb378/)

[2]$\wp$ stands for power set

(1)	A power set contains all subsets, hence also $(A \cap B)$
(2)	**T:** ...$\overline{(A \cup B) \cap (C \cup D)} = \overline{A \cup B} \cup \overline{C \cup D}$... **S:** de Morgan rule 2 applied to both complements
(3)	de Morgan rule 1 also holds for $\overline{C \cup D}$ de Morgan rule 2 means $\overline{A \cap B} = \overline{A} \cup \overline{B}$. In this case e.g. $\overline{A}$ = the term $\overline{A \cup B}$ $\overline{B}$ = the term $\overline{C \cup D}$. So $\overline{(A \cup B) \cap (C \cup D)} = (\overline{A} \cap \overline{B}) \cup (\overline{C} \cap \overline{D})$
(4)	$(A \cup B)$ must be in $\wp((A \cup C) \cap (B \cup C))$, since $(A \cap B) \in (A \cap B) \cup C$
(5)	$B \cup A \subseteq CB \cup A \subseteq D$. If A is a subset of C and B a subset of C, then both sets together must also be a subset of C. The same holds for D. $\overline{C \cap D} \cup \overline{A \cap B}$ applying the de Morgan rules.

Figure 1: Examples of students' utterances.

[the assumption]" vs. "[[$x \in B$] and so [$x \subseteq \overline{B}$]] and [[$x \subseteq \overline{A}$] given [the assumption]])" [3].

Semantic phenomena mainly involve imprecise, ambiguous, and/or informal references to domain relations, or formally incorrect, but metaphorically interpretable references. For example, "must be in" in (4) and "contains" in (1), can express the relation *element_of* or *subset_of*; "have common elements" refers to the non-empty intersection; "both sets together" in (5) may be interpreted as union or intersection; the relation "=", used mainly in equations or to indicate a value assignment, has been also used for term substitution in an axiom (in (3), $\overline{A \cup B}$ is to be substituted for $\overline{A}$, and $\overline{C \cup D}$ for $\overline{B}$). Another class of phenomena are references to structural parts of terms and formulae such as "the left side" or "the inner parenthesis" which are incomplete specifications: the former refers to a part of an equation, the latter, metonymic, to an expression enclosed in parenthesis. Moreover, structural parts of formulae constrain the search space for anaphor resolution (e.g. the left- and right-hand side of the equation in (2) for resolving "both complements"). Finally, generic and specific references can appear within one utterance (cf.(1), "a power set" is a generic reference, whereas "$A \cap B$" is a reference to a subset of a specific power set).

4 Ongoing work

Presently, we are implementing a formulae parser capable of (a) identifying formulae, (b) analysing their structural parts. Formulae analysis is a preprocessing step toward a semantic interpretation of utterances. Here, we are pursuing two approaches: (1) shallow recognition of derivation steps based on recurring keywords used to introduce them, and (2) syntactic and semantic analysis using a lexically-based grammar[4]. The input analysis is embedded within the context of implementing an Information State-based dialog manager that is to interact with a mathematical assistant in conducting a tutoring session (Benzmüller et al., 2003b).

[3] Square brackets illustrate alternative structural grouping of coordinated constituents

[4] We are using an open source CCG parser and realizer (http://openccg.sourceforge.net/)

References

C. Benzmüller, A. Fiedler, M. Gabsdil, H. Horacek, I. Kruijff-Korbayová, M. Pinkal, J. Siekmann, D. Tsovaltzi, B. Vo, and M. Wolska. A Wizard-of-Oz Experiment for Tutorial Dialogues in Mathematics. *Proc. of the AIED Workshop on Advanced Technologies for Mathematics Education*, Sydney, Australia, 2003a.

C. Benzmüller, A. Fiedler, M. Gabsdil, H. Horacek, I. Kruijff-Korbayová, M. Pinkal, J. Siekmann, D. Tsovaltzi, B. Vo, and M. Wolska. Tutorial Dialogs on Mathematical Proofs. *Proc. of the IJCAI Workshop on Knowledge Representation and Automated Reasoning for E-Learning Systems*, Acapulco, Mexico, 2003b.

A. Fiedler and M. Gabsdil. Supporting Progressive Refinement of Wizard-of-Oz Experiments. *Proc. of the 6th Inter. Conference on Intelligent Tutoring Systems Workshop on Empirical Methods for Tutorial Dialogue Systems*, San Sebastian, Spain, 2002.

A. Fiedler and D. Tsovaltzi. Automating Hinting in Mathematical Tutorial Dialogue. *Proc. of the EACL-03 Workshop on Dialogue Systems*, Budapest, Hungary, 2003.

J. Moore. What makes human explanations effective? *Proc. of the 15th Annual Meeting of the Cognitive Science Society*, Hillsdale, NJ, 1993.

M. Pinkal, J. Siekmann, and C. Benzmüller. Projektantrag Teilprojekt MI3–DIALOG:Tutorieller Dialog mit einem mathematischen Assistenzsystem. *Fortsetzungsantrag SFB 378–Ressourcenadaptive kognitve Prozesse*, Saarbrücken, Germany, 2001.

Tasks, Games and Domain Knowledge in Dialogue Management

Martin Beveridge, John Fox and David Milward
Cancer Research UK, London WC2A 3PX
{martin.beveridge, john.fox}@cancer.org.uk, david.milward@linguamatics.com

1 Problem

The inexorably rising cost of healthcare provision is a major problem worldwide. A cost-effective way of improving quality of care while containing costs is to manage clinical decision-making and workflow by means of computerized care pathways. These pathways are increasingly represented using machine-readable formalisms, and this creates important opportunities for developing flexible natural language interfaces for capturing clinical data (e.g. taking clinical histories) and giving advice (e.g. recommending clinical actions). Natural language interaction could also offer patients more control of their own care, by allowing them to obtain advice whenever and as often as they need it. Voice interaction seems a particularly suitable approach as speech is a natural way for people to communicate and does not require the patient to use any technology other than a telephone.

Spoken dialogue systems have great potential for delivery of healthcare services. However medical applications that provide advice to patients or to practitioners pose particular problems relative to standard information seeking dialogues such as flight booking or banking. In most current commercial dialogue systems, the dialogue designer specifies the exact interactions which can take place. Manual-coding allows precise control of what can occur within a dialogue, but is an expensive process, especially for complex dialogues. In the medical domain, the knowledge structures are particularly complex, and the system requires complex reasoning and decision-making to respond to the user. It therefore seems necessary to integrate technologies such as medical guidelines and advice systems directly with the dialogue system so that dialogues can be generated automatically to reflect user behaviour and changes in clinical context.

2 Solution

Cancer Research UK (CR-UK) has developed a dialogue system (Beveridge and Milward, 2003a; Beveridge and Milward, 2003b; Milward and Beveridge, 2003) as part of the EU HOMEY project (Home Monitoring through an Intelligent Dialogue System, IST-2001-32434). The first application of the system is intended for use by medical General Practitioners to determine whether a patient with suspected breast cancer should be referred to a cancer specialist (Bury et al., 2001). The dialogue system therefore needs to be closely integrated with medical domain knowledge, in this case in the form of an ontology for the breast cancer domain provided by Language & Computing n.v. (Ceusters et al., 2002), and knowledge of clinical tasks and processes, in this case the PRO*forma* process specification language (Fox et al., 2003).

In order to allow integration with these domain representations, our dialogue model is divided into high- and low-level representations. The low-level representation defines a finite-state network of communication acts represented by a VoiceXML specification. The high-level representation captures information regarding the intentional and informational structures underlying the dialogue, along with its current attentional state (Grosz and Sidner, 1986; Mann and Thompson, 1988). The information in the high-level representation is in turn derived from the underlying domain specification, with the *intentional* structure deriving from decisions, plans and other tasks in the medical care pathway and the *informational* structure deriving from the medical ontology. In order to make use of these representations in a practical system we have employed a multi-level architecture, similar to 3-layer hybrid agent architectures (Gat, 1998) as shown in Figure 1.

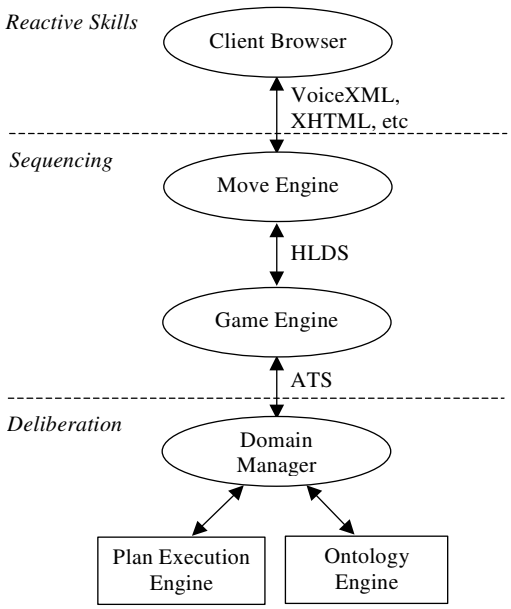


Figure 1. Dialogue System Architecture

The deliberative layer is provided by a domain manager that creates an abstract task specification (ATS) based on the outputs of the plan execution and ontology engines. The sequencing layer includes a Game Engine which determines the conversational games (Lewin, 2000) to be played in order to complete the tasks in the ATS. The Game Engine uses ontological knowledge to reorder games. For example, a game concerning a refinement of a concept will, if possible, be ordered to appear immediately after the concept has been introduced to ensure maximal dialogue coherence. The resulting High-Level Dialogue Specification (HLDS) is used by a Move Engine to generate the sequence of low-level communicative acts (moves) that can be made by either participant at the current point in the dialogue. The reactive layer interprets the low-level specification (VoiceXML) in order to complete the specified dialogue segment, and handles low-level reactive behavior required to support communication, e.g. repeating prompts, changing the speech volume etc. If an event occurs which the reactive layer cannot handle (e.g. the user taking the initiative) then it is passed back to the sequencing and deliberative layers to be processed.

The complete spoken dialogue system appears to be flexible and robust, resolving lexical and terminological ambiguities in real time using ontological knowledge, accommodating mixed initiative, and varied groupings of clinical data by the game engine (an animated demonstration of the system is available.) Formal evaluations of the technology are to be carried out in the domain of cancer referral decisions (described earlier) and in the collection of family history data for genetic risk assessment and management. We are also investigating the potential for the technology to scale up to a larger medical domain encompassing the diagnosis, treatment and long-term follow-up of patients with breast cancer.

References

Beveridge, M., and Milward, D. 2003a. *Definition of the High-Level Task Specification Language*. Deliverable D11, EU HOMEY Project, IST-2001-32434, http://www.acl.icnet.uk/lab/homey.

Beveridge, M., and Milward, D. 2003b. Combining Task Descriptions and Ontological Knowledge for Adaptive Dialogue. To appear in *Proc. Text, Speech and Dialogue (TSD)*, Ceske Budejovice, Czech Republic.

Bury, J., Humber, M., and Fox, J. 2001. Integrating Decision Support with Electronic Referrals. In R. Rogers, R. Haux, and V. Patel (eds). *Medinfo*, IOS Press, Amsterdam.

Ceusters, W., Beveridge, M., Milward D., and Falavigna, D. 2002. *Specification for Semantic Dictionary Integration*, Deliverable D9, HOMEY Project, IST-2001-32434, http://www.acl.icnet.uk/lab/homey.

Fox, J., Beveridge, M., and Glasspool, D. 2003. Understanding Intelligent Agents: Analysis and Synthesis. *AI Communications* (in press).

Gat, E. 1998. On Three-Layer Architectures. In D. Kortenkamp, R. Bonasso and R. Murphy (eds) *AI and Mobile Robots*, AAAI Press, Menlo Park, CA.

Grosz, B., and Sidner, C. 1986. Attention, Intention and the Structure of Discourse. *Computational Linguistics,* 12(3):175-204.

Lewin, I. 2000. A Formal Model of Conversational Game Theory, In *Proc. Gotalog-00, 4th Workshop on the Semantics and Pragmatics of Dialogue*, Gothenburg, Sweden.

Mann, W., and Thompson, S. 1988. Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. *Text*, 8(3):243-281.

Milward, D., and Beveridge, M. 2003. Ontology-Based Dialogue Systems. To appear in *Proc. IJCAI 3rd Workshop on Knowledge and Reasoning in Practical Dialogue Systems,* Acapulco, Mexico.

Building Spoken Dialogue Systems for Believable Characters

Johan Bos, Tetsushi Oka
ICCS, School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland, United Kingdom
{jbos,okat}@inf.ed.ac.uk

1 Introduction

Believable characters are defined in the arts as autonomous characters with personality, capable of showing emotion, being self-motivated rather than event-driven, adaptable to new situations, maintaining consistency of expression, and able to interact with other characters. These requirements impose a number of conversational capabilities on a spoken dialogue interface with a believable character: the character should react on its name, it should give subtle signals when it is listening or wants to take the turn, and it should allow interruptions while it is talking.

Implementing these conversational aspects of believability presupposes a flexible architecture capable of dealing with asynchronous processes. Most of today's spoken dialogue systems exhibit a pipelined architecture and even software agent-based architectures with the potential to function asynchronously work in practice as a pipeline approach. Spoken interaction with a believable character is, in the simplest case, replacing the speech synthesis component with a module that captures both the animation of the character plus speech synthesis. It is obvious that processing on a purely sequential basis wouldn't meet the requirements for believability—this paper introduces a dialogue system based on an asynchronous agent architecture developed within the MagiCster project, aiming to display the aforementioned believable aspects.

2 Face-to-Face Spoken Dialogue Systems

Our spoken dialogue system prototype is built around the embodied believable conversational agent Greta (see Figure 1) developed in MagiCster (www.ltg.ed.ac.uk/magicster/). This animation character includes detailed emotion modelling, within the application domain of giving advice to teenagers about eating disorders (Pelachaud et al., 2002).

In the framework of this system, we focussed on implementing several subtle but effective non-linguistic aspects of conversation, which we believe contribute to an improved sense of believability: allowing interruptions when Greta is speaking; system back-channelling (generating uhm's, nodding) and showing attention when the user is speaking; signalling awareness of the presence/absence of the conversational partner.

3 OAA for Spoken Dialogue Systems

Realisation of these capabilities in a spoken dialogue system requires a flexible asynchronous architecture. The Open Agent Architecture (OAA, www.ai.sri.com/~oaa/) is a framework for integrating software agents, possibly coded in different programming languages and running on different platforms. The OAA framework forms a piece of middleware allowing smooth integration of software agents for asynchronous dialogue systems in a prototyping environment.

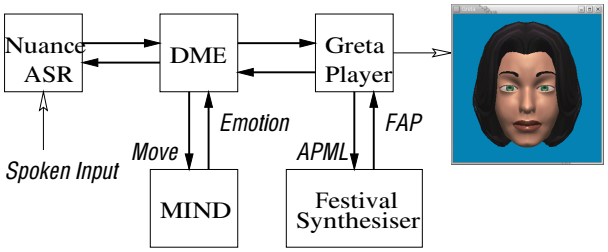


Figure 1: System Architecture

Figure 1 shows the components implemented as OAA agents. We adopted the grammar-based approach to language modelling that Nuance speech tools support (www.nuance.com). Using the slot-filling option of GSL, the value returned by the speech recogniser is the symbolic move of the utterance. The recognised string, the move, and an acoustic confidence value are sent as input to the dialogue move engine. Different speech grammars are loaded during different stages in the dialogue to increase performance of speech recognition.

The dialogue move engine (DME) controls all input and output relevant to the dialogue, interprets the user's move, and plans the system's next move. It can activate the ASR with a particular grammar, and receives information from the ASR (recognised moves, start of speech). It generates the system's move in the form of APML (Affective Presentation Markup Language), and sends this to the Greta Player. It is also able to tell the player to stop and it receives information when it is finished. The DME implements the information-state based approach to dialogue modelling.

The MIND component models emotions using dynamic belief networks, following the BDI (believes, desires, attentions) model (Pelachaud et al., 2002). Given

a symbolic move sent by the DME, MIND decides which affective state should be activated and whether certain emotions should be displayed and how.

Greta is the talking head that plays FAP files based on timing information received from the Festival synthesiser. Greta combines verbal and nonverbal signals in an appropriate way when delivering information, achieving a very rich level of expressiveness during conversation (Pasquariello and Pelachaud, 2001). Using the APML markup language, it is able to express emotions, and synchronise lip and facial movements (eyebrows, gaze) with speech.

The Festival synthesiser (`www.shlrc.mq.edu.au/festdoc/festival/`) takes as input an APML file and produces both a waveform and timing information structured according to syllables and words. This information is needed by the Greta player to determine gestures and other movements, including facial movements such us eyebrow lifting and movement of lips and other articulators (Pasquariello and Pelachaud, 2001).

4 Information States in DIPPER

The information-state approach to dialogue modelling (Larsson and Traum, 2000) combines the strengths of dialogue-state and plan-based approaches, using aspects of dialogue state as well as the potential to include detailed semantic representations and notions of obligation, commitment, beliefs, and plans. It allows a declarative representation of dialogue modelling and is characterised by the following components: (1) a specification of the contents of the information state of the dialogue; (2) the datatypes used to structure the information state; (3) a set of update rules covering the dynamic changes of the information state; and (4) a control strategy for information state updates.

Our implementation of the information-state approach, DIPPER, follows the TrindiKit (Larsson and Traum, 2000) closely, by taking its record structure and datatypes to define information states. However, in contrast to the TrindiKit, in DIPPER all control is conveyed by the update rules, there are no additional update and control algorithms, and there is no separation between update and selection rules. Furthermore, the update rules in DIPPER do not rely on Prolog unification and backtracking, and allow developers to use OAA-solvables in the effects (Bos et al., 2003).

Update rules specify the information state change potential in a declarative way: applying an update rule to an information state results in a new state. An update rule is a triple $\langle$*name*, *conditions*, *effects*$\rangle$, The conditions and effects are defined by an update language, and both are recursively defined over terms. The terms of the update language allows one to refer to a specific value within the information state, either for testing a condition, or for applying an effect. In DIPPER this is done in a functional rather than a relational way as implemented by the TrindiKit, effectively abstracting away from Prolog syntax and discarding the use of Prolog variables.

