EDILOG 2002

Proceedings of the sixth workshop on the semantics and pragmatics of dialogue

4 – 6 September 2002
The University of Edinburgh

Johan Bos, Mary Ellen Foster and Colin Matheson
Editors

Sponsors

The British Academy

SIRIDUS

M-PIRO

MagiCster

The University of Edinburgh

EDILOG is endorsed by SIGdial.

Castle image courtesy of http://www.edinburghguide.com/.

EDILOG 2002

Proceedings of the sixth workshop on the semantics and pragmatics of dialogue

4–6 September 2002
The University of Edinburgh

Johan Bos, Mary Ellen Foster, and Colin Matheson
Editors

Foreword

EDILOG 2002 is the sixth in a series of workshops that aims to bring together researchers working on the semantics and pragmatics of dialogues in fields including, but not limited to, artificial intelligence, formal semantics and pragmatics, computational linguistics, philosophy, and psychology. The founders of the series were Gerhard Jäger and Anton Benz, who at the time of the first meeting in 1997 were students at the CIS, University of Munich. MunDial 97 was followed by workshops in Twente (Twendial 98), Amsterdam (Amstelogue 99), Gothenburg (Götalog 2000), and Bielefeld (Bidialog 2001).

This year's workshop promises to be the largest yet, with over 60 submitted papers and around 100 registrations. Each paper was reviewed by at least 3 peers, and the resulting 24 accepted papers are printed here along with abstracts for the poster and demo sessions. The editors are very grateful for the swift and thorough reviews which we received from the program committee, and we trust that even the authors of rejected papers found the comments helpful for future work. The members of this year's committee were Johan Bos and Colin Matheson (joint chairs), Ellen Bard, David Beaver, Susan Brennan, Jean Carletta, Robin Cooper, Mark Core, Nissim Francez, Claire Gardent, Jonathan Ginzburg, Joris Hulstijn, Jörn Kreutel, Staffan Larsson, Alex Lascarides, Oliver Lemon, Ian Lewin, Bill Mann, David Milward, Johanna Moore, Manfred Pinkal, Paul Piwek, Massimo Poesio, Rob van der Sandt, Frank Schilder, Keith Stenning, David Traum, Enric Vallduví, Bonnie Webber, and Henk Zeevat.

One of the most difficult problems facing local organisers is the choice of name for the workshop. Given the algorithm of using some form of *dial* or *log* as an affix to part of the name of the host city, various options presented themselves. We rejected DIALBORO on the grounds that there are lots of 'boroughs', and EDDIAL or EDIDIAL for problems of pronunciation (and just plain awkwardness). DIALED looked promising, but we felt that an impracticable diacritic was necessary on the final syllable. Thus we arrived at EDILOG.

We trust that readers who attend the workshop enjoy their visit to Edinburgh, and we are confident that the accepted papers represent an impressive contribution to knowledge and provide a stimulating basis for future discussions. We would like to acknowledge the hard work and enduring cheerfulness of the secretarial staff in the Human Communication Research Centre, in particular Margaret McMillan, and we would also like to thank Mary Ellen Foster for her extremely professional help in preparing these proceedings.

Finally, we would like to thank the various sponsors of the workshop, without whom there would be no meeting: the University of Edinburgh, the British Academy (grant number BCG-34668), and the M-PIRO (IST-1999-10982), SIRIDUS (IST-1999-10516), and MagiCster (IST-1999-29078) projects.

August 2002

Johan Bos and Colin Matheson
EDILOG 2002 Programme Chairs

Programme committee and Referees

Johan Bos and Colin Matheson *(joint chairs)*

Ellen Bard
David Beaver
Jean Carletta
Robin Cooper
Mark Core
Nissim Francez
Claire Gardent
Jonathan Ginzburg
Joris Hulstijn
Jörn Kreutel
Staffan Larsson
Alex Lascarides
Oliver Lemon
Ian Lewin
Bill Mann
David Milward
Johanna Moore
Paul Piwek
Massimo Poesio
Rob van der Sandt
Frank Schilder
Keith Stenning
David Traum
Bonnie Webber
Henk Zeevat

Invited speakers

Susan Brennan	State University of New York at Stony Brook
Stanley Peters	CSLI Stanford
Manfred Pinkal	University of the Saarland
Enric Vallduví	Universitat Pompeu Fabra, Barcelona

Contents

Invited talks

Papers

Posters and demonstrations

Audience Design and Discourse Processes: Do Speakers and Addressees Really Adapt to One Another in Conversation?