We took the update rules of the GODIS system (Larsson and Traum, 2000) as a basis for our dialogue system for Greta. To implement the aspects of conversational believability in our system, several update rules were added, working mostly on the attentional state such as user-awareness and back channelling.

5 Conclusion

Implementing complex spoken dialogue systems—systems that go beyond the rather straightforward pipelined architectures—poses several requirements on the flexibility on dialogue modelling and the underlying architectures. We presented a system involving spoken interaction with an embodied character showing several believable characteristics of conversational behaviour, using the DIPPER framework for building spoken dialogue systems (`www.ltg.ed.ac.uk/dipper/`). We argued that OAA combined with an information-state theory of dialogue modelling is a good way of managing asynchronous processes that are imposed by these phenomena. We further claimed that an optimisation of the TrindiKit, where all dialogue controlled is specified by update rules, in combination with a language for update rules that abstracts away from Prolog and is tighter integrated with OAA, improves the facilities in the developer's workbench for complex spoken dialogue systems.

References

Johan Bos, Ewan Klein, Oliver Lemon, and Tetsushi Oka. 2003. Dipper: Description and formalisation of an information-state update dialogue system architecture. In *4th SIGdial Workshop on Discourse and Dialogue*. Sapporo, Japan.

Staffan Larsson and David Traum. 2000. Information state and dialogue management in the trindi dialogue move engine toolkit. *Natural Language Engineering*, 5(3–4):323–340.

S Pasquariello and C. Pelachaud. 2001. Greta: A simple facial animation engine. In *6th Online World Conference on Soft Computing in Industrial Appications*.

C. Pelachaud, V. Carofiglio, B. De Carolis, F. de Rosis, and I. Poggi. 2002. Embodied contextual agents for information delivery applications. In *Proceedings of AAMAS'02, Bologna*.

Subdialogs and Attentional State

Donna K. Byron*
Dept of Computer and Info. Science
The Ohio State University
byron.18@osu.edu

Sarah Brown-Schmidt
Dept of Brain and Cognitive Sciences
University of Rochester
sschmidt@bcs.rochester.edu

1 Introduction

Assigning meanings to Noun Phrases (NPs) is crucial to the language understanding process. Previous research demonstrated that many sources of information are used in this process, e.g. task pragmatics (Brown-Schmidt et al., 2002), expectations introduced by disfluencies (Arnold et al., In press), and the task-relevant properties of objects (Chambers et al., 2002). (Grosz and Sidner, 1986) (henceforth G&S) showed that the relative embedding of discourse segments also affects the meaning of NPs. This report presents the initial findings of a psycholinguistic experiment to examine whether subjects are able to utilize this high-level property, discourse structure, in the earliest moments of NP interpretation.

The model reported in G&S organizes the entities in a discourse as a set of *focus spaces*, one per discourse segment, that are managed as a stack. This stack, called the *attentional state*, are searched to find referents for anaphoric NPs. As a discourse segment unfolds, its focus space accumulates discourse referents. When the discourse segment ends, its entities are removed from the stack. In the G&S model, 'only the attentional state can constrain the interpretation of referring expressions directly'[G&S, pg. 180].

G&S is not intended as a cognitive model; however, the theory leads to an intriguing question for human processing. After the close of a subdialog that interrupts a discourse segment, are entities from that subdialog considered, if only briefly, as the referents for subsequent NPs or are they 'popped off the stack' and not considered?

*This material is based upon work supported by the National Institutes of Health under award number NIH HD-27206 to M.K. Tanenhaus.

2 Experimental Design

To examine this question, we conducted a preliminary experiment using the visual-world paradigm, in which we obtained a record of the participants' eye-fixations with the use of a head-mounted eye-tracker. Approximately 200ms after the onset of an NP, a participant will look at items he is considering as the referent (Tanenhaus et al., l995). The results we describe here must be considered to be preliminary since the experiment included only one trial. Data was collected from 36 undergraduate participants[1] from the University of Rochester.

Participants sat at a table with supplies for dying eggs, including a carton containing either 1 egg & 1 non-egg competitor object or 2 eggs. The participants then heard prerecorded instructions to dye an egg, shown in Table 1. The instructions for each condition include an initial instruction (segment A), a subdialog (segment B) directing subjects to prepare a bowl of dye and arrange objects on the table, and a final segment after the subdialog closes that returns to the main task of dying the egg (segment C).

Subjects were randomly assigned into one of four conditions. Conditions varied along the following two dimensions: 1) the competitor object is an egg or a non-egg (a small sheet of stickers used to decorate eggs), 2) the competitor object is or is not mentioned in the subdialog. The non-egg competitor condition serves as a baseline for comparing the proportion of looks to the competitor egg (conditions C1 and C2) with the number of looks to the stickers (conditions C3 and C4).

The critical point in the experiment is the

[1] 10 in conditions C1 and C3; 9 in conditions C2 and C4. Audio for 1 person in each of C2 and C4 was lost.

Table 1
Instruction Sequences

C1) Competitor egg is mentioned in subdialog	**C3) Competitor stickers are in subdialog**
A1) In this experiment you're going to be dying an easter egg.	A1) In this experiment you're going to be dying an easter egg.
A2) Pick which egg you want to paint.	A2) *Put the egg in the egg stand so that you can paint it.*
A3) *And put it in the egg stand so you can paint it.*	
B4) Now before we can proceed we need to prepare the dye.	B3) Now before we can proceed we need to prepare the dye.
B5) Fill up the empty bowl with water.	B4) Fill up the empty bowl with water.
B6) Pick which color of dye you want to use.	B5) Pick which color of dye you want to use.
B7) Add a couple of drops of dye to the water.	B6) Add a couple of drops of dye to the water.
B8) *Move the styrofoam carton with* **the egg** *in it over by the towel and out of the way.*	*B7) Move the styrofoam carton with the stickers in it over by the towel and out of the way.*
C9) Great! Now that's all done so we're ready to dye.	C8) Great! Now that's all done so we're ready to dye.
C10) Use the dipper to pick up **the egg** and dip it in the dye.	C9) Use the dipper to pick up **the egg** and dip it in the dye.
C2) Competitor egg is not mentioned in subdialog	**C4) Competitor stickers are not in subdialog**
A1) In this experiment you're going to be dying an easter egg.	A1) In this experiment you're going to be dying an easter egg.
B2) First we need to prepare the dye.	B2) First we need to prepare the dye.
B3) Fill up the empty bowl with water.	B3) Fill up the empty bowl with water.
B4) Pick which color of dye you want to use.	B4) Pick which color of dye you want to use.
B5) Add a couple of drops of dye to the water.	B5) Add a couple of drops of dye to the water.
B6) *Pick which egg you want to paint and put it in the egg stand so you can paint it.*	B6) *Put the egg in the egg stand so that you can paint it.*
C7) Great! Now that's all done so we're ready to dye.	C7) Great! Now that's all done so we're ready to dye.
C8) Use the dipper to pick up **the egg** and dip it in the dye.	C8) Use the dipper to pick up **the egg** and dip it in the dye.

mention of 'the egg' in the last instruction. At the onset of this instruction, both the task and the discourse structure indicate a return of attention to the target egg. In C1 and C2, there are two eggs that could be the referent of the critical NP. In C1, the competitor egg (e.g. the egg the participant is not dyeing) was mentioned most recently in the subdialog, but in C2, it was never mentioned. If entities in the subdialog have been 'popped off the stack' in a manner analogous to the G&S model, subjects in C1 should be no more likely to look at the competitor egg than the stickers in C3. If participants consider the competitor egg as a possible referent, we expect to see a larger proportion of fixations to the competitor egg than to the stickers.

3 Results

After hearing the critical instruction, each participant quickly selected the target egg and completed the task. This is consistent with the predictions of G&S, that closing the subdialog eliminated the competitor egg as a plausible referent of the critical NP.

Eye-fixations suggest that closing the subdialog largely (but not entirely) eliminates referents introduced in the subdialog as possible referents of upcoming NPs. After the critical NP "the egg", when the competitor object was the stickers (C3 and C4), participants looked primarily at the target egg and the instrument they are about to use (the dipper), with minimal looks to the stickers. Likewise when the competitor object was an egg, but it was not mentioned (C2), eye fixations followed this same pattern.

Crucially, in C1 we did observe a few looks to the competitor egg. The lack of fixations to the competitor egg in C2 suggests that these fixations were a result of the competitor egg being mentioned in the subdialog. This suggests that entities mentioned in subdialogs may be maintained as potential referents for upcoming NPs.

References

J.E. Arnold, M. Fagnano, and M.K. Tanenhaus. (In press). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research.*

S. Brown-Schmidt, E. Campana, and M. Tanenhaus. 2002. Reference resolution in the wild. In *24th Annual Meeting of the Cognitive Science Soc.*, Fairfax, VA, August.

C. Chambers, M. Tanenhaus, K. Eberhard, G. Carlson, and H. Filip. 2002. Circumscribing referential domains in real time language comprehension. *J. Memory and Language*, 47:30–49.

B. Grosz and C. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

M.K. Tanenhaus, M.J. Spivey-Knowlton, K.M. Eberhard, and J.E. Sedivy. l995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632–1634.

A Tool to Simulate Affective Dialogs with an ECA

A.Cavalluzzi, V.Carofiglio, G.Cellamare and G.Grassano
Department of Informatics
University of Bari
{cavalluzzi,carofiglio,grassano}@di.uniba.it

Abstract

We describe a testbed for simulating affective dialogs with an Embodied Conversational Agent (ECA). We show how the dialog is influenced by the social context and the agent's emotional state and how this state is, in its turn, dynamically influenced by the dialog.

1 Introduction

Psychologists agree in claiming that cognition influences emotions and vice versa. According to some authors, activation of emotions in artificial agents is due to variation in their beliefs and high-level goals. On the other hand, affective states produce changes in active beliefs, in goal activation and priority and in the reasoning skills; they consequently influence learning, decision making and memory (Castelfranchi (2000), Forgas (2000), Picard (1997)). Simulating dialogues in domains in which affect plays a relevant role requires modeling these dynamic phenomena and building agents that are able to show a reactive behavior during the dialogue, as far the situation evolves. To this aim, models of emotion activation have to be built (Ortony et al. (1998)) and connected to the reactive component of the dialog planner. Emotions must drive reasoning behind the dialog and regulate it. If the dialogue occurs between the user and an Embodied Conversational Agent (ECA), the influence of emotional factors must also be visible in its 'body': this requires implementing the agent's ability to express emotions through face, gesture and speech (Cassell et al, 2000). Simulating affective dialogs therefore requires investigating which emotions the agent may feel during the dialog, as a consequence of exogenous factors (the user moves) or endogenous factors (the agent's own reasoning). It also requires studying how every emotion influences the dialog course: in particular, priority of communicative goals and dialog plans (de Rosis et al, 2003).

A testbed may enable designers to evaluate the role of the main variables involved in this simulation and to assess how they influence the dialog course. Examples of these variables are the social context in which the dialog occurs, the role played by the ECA, its personality, its relationship with the user and the environment in which interaction takes place. In this demo, we will show the testbed that we developed in the scope of Magicster. The system is driven by a Graphical Interface, which interacts with the user and coordinates activation of various modules and exchange of information among them:

- *Mind* initially receives information about the setting conditions and selects "personality", "context" and "domain" files accordingly. It subsequently receives an interpreted user move and sends back a list of "emotion intensities" that this move activates in the agent;
- *Dialog Manager* receives initial information about the dialog conditions. At every user turn, it receives an interpreted user move with a description of changes produced in the agent's affective state. It sends back an agent move, which is displayed in natural language. This move is annotated with an 'Affective MarkUp Language" (De Carolis et al, in press) and is stored as an XML file;
- *Body* reads this file and generates the ECA, which is displayed in the Interface. Due to the mind-body independence of our tool, several Embodied Agents may be employed to express the agent move. So far, we integrated a 3D-realistic character in a DLL of the Interface (face animation by (Pelachaud et al, 1996) and speech by (Festival, website)) and we developed a wrapper for MS-Agents (website).

Designers may employ this tool to simulate dialogs in various conditions. At the beginning of interaction, they set the simulation conditions: agent's personality, its relationship with the user, its 'body' and the application domain. They then input the user moves in natural language. The interface enables following the dialog in natural language and with the selected Embodied Agent by showing (in graphical form) how the agent's emotional situation evolves during the dialog.

The dialog is goal-driven: every goal, with a given priority, is linked by an application condition to a plan that the agent can perform to achieve it. Some goals are initially 'inactive': they may be activated if an emotional situation occurs or if the user needs to be persuaded to follow some suggestion In the first case, goals are activated with a priority which depends on the emotion felt. For example, if something undesirable occurs to the user and the agent is in an 'empathic' relation with her, it will feel *sorry-for* and will activate the goals enabling to show this emotion in verbal and nonverbal forms. These goals will take the priority over the current goals. If, on the contrary, the user rejects some suggestion, the high-priority goal becomes to persuade her to accept it: this requires activating the ability to provide burdens of proof and dialectical arguments (Carofiglio et al (in press)).

We employed our testbed to adjust the system components after evaluating their behavior in various situations. We tested the role of context and personality in the activation of multiple emotions, upgraded the dialog strategy and the plan library, revised interpretation of the user moves and improved rendering of the agent moves. The Interface was implemented with the Visual C++, while the sockets insuring the communication among the different processes are built-in classes of the Interface code. The dialog manager is implemented with TRINDIKIT (website); emotion activation and argumentation strategies are modeled with belief networks and are implemented with HUGIN APIs (website).

After refining the individual modules with the aid of our testbed, we plan to implement the final prototype within a mobile technology framework. The agent will be displayed on a high-resolution screen; the server side will process the agent's mind and body and the dialog manager by communicating, via socket, with a mobile device (a PDA). The main tasks of the PDA will be to handle the user input and to update a user model which will include affective aspects. We will assess advantages and disadvantages of developing a speech or graphical interaction through the PDA.

Acknowledgements: This research was financed from the EU Project MagiCster (IST 1999-29078). We thank B. De Carolis, F. de Rosis and S. Pizzutilo for guiding us in the design of this system and C. Pelachaud and M. Steedman for enabling us to employ a prototype of their ECA, in the scope of MagiCster.

References

Castelfranchi C (2000). "Affective Appraisal Versus Cognitive Evaluation in Social Emotions and Interactions". In A. PAIVA, (Ed). *Affective Interactions.* Springer LNAI 1814, Berlin: 76-106.

Carofiglio V and de Rosis F. (in press). "Combining Logical with Emotional Reasoning in Natural Argumentation". http://aos2.di.uniba.it:8080/papers/um03-affect.zip

Cassell J, Sullivan J, Prevost S, Churchill E (2000). *Embodies Conversational Agents*. The MIT Press.

De Carolis B, Pelachaud C, Poggi I, Steedman M (in press). "APML, a markup language for believable behavior generation". In Prendinger H (Ed.). Life-like Characters. Tools, Affective Functions and applications

de Rosis F, Pelachaud C, Poggi I, Carofiglio V, De Carolis B (2003). "From Greta's Mind to her Face: Modeling the Dynamics of Affective States in a Conversational Embodied Agent". *International Journal of Human-Computer Studies*. 49.

FESTIVAL: http://www.cstr.ed.ac.uk/projects/festival

Forgas J P (2000). Feeling and thinking. The role of affect in social cognition. Cambridge University Press.

HUGIN: http://www.hugin.com/

Ortony A, Clore G L, Collins A (1988). The cognitive structure of emotions. Cambridge University Press, Cambridge, MA.

Pelachaud, C, Badler, N I, Steedman, M (1996): "Generating Facial Expressions for Speech". *Cognitive Science*, 20, 1. 1-46.

Picard R W (1997). Affective Computing. The MIT Press.

TRINDIKIT Manual:
http://www.ling.gu.se/research/projects/trindi

Communicative Effectiveness in Multimodal and Multilingual Dialogues

Erica Costantini
ITC-irst
Trento, Italy
costante@itc.it

Susanne Burger
Carnegie Mellon University
Pittsburgh, PA, U.S.A
sburger@cs.cmu.edu

Fabio Pianesi
ITC-irst
Trento, Italy
pianesi@itc.it

1 Introduction

Multilingual communication enabled by a multimodal speech-to-speech translation system may differ from 'ordinary' monolingual conversation in the conversational structure and in the way gestures are integrated in speech. We describe the second of two user studies conducted within the NESPOLE![1] project investigating these issues. NESPOLE! exploited a client-server architecture to allow an English, French or German-speaking user, while browsing through the web pages of a service provider on the Internet, to connect to an Italian-speaking human agent. Speech-to-speech translation (STST) is provided so that both speakers can use their own native languages.

2 NESPOLE! User Study2: Method

The second NESPOLE! user study (Burger et al., 2003) was designed to deeper investigate certain results of the first study (Costantini et al., 2002). Multilingual dialogues (English/ Italian, using the STST system as translation) were compared with monolingual (Italian/Italian) dialogues, using the system with and without push-to-talk mode (PTT). We devised three experimental conditions:

- **STST condition**: multilingual, PTT mode;
- **PTT condition**: monolingual, PPT mode
- **Non-PTT condition**: monolingual, free talk

We expected the multilingual condition to be different from the monolingual conditions with respect to dialogue length, spoken input, dialogue structure and speech-gesture integration patterns. The PTT mode would also play a role, resulting in differences between the two monolingual conditions.

The scenario featured a customer connecting with a human agent to find information about winter holidays. She had to choose a destination and a tourist package in compliance with a given specification, while the agent had to provide the explicitly requested information. We recorded 7 dialogues for the STST condition and 16 monolingual dialogues, half in PTT condition and half in Non-PTT condition. The interface allowed speakers to see each other, to share images and to point at portions of the image by pen-based gestures. The recorded dialogues were transcribed according to the VERBMOBIL conventions[2], and included annotations for gestures. Special annotations were added following an extended version of the Dialogue Structure Coding Scheme (DSCS) from the HCRC research group[3].

DSCS was developed for the Map Task Corpus (Carletta et al. 1997). It classifies single utterances according to their discourse goals and captures the higher-level structure in terms of *games*. Conversational *games* are associated with mutually understood conversational goals, e.g. obtaining information. *Games* consist of conversational *moves* which are different kinds of initiations and responses classified according to their purposes.

Table 1 displays the modified annotation schema; a star marks the newly added moves. The *proposal, disposition, action* and *information* moves are subclasses of the DSCD's *information* move.

[1] Project web-site at http://nespole.itc.it

[2] http://www.is.cs.cmu.edu/trl_conventions/

[3] http://www.hcrc.ed.ac.uk/Site/

Initiation Moves	
Align	checks transfer successfulness
Check	checks confirmation
Query-yn *Query-w*	yes/no questions (yn) open questions (w)
Request	requests (former *instruct* move)
Proposal	proposal or offer
Disposition	needs or interests
Action	description of actions
Information	spontaneous information, not elicited
Response Moves	
Acknowledge	confirming
Reply-y *Reply-n* *Reply-w* *Reply-amp*	yes/no answers, answers to open questions (w), answers adding not requested information (-amp: former *clarify* move)
**Problem*	negative feedback (notification of non-successful communication)
**Other*	speaker misunderstood the question, talked about different things
Other Moves	
Preparation	expressing readiness to start
**Comment*	out of domain comments
**Noise*	turns without linguistic content

Table 1. Move Annotation Schema

3 Results

The results for all three conditions reported in Table 2 show that the dialogues in STST condition lasted longer, but had an even lower percentage of actual dialogue contributions. 87% of the time was taken by the STST system's delays, transfer, translation and PTT mode (the PTT condition still shows 30% of non-speech part compared with the Non-PTT condition). The STST condition is also characterized by more repetition turns. Analyzing the involved moves ascribes these repetitions to meta-communicative concepts supposed to resolve misunderstandings. The system failed to translate these. Furthermore, the STST dialogues show: shorter dialogue games, fewer nested games; more questions, more replies, less spontaneously provided, non-elicited information and fewer *acknowledgment* moves. In STST dialogues the speakers focused on 'essential' information, reduced the dialogue complexity and tried to adhere to a question/answer pattern. The number of gestures was similar in all conditions, but in the STST condition, gestures were performed before and more frequently after talking. This suggests that the speech-gesture integration can be lost as soon as the interaction becomes more complex, when more tasks such as PTT, translation and drawing must be handled in parallel. The results for the PPT condition were usually intermediate between those of the Non-PTT condition and those of the STST condition, proving that PTT has an additional effect on STST condition.

Measures	**STST**	**PTT**	***NonPTT***
Dialogue length (min)	23	9.85	8.87
% non-speech partition	87%	49%	19%
% repetition turns	24%	6%	1.3%
Moves per game	4.6	4.6	5.6
% of nested games	10%	26%	23%
% of questions	35%	23%	14%
% of replies	24%	21%	16%
% of *information*	8%	12%	15%
% of *acknowledge*	11%	17%	33%
Gestures during speech	14%	61%	96%

Table 2. Results for all three conditions

4 Conclusions

The results show the existence of adaptive communication strategies to the different contexts of communication. Using Dialogue Structure Analysis seems to be a sufficient method of discovering, understanding and clarifying the phenomena. The revealed communicational structures should be of great interest to the STST research community, both, for evaluation of dialogue effectiveness, but also for the design of appropriate scenarios and choice of training materials covering the linguistic phenomena which are expected to be found during the interaction with an actual translation system.

References

Erica Costantini, Fabio Pianesi & Susanne Burger. 2002. The Added Value of Multimodality in the NESPOLE! Speech-to-Speech Translation System: an Experimental Study. In *Proceedings of ICMI'02*, Pittsburgh, PA.

Jean Carletta et al. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 23 (1) 13-21.

Susanne Burger, Erica Costantini & Fabio Pianesi. 2003. Communicative Strategies and Patterns of Multimodal Integration in a Speech to Speech Translation System. To appear in *Proceedings of MT-summit*, New Orleans, LA.

DiaMant: A Tool for Rapidly Developing Spoken Dialogue Systems

Gerhard Fliedner, Daniel Bobbert
CLT Sprachtechnologie GmbH
Stuhlsatzenhausweg 69
D-66123 Saarbrücken
{fliedner,bobbert}@clt-st.de

1 Introduction

Following the recent advances in speech recognition, now allowing the reliable speaker-independent recognition of one to two thousand words, an increasing number of natural language dialogue systems has been developed. These aim at two main areas: information access dialogues (e.g. travel information and bank account management) and machine control (e.g. controlling car functions, VCRs, and robots). There is a growing demand for tools that allow the rapid development of such dialogue systems. The users increasingly expect systems that allow a free dialogue without the need to learn a special command vocabulary and a menu-oriented hierarchical dialogue structure.

At CLT Sprachtechnologie, we have developed DiaMant (Dialogue Management Tool), a tool that allows the rapid development of dialogue systems based on extended finite state dialogue models. The development is done inside a graphical user interface that intuitively shows the dialogue model as an automaton in a graph representation. The dialog flow can be edited without the need to write a single line of code. The tool has already been successfully used in a number of applications.

As an important feature, the tool allows setting up Wizard of Oz experiments with one mouse click, where a human 'wizard' can replace input devices such as a speech recognizer in early development phases. This has proven especially important in testing and improving dialogue systems, leading to a higher acceptance of the final systems.

2 Dialogue Modeling Using FSA

Finite state automata (FSA) are an intuitive way of modelling dialogues (cf. e.g MacTear 99). An FSA can be represented as a graph with nodes corresponding to FSA states and edges to transitions.

In this graph representation of an FSA-based dialogue model, then, an automaton state implicitly represents a dialogue state (e.g. a state were a certain information has just been provided by the user). The user's dialogue moves are represented by the edges of the graph. Executing a dialogue in this model can then be thought of as traversal of the graph, selecting transition edges according to the user's input in the corresponding state.

This basic model, however, is only suited for modelling very simple dialogues. For more complex dialogues, it is useful to add a possibility of explicitly storing information in slots (variables). Such an enhanced FSA with slots forms the basis of our dialogue development tool.

3 DiaMant: Overview

DiaMant has been implemented in Java. It is therefore platform independent, running on all operating systems for which a Java VM is available (including Windows, Linux, Solaris, and MacOS). Hardware requirements for the system itself are modest (Pentium II @ 200 MHz, 64 MB or comparable).

Client modules (among them speech recogniser and TTS systems but also databases or other devices) can be integrated using a simple-to-use interface. Since the interface is TCP/IP based, clients can even be running on other machines. For a number of commercial speech recognisers and TTS systems, the required modules already exist. Others can easily be added via a simple Java wrapper provided by CLT that allows users to develop new modules simply by extending a Java class without the need to deal with network and protocol details.

Developing a dialogue model is done with a graphical interface. The dialogue model is repre-

sented as a graph. A number of different types of nodes is available: input/output and slot manipulation nodes, sub-automata calls, and others.
Each of the different node types allows a suitable parametrization. An output node, e.g., allows the user to specify an output device (e.g. a TTS system or a display window) and a text string or data record to send to it. Input nodes allow giving any number of different possible input values to be matched against the input of a specified device. The input to be matched is given as a regular expression, allowing to match natural language input, especially from a speech recogniser, with a high degree of flexibility. For each possible input value, an outgoing edge is added to the input node that can be connected to other nodes by drag and drop.
Input can be stored in typed slots that can hold, e.g., numbers or strings. These slots can be accessed anywhere in the dialog, in order to branch depending on a slot value or to produce context sensitive output. Slots and constant values can be combined and evaluated in Boolean and arithmetic expressions.
Recurring dialogue tasks can be stored in sub-automata. These sub-automata have their own, local dialogue slots. When calling a sub automaton, parameters can be passed and resulting values can be returned. Thus, sub-automata can be readily compared with sub-routines and functions in programming languages, allowing to modularise the dialogue and keep it compact.
This intuitive way of dialogue modelling within an easy-to-use graphical user interface allows designing simple dialogs within a matter of hours.

4 Wizard of Oz Experiments

An important point in dialogue development is providing a possibility for user experiments (Muntteanu and Boldea 2000). These experiments should, in general, be employed as early as possible in the dialogue development cycle: Very often, small matters decide on the user acceptance of a system, sometimes as little as the wording of a system prompt. We have found it useful in dialogue development to be able to directly use the current dialogue model in a user experiment.
DiaMant provides a convenient way of setting up Wizard of Oz-type experiments. Here, one or more input devices are simulated by a 'wizard', one or more persons, possibly identical with the experimentator. This allows testing a system early, when the full functionality is not necessarily present yet.
When run in the 'Wizard mode', the system uses a window on the wizard's machine to show information about the dialogue state to the wizard. Whenever an input is expected, the different possibilities provided for in the dialogue model are presented to the wizard to choose from (by mouse or keyboard).
User experiments are extensively logged (including exact time stamps), allowing the easy identification of problematic areas. Using the log file, an experiment can be replayed, helping with a detailed analysis.

5 Example Systems

Several example systems have been built using DiaMant. One of them is a speaking elevator that allows the user entering an elevator carriage to speak their destination floor aloud (including just naming a person, whose office is on that floor).
In a recent seminar (winter term 2002/2003) at the Saarland University's Computational Linguistics dept., DiaMant has been successfully used by student groups to implement dialogue models for controlling robots built with LEGO Mindstorms kits. A similar course is now taking place at the University of Kassel (summer term 2003).
DiaMant was also used successfully for user experiments while developing a complex information seeking dialogue for a flight information system.

6 Further Development

DiaMant has been successfully employed in a number of tasks, commercially as well as for educational purposes. We currently plan to make the tool available for interested researchers/developers.
Important goals are the adaptation of the tool to process standard data formats, especially import and export Voice XML, and a tighter integration of grammar development for the speech recogniser.

7 References

Munteanu, C. and Boldea, M. 2000. "MDWOZ: A Wizard of Oz Environment for Dialog Systems Development." *Proceedings of LREC*.

McTear, Michael F. 1999. "Software to Support Research and Development of Spoken Dialogue Systems." *Proceedings of the Eurospeech.*

A Framework for Information-State based Dialogue

Gerhard Fliedner, Daniel Bobbert
CLT Sprachtechnologie GmbH
Stuhlsatzenhausweg 69
D-66123 Saarbrücken
{fliedner,bobbert}@clt-st.de

1 Introduction

For the development of flexible and user-adaptive natural language dialogue systems, one needs a powerful dialogue engine. This engine must allow representing the dialogue in a dialogue model and executing this model. One framework for dialogue description that has evolved over the past years is based on the notion of information states and their update via rules.

At CLT Sprachtechnologie, we have developed a tool based on information state update (ISU). The tool allows the user to model dialogues using an information state, update rules and static plans. The execution of these models is based on an interaction of input processing and plan execution. The system provides the user with a dialogue modelling framework that is well-integrated into an execution system, obviating the need of worrying over basic implementation details, yet flexible enough to allow for different sorts of dialogues. Input and output devices (speech recognizer, TTS, screens, buttons, etc.) as well as databases can be integrated over an easy-to-use client interface.

Our system has been successfully used in developing, among others, an airport flight information system, giving e.g. information on departure times, gates, and flight status, with a high degree of flexibility in user input. However, for accommodating novice users the system automatically switches into a more system-guided mode asking the user for their flight information one step at a time. In usability tests, this dialogue system performed very well, with usability scores around 70%. This shows that the underlying system is well-suited for flexible dialogue systems. As the demand increases with the recent advances in dialogue systems, we expect a growing interest in such tools.

2 Information State Update

The idea of information state update for dialogue modelling is centred around the information state (IS). Within the IS, the current state of the dialogue is explicitly represented. Dialogue moves involve an update of the information state. Thus, user inputs are matched against a set of possible update rules that change the IS in the appropriate places (e.g. a new value is entered into a slot).

ISU or related frameworks have been used in a number of academic work recently. However, this work has often relied on implementing tailor-made solutions. General solutions (most prominent among them TrindiKit, Traum et al. 99) have provided a flexible framework, but left much of the actual implementation details to the user.

When developing our system, we have strived for a solution that is, on the one hand, flexible enough to accommodate a wide range of different dialogue types and I/O modalities, on the other hand, has a full-fledged execution system allowing a rapid development of dialogue systems.

3 System Overview

Our system has been implemented in Java. It is thus platform independent and runs on any computer system for which a Java VM is available (among others, Windows, Linux, Solaris, and MacOS). Hardware requirements for the system itself are modest (Pentium II processor @ 200 MHz, 64 MB main memory or comparable), plus hardware requirements for devices (e.g. speech recognisers).

In our system, the information state is realised as a typed, attributed structure. This structure can be freely defined by the user for the task in hand. The systems offers a full type system, including basic

types such as integers, strings, and Booleans, and complex types such as lists and records of these basic types. Attributes allow, e.g., the easy representation of meta information (such as grounding).
User input can be pre-processed by a built in context-free grammar parser with semantic tags to allow a semantic interpretation of the user utterance. Input is then passed to a set of rules, which try to match the input with patterns for the current information state, the input device and the actual input. Patterns can be simple value tests but also structural patterns or regular expressions, allowing the input or part of it to be bound to variables for further processing. The first matching rule's body is then executed. A rule can update the information state, generate system output, and push new plans that try to fulfil the systems goals. These plans are defined statically within the dialog model.
Plans are written in a procedural programming language, closely resembling JavaScript. The power of this language combined with the parameterisation of plans enables the generation of context sensitive, user adaptive dialog moves. An important additional feature of our plans is that they can carry information about their own 'fulfilment'. If a plan for eliciting some information from the user (e.g. a time) has been successful (i.e., the user has given the relevant time in answer) is called a second time, it would find that the information is already present and therefore not ask the user again to provide it. This has proven especially important in cases when there is more than one way of eliciting a certain information (e.g. if the user provides a flight number, asking for the departure time becomes unnecessary).
Closely connected with this is the notion of plan failure: The system maintains a stack of currently executed plans. Whenever a plan signals its failure, a mechanism resembling exception handling in many programming languages is used to identify a higher plan that is 'willing' to take over control.
Every input is usually handled in its own thread allowing for example barge-in (if the speech recogniser supports this). On the other hand, plans can specify that they need to wait for user input in order to be able to proceed. This turn-taking between system and user (corresponding to rules and plans) allows for a very flexible dialogue management.
The execution system gives a graphical representation of the current information state and client communication, allowing the in-place modification of values and attributes. Simulation of specific dialog situations for testing purposes therefore becomes very easy.

4 One Example System

As an example system, we have used our dialogue engine to implement an information seeking dialogue system in German based on a database with typical flight information of an airport. The user can enquire about a number of different flight details (ETA/ETD, delays), but also airport information (departure gate, parking facilities).
The system lays great stress on flexibility (i.e., the user can give information in any sequence, give more than one piece of information in one turn, use a wide variety of different words and sentence constructions). This, of course, necessitates various different manners of grounding (implicit to explicit, including clarification). An especially important point for novice users is that the system can switch to a more system-initiated mode where it asks the user for pieces of information. Even then, however, the user is always free to give information that the system had not requested or to correct a piece of information given earlier.
We have found, that our dialogue management tool has been very helpful in implementing this free example dialogue system. This is especially due to the fact that the rules allow the matching of the users' inputs even in places where the system did not actually expect them, combined with the plans carrying information about their own fulfilment

5 Further Development

After the recent, very promising experiences from our example systems, we plan to offer our development tool to interested users in the community.
As an important addition, we plan to enhance our system by adding a graphical debugging interface that allows the close inspection of the running system in addition to the existing logging facilities. Also, we plan to replace our custom plan description language with ECMA-Script, making the software more standards compliant.

6 References

David Traum et al. 1999. *A model of dialogue moves and information state revision, Trindi Deliverable D2.1*, http://www.ling.gu.se/research/projects/trindi/

Developing a Typology of Dialogue Acts: Tagging Estonian Dialogue Corpus

Tiit Hennoste
Department of Estonian and Finno-Ugric Linguistics
University of Tartu
hennoste@ut.ee

Mare Koit
Institute of Computer Science
University of Tartu
koit@ut.ee

Andriela Rääbis
Department of Estonian and Finno-Ugric Linguistics
University of Tartu
andriela@ut.ee

Krista Strandson
Department of Estonian and Finno-Ugric Linguistics
University of Tartu
ks@ut.ee

Maret Valdisoo
Institute of Computer Science
University of Tartu
maret@ut.ee

Evely Vutt
Institute of Computer Science
University of Tartu
nurm@ut.ee

1 Introduction

Estonian dialogue corpus includes recordings of spoken conversations, among them 114 calls for information and/or to travel bureaus.[1] All the recordings were transliterated using conversational analysis (CA) transcription. We have worked out a typology of dialogue acts and are using it for tagging the corpus (Hennoste et al. 2003). All the dialogues were tagged by two people and then unified. Our tagged corpus (114 dialogues) includes 5815 dialogue act tags, among them 633 questions and 1081 answers, 308 first and 258 second parts of directive adjacency pairs (AP). The kappa value is between 0.59 (for some travel bureau dialogues) and 0.79 (calls for information).

2 Typology of Dialogue Acts

There are several well-known typologies (Sinclair, Coulthard 1975, Stenström 1994, Mengel et al. 2001). Nevertheless, we have decided to develop our own dialogue act system because the categories used by most of the typologies are too general in our opinion. The principles underlying our typology are the same as for other coding schemes (Edwards 1995). Our typology departs from the point of view of CA that focuses on the techniques used by people themselves when they are actually engaged in social interaction. This is an empirical, inductive analysis of conversation data (see e.g. Hutchby, Fooffitt 1998).

The departing point of the CA is that a dialogue participant always must react to previous turn regardless of his/her own plans and strategies. This is the reason why we do not start our typology with determination of forward- and backward-looking functions but distinguish AP relations from non-AP ones. The computer as a dialogue participant must follow the norms and recognize signals of violations of the norms by the partner.

Secondly, acts used in dialogue are typically divided into two groups: information acts and dialogue managing acts. The last acts must be divided into 1) fluent conversation managing acts and 2) acts for solving communication problems. The computer must be able to differentiate a problem solving act from an information act or fluent interaction. It is essential because some information acts and repair acts have similar form (e.g. almost all initiations of repairs are questions in Estonian).