Susan E. Brennan
State University of New York at Stony Brook
susan.brennan@sunysb.edu

Audience design is the notion that speakers formulate their utterances in ways that appear to accommodate and benefit addressees. Audience design is often assumed to represent an adaptation made to a specific partner—made *for* that partner's needs. However, as Brown and Dell (1987) pointed out, it is possible that speakers simply make choices that are easiest for themselves, that just happen to benefit addressees. Take the process of referring. When two people in a conversation discuss the same objects repeatedly, they tend to converge on the same terms, which become more efficient and stable upon re-referring (and, at the same time, diverge from terms chosen by other speakers in other conversations about the same objects). Coordinating language behavior with a partner need not require maintaining a detailed model of the partner's knowledge or needs. Speakers could simply continue to use the same referring expressions that worked on the previous occasion, following what Garrod and Anderson (1987) have called an *output-input coordination principle.* Alternatively, they could track and represent certain partner-specific information (Brennan & Clark, 1996). It can be difficult to disentangle these possibilities (Keysar, 1997).

Understanding audience design is hampered by discussions that ignore basic findings from cognitive psychology. People adapt to their partners within the constraints of ordinary cognitive processing, as resources permit (Bard & Aylett, 2000; Horton & Gerrig, in press; Schober & Brennan, in press). I will present evidence about several sorts of potential adaptations—instrument mention (Lockridge & Brennan, in press), interpretation of referring expressions (Metzing & Brennan, 2002), and prosodic disambiguation (Kraljic & Brennan, 2002)—and discuss how these adaptations may emerge.

References

Bard, E.A., & Aylett, M.P. (2000). Accessibility, duration, and modeling the listener in spoken dialogue. *Götelog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue.* Götalog, Sweden: Gothenburg University.

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

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

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

Horton, W. S., & Gerrig, R. J. (In press). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language.*

Kraljic, T., & Brennan, S. E. (2002). Use of prosody and optional words to disambiguate utterances. *Unpublished manuscript.*

Lockridge, C. B., & Brennan, S. E. (In press). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review.*

Metzing, C., & Brennan, S. E. (2002). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Manuscript under review.*

Schober, M. F., & Brennan, S. E. (In press). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes.* Hillsdale, NJ: Lawrence Erlbaum.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 1.

Invited talk: Modeling Dialogue Context for Collaborative Activities

Stanley Peters
CSLI Stanford

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 2.

Invited talk: Semantics, Pragmatics, and Dialogue Applications

Manfred Pinkal
University of the Saarland

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 3.

Information packaging and dialog

Enric Vallduví
Universitat Pompeu Fabra, Barcelona

The idea that information packaging (theme-rheme articulation) has to do with the dynamics of linguistic communicative interaction is an old one. It is present, more or less explicitly, in some of the early work and it is preserved as a defining element in some recent proposals. This view, however, is not universally agreed upon. This talk will attempt to show that information packaging is indeed tied to the dynamics of context change in communicative interaction and that dialog is precisely where the job carried out by the informational organization of utterances is the most visible. A full understanding of how information packaging operates is not possible without a detailed analysis of the everchanging structure of context during communicative interaction.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 4.

Cooperation and Collaboration in Natural Command Language Dialogues

J. Gabriel Amores and **José F. Quesada**
University of Seville
jgabriel@us.es jquesada@cica.es

Abstract

This paper discusses cooperative and collaborative behaviour in Natural Command Language Dialogues (NCLDs). We first introduce Natural Command Language Dialogues and then briefly compare them with other types of dialogue. Cooperation and collaboration in NCLDs is then analyzed. Finally, a typology of conflicts in NCLDs is proposed, for which a solution has been implemented in two spoken dialogue prototypes. Sample dialogues are taken from research carried out under the Siridus and D'Homme European projects.

1 Natural Command Language Dialogues

A Natural Command Language (NCL) is a command language expressed through the medium of natural language. We take NCLs as the set of input and output natural language expressions which are acceptable in a given application domain. This domain is semantically defined by the functions (commands) known by the user and the system, and the natural language vocabulary which may be used to express those commands. In addition, NCLs should contain metalinguistic patterns and expressions typical of human–like interaction. Natural Command Language Dialogues are artificially constructed models of action–oriented dialogues (including knowledge representation and reasoning) able to guide the interaction between the different parts involved in a dialogue based on a NCL. A NCLD should allow the following kinds of phenomena:

- **Multiple Task NCL:** In contrast to Task–Oriented Dialogue models, NCLDs must be able to manage different tasks. Thus, one of the main functions of NCLD systems will be task detection, as will be explained below.
- **Context Dependency:** Only at the dialogue level is it possible to understand anaphora, ellipsis and other context dependent constructions. From the dialogue system design persepective, the treatment of these discourse phenomena will imply the representation and storage of the whole dialogue history. For an illustration, see (Quesada and Amores, 2002) in this volume.
- **Man–Machine Interaction:** An adequate level of naturalness and flexibility in the flow of interactions (restricted to the linguistic limitations of the underlying NCL), relevance and adequacy of the outputs, and consistency (order of arguments in commands) should be accomplished.
- **Interface with External Functional Components:** NCLs are aimed at the specification of commands belonging to a command language. The user's goal is to execute the command(s). Therefore, the dialogue level must not only understand the naturally expressed commands, but also execute them. In fact, this implies the definition and use of another Command Language between the NCLD System and the external functional components in charge of the execution of the commands.

This conceptualization of NCLs has been applied in two spoken dialogue systems under the Siridus and D'Homme European Projects.

1.1 Siridus and D'Homme

In Siridus (Siridus, 1999), we are building an Automatic Telephone Operator System in Span-

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 5-11.

ish jointly with Telefónica I+D. The user should be able to naturally issue commands in order to perform the following tasks: call an extension by name or office; redial a number; transfer incoming calls to another extension or office; cancel transference; set up a conference call; look up an office in the company's directory; and look up an e–mail address in the company's directory.

The benefits of a natural command language system in this domain is evident. For example, functions such as transferring incoming calls, placing conference calls, transferring ongoing calls etc, are typically performed by dialling several codes in a pre–defined sequence. However, since these codes are difficult to remember, users tend not to use them. Automatic dialogue systems seem to be very appropriate for these circumstances since they could allow users to perform some or all of these functions using their voice and natural language. Similar systems have been built by Bell Labs, AT&T and Wildfire Communications. For a comparison with our system, and a discussion about the benefits of automatic telephone systems, see (Torre and others, 2001)

In the D'Homme project (DHomme, 2001), the implemented system was able to perform the following functions in a home environment: switch on/off a device or set of devices; dim or bright a device or set of devices; and consult the state or location of a device or set of devices. A spoken interface offers several benefits to users, especially the elderly and disabled: we can query multiple devices with a single command, we can control devices, we can also use language to program devices to interact with each other, etc. Remote control via the telephone is a natural extension to home deployed systems. We can phone up the house to see if we forgot to turn off some device, or we could ask our house to turn on the heating on our way home after some days of absence.

In addition to these domains, there are some other dialogue systems implementing some version of a command language, such as the WITAS system (Doherty and others, 2000).

2 Comparing NCLDs with other types of dialogue

Before we analyze the cooperative and collaborative behaviour, it is worth pointing out some characteristics of NCLDs.

As a consequence of their own nature, NCLDs involve just two participants: the user and the machine. One of the first decisions concerns whether we should model the participants' internal beliefs, or more external aspects of the dialogue. Since the goal of this type of dialogues is that the user have control over the execution of one or more commands by the machine, most dialogues exhibit a marked functional or operational tendency, i.e. are action–oriented dialogues geared toward the execution of some non–communicative action. So, it seems reasonable to focus our model on the external aspects of dialogue. That is, it should be based more on what was said than what was in the minds of the participants when their interactions were produced.

NCLDs are different from other types of dialogue such as Information Seeking and Negotiative Dialogues. In Information Seeking dialogues, one participant requests information from the other. In Negotiative Dialogues, the goal of both participants is to come to an agreement about some conflict of interests.

Another aspect which differentiates NCLDs from Information Seeking Dialogues is the dynamic nature of the knowledge bases involved. In an information seeking dialogue, in which the system as a whole is viewed by the user as a repository of knowledge, knowledge bases have a clear static character from the point of view of the user. That is, the data in these resources may be updated, but not by the user during its interaction with the system. On the contrary, one of the main features of NCLDs is the presence of a command execution which is capable of dynamically modifying the contents of external resources to the dialogue, such as the knowledge bases associated to the domain.

2.1 Functional Embedding

An important aspect of NCLDs is that they usually exhibit *functional embedding* (Walton and Krabbe, 1995; Reed and Long, 1997). Functional embeddings occur when the goal of a subdialogue shifts to another dialogue type.