We differentiate 8 groups of dialogue acts in our typology, the first 7 groups include acts that

[1] http://sys130.psych.ut.ee/~linds/english/index.html

form APs: 1) Conventional (greeting, thanking etc), 2) Topic change, 3) Contact control, 4) Repair, 5) Questions and answers, 6) Directives and reactions (request, etc), 7) Opinions and reactions (assertion, argument etc), 8) Non-AP acts. The overall number of dialogue acts is almost 130. Every type in our typology contains a subtype 'other' which is used for annotating the things we are not interested in at the moment, or are not able to determine exactly. This gives us a possibility to extend our typology in future.

3 Questions and Directives in Estonian Information Dialogues

Sometimes questions and directives are differentiated on the basis whether the user needs some information (question) or (s)he wants to influence the hearer's future non-communicative actions (directive). Our departing point is another: it is not important for dialogue continuation whether the hearer must to do something outside of current dialogue or not. (S)he must react to both a question and a directive because both are the first parts of APs. The second part of AP can be verbal or non-verbal (some action). It can come immediately after the first part of AP or later. The main difference between directives and questions is formal – questions have special explicit form in Estonian (interrogatives, intonation, specific word order) but directives do not have it.

There are three types of questions: 1) questions expecting giving information: open (wh)question, open yes/no question, 2) questions expecting agreement/refusal: closed yes/no question, question that offers answer, 3) questions expecting the choice of an alternative. The suitable answers are: giving information / missing information, closed answers: yes / no / agreeing no / other yes/no-answer, alternative answers: one / both / third choice / negative / other alternative answer. Open and closed yes/no questions have similar form but they expect different reactions from the answerer (e.g. *Are you open in winter?* expects the answer *yes* or *no*, but by asking *Is there a bus that arrives in Tallinn after 8 p.m.?* the questioner wants to know the bus times).

The first parts of directive APs are 1) request, 2) proposal and 3) offer. The second parts are fulfilling directive: giving information / missing information / action, agreement with directive, refusal of directive, postponing the answer of directive, restricted fulfilling of directive, restricted agreement with directive.

Fulfilling of request is obligatory, fulfilling of proposal or offer is optional. Requests are similar to wh-questions – they expect giving information and not yes/no answer or choice of an alternative. Proposals and offers differ from requests because they expect the different second part. They are similar to closed yes/no questions. The suitable reactions are agreement or refusal. Offer must be distinguished from proposal. In the first case, the action originates from the speaker (offer: *I'll send you the programme*), in the second case from the partner (proposal: *please come tomorrow, call me later*).

Our next aim is to develop a programme which will implement statistical learning methods for recognising dialogue acts.

Acknowledgement

The work was supported by Estonian Science Foundation (grant No 4555) and Estonian Ministry of Education (state program Estonian language and national culture).

References

Edwards, J. 1995. Principles and alternative systems in the transcription, coding and mark-up of spoken discourse. Geoffrey Leech, Greg Myers, Jenny Thomas (eds.), *Spoken English on Computer. Transcription, Mark-up and Application*. London: Longman, 19-34.

Hennoste, T., Koit, M., Rääbis, A., Strandson, K., Valdisoo, M., Vutt, E. 2003. Developing a Typology of Dialogue Acts: Some Boundary Problems. *Proc. of 4th SIGdial workshop,* Sapporo, 226-235.

Hutchby, I., Wooffitt, R. 1998. *Conversation Analysis. Principles, Practices and Applications*. Cambridge, Polity Press.

Mengel, A.; Dybkjær, L.; Garrido, J. M.; Heid, U.; Klein, M.; Pirrelli, V.; Poesio, M.; Quazza, S.; Schiffrin, A.; Soria, C. 2000. *MATE Dialogue Annotation Guidelines. Dialogue acts*. http://www.ims.uni-stuttgart.de/projekte/-mate/mdag/da/da_1.html (28/07/2003).

Sinclair, J.M.; Coulthard, R.M. 1975. *Towards of Analysis of Discourse: The English used by Teachers and Pupils*. London: Oxford UP.

Stenström, A.-B. 1994. *An Introduction to Spoken Interaction*. London and New York: Longman.

A Multimodal Interaction System for Navigation

Dennis Hofs, Rieks op den Akker, Anton Nijholt & Hendri Hondorp
Centre of Telematics and Information Technology
University of Twente, Enschede, the Netherlands
{infrieks,anijholt}@cs.utwente.nl

1 Introduction

To help users find their way in a virtual theatre we developed a navigation agent. In natural language dialogue the agent assists users looking for the location of an object or room, and it shows routes between locations. The speech-based dialogue system allows users to ask questions such as "Where is the coffee bar?" and "How do I get to the great hall?" The agent has a map and can mark locations and routes; users can click on locations and ask questions about them.

In an earlier version (Luin et al., 2001) no underlying dialogue model was used. Now we have a generic architecture and dialogue model allowing multimodal interactions, including reference modelling and backward/ forward looking tags to determine and model dialogue structure.

The architecture of the system can be conceived as a box containing a dialogue manager (DM) and world knowledge, to be connected with input and output processors. The processors can be plugged into different dialogue systems. E.g., our speech processor and generator have been used with some adjustments for both a navigation agent and a tutor agent application. The implementation supports streaming at the recording side as well as the playback side to decrease the delay between user utterances and system responses. The architecture enables the user to interrupt when the system is speaking.

2 The Multimodal Dialogue Manager

Speech input is sequentially processed by the dialogue act determiner (that selects a dialogue act), the parser, the reference resolver and the action stack. The dialogue management module itself is not implemented as a system of asynchronous distributed agents: there is a strict logical order in the execution of the updates of dialogue information, the selection of goals and the execution of actions after the system has received user input. The mouse input reaches the DM through a map, which is a visual 2D representation of the world (the virtual theatre). The user can point at objects or locations on the map and the world notifies the DM of these events.

In our dialogue and action management module we allow mixed-initiative dialogues and several types of subdialogues. An Action stack stores the system's actions that are planned and a subdialogue stack keeps track of the current dialogue structure. All dialogue acts are also kept in a history list to be retrieved for later use.

The input queue of the DM receives the user's utterances in the form of lists of possible acts. The relation between word sequences and acts is specified in the grammar for the speech recogniser. A dialogue act contains the original sentence and a forward and backward tag, based on the DAMSL scheme. In addition, a dialogue act may have a domain-specific argument.

When the DM gets a list of acts from the input queue, it passes it to the dialogue act determiner together with the history. Information in the history that the dialogue act determiner uses, are the forward tags of the last utterance in the current subdialogue and the last utterance in the underlying subdialogue, if present. For every possible forward tag, the dialogue act determiner holds an ordered list of preferred backward tags that can follow it. The dialogue act determiner selects the preferred dialogue act and returns it to the DM.

Our dialogue act determiner also helps to determine the dialogue structure with respect to subdialogues. If the user could end the current subdialogue – that is if the last dialogue act in the underlying dialogue was performed by the system and the user started the current subdialogue – the dialogue act determiner will always try to end the current subdialogue by connecting the user's dialogue act to the underlying dialogue.

3 Parser and Reference Resolver

The DM can start processing the selected dialogue act. It starts with parsing the phrases that occur in parameters in the dialogue act's argument. The feature structure that is obtained is stored together with the original parameters in the dialogue act. The next step is to bind the parameter to a real object in the navigation agent's world. That is where the reference resolver comes in. The reference resolver is used for all references to objects in the world. References can be made by the user or the system, by talking about objects or by pointing at them. The reference resolution algorithm is a modified version of Lappin and Leass's algorithm that assigns weights to references, based on a set of salience factors. Although language used in the dialogue system is rather simple, compared to complex structures in written texts, resolving referring expressions in multimodal interaction is far from trivial. The modification makes the algorithm suitable for multimodal dialogues. For details of the adapted algorithm see (Hofs et al, 2003).

After resolving references, the set of objects found is added to the parameter in the dialogue act where the reference occurred. So now we have a dialogue act that consists of a forward tag, a backward tag and an argument. The parameters in the argument have a value that was taken directly from the recognition result as well as a feature structure received from the parser and a set of objects received from the reference resolver.

4 Dialogue and Action Stacks

The history contains all dialogue acts that occurred during the dialogue. Besides the user's dialogue acts, it also contains the system's dialogue acts. A subdialogue stack is used for the currently running subdialogues. The action stack contains the actions that the system still needs to execute, but the stack also creates those actions, when it is provided with the user's dialogue acts after the parameters have been parsed and references were resolved. Actions are specified in templates and are part of the action stack.

When the action stack receives a user dialogue act, it will try to find an action template with matching forward tag and argument. It creates actions based on the action names in the template and it extracts the action arguments from the user's dialogue act. The new actions are put on top of the stack and when it is the system's turn, it will take an action and execute it. If a dialogue act is performed within the action, the reference resolver's dialogue model should be updated and the dialogue act should be added to the history. Like a user dialogue act, a system act consists of a forward tag and backward tag. It does not have an argument. Instead it contains the action within which the act was performed.

5 Conclusions

We presented a generic dialogue model for spoken multimodal interaction. One of our applications, discussed here, is a navigation agent that helps users to find their way in a virtual environment. The system uses Dutch speech recognition and synthesis models. References to objects in the environment can be made by user or system. The resolution algorithm is a multimodal version of the well-known Lappin and Leass's algorithm. Backward and forward-looking tags according to the DAMSL scheme are used to model the dialogue structure. Until now we used our system to implement a navigation agent in a virtual environment and, using the speech architecture module, a tutor agent. In progress is an application where we integrate speech and haptics in a virtual nurse education environment.

References

D. Hofs, R. od Akker & A. Nijholt. A Generic Architecture and Dialogue Model for Multimodal Interaction. *Nordic MUMIN* Workshop, Denmark, 2003.

S. Lappin & H. Leass. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535-561, 1994.

J. van Luin, A. Nijholt & R. op den Akker. Natural Language Navigation Support in Virtual Reality. *ICAV3D* Workshop Greece, 2001, 263-266.

The Critical Agent Dialogue (CrAg) Project

Amy Isard Carsten Brockmann Jon Oberlander Michael White
ICCS, School of Informatics
University of Edinburgh
{Amy.Isard,Carsten.Brockmann,
J.Oberlander,Michael.White}@ed.ac.uk

Abstract

The aim of this project is to build and evaluate a simple, adaptable natural language generation system which can generate dialogue incorporating relatively subtle linguistic features which reflect dimensions of personality such as extraversion and neuroticism. The system will then be evaluated to investigate the impact on user impressions of altering personality parameters.

1 Introduction

The aim of the Critical Agent Dialogue (CrAg) project[1] is to build and evaluate a simple natural language generation system which can produce dialogue involving relatively subtle language features reflecting dimensions of personality. The generation model will also be informed by the Interactive Alignment Model put forward by Pickering and Garrod (2003).

The system will be demonstrated via a pair of 'artefact critical agents', who will play the part of movie reviewers, putting forward opinions and arguing about recent movie releases. The dialogues are intended to resemble those from popular television shows, such as Ebert and Roeper in the US, in which film critics discuss new releases. (We also hope to be able to model Statler and Waldorf, the grumpy old men from the Muppet Show, once we have determined their personality types.) The goal is for the personalities of the critics to be clearly identifiable through their use of language, and for the interaction between them to be believable and engaging for both researchers and the general public.

2 Personality Features and Alignment

The generated dialogues will vary according to the personality type assigned to the characters, based on recent research into different vocabulary, syntax and dialogue strategies exhibited according to personality type (Gill and Oberlander, 2002). This research uses Eysenck's three factor model (Eysenck and Eysenck, 1991), in which personality is described in terms of the three dimensions Psychoticism, Extraversion, and Neuroticism, each of which can separately influence language production.

Each character's utterances will also vary in reaction to the utterances of the other participant in the dialogue. Garrod and Pickering's Interactive Alignment Model (Pickering and Garrod, 2003) argues that common ground (Wilkes-Gibbs and Clark, 1992) need not be explicitly computed during dialogue, but that it arises as a by-product of intra- and inter-personal priming processes, by which dialogue participants align their representations at every level, including lexical, semantic, and syntactic.

Some example hypotheses about the way in which personality interacts with dialogue behaviour are: high Psychotics are individualistic; they are less worried about others' opinions and are less likely to align. High Extraverts want to gain and retain the conversational floor. Their utterances are longer, and they tend to align with their dialogue partner. High Neurotics are likely to talk about themselves and choose negative content.

User evaluations will assess subjects' judge-

[1]This project is funded by Scottish Enterprise as part of the Edinburgh-Stanford Link.

ments concerning the distinctiveness, friendliness, trustworthiness, continuity, motivation, engagement and sociability of the individual agents in the pair, and the pair taken together. This will allow us to test hypotheses about how users' personality types affect the way in which they react to different agents (Nass and Moon, 2000).

3 System Design

The dialogue system will use the Stanford Open Agent Architecture (OAA) (Cheyer and Martin, 2001) and the Edinburgh DIPPER architecture (Bos et al., 2003) as a framework for the dialogues. We are considering the use of Information States (Traum and Larsson, 2003) to manage dialogue moves. The generation will be done using the OpenCCG Realizer (White and Baldridge, 2003), with an extended grammar to cover the movie domain.

Initial data on the subject and style of movie reviews is being gathered to extract domain-specific vocabulary and to aid in compiling a list of topics which commonly occur. Topics found so far include: dialogue, action, humour, special effects, story, ending, protagonist, fight sequences, plot holes, directing style, cinematography style, director, music, character development. A corpus of spoken dialogues where the participants exchange views on a given movie is in preparation. The personality types of the agents and the data about the movies under review will be stored as XML documents. The design will be informed by the formats used by the NECA project (Piwek, 2003).

We intend to use a mixture of deep and surface generation, so that utterances generated from scratch can be combined with longer pieces of text from the database describing an aspect of one of the topics mentioned above. This strategy follows the one used successfully by the M-PIRO project (Isard et al., 2003).

4 Conclusions

We hope, by altering personality parameters, that we can create agents which have noticeably different characteristics and different dialogue strategies. We will be attempting to model this behaviour and later to assess whether human observers find dialogues which contain alignment to be more realistic. In the process, we hope to generate dialogues that are intrinsically interesting and entertaining to observe.

References

Johan Bos, Ewan Klein, Oliver Lemon, and Tetsushi Oka. 2003. DIPPER: Description and formalisation of an information-state update dialogue system architecture. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan.

Adam Cheyer and David Martin. 2001. The Open Agent Architecture. *Journal of Autonomous Agents and Multi-Agent Systems*, 4(1):143–148.

H. J. Eysenck and S. B. G. Eysenck. 1991. *The Eysenck Personality Questionnaire-Revised*. Hodder & Stoughton, Sevenoaks.

Alastair Gill and Jon Oberlander. 2002. Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, Fairfax, VA, USA.

Amy Isard, Jon Oberlander, Ion Androutsopoulos, and Colin Matheson. 2003. Speaking the users' languages. *IEEE Intelligent Systems*, 18(1):40–46. Special Issue: Advances in Natural Language Processing.

Clifford Nass and Kwan Min Moon. 2000. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, pages 171–181.

Martin J. Pickering and Simon Garrod. 2003. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*. To appear.

Paul Piwek. 2003. A flexible pragmatics-driven language generator for animated agents. In *Proceedings of EACL-03, Research Notes*, Budapest, Hungary.

David Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In Smith and Kuppevelt, editors, *Current and New Directions in Discourse and Dialogue*. Kluwer. To appear.

Michael White and Jason Baldridge. 2003. Adapting chart realization to CCG. In *Proceedings of the 9th European Workshop on Natural Language Generation*, pages 119–126, Budapest, Hungary.

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

Animated User Representatives in Support of Asynchronous Group Decision Making: Challenges for Dialog Processing

Anthony Jameson*
DFKI, German Research Center for Artificial Intelligence
and International University in Germany
jameson@dfki.de

Abstract

The goals of this demo are (a) to illustrate a novel application area for natural language dialog systems and (b) to discuss some of the issues that arise in the design of systems for this area.

1 Dialog Systems as Representatives of Absent Group Members

The system demonstrated, called the TRAVEL DECISION FORUM (see Jameson, Baldes, & Kleinbauer, 2003) helps three members of a group to agree on a single set of criteria that are to be applied in the making of a particular decision (e.g., what their planned joint vacation should be like). The system is intended for use in situations in which the group members cannot communicate face-to-face or with synchronous communication media.

When one group member interacts with the system, he or she sees two animated characters that represent the other two group members, who are not currently on-line (see Figure 1). These *representatives* respond with natural language and gestures to proposals made either by the *mediator* character or by the current user. Each proposal concerns a set of joint preferences concerning one aspect of the decision to be made. Each representative has a certain degree of authority to accept proposals on behalf of the corresponding real user.

* This research was supported by the German Ministry of Education and Research (BMB+F) under grant 01 IW 001 (project MIAU).

A representative's responses to a proposal are based on the domain-specific *preferences* of the corresponding real group member (e.g., how important it is to have access to a beauty farm), which that member has previously specified by filling in electronic forms (see the lower right-hand corner of Figure 1).

The dialog behavior of a representative is also determined by a number of more general parameters that the real group member has set. These parameters determine, among other things, the way in which the representative takes the preferences of other group members into account when evaluating a proposal. For example, in Figure 1, Ritchie's representative is concerned about whether a proposal is more favorable for Tina than for Ritchie. The real user can specify not only how his or her representative *actually* evaluates proposals but also how it *appears* to evaluate them (cf. Jameson, 1989). For example, instead of allowing the rather childish-looking behavior shown in Figure 1, Ritchie might specify that his representative should argue as if Ritchie were concerned about the overall utility of each proposal for all group members.

By interacting with animated representatives of absent group members, the current user can enhance her awareness of the preferences and motivation of these group members. She can then generate proposals that have a good chance of being accepted by the other members and/or by their representatives. The current user can also change the specification of her own preferences and motivation so as to facilitate the reaching of consensus.

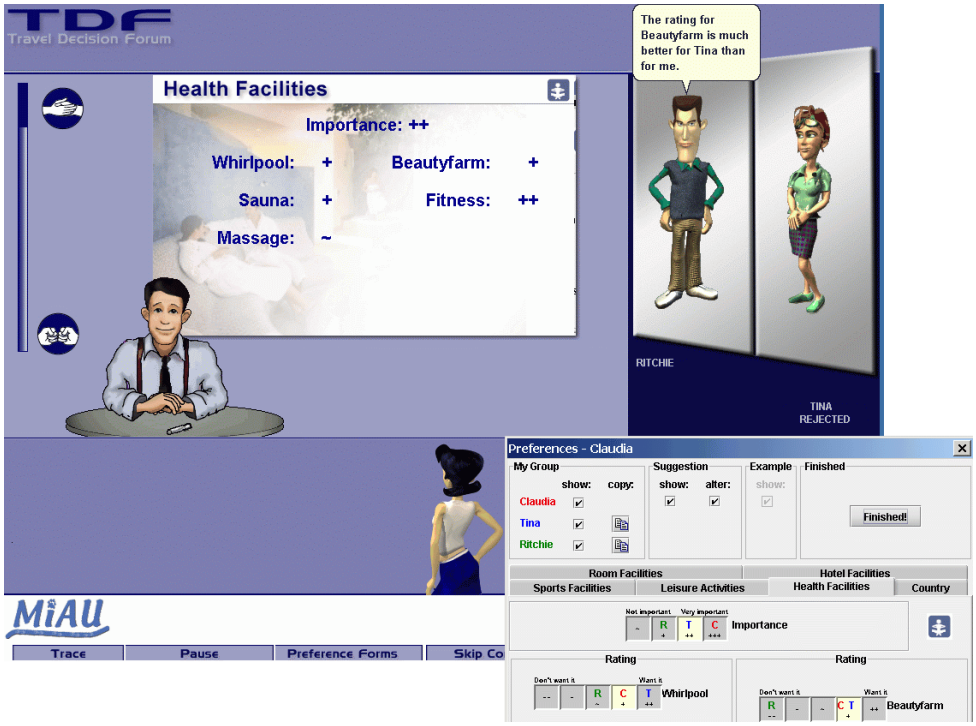


Figure 1. *Snapshot of an interaction in the Travel Decision Forum.*
(The proposal generated by the mediator (left) for the dimension Health Facilities *is shown on the screen, as well as in the preference form at the bottom right. The representative of Tina has already explained why the proposal is unacceptable to Tina. After the representative of Ritchie has finished commenting on the proposal, the current user, Claudia, will decide how to respond to to it.)*

2 Issues Raised by Representatives

This type of application scenario raises a number of general issues concerning dialog system design. In particular, it is not clear what the best approaches to the following two problems are, even though there exists a good deal of relevant previous research:

Avoiding monotony. The comments of the representatives tend to be similar in form, since they all concern responses to proposals. But if the representatives' presentations are monotonous, the acceptance of the entire system is endangered. Possible partial solutions to this problem include (a) keeping utterances as short as possible after an initial period of familiarization; (b) introducing more or less random variation in the formulations; and (c) generating utterances that refer back to previous utterances (e.g. "I feel the same way as Tina does, except that . . . ").

Handling conflicts between actual and ostensible values of dialog parameters. The presentation of insincere arguments raises interesting challenges for natural language generation on the pragmatic level. For example, what should a representative say if the current proposal is clearly acceptable according to its ostensible motivation but unacceptable according to its real motivation? Should the representative accept such an undesirable proposal simply in order to stay consistent with its ostensible motivation?

References

Jameson, A. (1989). But what will the listener think? Belief ascription and image maintenance in dialog. In A. Kobsa & W. Wahlster (Eds.), *User models in dialog systems* (pp. 255–312). Berlin: Springer.

Jameson, A., Baldes, S., & Kleinbauer, T. (2003). Enhancing mutual awareness in group recommender systems. In B. Mobasher & S. S. Anand (Eds.), *Proceedings of the IJCAI 2003 Workshop on Intelligent Techniques for Web Personalization.* Menlo Park, CA: AAAI. (Available via http://dfki.de/ jameson/.)

Marked Informationally Redundant Utterances in Tutorial Dialogue

Elena Karagjosova
Computational Linguistics
Saarland University
elka@coli.uni-sb.de

Abstract

The paper presents a preliminary corpus study of marked Informationally Redundant Utterances (IRUs) in tutorial dialogues. The study suggests that marked IRUs have the function to make non salient propositions salient or to maintain the saliency of propositions in dialogue. This is interpreted as an indication that marked IRUs may have a greater learning effect, since students are able to recognize old information, which may help activate related knowledge and thus promote active learning.

1 Introduction

We present ongoing research on marked Informationally Redundant Utterances (henceforth IRUs) in tutorial dialogues. The work is part of the project DIALOG whose goal is to investigate the use of flexible NL dialog in tutoring mathematics and to develop a prototype gradually embodying the empirical findings (Pinkal et al., 2001).

One aspect of this goal is exploring tutoring techniques and their NL realization and implementation. In tutorial dialogues, it may be beneficial for the student (S) that the tutor (T) presents information that has been already mentioned, as S will e.g. easier memorize it. However, S will not always be able to recognize old information, unless it is marked as such. One disadvantage for S in not recognizing old information is that he may not be able to distinguish it from new material as pointed out in (Moore, 2000). Also, recognizing old information may have a greater learning effect on S as it may promote active learning, i.e. help S activate it as well as further relevant knowledge.

In the next section we describe Walker's theory of IRUs (Walker, 1993) on which the present work is based. Section 3 describes the results of our corpus study and Section 4 provides a summary.

2 Background

Walker (Walker, 1993) defines an utterance u_i as informationally redundant in a discourse situation S if u_i expresses a proposition p_i, and another utterance u_j that entails (repeats, paraphrases), presupposes or implicates p_i has already been said in S. The utterance u_j that originally added the propositional content of the IRU is the *antecedent* (ANT) of the IRU u_i. The *salience status* of ANT can be *adjacent* (ANT in previous turn), *same* (IRU in same turn as ANT), *last* (ANT in last turn of current speaker), and *remote*.

Walker considers three types of IRUs according to their *communicative function*. **Attitude IRUs** demonstrate the hearer's attitude (accept/reject) to an utterance just contributed by a speaker to the discourse situation. **Attention IRUs** make a proposition salient and are of three types. *Deliberation IRUs* make a proposition salient in order to provide a premise for an argument.[1] *Open Segment IRUs* signal a return to an earlier discussion/segment. *Close Segment IRUs* include important information, e.g. by repeating main points or making explicit causal relations. **Consequence IRUs** like *Inference-Explicit IRUs* make an infer-

[1] Deliberation IRUs can be *Warrant IRUs* which support an intention, and *Support IRUs* which support another proposition.

ence (e.g. entailment, implicature) explicit.[2]

Walker does not account of when and how IRUs can be marked as such.

3 Corpus study: Results and outlook

We investigated a corpus of German tutorial dialogues collected in a Wizard-of-Oz experiment that simulated a system for teaching proofs in naive set theory (Benzmüller and al., 2003). In the experiment, the use of NL expressions was left to the wizard's own discretion.

So far, we have identified 10 marked IRUs out of 42 IRUs in 67 dialogues. 2 IRUs were marked by *past tense* of the finite verb, 3 by means of expressions like *Der letzte Schritt war: F* (*The last step was: F*), where *F* is a set-theoretical formula. 1 IRU was marked by the expression *Wie wir eben festgestellt haben, [...]* (*As we have just noticed, [...]*), 2 by *wie gesagt* (*as already said*), and 2 by means of the modal particles *eben* and *ja*. All marked IRUs were uttered by T. 7 of them had an ANT uttered by T, 3 by S. None of the ANTs were adjacent or same, 6 were remote and 4 last. 3 of the IRUs were Repetitions, 7 Paraphrases, none were Inferences. All 3 Repetitions had a remote ANT, the Paraphrases of 4 ANTs were last and 3 remote. All 10 were Attention IRUs: 4 Deliberation (Support), 3 Open and 3 Close Segment.

The fact that all marked IRUs were Attention IRUs whose function is to make a non-salient proposition salient or to maintain the saliency of a proposition confirmed our expectation that marked IRUs will be used by T to provide old information that is not salient and that S may not be currently aware of but that is relevant for solving the task. It remains to be examined in what way the marked IRUs differ from the ones that are not marked.

The observation that all marked Repetition IRUs were remote runs against Walker's findings. This may have to do with the mathematical domain where repetition is natural as precision is essential, as well as with the tutorial dialogue genre where repeating given information rather than referring to it may help avoid ambiguities which could lower the learning effect. It remains to be seen whether this depends on the type of IRU (marked vs non-marked) or it is domain specific.

Another finding that seems to be genre specific is that we haven't come across marked Open Segment IRUs that are uttered by S.[3] We also didn't find Close Segment IRUs uttered by S. This may have to do with the fact that it is usually T who is in control of the dialogue and decides when a topic is closed or a (sub)task completed.

The study also suggests that some markers may only be used with IRUs that have a particular relation to their ANTs, e.g. *Der letzte Schritt war F* is only used to mark IRUs that are Repetitions.

Further work will explore possible correlations between IRU-markers, type of relation to ANT and communicative function. The claim that marked IRUs increase the learning effect also will be tested in subsequent experiments.

4 Summary

We presented the preliminary results of a corpus study that has the goal to explore the role of marked IRUs in tutorial dialogues. The study yielded that all marked IRUs were Attention IRUs whose function is to make a proposition salient thus confirming the intuition that marked IRUs will be used by T to provide old information which is relevant for solving the task at hand and which S may not be currently aware of.[4]

References

Christoph Benzmüller and al. 2003. A wizard of oz experiment for tutorial dialogues in mathematics. In *Proceedings of the AIED*, Sidney.

J. Moore. 2000. What makes human explanations effective? In *In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*.

Manfred Pinkal, Jörg Siekmann, and Christoph Benzmüller. 2001. Projektantrag Teilprojekt MI3 — DIALOG: Tutorieller Dialog mit einem mathematischen Assistenzsystem. Universität des Saarlandes.

Marilyn Walker. 1993. *Informational redundancy and resource bounds in dialogue*. Ph.D. thesis.

[2] See (Walker, 1993) for more detailed descriptions.

[3] We have found however unmarked cases where S used an Open Segment IRU to return to a point that he needs T to clarify.

[4] This work is supported by SFB 378 at Saarland University, Saarbrücken. Thanks to two reviewers for suggestions.

Information Seeking Dialogues with Automatically Acquired Domain Knowledge

Udo Kruschwitz
Department of Computer Science, University of Essex
Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom
udo@essex.ac.uk

1 Overview

Improving Web search technology is a hot topic. One aspect that makes it so interesting is the fact that Web documents are typically not plain text files - instead, they contain a tremendous amount of implicit knowledge stored in the markup of the documents. Much of this need not be used in general Web search, because the search engine doesn't need to understand the documents it is accessing. But what if the document collections are very domain-specific or limited in size? This type of data source is everywhere, from corporate intranets to local Web sites. To help a user search such collections we created a search system based on a generic framework. The system incorporates a simple domain-independent dialogue manager and an automatically created model of the domain.

2 The Domain Model

The simple graph in figure 1 is an actual example from one of our sample domains that gives an idea of how a domain model may be structured. We see a simple tree of related terms that can be used to either assist a user in the search process, perform some automatic query refinements or allow the user to browse the collection. Note, that the *types* of relation in the sample tree are not formally specified. This is significantly different from formal ontologies or linguistic knowledge sources such as *WordNet*.

However, typically a domain model is not available. We construct a domain model automatically by exploiting the markup of documents (see (Kruschwitz, 2001) for detailed information on this process as well as links to related work).

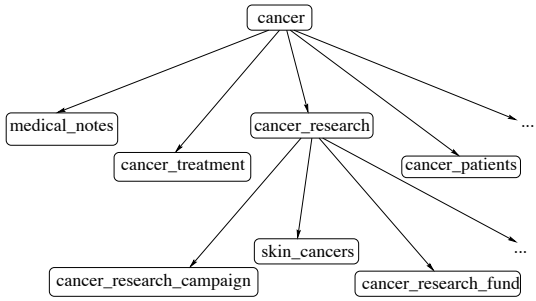


Figure 1: Partial domain model

Having built the model it can now easily be incorporated into our generic framework which is an online search system that has access to a standard search engine and the domain model (Kruschwitz, 2003). This search system is a specialized dialogue system that offers and makes choices about search through the set of documents based on the domain model. This dialogue system is unlike *single shot* systems such as standard search or question answering in that it interacts with the user by offering options to refine or relax the query. With a new document collection coming along we just have to press a button to create the appropriate domain model. The rest of the framework can be left unchanged. Note, that the dialogue manager itself is domain-independent.

3 The Dialogue Manager

We understand dialogue roughly as a movement in the space of document descriptions. That means there are no strictly defined dialogue states as they

are commonly used in dialogue systems. Our dialogue states are calculated automatically, and they are represented as a tuple containing a number of different types of information as we shall see further down.

The dialogues we describe are system initiated. Although the user has some freedom to navigate through the dialogue, it is basically an information seeking task that needs to be performed and the system is merely an assistant to help the user get to the right set of answers. As such the focus of attention is a system that presents results alongside possible choices the user might want to consider to continue the search task.

How do we move from one state to the next one? The general idea is to analyze the current dialogue state and the user input in order to generate a new dialogue state. A dialogue state is characterized by: (1) a dialogue history; (2) a formalized user query (called *goal description*); (3) the set of results that matches this goal description; (4) a set of *potential user choices* (options to relax, refine or change the query in other ways).

The list of choices is selected from a set of possible modifications to the goal description taking into account their effect on the result set. This may sound very general. The reason is that we are able to describe a variety of dialogue strategies depending on how those possible choices are to be selected by the system.

Relaxation and refinement steps are the essential building blocks that define the choices a user can select from to continue the dialogue. Ideally, our dialogue system would explore all relaxation and refinement possibilities and select the best ones. However, the reasons for not doing so include the complexity involved and our experience that users are generally not very happy if the system performs complex actions automatically, e.g. automatic query expansion using hypernyms *and* synonyms. Therefore, each of the choices represents a *single* refinement or relaxation step. These choices are derived from the domain model. Customizations of the dialogue manager allow the incorporation of other knowledge sources.

The interaction between user and search system serves the purpose of navigating the user from some initial search request to a satisfactory set of answers. In other words, the *goal description* is constantly being updated in a sequence of simple *dialogue steps* until it matches a set of documents the user is happy with. A dialogue step can be seen as a number of smaller steps to be performed in this order: (1) evaluate the user input; (2) calculate the new dialogue state; (3) perform all actions corresponding to the state transition (e.g. display results and choices).

A *goal description* is a set of attribute-value pairs where each attribute is a document property or a system property and its value is an element from the domain of that property. A goal description is something like a formalized user query. At each stage of a dialogue we can have a look at the goal description and see the current user request. In the simplest case we would just have a list of keywords. In a more advanced search system we have a number of properties that can be tuned and changed during the dialogue.

4 Results and Next Steps

UKSearch is an implemented prototype of our dialogue based search system. Results of a detailed task-based evaluation (using TREC's interactive track guidelines) have shown that users prefer this system over a baseline approach - a standard search system (detailed results to be published shortly). The focus is now on further evaluation steps. An EPSRC grant has been awarded for a 15-month research project entitled *"Investigating the Usefulness of Markup-Based Knowledge Extraction"*[1]. The project has started in June 2003 and will include larger scale evaluations of the search framework for a number of document collections.

References

U. Kruschwitz. 2001. A Rapidly Acquired Domain Model Derived from Markup Structure. In *Proceedings of the ESSLLI'01 Workshop on Semantic Knowledge Acquisition and Categorisation*, Helsinki.

U. Kruschwitz. 2003. An Adaptable Search System for Collections of Partially Structured Documents. *IEEE Intelligent Systems*, July/August:44–52.

[1]EPSRC grant number GR/R92813/01

Functions and Patterns of Speaker and Addressee Identifications in Distributed Complex Organizational Tasks Over Radio*

Bilyana Martinovski, David Traum, Susan Robinson, Saurabh Garg
Institute for Creative Technologies, University of Southern California
13274 Fiji Way, Marina del Rey, CA 90292 USA

1 Introduction

In multiparty dialogue speakers must identify who they are addressing (at least to the addressee, and perhaps to overhearers as well). In non face-to-face situations, even the speaker's identity can be unclear. For talk within organizational teams working on critical tasks, such miscommunication must be avoided, and so organizational conventions have been adopted to signal addressee and speaker, (e.g., military radio communications). However, explicit guidelines, such as provided by the military are not always exactly followed (see also (Churcher et al., 1996)). Moreover, even simple actions like identifications of speaker and hearer can be performed in a variety of ways, for a variety of purposes. The purpose of this paper is to contribute to the understanding and predictability of identifications of speaker and addressee in radio mediated organization of work.

2 Corpus and Annotation

The data set used in this study consists of a 1/2hr fragment of a 1hr40min simulation exercise involving trainees in helicopter flight simulators, involved in a coordinated mission with a command post and semi-automated simulated forces. A small excerpt is shown in Figure 1. There are over 30 speaking participants communicating over multiple radio frequencies; simultaneous speech is not necessarily interrupting.

*The work described in this paper was supported by the Department of the Army under contract number DAAD 19-99-D-0046.

N	Freq	Spkr	Addr	Said
1	45	R06	STW	STEEL tower .
2	45	R06	STW	rogue 0 6 ,
3	42	STW	R06	0 6 ,
4	42	STW	R06	tower ?
5	45	R06	STW	tower .
6	45	R06	STW	this is rogue 0 6 ,
7	42	STW	R06	0 6 taxi for departure ?
8	45	R06	STW	rogue 0 6 .
9	42	R06	R07	rogue 0 6 ,
10	42	R06	R07	0 7 ,
11	42	R06	R07	0 6 is up at this time
12	8, 43	R7-B	DO	a:nd dragonops ,
13	8, 43	R7-B	DO	rogue 0 6 and rogue 0 7 are alpha at this time ,

Figure 1: Example of Radio Talk (IDs are underlined, IDAs in italics).

We segment the data into several levels of interaction structure. A **transmission** is a communication unit delivered over the radio. At a finer-grained level, we segment communication units into **utterance units** (Gross et al., 1993). Sequences of utterance units by multiple speakers are clustered into **episodes** of sub-activities, each taking place within a contiguous block of time and centered on a single purpose. Episodes typically have three chronologically organized phases: *beginning*, *action*, and *closure*. They include on average 10-20 utterance units and between 4 and 15 transmissions. In addition to segmentation and physical communication factors such as time and frequency of transmission, we also annotated several dialogue functions, including the addressee of each utterance unit, any indications of call signs identifying the speaker **(ID)** or the addressee **(IDA)**, see Figure 1.