As pointed out above, task identification subdialogues are possible in NCLDs. As opposed to other dialogue types in which the overall task of the dialogue is predefined (seeking information about flights in a travel agency, etc.), the system in a NCLD does not know *a priori* which function the user desires to perform. The first task, therefore, involves identifying the speaker's intentions. Once the task has been identified, the corresponding plan may be unfolded. Task identification may be considered a sort of *deliberation* subdialogue in the sense of Walton and Krabbe (1995). In deliberation dialogues, the participants jointly aim to reach a decision or form a plan of action. Deliberation is goal–directed, in contrast to the more abstract reasoning characteristic of negotiation.

Clarification subdialogues are also common in NCLDs, when the name of the destination of a call or some other parameter required for the successful completion of a command was misunderstood or missing. If some portion of the desired action was understood, it will be used as indirect feedback to the user:

- **U(1)**: Please, transfer my calls to Meeting room E–30
  (function was understood, but not the destination)
- **S(1)**: Could you repeat where I should transfer your calls?

Clarifications may be viewed as negotiation stages within the overall dialogue. Another instance of negotiation in NCLDs occurs when the system proposes alternatives in cases of conflict. This will be discussed in more detail below.

Information seeking subdialogues are common in NCLDs when the user wishes to know the state of specific devices, or consult the details (e–mail, office number) of another user, as pointed out above.

Figure 1 below shows a possible cascade of functional embedding in NCLDs.

As has become apparent in this section, NCLDs exhibit much more complexity than one could originally envisage given the apparent simplicity of the task at hand.

3 Cooperative behaviour in NCLDs

Let us now turn to the question of cooperative behaviour.

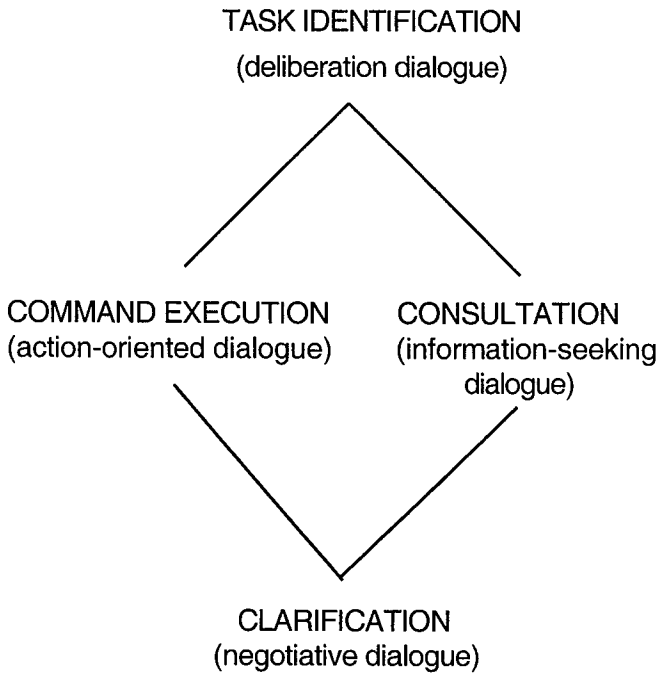


Figure 1: Cascade of functional embeddings in NCLDs

The linguistic definition of cooperation was first proposed by Grice (1975). Reed and Long (1997) propose a notion of cooperation which is similar to that of Grice, but acts at a higher level. Gricean cooperation is speaker–centered, whereas in their view, cooperation is more objectively discourse–centered. In their opinion all dialogue is inherently cooperative since any instance of true dialogue involves the participants accepting a common goal and working towards that goal within a given set of rules. This concept is also similar to the *Conversational Contract* of Fraser (1990).

Another relevant aspect regarding cooperation is that proposed by Allwood (1976). According to Allwood, two or more parties interact cooperatively to the extent that they:

1. take each other into cognitive consideration,
2. have a joint purpose,
3. take each other into ethical consideration,
4. trust each other to act in accordance with 1–3.

Of these, the most relevant requirement as regards NCLDs is 4 since, in fact, one of the

agents (the machine) in NCLDs is just *pretending* to act according to the principles. This idea reinforces our previous analysis of NCLDs as an inherently cooperative activity since, as long as the agents trust each other to act according to a set of principles, communication will be cooperative.

However, there is a simple fact about NCLDs which may change our perspective. Namely, the system in NCLD applications is usually *dispensable*. Its role is to make our lives easier. If it fails to achieve that goal, the user may just ignore it and proceed to perform the desired function in the ordinary way (by pressing a sequence of digits in the telephone pad, using light switches, etc.).

In this kind of scenario the system should be as unobtrusive as possible, trying to avoid a situation of frustration on the other participant which would indeed lead to interrupting any communication between them. So, what is usually modeled in NCLD systems is a *collaborative* behaviour in which the participants are in some sense working together to reach the desired outcome. The presence or absence of conflict will determine the level of collaboration in a dialogue. The next section describes the types of conflict which may arise in NCLDs and the expected level of collaboration which should be incorporated.