3 Analysis

We analyze several types of factors that play a role in the interpretation of the data, including **Linguistic:** location of ID/IDA in the episode and the transmission; other speech acts or military expressions in the same transmission as IDA/ID; and **Social:** the role of the speaker; the relationship (including military rank) between participants; the episode.

We grouped transmissions into seven patterns with respect to occurrence of ID, IDA and other acts: (1) Basic IDA-ID pair (2) Reversed ID-IDA pair (3) Single ID or IDA (4) Correction of ID or IDA (5) ID or IDA as 3rd person reference in a core act (6) ID or IDA with other core acts or military expressions (7) no ID or IDA.

There are three types of sequences of transaction patterns: *standard truncated*, and *extended*. The standard pattern (e.g., 1-7 in Figure 1) follows the exact instructions in the call sign protocols for the Army, involving at least three transmissions, each of which contains pattern (1): IDA ID. However, the standard sequence is very rare, only 4 sequences are of this type and they are all initiations of contact between an entity and a command site, i.e. an inter-team call. An Example of the truncated sequence pattern is lines 12-13 in Figure 1 where the pattern is: `a:nd` IDA, ID as subject in 3rd person. The truncation is expressed in several ways: the initial conjunction indicates the existence of previous contact, call sign of the addressee is the short name of the entity, not the full name, Dragon Operations (compare to the full name used in the initial contact with Steel Tower on line 1), and finally the self-identification, although full, is used as the subject of a statement informing the coordinating center of the activities of both helicopters. In this sense the identification function of the ID on line 13 is assumed. The extended sequences involve searching for an entity with multiple calls, achieving a response only after subsequent calls, if at all.

The call signs are pervasive in radio talk: 42.3% of all utterance units (from a corpus of 977 utterance units, total) consist of ID or IDA, of which 20.1% are IDs and 22.2% IDAs. The position in the episode is the most significant factor influencing the frequency of the IDA/ID: 75.3% of all ID/IDAs are in beginning phase of an episode, regardless of their initiating or responding function.

67.9% of all IDs and 71.9% of all IDAs occur in communication units which initiate a speech act. That means that both IDs and IDAs are used mostly for initiation, both on utterance function level and on episode level. In addition, the prescribed format for use of call signs demands the first mention of IDA and second mention of ID, which is the dominating pattern in the data. The cases in which the call sign contact follows the most explicit 3-step pattern are all in the first time establishment of a contact at the beginning of an episode, thus function as activity or topic management as found for other kinds of information dialogues (Rats, 1996).

The roles, ranks, and relationships between speakers and entities are a strong influencing factor. As expected, exchanges between entities representing different task teams, i.e. "inter-team" exchanges are typically much more formal and use more IDA/IDs whereas exchanges between members of the same unit are typically less explicit. 75% of inter-team transmissions include an ID or IDA, while this is true of only 50% of intra-team transmissions.

The activity type also influences the use of IDA/IDs. Most of the IDA/IDs are used in achieving a task, information gathering, and status report activities. Communication checks and single calls consist mainly of IDA/IDs. It is expected that activities which require more changes of addresses and speakers, such as gathering information from different entities may be, will have higher number of call signs, simply because they will be also different call signs and different contacts.

References

C. Churcher, D. Souter, and E. Atwell. 1996. Dialogues in air traffic control. In *Proceedings of the Twente Workshop on Langauge Technology: Dialogue Management in Natural Language Systems (TWLT 11)*.

Derek Gross, James Allen, and David Traum. 1993. The TRAINS 91 dialogues. TRAINS Technical Note 92-1, Department of Computer Science, University of Rochester, July.

M Rats. 1996. *Topic Management in Information Dialogues*. Ph.D. thesis, Katholieke Universiteit Brabant.

Initiative in Health Care Dialogues

Mary McGee Wood, Richard Craggs Department of Computer Science University of Manchester

Ian Fletcher, Peter Maguire Psychological Medicine Group University of Manchester

{mary,rcraggs}@cs.man.ac.uk{ian.fletcher,peter.maguire}@man.ac.uk

Abstract

"Initiative", or "control over the flow of conversation", can be superficially mapped to the distinction between "open" and "closed" questions developed by psychiatrists in the analysis of human-human cancer care dialogues. We explore here the linguistic patterns and mechanisms of initiative found in a corpus of health care dialogues [1], and the significance of this data for more general research. The dialogue functions of individual utterances, especially questions, are often ambiguous until seen in the context of a longer sequence.

1.1. "Initiative" in dialogue analysis

Walker & Whittaker (1990) distinguish four utterance types in terms of shifts in control:

assertions, commands, questions - maintain speaker control;

prompts ("utterances which do not express propositional content") - pass control to the hearer.

"Prompt" covers a wider range of syntactic utterance types than indicated: backchannels are the clearest case, but repetitions, reformulations, most types of questions, and even completions can be used as prompts. Our corpus also shows many types of questions. Both wh- and yes-no questions can be used either to maintain or transfer initiative, depending on their semantic content and discourse context.

[1] Our corpus consists of 37 dialogues between Macmillan Cancer Care nurses and patients, each comprising 200-1200 utterances (mostly 300-600). They were collected in a study undertaken by the Psychological Medicine Group, University of Manchester, funded by Cancer Research UK. See Wood 2001, Wood et al 2002 for details.

1.2. Dialogue analysis in cancer care

Cancer patients' mental health improves when they discuss freely whatever they may be worried about. Doctors and nurses aim to speak in a way which encourages this. We thus find a higher than usual proportion of grounding moves to ensure mutual understanding and trust; repetitions, paraphrases, and summaries ensuring factual clarifications; backchannels, and a wide range of other prompting utterance types (Wood et al 2002). Psycho-oncology analyses distinguish questions primarily as "open" or "closed": opening out, or restricting, the range of appropriate answers available to the patient (Maguire et al 1996).

These dialogues are also unusual in combining elements of "tutorial" and "collaborative" style, often in distinct phases within a single dialogue. The nurses' goals are both to learn about the patients' condition, and to inform the patients about their diagnosis and treatment. It is thus no surprise to find patterns of initiative shifting.

1.3. "Initiative" in cancer care dialogues

The concept of "initiative" is obviously valuable in describing the dialogue patterns of interest in psycho-oncology. Some mappings between the two fields of study are fairly clear. In particular, "prompts", relatively rare in well-studied dialogue types, are common here.

The simplest form of prompt is the backchannel (Whittaker et al's paradigm case). Typically, a nurse uses backchannels to encourage the patient to keep talking, actively promoting hearer control. Backchannels are easily recognised, and good indicators of

initiative, but perhaps even more significant as embedding contexts for Dialogue Acts which in isolation could be ambiguous, such as certain question types.

2. Questions and control

Questions can be classified along many distinct parameters, including syntactic, semantic, and pragmatic. (These should be kept distinct: confusion and inconsistency are likely to arise in annotation schemes where they are not.) It is tempting to align wh-questions with open questions and hearer control, yes-no questions with closed questions and speaker control, as claimed in Wood (2001), where the utterance `Why did you say that about selectron? Have you had it before?` is analysed as an open question followed by a closed question. Many clear-cut examples do exist, but all four mappings are found in our corpus. Where wh-questions are closed or yes-no questions are open, the function can be determined either by semantic content or by dialogue context. A sequence of specific wh-questions can be highly directive, while a yes-no question in a context rich in prompts can give the hearer control.

Open wh-questions

Wh-questions are commonly open-ended and offer initiative to the hearer. Some are clear even out of context:

`N: What sort of things are frightening you?`

`N: Well what else is remaining to be sorted?`

They are often surrounded by backchannels, making the prompting nature of the question even clearer.

Closed wh-questions

On the other hand, some wh-questions are requests for specific information:

`P: I don't take the what are they called?`

`N: Paracetamol.`

Sequences of wh-questions keep tighter control, although the hearer still has the opportunity to take the initiative by adding more information than is asked for.

Closed yes-no questions

Yes-no questions commonly keep control with the speaker:

`N: Have you had to use your inhaler?`

`P: He's a specialist isn't he?`

Open yes-no questions

Yes-no questions can, however, be open. Again, sometimes this is clear from the semantics of the utterance in isolation:

`N: Is there anything that's worrying you?`

`N: Can you bear to tell me what in particular?`

Other examples show repetition or recapitulation, almost resembling tag-questions. While these do drive the conversation, they drive it in a direction established by the patient, and by indicating empathy encourage him/her to continue. Open yes-no questions are also often embedded in a general hearer-control context of prompts.

3. Conclusions

Initiative in our cancer care dialogues arises from a number of different linguistic features. Syntactic classification of utterances does not support a one-to-one mapping to dialogue control strategies. The semantic content of an utterance is also important in determining its initiative function.

Dialogue context is also essential: extended sequences can be characterised far more accurately than isolated utterances. For example, any question type, even if normally associated with speaker control, can function as a contribution to hearer control if embedded in an environment of backchannels.

Although our dialogue type is more complex than straightforward tutorial or collaborative dialogue, we predict that these findings will be true there too.

4. References

Maguire et al. 1996. *Helping Cancer Patients Disclose Their Concerns.* European Journal of Cancer 32A(1):78-81.

Walker, M. & Whittaker, S. 1990. *Mixed Initiative in Dialogue.* ACL 28.

Wood, M. M. 2001. *Dialogue Tagsets in Oncology.* SIGdial2.

Wood et al. 2002. *Rare Dialogue Acts Common in Oncology Consultations.* SIGdial3.

Software components for a dialogue multiagent system

Maxime Morge
PhD Candidate
ENS Mines Saint Etienne (France)
morge@emse.fr

Steven Collins
Research Engineer
ENS Mines Saint Etienne (France)
collins@emse.fr

Abstract

This work proposes a software component approach to design a dialogue multiagent system.

1 Introduction

Interaction is widely recognized as the most important issue to design complex software (Singh, 1997). In a software engineering viewpoint, a MAS software system is made up of multiple independent and encapsulated loci of control (i.e. agents) interacting with each other in the context of a specific application.

MAST (MultiAgent System Toolkit) is a project developed in our laboratory (Vercouter et al., 2003). The goal of this project is to provide a software framework to simplify the implementation of agent-based applications. MAST is composed of different modules :

- DeMas : agent platform for distributed execution providing low level services like registry (`Agent Identifier AID`) or message transport ;
- AdMas : an administrator's tool to supervise the deployment and the execution through observation of DeMas services ;
- GeMas : a library of software components providing models and tools coming from academic works, to build applications ;
- MeMas : an Integrated Development Environment (IDE) for building applications based on GeMas.

2 MASTComponents

This library of MASTComponents is based on the vowels approach where : (1) facet "agent" provides several models of reasoning, (2) facet "environment" is a model for perception and action, (3) facet "interaction" is a dialogue system, (4) facet "organization" : an organizational model for multiagent systems and (5) facet "user" : a Graphical User Interface (GUI) for agents.

Figure 1 shows the UML class diagram that explains the relationships between the classes that represent components, roles and events. A MASTComponentRole, an abstraction of services provided by a component, specifies behaviours implemented by a MASTComponent. A MASTComponentRole is a member of one or several Facets. The MASTEvents link MASTComponentRoles in the same Facet.

This paper focuses on the InteractionFacet.

3 InteractionFacet

The dialogue system proposed is for an argumentation system (Parsons et al., 2002), a generic model of reasoning. Figure 2 reports the UML object diagram that explains the relationships among the objects that represent MASTComponentRoles in the InteractionFacet, and InteractionEvents between them.

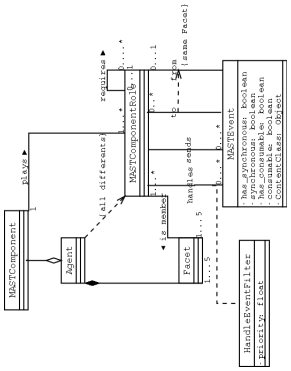


Figure 1: UML Class Diagram of a MASTComponent

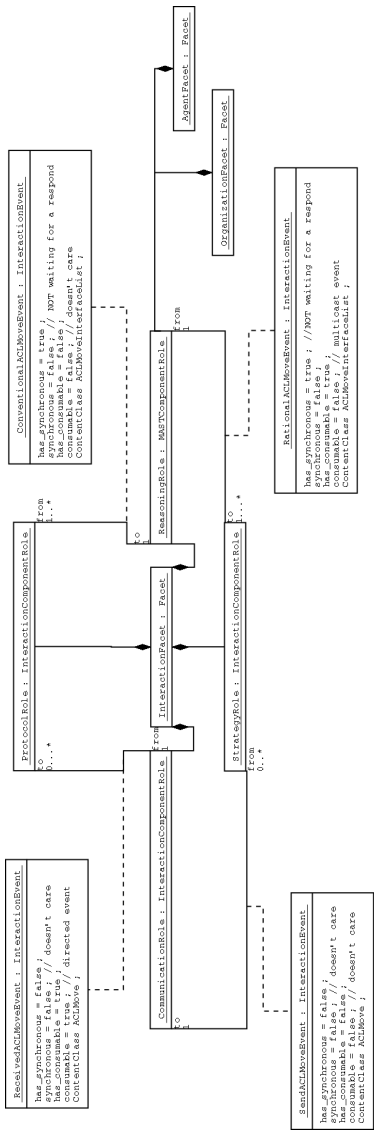


Figure 2: UML Object Diagram of InteractionFacet

an ACLMove is identified by a Move Identifier (MID) which can be referred to later, and contains : (1) a speaker (AID), (2) some hearers (AIDs), and (3) a reply field : a MID or empty if the ACLmove is the first one. an ACLMove follows (4) a conversational protocol (ProtocolRole) and contains (5) an illoctionary act made up of a performative and a content.

An InteractionComponentRoles don't necessarily have all the fields of their ACLMove specified. They exchange ACLMoveInterfaces or a ACLMoveInterfaceList (a collection of ACLMoveInterfaces). A CommunicationComponentRole receives (resp. sends) ACLMoves coming from (resp. to) other agents. A ProtocolComponentRole manages a conversation protocol. It handles the latest received ACLMove and computes a collection of conventional answers. A ReasonningComponentRole handles this collection and selects a subset of chosen ACLInterfaceMove. A StrategyComponentRole chooses an ACLMove, the best depending on the strategy it implements.

4 Conclusion

The goal of this work is to provide a generic way to specify a dialogue multiagent system. Instantiation of this software components is a work in progress.

References

[Parsons et al.2002] Simon Parsons, Michael Wooldridge, and Leila Amgoud. 2002. An analysis of formal inter-agent dialogues. In *Proc. of the first international joint conf. on autonomous agent and multiagent systems*, volume part 1, pages 394–401.

[Singh1997] Munindar Singh. 1997. Interaction-oriented programming for the web. In *Proc of the international conference on Cooperative Information Systems (CoopIS)*, May.

[Vercouter et al.2003] L. Vercouter, O. Boissier, S. Collins, M. Morge, C. Carabelea, B. Bratu, G. Muller, A. Mounier, P. Beaune, and F. Gaultier. 2003. Mast : a multi-agent system toolkit. Technical report, ENS. Mines St Etienne.

A Tool for Multi-Level Annotation of Language Data

Christoph Müller and **Michael Strube**
European Media Laboratory GmbH
Villa Bosch
Schloß-Wolfsbrunnenweg 33
69118 Heidelberg, Germany
{Christoph.Mueller, Michael.Strube}@eml.villa-bosch.de

1 Motivation

We present MMAX, an XML-based Java tool for the annotation of language data on multiple levels of linguistic description.[1] Most currently available annotation tools are restricted to single levels of linguistic description, e.g. coreference, dialog acts, or discourse structure. Annotations produced for individual levels cannot easily be combined or applied to the same language data. This, however, would be highly desirable because it would allow for simultaneous browsing and annotating on several linguistic levels. In addition, annotation tasks could be distributed to several research groups with different expertise, with one group specializing in e.g. dialog act tagging, another in coreference annotation, and so on. After completion of the individual annotation tasks, the annotations could be combined into one *multi-level* annotation that a single group could not have produced. The annotation tool MMAX is intended as a lightweight and flexible implementation of multi-level annotation of potentially multi-modal corpora. It is based on a simplification of annotations to sets of markables having attributes and standing in relations to each other. Due to its simplicity, the tool is fast, robust, and highly usable.

2 Underlying Concepts

1. Base Data We use the term *base data* to refer to the language data to which annotation is added. MMAX supports the annotation of both written text and (transcribed) spoken dialog. Written text is simply modelled as a sequence of sentence elements, each of which spans a sequence of word elements. For spoken dialog, turns are used instead of sentences.[2] Turn resp. sentence and word elements are stored in separate files and linked by means of *span* attributes, which serve as pointers:

```
<turn ID="turn_12"
  span="word_163..word_169" speaker="A"/>

   <word ID="word_163">What</word>
   <word ID="word_164">'d</word>
   <word ID="word_165">you</word>
   <word ID="word_166">do</word>
   <word ID="word_167">with</word>
   <word ID="word_168">them</word>
   <word ID="word_169">?</word>
```

2. Annotation The MMAX tool is based on the assumption that annotations can be simplified to sets of markables having attributes and standing in certain relations to each other.

A **markable** serves as the carrier of information. It aggregates an arbitrary set of elements from the base data. In the case of coreference annotation, markables would represent *referring expressions*, in the case of dialog act tagging, markables would represent *utterances*, and so on. In order to add information to the base data, markables have to associate some **attributes** with them. In MMAX, markables can have arbitrarily many name-value pair attributes. In dialog act tagging, when markables represent utterances, one relevant attribute could be *dialog_act*, with possible values like *initiation*, *response*, and *prepa-*

[1] This abstract is based on a longer version in (Müller & Strube, 2003). The current release version of MMAX can be downloaded at http://www.eml.org/nlp.

[2] For multi-modal dialogs, turns can contain not only words but also gestures.

ration. While markables and their attributes are sufficient to add information to sequences of base data elements, they cannot *relate* these to each other to model structural information. For this, **relations** between markables are required. *Member* and *pointer* are among the relations currently supported by MMAX: member relations express undirected relations between arbitrary many markables. For coreference annotation, a member relation like *coref_class* could be used to mark sets of coreferring markables. Pointer relations express directed (1-to-1 or 1-to-n) relations. The following is an example markable from a coreference annotation. Not all actual attributes are shown.

```
<markable ID="markable_75" span="word_165"
   coref_class="set_7" npform="prp" ... />
```

There is one set of markables per annotation level in a MMAX document. Different annotation levels are kept in separate markable files. Multi-level annotation is made possible because markables are not directly embedded into the base data, but reference base data elements in a *stand-off* fashion (Ide & Priest-Dorman, 1996) by means of their *span* attribute. Since markables on different levels are related only indirectly by virtue of shared base data elements, issues like overlap or discontinuous elements do not arise.

3. Annotation Scheme The annotation scheme contains the user's specification of which (user-defined) attributes and relations are permitted on a markable under which circumstances. These specifications are enforced during the annotation process, thus ensuring annotation consistency and at the same time guiding the human annotator. In a multi-level annotation, markables on different annotation levels (e.g. referring expressions and utterances) require different attributes and relations. Therefore, one separate annotation scheme can be defined for each annotation level.

3 The Tool

For performance reasons, the MMAX tool has a text-only display, but it can visualize relations between markables graphically by means of lines drawn on the text. Installing the tool (under Windows or Linux) is done by simply extracting a directory structure to the local hard disk; no further installation is required. A MMAX project file contains references to all files comprising a MMAX document. For each annotation level, there is one markable file and one annotation scheme file. The project file also contains a list of customizable XSL style sheets which allow different *views* of the same document during a MMAX session. Among other things, these style sheets support the insertion of *markable handles* (usually brackets) in the display, which allow the direct selection of a markable even in cases of multiple embedding. Annotation levels can be activated or deactivated. If required, a popup menu is displayed containing all active markables in a clicked display position. The attributes of the currently selected markable are displayed (and can be modified) in a separate window. If the markable has relations to other markables, these relations are visualized graphically. Creating and deleting markables and relations between them is done by means of context-dependent popup menus. After each modification, the display is refreshed in order to reflect changes to the selected markable's attributes.

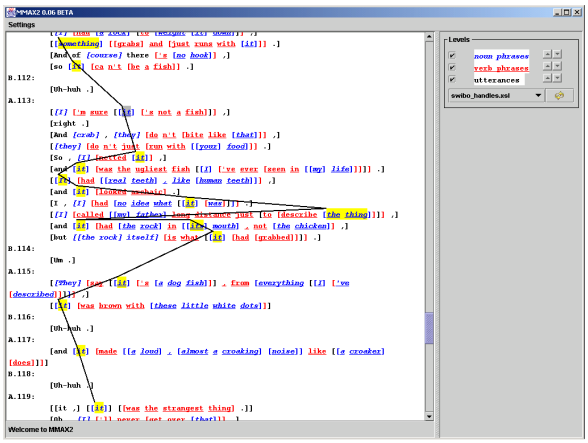


Figure 1: Selected coreference set in MMAX

References

Ide, Nancy & Greg Priest-Dorman (1996). *The Corpus Encoding Standard*. http://www.cs.vassar.edu/CES.

Müller, Christoph & Michael Strube (2003). Multi-Level Annotation in MMAX. In *Proceedings of 4th SIGDial Workshop on Discourse and Dialogue*, Sapporo, Japan, 5-6 July 2003, pp. 198–207.

Generic manager for spoken dialogue systems

Hoá NGUYEN
Laboratory CLIPS-IMAG
University of Joseph Fourier - France
Ngoc-Hoa.Nguyen@imag.fr

Jean CAELEN
Laboratory CLIPS-IMAG
University of Joseph Fourier - France
Jean.Caelen@imag.fr

Abstract

Based on three principal elements: dialogue acts, strategy, and dialogue goal, this paper presents a management model of human-machine spoken dialogue.

1 Introduction

In the context of company's voice portal (PVE[1]) project, our analysis of use shows that the voice service is very useful for applications such as information requests, confirmation of a request, secretary works. Spoken dialogue in these situations is normally short but contains complex statements.

However, it is clearly necessary to have a generic dialogue model that increases as much as possible the independence of the task. In this paper, we propose a generic architecture for a spoken dialogue system and in more details we concentrate on the dialogue management that permits this problem to be resolved.

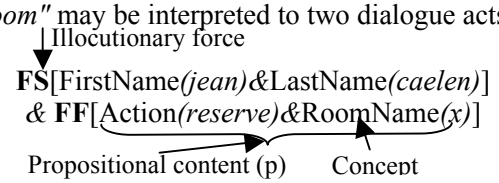
2 Basic principles in dialogue management

2.1 Dialogue act

Austin (1962) considers all utterance as an act of communication called a speech act. For each speech act, by combining with the illocutionary logic notion, Vanderveken (1990) defined the illocutionary force of a speech act. And then, as Caelen (1997), it is useful to retain the following illocutionary forces in the human-machine dialogue domain:

Act	Signification
F	Do or execute an action.
FF	Ask the hearer to perform an action.
FS	Communicate information.
FFS	Ask for information.
FP	Give a choice, make an invite.
FD	Oblige to do without giving an alternative

A *dialogue act* is a speech act enhanced by the illocutionary force. In our dialogue management, a dialogue act is represented by an illocutionary force that specifies what the speaker wishes to achieve, and a propositional content represented the semantic schema of statement. Each utterance can contain one or more than one dialogue act. For example, the utterance "*Jean Caelen here, I would like to book a conference's room*" may be interpreted to two dialogue acts:



We consider that the input of the dialogue manager has always dialogue acts structured like this schema.

2.2 Dialogue goal

A goal is generally a state of the task or a mental state that one wants to reach. A *dialogue goal* is the goal that is sustained during an exchange.

In our model of dialogue management, a dialogue goal is determined by the abstraction of dialogue act with the help of the dialogue plan (which is specified in the task model by elementary goals, called goal of task, and managed by the task manager). For example: with the utterance "I would like to reserve a room please" and the plan for booking a room like **RESERVED**[RoomName:Size:Material], we have the dialogue act as **FF**[Action(*reserve*)&RoomName(*x*)], and the dialogue goal as **ToRESERVE**(*new*) (*new*: new goal stated by the user).

Once the dialogue manager formed dialogue goal, it sends this goal to the task manager to know if this goal is either reached, impossible to reach, or miss information (states concerning tasks). Then, the dialogue manager must resolve itself the other states like satisfied, awaited or left.

2.3 Dialogue strategy

The dialogue strategy is the way to handle the talking turns between speakers to lead a dialogue goal. The strategy aims at choosing the best direction of fit of the goals at a given moment. It decides directly to the dialogue efficiency calculated by the speed of convergence of the dialogue acts towards the final goal. As Caelen (1997), the typology of dialogue strategy is: directive strategy, reactive strategy, constructive strategy, cooperative strategy and negotiated strategy.

3 Dialogue management

For developing spoken dialogue system, we used a generic architecture (Nguyen 2003) as figure 1:

[1] PVE - www.telecom.gouv.fr/rnrt/projets/res_01_5.htm

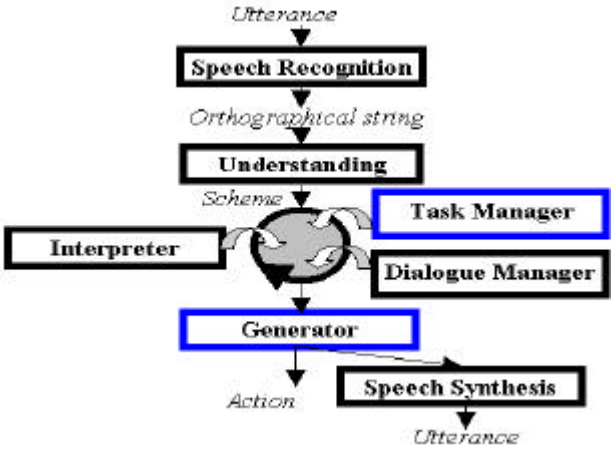


Figure 1: Generic architecture for a spoken dialogue system

In our dialogue management, the *dialogue manager* with the dialogue act passed by *Interpreter* takes generally the role of driving the dialogue cycle, determining the dialogue goal, calculating the adequate strategy, and producing the dialogue act of the machine. Our approach of the dialogue management is based on these three main elements: dialogue act, dialogue strategy, and dialogue goal.

As in the previous section, we could see that:

- The dialogue act, generated by the *Interpreter* module, is independent to the task model.
- The dialogue goal delegated to the task model.
- The dialogue strategy, specified in the dialogue manager, does not related to the task model.

So by these above reasons, our dialogue management is obviously very generic and independent to task model. This remark is very important and it gives us many advantages while building spoken dialogue systems, for example the reusability of dialogue manager, the openness of whole system

By using three elements, a spoken cycle in the dialogue manager happens as the figure 2.

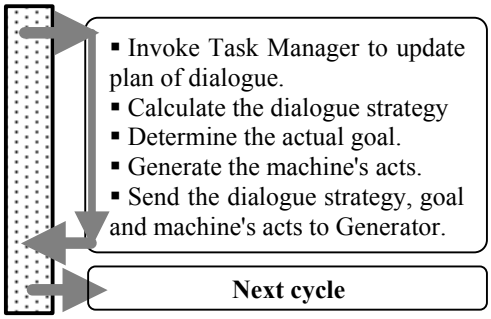


Figure 2: Handling steps in dialogue manager

4 Simulation and validation

By using the principles and the model mentioned above, we have implemented the core of our spoken dialogue system, with the dialogue manager and the task manager, which aims to replace the secretary to realize tasks concerning the reservation of a conference room.

The following example illustrates a dialogue to reserve a room: the system takes the dialogue acts of user's utterance as the input and produces the dialogues acts of the system, dialogue goal and strategy as the outputs.

	Acts	S	Goals
U1	*hello, it is jean caelen* FS[FirstName(jean)]&FS[LastName(caelen)]	D	Role *reached*
M1	FS[Role(directeur du laboratoire)] & FS[Sexe(0)] *Good morning mister director, please to hear you*		
U2	*I would like to reserve a room* FF[Action(réserver)]&FS[RoomName(x)]	C	(Role *satisfied*) Time *new*
M2	FFS[Time(x)] *For what date, please?*		
U3	*What are the possibilities?* FFS[Date(*)]	C	Room *new*
M3	FFS[RoomName(x)] *Do you have any room preference*		
U3	*No, but I need a room for twenty people* FS[Size(+20)]	C	Room *reached* Time *reached*
M3	FS[RoomName(lafayette)] & FS[Date(20030323)] & FS[Hour(480)] & FS[Duration(600)] *The Lafayette room will be available next Friday*		
U4	*That's fine and notify the pve's member* FS[RoomName(OK)]&FS[Time(OK)]& FF[Inform(x)]&FS[Members(PVE)]	R	(Time, Room, *satisfied)* Mode *new*
M4	FFS[Mode(email)] *Shall I notify them by email?*		
U5	*OK,cool, goodbye* FF[Mode(OK)]&FF[Close(OK)]	R	Mode *satisfied* Close *satisfied*
M6	FS[Close(OK)] *Goodbye*		

Table 1. Example of conference room reservation

5 Conclusion

In many actual spoken dialogue systems, the dialogue manager still depends on the task model. The biggest advantage of our dialogue management model is that it is very generic and independent in relation to the task model. Therefore, this module is completely reusable for another spoken dialogue systems.

References

Austin, J. L. 1962. *How to Do Things with Words*, Harvard University Press.

Caelen J. 1997. *Interaction Verbale*, Editions CEPADUES.

Nguyen H. 2003, *Vers une architecture générique de système de dialogue oral homme-machine*. RECITAL.

Vanderveken D. 1990. *Meaning and Speech Acts*, Volumes 1 and 2, Cambridge University Press.

Prototyping Dialogues for Information Access

C.J. Rupp
Computerlinguistik
Universität des Saarlandes
D-66041 Saarbrücken
cj@coli.uni-sb.de

1 Introduction

This project note describes work conducted in the larger COLLATE project, concerned with fundamental issues in the development of information systems. The focus of this part of the project is the conjunction of Question Answering and Dialogue Modelling in several different ways.

- Comparing the notion of question answering in information retrieval with the use of questions as an organisational construct in an information state model for dialogue.
- Extending natural interaction with information systems beyond the question answering task.
- Examining the type of dialogues that arise between users and complex information systems.

Since the use of dialogue to access information services is not yet an accepted technology, suitable dialogue data is in short supply. To counter this, we prototype a partial functionality both as a technical infrastructure into which extended components can be slotted and as a research tool for data collection.

2 Questions and Information Access

Research in Question Answering has been stimulated by the competitive framework of TREC which has spawned a variety of question answering systems. However, most existing systems exhibit a functionality that lies somewhere between that of a web search engine and a human response to a direct question. This is usually a one shot activity, similar to answering quiz questions for humans, a highly stylised and minimal form of dialogue.

A more natural dialogue interaction in response to questions allows for clarification. An answer is not a fact in isolation but part of a system of justified beliefs to which the answerer has committed himself. Therefore, he must be prepared to offer a justification and it is, similarly, natural for an answer to trigger further related questions.

In addition, an information access dialogue may address aspects of how information is obtained. This is a kind of meta-level subdialogue, concerning the process of fulfilling the task. Thus, the system plays two roles, as an expert on information resources, as well as the agent who holds the information that is the answer to a question. There is no natural parallel to this dual role in human dialogue.

3 Implicit Questions

Recent development in dialogue modelling has also been concerned with the relationship between questions and answers as the dynamic driving dialogue interactions. However, such *questions under discussion* (Ginzburg, 1994; Ginzburg, 1996), are more likely to be implicit questions that are raised and answered in the course of a dialogue. Explicit questions enter this representation domain, but in so doing they tend to raise further implicit questions, e.g. as a prototype method for arriving at an answer.

Much of this work has been concentrated within the ISU (Information State Update) approach to dialogue modelling, based on an information state, as a single representation of all the relevant context information for the current point in the discourse, and a system of update rules which map one state to another. The ISU approach provides us with a general, high level tool for prototyping dialogue systems. TRINDIKIT (Larsson et al., 2002) combines a framework for the definition of information states and update rules with the ability to integrate an ISU dialogue model and external resources, such as parsers, speech recognition, and information resources.

4 Available Resources

Dialogue is, characteristically, a complete application for language technology. In principle, resources are required at all levels of analysis, in addition to the application domain. In practice, compromises are possible, provided that the linguistic domain can be adequately restricted. Prototyping also allows the development of

different components at different speeds. In the context of COLLATE we have a number of resources available:

- A corpus of business reports from a newspaper, hand annotated with relational templates,
- A partial parser, SPROUT (Becker et al., 2002), based on finite state technology,
- A grammar of German, defined for named entity extraction in SPROUT over a business domain.

To obtain realistic data for information access dialogues, we set about combining these resources in a prototype dialogue system for closed domain question answering.

5 Our Prototype System

To construct an application from the available resources, a number of additional steps were required:

- The relational templates in the corpus annotation were compiled into a database.
- A robust parser for question constructions was defined, using technology developed for parsing spoken language (Pinkal et al., 2000; Worm, 2000). Fragmentary analyses from the SPROUT parser were taken as input.
- The SPROUT parser, robust parser and business database were encapsulated as modules for TRINDIKIT.

For the remaining modules where we had no specific resources we followed standard practice in TRINDIKIT prototypes, e.g. defining template generation for system outputs.

The dialogue model divides the core task into three phases:

1. obtaining information to consult the database,
2. interaction with information resources,
3. the appropriate presentation of results.

The first and third phases are conducted as dialogue, including clarification subdialogues. The order in which information is exchanged in each phase can be left relatively free, because of the ISU model. In the presentation phase, follow-up questions that can be handled with information already present in the information state are regarded as clarifications, but a question requiring new information from the database is treated as a new query. However, follow-up questions are interpreted in the existing context, thus considerably shortening the new initial phase.

6 Usage

The purpose of constructing this prototype was to get a foothold on the phenomena of realistic information access dialogues, using a restricted system as an environment for eliciting dialogue data. There are two parallel goals here: to conduct usability experiments to see how natural interactions in such an environment can be, and to obtain linguistic data to improve the linguistic components. Since these results feed into subsequent versions of the system, our strategy is effectively one of bootstrapping by means of prototyping. To some extent, this replicates on a small scale the development process effectively employed in Verb*mobil* (Wahlster, 2000).

Acknowledgement

This work is part of the COLLATE project funded by the German Federal Ministry of Education and Research (BMBF) under grant 01INA01B.

References

Markus Becker, Witold Drozdzynski, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. 2002. Sprout - shallow processing with typed feature structures and unification. In *Proceedings of the International Conference on NLP (ICON 2002). December 18-21*, Mumbai, India.

Jonathan Ginzburg. 1994. An update semantics for dialogue. In Harry Bunt, Reinhard Muskens, and Gerrit Rentier, editors, *1st International Workshop on Computational Semantics*, Tilburg, The Netherlands. ITK.

Jonathan Ginzburg. 1996. Dynamics and the semantics of dialogue. In Jerry Seligman and Dag Westerstaahl, editors, *Language, Logic and Computation, Volume 1*, CSLI Lecture Notes. CSLI, Stanford.

Staffan Larsson, Alexander Berman, Leif Grönqvist, and Fredrik Kronlid. 2002. Trindikit 3.0 manual. Siridus Report 6.4, Göteborg University, Department of Linguistics, Göteborg, Dezember.

Manfred Pinkal, C. J. Rupp, and Karsten Worm. 2000. Robust semantic processing of spoken language. In W. Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, Artificial Intelligence, pages 322–336. Springer, Berlin.

Wolfgang Wahlster, editor. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag, Berlin.

Karsten L. Worm. 2000. *Robust Semantic Processing for Spoken Language*. Ph.D. thesis, Universität des Saarlandes, Saarbrücken, Germany, June.