4 Sources of conflict in NCLDs

Conflict arises from different sources in NCLDs.

First, the user may just not know exactly what functionality is available, or the parameters which they require. Providing help is one way of solving this conflict. Two *Dialogue Moves* take care of providing help in our systems. The first one provides general help on *what* commands are available. The second instructs the user about *how* to carry out a specific command, i.e. what parameters are required.

Second, given the dynamic nature of the domains in which we are working, one of the agents (the user) may just ignore the exact state of the world at that precise moment. Requesting information is then a collaborative activity which the system should model. Several dialogue moves have also been incorporated to take care of this kind of interaction.

A third type of conflict arises when one of the agents wrongly believes that a specific goal is possible, but the state of the world does not permit its accomplishment. Solving this conflict involves proposing alternatives, as the following dialogue illustrates:

- **U(1)**: Call Mr Fernández
- **S(1)**: This extension does not respond. Do you want to leave a message or shall I try his mobile?

A more cooperative scenario is proposed in the following example:

- **U(1)**: Call Mr Fernández at extension 311 (Extension and name do not match in the database)
- **S(1a)**: Mr Fernández does not have this extension; shall I put you through his current extension?
  (or)
- **S(1a)**: Extension 311 belongs to Mr González; do you want him instead?

In some other contexts, the alternatives proposed by the system may be based on what the user *usually* does, that is, her most frequent destinations or actions in the past. Behaving cooperatively in this context occurs, for example, when the user underspecifies a destination of a call by a last name, and that last name turns to be ambiguous in the telephone directory. In such an event, the system may consult the most common destinations (chosen in the past by the user) and propose one who matches the last name provided.

- **U(1)**: Call Fernández
  (after spotting a frequent Pedro Fernández in the user's profile)
- **S(1)**: Do you mean Pedro Fernández?
- **U(2)**: Yes, please

This form of user adaptation is related to advanced modes of cooperation to be discussed in section 5 below.

Reaccommodation may be seen as a kind of cooperative behaviour as well since it involves solving an elliptical utterance. Question/Issue Reaccommodation is defined by Larsson (2002,

162) as "adjustments of common ground required to understand an utterance addressing an issue which has been resolved and does not influence any other resolved issue, influences another resolved issue or concerns grounding of a previous utterance." Two instances of reaccommodation have been implemented in our system. The first example exemplifies a case in which there is a previous history in the dialogue, and the user issues an incomplete command.

- **U(1)**: Switch on the light in the kitchen (the light in the kitchen goes on)
- **U(2)**: Now the one in the living room

A more complex type of reaccommodation occurs when, after a function has been executed, the user retracts and wishes to undo the previous action:

- **U(1)**: Switch on all indoor lights (all indoor lights go on)
- **U(2)**: No, the one in the garage (the system must undo the previous action first)

The last (and more productive) source of conflict arises from misunderstandings stemming from misrecognition. In our systems, some degree of cooperative behaviour has been implemented, depending on the consequences of adopting an aggressive attitude or whether explicit confirmation was seen as necessary.

- **U(1)**: I want to call 1 2 3 4 collect
- **S(1)**: Do you really want to call 1 2 3 5 collect?
- **U(2)**: no, 1 2 3 4.
- **S(2)**: OK, shall I call 1 2 3 4 collect?
- **U(3)**: yes, please.

The table below summarizes the types of conflict which may arise in NCLDs, the related Dilaogue Move and the associated cooperative/collaborative action proposed for each case.

5 Advanced Cooperation

This section outlines some modes of advanced cooperation in NCLDs. Some of them may be considered cases of advanced flexibility or even robustness, which, in some sense, may be viewed as other aspects of cooperative behaviour. This functionality has not been fully implemented in our systems yet, partly because it relies on the technical limitations of the specific hardware being used.

- **User profiling:** In the automatic telephone domain, the command *Call my wife at her mobile* is only possible if a personal directory for each person making use of the system is made available, in addition to the general one in which all the personnel is stored.

  In addition, the system might allow the user to set *modes* of behaviour. For example, in the home domain, it could be possible to issue the command *set night mode*, and automatically the system would perform a series of tasks such as setting some external lights on, all indoor lights off, switching the alarm on, etc.

  Similarly, in the telephone domain, the user might want to set a *non–disturb* mode, and the system should transfer all incoming calls to the answering machine or to the secretary.

- **Default** reasoning may be considered a form of cooperative behaviour, at least in the telephone scenario. In our system a default action has been specified, whereby if no other action has been recorded in the dialogue history, a *PhoneCall* action will be executed. This is useful in the (frequent) cases in which the user just picks up the telephone and utters,

  – **U(1)**: Carlos García, please

- **Proactive Behaviour** is the most extreme case of cooperativeness. In the home environment, proactive behaviour has already been proposed for cases of fire, gas or water alarm, but they rely more on the technical capabilities of the specific setting than on dialogue capabilities.

  In the telephone scenario, however, a proactive cooperative behaviour stemming

Type of Conflict	Related Dialogue Move	Proposed Action
user ignores overall functionality	askHelp	provide general help
user ignores how to carry out a specific command	askHelp	provide specific help
user ignores current state of the world	requestQuantity request Exist requestLocation requestDevice requestState	provide requested information
desired command cannot be accomplished	informExecution	Propose alternatives
elliptical command	errorRecovery	Try reaccommodation
Command and/or parameter misunderstanding	askRepeat askConfirmation	Clarification subdialogue

Table 1: Types of conflict in NCLDs and cooperative action proposed

from the dialogue occurs when the destination is busy, and the system proposes to trigger a call–back when she is done:

- **S(1)**: Calling Juan Fernández ...
- **S(2)**: Mr. Fernández is busy. Shall I have him call you back when he is done?
- **U(1)**: Yes, please.

6 DISC Guidelines

Finally, let us briefly examine to what extent our systems comply with the DISC guidelines for cooperative dialogue (Consortium, 1999). In particular, we will focus on the *specific guidelines* proposed.

- **SG1** *Summarising feedback*: is achieved through direct or indirect confirmation *Shall I transfer your calls to 0 1 2 3?*
- **SG2** *Provide immediate feedback*: in some tasks, such as conference calls, it is better to confirm one destination party at a time, given the ambiguities that would result from the combinations of first names, last names, second last names, etc.
- **SG3** *Ensure uniformity*: does not apply in the home environment since no linguistic feedback is generated for most command. In the telephone scenario this is achieved through 'canned expressions' in the synthesizer.
- **SG4** *State your capabilities*: is achieved through general help. A non recognized command will only turn into *I cannot do that* or *that's beyond my current capabilities* if the expression was understood as a command, and this command is not in the set of those supported by the current system. Otherwise, misinterpretation will arise.
- **SG5** *State how to interact*: is achieved through specific help.
- **SG6** *Be aware of user inferences*: is achieved through profiling behaviour.
- **SG7** *Adapt to target group, novice/expert users*: some version of this behaviour may be achieved in the telephone scenario since the recognizer may take input even before the system finished to provide instructions.
- **SG8** *Cover the domain*: some degree of flexibility is achieved trhough advanced cooperative functions such as user profiling.
- **SG9** *Enable system repair*: yes.
- **SG10** *Enable inconsistency clarifications*: yes, for example in the home domain, as in *The light in the bathroom is already on.*

- **SG11** *Enable ambiguity clarification*: yes, through anaphora resolution, default reasoning and user profiling.

7 Conclusion

This paper has analyzed the complexity of NCLDs, and the different perspectives from which such dialogues may be cooperative and collaborative. In particular, several types of conflict have been identified, for which an adequate solution has been implemented in two spoken dialogue prototype systems under the D'Homme and Siridus projects. As we have shown, the systems comply with most of the DISC guidelines for cooperative dialogue.

References

Jens Allwood. 1976. Linguistic communication as action and cooperation. Gothenburg Monographs in Linguistics 2, Department of Linguistics, University of Göteborg.

The DISC Consortium. 1999. Disc dialogue engineering model. Technical report, DISC, http://www.disc2.dk/slds/.

DHomme. 2001. http://www.ling.gu.se/ projekt/dhomme/.

Patrick Doherty et al. 2000. The witas unmanned aerial vehicle project. In W. Horn, editor, *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000)*, pages 747–755.

Bruce Fraser. 1990. Perspectives on politeness. *Journal of Pragmatics*, 14(2):219–236.

Herbert Paul Grice. 1975. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Speech Acts*, volume 3 of *Syntax and Semantics*, pages 41–58. Seminar Press, New York.

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

José F. Quesada and J. Gabriel Amores. 2002. Knowledge–based reference resolution for dialogue management in a home domain environment. In *Proceedings of Edilog*, Edinburgh.