A Tool for Semi-Automatic and Interactive Annotation of Dialogue Utterances with Information States

David Schlangen and **Alex Lascarides** and **Jason Baldridge**
School of Informatics
University of Edinburgh
{das|alex|jmb}@cogsci.ed.ac.uk

1 Introduction

The availability of linguistically annotated corpora like the Penn Treebank has long proven beneficial for computational linguistics research. However, attempts have begun only recently to provide corpora with *discourse* information (c.f. e.g. URML (Reitter and Stede, 2003) or the Penn Discourse Treebank (PDTB) project at the University of Pennsylvania[1]). Here we report on a project whose goal is to generate a corpus annotated with deep discourse semantic information which can then be used to train statistical models of semantic interpretation. We introduce our general methodology and describe how it is instantiated in a purpose-built tool that supports interactive, semi-automated annotation. The novel feature of this tool is its use of a reasoning engine which implements a semantic theory of discourse interpretation to suggest annotations to the user.[2]

2 Overview of the System

At the core of our annotation system lies a module which provides an infrastructure for interactively annotating dialogues. It provides routines to read (XML-formatted) dialogues, pass them to a processing engine, let users edit and possibly reprocess the results of that step, and finally store the finished annotation (also XML-formatted). This infrastructural module is called ANT, for *Annotation Tool*.

This part of the system is independent from any particular annotation task; the modules that make use of the infrastructure on the other hand have to be tailored to the task at hand. For example, for any given task the exact input and output format has to be specified, and a discourse-processing module which computes suggestions for annotations has to be provided. Figure 1 shows an instantiation of the architecture where the discourse processing is done by a system called RUDI (*Resolving Underspecification using Discourse Information*, cf. (Schlangen and Lascarides, 2002)) and the input consists of semantic representations computed by a wide-coverage HPSG (the *English Resource Grammar* (ERG); disambiguated parses of the input utterances are provided by the REDWOODS treebank (Oepen et al., 2002)). This instantiation of the system is called RANT (for RUDI-*based* ANT).

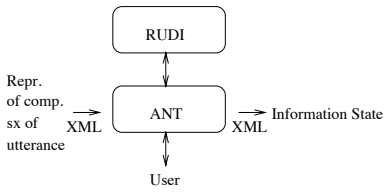


Figure 1: Schematic Overview of RANT

In general terms, an annotation cycle in ANT proceeds as follows:
(a) First, one basic unit of the data that is to be annotated is read in; this could be an orthographic transcription of an utterance in a dia-

[1] C.f. http://www.cis.upenn.edu/∼dltag/

[2] Thus the development of the tool also allows us to evaluate and improve the symbolic theory underlying the computation of suggestions by exposing it to large sets of "real-world" data.

logue, together with a representation of its syntactic and semantic structure. (b) This information is passed to the discourse processing module, together with previously collected information (the *context*). The system expects back from that module an *information state* (IS) representation, which might for example contain information about performed speech acts, or about changes in the common ground of the dialogue participants. In general, any information state theory of dialogue could be utilised here. (c) The IS is presented in a GUI to the user/annotator, who can then edit it. This process is interactive, because at any point the revised information state can be sent back to the processing module in order to compute consequences of the changes. (d) Once the annotator is satisfied with the IS, it is stored, and a new annotation cycle begins with the current IS as part of the *context*.

3 Extending the REDWOODS Treebank

We use RANT, an instantiation of this architecture, to create a treebank annotated with discourse information. The basis of this is the REDWOODS treebank (Oepen et al., 2002), which provides disambiguated parses for approximately 7000 dialogue-utterances from the domain of appointment scheduling. The kind of discourse information we are interested in and for which suggestions are computed by RUDI can be illustrated with the following example. (The h_i in parentheses are the *labels* of the utterances.)

(1) (h_1) A: What is a good time for you?
(h_2) B: After 2pm on Monday...
(h_3) ...and I'm also free on Wednesday.

The IS with which we annotate the utterances consists of two main elements: a logical form for the dialogue in terms of a *discourse structure*, and (parts of) the *model* that satisfies that structure. This discourse structure consists of the following elements: (i) the grouping of the utterances into larger discourse units (e.g., utterances h_2 and h_3 in (1) are grouped together); (ii) rhetorical relations connecting these segments (e.g., h_3 is a *continuation* of h_2, and the resulting segment provides an indirect answer to question h_1); and (iii) resolutions of (some) underspecification in the logical forms, namely that arising from the use of fragments (e.g., h_2 is resolved to something paraphraseable by "After 2pm on Monday is a good time..."), and that arising from the need to bridge definites to the context (e.g., "Wednesday afternoon" in h_3 is resolved to be the next Wednesday afternoon after the time of utterance). These three elements of discourse structure are logically dependent; for example, a particular rhetorical relation can have truth conditional consequences which constrain the values of bridging relations among temporal referents.

The IS also records certain parts of models of the discourse structure, namely the denotations of temporal referents—defined as intervals on the calendar—that make the discourse structure true, given knowledge about when the conversation took place. Finally, it also records the purpose behind each utterance, insofar as the overall goal of finding a time to meet is concerned.

4 Related Work

While we use the system with a particular dialogue theory as background, we expect ANT to be useful for evaluating and developing all kinds of IS-based theories. In contrast to URML and PDTB mentioned above we aim to annotate our corpus with much more detailed semantic information such as goals and resolutions of semantically underspecified information described in Section 3, expecting that this additional information will prove useful for training statistical models.

References

S. Oepen, D. Flickinger, K. Toutanova, and C.D. Manning. 2002. LinGO redwoods. a rich and dynamic treebank for HPSG. In *Proceedings of the 1st Workshop on Treebanks and Linguistic Theories*, pages 139–149, Sozopol, Bulgaria.

D. Reitter and M. Stede. 2003. Step by step: Underspecified markup in incremental rhetorical analysis. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora*, Budapest.

D. Schlangen and A. Lascarides. 2002. Resolving fragments using discourse information. In J. Bos, M.E. Foster, and C. Matheson, editors, *Proceedings of EDILOG 2002*, pages 161–168, Edinburgh, September.

Conveying spatial information in linguistic human-robot interaction

Thora Tenbrink
FB10: Faculty of Linguistics and Literary Sciences
University of Bremen, Germany
tenbrink@sfbtr8.uni-bremen.de

Abstract

An overview is presented of the range of variability employed by speakers with regard to the specificity of information in spatial communication, focussing on the significance of the context. This variability is described in terms of two underlying dimensions, one of which reflects a dichotomy of underdeterminacy and redundancy, while the other concerns vagueness and precision. Relevant results of several HRI experiments are used for exemplification.

1 Introduction

How does the given context influence the richness and unambiguousness of information or detail in spatial communication? While it is well established that redundancy is pervasive in linguistic communication, an impressive amount of literature deals with the ways in which utterances can be underdetermined or vague, establishing that most language is neither precise nor unequivocal. Linguistic spatial communication is specifically problematic. In any discourse, many aspects of context contribute to understanding the processes enhancing or hampering the success of the communicative goals. No matter how much theoretical ambiguity exists for an utterance, in most cases speakers manage to convey enough information for a human listener to infer what is meant. Modelling these processes in order to enable automatic language understanding systems to achieve similar communicative success is a major challenge for research in NLP.

Experiments in human-robot interaction offer an optimal research area in this regard, since the discourse context (including linguistic elements such as the robot's output and the discourse history as well as situational factors such as spatial configurations) can be controlled and modified to an exceptionally high degree. Leaving the test persons ignorant with respect to the research aims, the density and specificity of information conveyance can be tackled on the basis of natural language data where the circumstances of production are specifiable in great detail.

One useful distinction (Pinkal 1985) adopted in the present analysis is that between *vague* utterances that leave room for a *continuous* spectrum of precisification options, and *ambiguous* ones that offer *discrete* options for precisification. While Pinkal addresses these phenomena principally on the lexical level, they can occur on all levels of linguistic communication, such as grammatical, pragmatic, or conceptual.

2 Method

Natural language data were collected in several studies of human-robot interaction of two different kinds. In each study, human users were asked to instruct an unfamiliar robot to perform a task with regard to several objects placed on the floor together with the robot. The tasks to be performed were either to move to one of the objects, or to measure the distance between two of them. The robot variably occupied different positions. The configurations differed with regard to complexity, and while in the 'move' tasks the robot understood specific kinds of utterances (which the users were not told about), the 'measure' tasks were faked. Here, the computer into which the users typed was not connected to the robot, but produced answers in a predetermined fashion. Thus, different kinds of 'dialogues' emerged. Together, the studies con-

stitute a data pool of natural language instances of spatial reference in human-robot interaction scenarios in which (in contrast to most other work in spatial human-robot interaction) the users were not informed about the robot's functionalities.

3 Results and Discussion

Utterances in which some piece of information necessary for identification of the goal object is missing and needs to be inferred are classified as *underdetermined*, while in cases of *redundancy* some information is provided in more than one way. These phenomena occurred in the following areas in our data:

- **Perspective.** Since there are three kinds of possible perspectives, some specification is necessary but mostly not provided by users.
- **Reference systems**. Some spatial expressions are capable of reflecting different kinds of reference systems: 'rechts' can be used in relative and intrinsic reference systems, which in some configurations may be a source for ambiguity.
- **Directions and angles**. Some linguistic expressions (e.g. 'diagonal') presuppose a vector, needing a starting point as well as an end point for interpretation. Utterances containing less information are therefore underdetermined. Utterances indicating an angle but leaving out information about the underlying reference frame are underdetermined, as in "minus 10 Grad" (minus 10 degrees).
- **Configuration and figures.** Some expressions presuppose knowledge about the spatial setting for interpretation. Implicit reference to groups (as in the adjectival use of projective terms) belongs in this category as well as all expressions containing superlatives ('farthest', 'nearest'). Other expressions reflect the conceptualisation of a specific figure, such as a triangle or a square. Utterances containing no explicit definition of such a group or figure are underdetermined in this regard.
- **Linguistic underdeterminacy.** Viewed in isolation, all utterances containing phoric (i.e., anaphoric, cataphoric or exophoric) elements are inherently underdetermined, since they depend on the – situational or textual – context and require the recipient to infer the intended referents. In spatial communication, it is natural to refer (deictically) to the perceived spatial surroundings, and exophoric elements are therefore pervasive in the data. For successful communication it is essential that the interlocutors' perception is compatible or can be matched via the linguistic representations.

Often, users failed to provide information with regard to one of these categories, while giving elaborate and redundant information on a different aspect. This variability may reflect the user's current focus of attention in the given interaction.

On another dimension, instructions vary between *vagueness* and *precision*. Many linguistic expressions specify spatial relations in a 'qualitative' fashion, i.e., do not require the speaker to provide explicit information about quantitative measures. Overwhelmingly, in the data the qualitative information given by spatial expressions was not specified further. Hedges and modifications as well as combinations of several reference systems were used to indicate the speakers' awareness of the vagueness of their utterances, as in "das vordere Objekt etwas links von der Mitte" (the front object a little to the left of the middle). Thus, users attempt to render the instruction more precise by narrowing the range of applicability. This occurred most often in more complex scenarios.

The classic difference between 'continuous' and 'discrete' kinds of specification possible in cases of indefiniteness is reflected in spatial language and can be applied to conceptual as well as linguistic phenomena. Underdetermined utterances contain expressions presupposing elements that need to be inferred by the context, while vague utterances need not be presuppositional, but contain inherently 'fuzzy' expressions that represent spatial phenomena qualitatively rather than quantitatively.

The phenomena identified in our data can be interpreted via the notion of **contrast**: Users' choices on the scales between redundancy and underdeterminacy as well as vagueness and precision reflect their aim at referring to the goal object(s) in a way that is sufficiently distinct to any competing objects, rendering the interpretation highly dependent on a reliable mapping between the user's and the system's conceptualisations of the current scenario.

Reference

Pinkal, M. 1985. Logik und Lexikon: Die Semantik des Unbestimmten. Berlin: de Gruyter.

The BE&E Tutorial Learning Environment (BEETLE)*

Claus Zinn, Johanna D. Moore, Mark G. Core,
Sebastian Varges, and Kaśka Porayska-Pomsta
School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK
[zinn|jmoore|markc|varges|kaska]@cogsci.ed.ac.uk

Abstract

We describe the architecture and the components of BEETLE, a dialogue-enhanced Basic Electricity and Electronics Tutorial Learning Environment.

BEETLE

There is mounting evidence from cognitive science and intelligent tutoring systems research that the most effective tutors are those that prompt students to construct knowledge for themselves. Natural language dialogue offers an ideal medium for eliciting knowledge construction, via techniques such as co-construction of explanations and directed lines of reasoning. These strategies unfold over multiple turns and require a dialogue system to be flexible enough to deal with an unexpected response, interruption, or failure of tactics. To provide these capabilities, we have developed a dialogue management framework, inspired by the three-level architectures used in robotics.

Fig. 1 displays BEETLE's underlying generic and modular architecture for the management of tutorial dialogue. It is divided into four major parts (from left to right): external knowledge sources, the information state, the update engine, and the three-layered planning and execution architecture. BEETLE's architecture emphasises the importance of clearly separating the knowledge sources involved in tutorial dialogue, and thus minimises the re-representation of knowledge in different parts of the tutor system. Moreover, this design emphasises and encourages the reusability of components. Our system is composed of the following modules, which interact using the Open Agent Architecture (Martin et al., 1999):

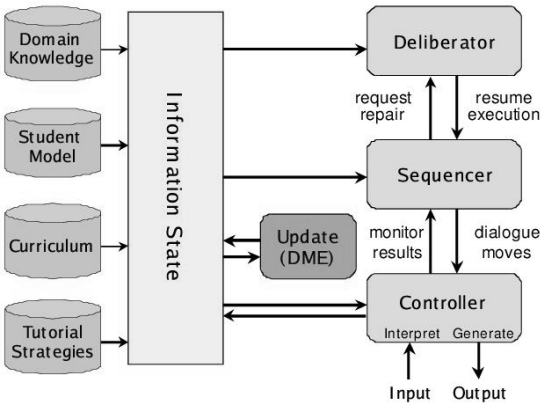


Figure 1: BEETLE's 3-level Architecture.

O-Plan, a deliberative planning and execution monitoring module implemented using the Open Planning Architecture (Currie and Tate, 1991). O-Plan is capable of performing on-the-fly plan repair when tutorial situations occur that have not been anticipated during the planning of the dialogue.

NUBEE, an NLU module which translates the user's typed English input to a logical form. This is built using the CARMEL NLU toolkit (Rosé, 2000), which includes a robust parser and a wide coverage grammar. We augmented CARMEL by

*This research is supported by Grant #N0001402IP20005 from the Office of Naval Research, Cognitive and Neural Sciences Division.

adding a reference resolution module, and an interface to WordNet (Miller, 1990) that can look for known synonyms of unknown words.

BEETLEGEN, a generation module which synthesises English text from logical forms for each system turn. BEETLEGEN is a hybrid generator combining a template-based approach with grammar-based processing within a pipeline of XML document transformations. BEETLEGEN's built-in intelligence includes reasoning about implicit speech-acts, NP and VP pronominalisation, and lexical choice. BEETLEGEN uses standard XSLT stylesheet processors (Clark, 1999).

TIS, a dialogue manager developed using the TRINDIKIT dialogue system shell (Larsson and Traum, 2000). This module maintains the system's blackboard, including the dialogue history, and encodes rules of conversation (*e.g.,* listeners are obliged to address questions). Note that knowledge of how to converse is kept separate from the planner's knowledge of how to tutor, and knowledge of the domain being tutored.

BEER, a Basic Electricity and Electronics Reasoner which encodes all of the system's knowledge about the subject matter to be tutored. BEER is able to compute answers to student questions and to simulate the execution of student lab actions. The domain reasoner is written in LOOM, a description-logic-based knowledge representation and reasoning engine (MacGregor and Bates, 1987).

CURBEE, a curriculum agent which encodes BEETLE's representation of the teaching material. The material is annotated for *learning goals, level of difficulty*, and *allocated time*.

BEESM, a situational modeller which infers relevant information about the immediate tutorial situation based on the student's interaction with the system (time elapsed, amount of material covered, student's answer correctness, *etc*.). BEESM derives values for the *autonomy* and *approval* to be given to the student, which inform (i) the tutorial planning component that selects the next appropriate tutorial strategy, and (ii) the text generator BEETLEGEN that generates the appropriate linguistic realisations.

BEEGLE, a Tcl/Tk-based GUI displaying hypertext lessons, multiple choice questions, a chat interface, and the circuit simulation environment. A screenshot of BEETLE is depicted in Fig. 2.

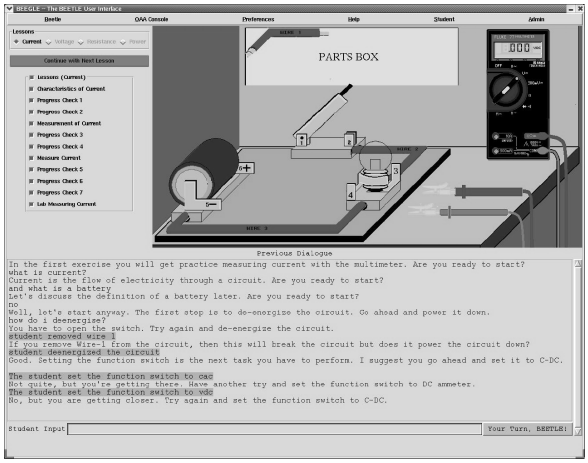


Figure 2: A Screenshot of BEETLE.

Currently, BEETLE has a series of hypertext lessons and multiple choice questions covering the topics of current, voltage, resistance, and power. The practical exercise *measuring current* is supported with natural language tutorial dialogue.

References

J. Clark. 1999. XSL Transformations (XSLT), Version 1.0, W3C Recommendation. http://www.w3.org/TR/xslt.

K. Currie and A. Tate. 1991. O-Plan: the open planning architecture. *Artificial Intelligence*, 52:49–86.

S. Larsson and D. Traum. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6(3–4):323–340.

R. MacGregor and R. Bates. 1987. The Loom knowledge representation language. Technical report, USC Information Sciences Institute, Marina del Rey.

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

George Miller. 1990. WordNet: An on-line lexical database. *Intl. Journal of Lexicography*, 3(4).

Carolyn P. Rosé. 2000. A framework for robust semantic interpretation. In *Proc. of the 1st Annual Meeting of the North American Chapter of the ACL (NAACL-00), Seatle*.