Chris A. Reed and Derek P. Long. 1997. Collaboration, cooperation and dialogue classification. In K. Jokinen, editor, *Working Notes of the IJCAI97 Workshop on Collaboration, Cooperation and Conflict in Dialogue Systems, IJCAI 97*, pages 73–78, Nagoya, Japan.

Siridus. 1999. http://www.ling.gu.se/projekt/ siridus/.

Doroteo Torre et al. 2001. User requirements on a natural command language dialogue system. Deliverable 3.1, Siridus Project.

Douglas N. Walton and Erik C. W. Krabbe. 1995. *Commitment in Dialogue*. State University of New York, New York.

Referential form and word duration in video-mediated and face-to-face dialogues

Anne H. Anderson
Multimedia Communications Group
Department of Psychology
58 Hillhead Street
University of Glasgow
Glasgow, G12 8QB
anne@mcg.gla.ac.uk

Barbara Howarth
Multimedia Communications Group
Department of Psychology
58 Hillhead Street
University of Glasgow
Glasgow, G12 8QB
barbara@mcg.gla.ac.uk

Abstract

It is widely believed that speakers adapt their speech to meet the comprehension needs of their listener. Yet recent work in face-to-face communication has shown that word articulation and the design of referring expressions are insensitive to certain aspects of listener knowledge (Bard et al., 2000; Bard & Aylett, 2001). Word articulation and referential form were investigated in two studies. Study 1 examined the duration of words forming the names of landmarks on a map in video-mediated dialogues. Study 2 explored the impact of cognitive load (as illustrated by time pressure) on word duration and referential form in both face-to-face and video-mediated communicative contexts. It was found that second mentions of words were articulated more quickly than first mentions regardless of which interlocutor introduced the names of the landmarks (study one). Study two showed that speakers responded to time pressure by shortening the names of landmarks on a map but only when the task was video-mediated. Speakers did not, however, resist the pressure to shorten second mentions of words relative to first mentions under time pressure. The implications of these findings are discussed in terms of the Dual Process Model of speech processing in spoken dialogue proposed by Bard et al.,(2000).

1 Introduction

Spoken dialogue presents a particularly challenging area of research since speakers must produce utterances in real-time, interact with another person, and cope with the demands of the specific task. Nowadays the onslaught of computer-based technologies, such as videoconferencing systems, provides novel communicative contexts in which a dialogue may take place. It is possible, that communicating in such a setting places additional demands on the speaker. This raises the question of whether spoken output will be affected by communicative context in which a dialogue takes place.

2.1 Variation in spoken output in face-to-face communication

In face-to-face communication, referring expressions have been shown to vary in terms of the way expressions referring to entities within a discourse are chosen and in terms of the way words forming referring expressions are articulated. Studies have demonstrated that the articulatory quality of words varies with factors such as sentence context (Hunnicutt, 1985; Lieberman, 1963) and repeated mention (Fowler & Housum, 1987; Robertson & Kirsner, 2000). Fowler and Housum (1987), for example, demonstrated that repeated, or second mentions, of word tokens were articulated more quickly and less clearly than introductory mentions of words. Mere repetition was not enough to induce the effect. Rather, both words had to refer to the same entity (Fowler, 1988).

With respect to referential form, it has been shown that speakers vary the descriptive content of referring expressions depending on factors such as; assumed listener knowledge (Fussell & Krauss, 1992; Isaacs & Clark, 1987); the typicality of an object (Dell & Brown, 1991); and the cognitive load associated with a given task (Horton & Keysar, 1996; Rosnagel, 2000). Generally speaking referring expressions tend to be shorter and less explicit if the speaker can assume that; the listener is familiar with the object of conversation, the object being referred to is typical given the context, or the cognitive load associated with the task is increased. Cognitive load can be thought of as the "mental

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 13-20.

energy" required to process a given amount of information (Sweller, 1988) and can be increased by imposing a time pressure or increasing the difficulty of a task. Rosnagel (2000), for instance, showed that speakers used shorter descriptions of component parts of a model such as a technical term alone (rather than a term plus a supplementary description) when the difficulty of the task was increased.

Traditionally, differences observed in articulatory quality and referential form have been interpreted in terms of the belief that speakers adapt their speech in order to produce utterances that will be comprehensible to the listener (Clark & Clark, 1977). The finding that speakers varied referring expressions depending on what the listener could be assumed to know (Fussell & Krauss, 1992; Isaacs & Clark, 1987) has been taken as evidence in support of this claim. The use of shorter, less explicit expressions under conditions of high cognitive load has also been interpreted in terms of adaptation to the listener. Rosnagel (2000) suggests that choosing a referential expression must involve making a complex assessment of how readily a listener will be able to interpret a given referring expression. The use of less explicit referring expressions under conditions of high cognitive load is taken as an indication that the speaker has sacrificed adjustment to the listener's perspective.

The view that speaker's adapt their speech to the comprehension needs of the listener has also been applied to the articulatory shortening of second mentions of words relative to first. Fowler and Housum (1987) showed that listeners could make use of differences in articulatory quality to distinguish *old*, or *given*, information from "*new*" information. *Given* information has been described as that "which the speaker assumes is already in the addressee's consciousness" (Chafe, 1974). It has been argued that if a word is partially specified by the Given status of the entity to which it refers, the speaker may assume that the listener would not require as clear a signal. Thus words referring to given entities (by virtue of previous mention, for instance) are articulated more quickly and less clearly than words referring to new entities.

Several researchers have questioned the position that variation in spoken output is tailored to the comprehension needs of the listener. Dell and Brown (1991), for example, reasoned that certain within-clause structures could be adequately accounted for in terms of generic language processing mechanisms. Horton and Keysar (1996) suggest that certain stages of speech production, such as initial planning, may be dependent on the speaker's own knowledge and that models of what the listener can be assumed to know are implicated in the later stages as part of a correction mechanism. Bard et al., (2000) showed that speakers failed to mitigate the attenuation of repeated mentions of words in cases where the entity being referred to could not be taken as given by the listener. If speakers adapted their speech to the listener, one would expect the speaker to resist the pressure to attenuate repeated mentions of words in such cases.

2.2 The Dual Process Model of speech production in spoken dialogue

In order to account for their results Bard et al., (2000) proposed a Dual Process Model of speech production. The model is based on that proposed by Dell and Brown (1991); (Brown & Dell, 1987) and holds that two basic types of processes underpin speech production in spoken dialogue. On the one hand articulatory effects, such as the attenuated articulation of repeated mentions of words relative to introductory mentions of words, are attributed to priming processes which are fast, automatic and dependent on the sole experience of the speaker. Priming can be triggered by given status and result in faster articulation and reduced articulatory detail. The claim that *givenness* per se triggers priming, would account for the observation that introductory mentions of words that can be taken as given by virtue of physical presence tend to be articulated more quickly and less clearly than mentions referring to new entities (Bard & Anderson, 1994). This would also explain why second mentions of words were attenuated relative to first mentions, regardless of whether both mentions were uttered by the same speaker or by different

speaker. Since priming is triggered by given status it is unimportant which interlocutor introduces an item into the dialogue.

In contrast to priming processes, complex processes, such as decision making and maintaining models of the listener, are slower and are subject to the demands on the speaker's time and attention. If the demands on the speaker's attention are high, or time is short, there may not be sufficient time to assimilate feedback from the listener or interpret what the listener can be assumed to know. These type of processes offer an account of why a speaker's control of the word articulation appears to be insensitive to these aspects of listener knowledge (Bard et al, 2000; Bard and Aylett, 2001). The processes involved in interpreting feedback or drawing inferences about what the listener knows are deemed too slow to precede articulatory production on a word-to-word basis. Consequently, priming is not mitigated in these circumstances.

The Dual Process Model predicts articulatory control and the design of referring expressions may function differently in spoken dialogue. In support of this claim, Bard and Aylett (2001) showed that both word articulation and referential form were insensitive to subtle aspects of listener knowledge such as indications about what the listener could or could not see. However, referential form was more sensitive than word duration to gross aspects of listener knowledge, such as listener identity, and to aspects of speaker knowledge, such as what the speaker could or could not see. Thus the design of referring expressions did not appear to be based on the same criteria as word articulation. More specifically, articulation must be designed without regard to costly information such as listener-knowledge. The design of referential form, on the other hand, may be sensitive to such knowledge depending on the demands on the speaker and the time available.

3 Video-mediated and face-to-face communication

Much of the research in video-mediated communication has tended to focus on aspects of communication such as dialogue length, turn-taking and interrupting. Thus it is not clear how speakers will articulate words or choose referring expressions in a video-mediated context compared with a face-to-face context. There are, however, certain features associated with video-mediated communication which suggest that speakers may behave differently in this context. For example, it has been suggested that participants may view the technology as novel (Doherty-Sneddon et al., 1997). While (Clark, 1996) suggests that certain conversational settings may restrict the language used if the participants are not co-present. Social presence theorists suggest that the remoteness of the communicative situation may impose a sense of social distance, or lack of salience of the other person and their common surroundings (Short, Williams, & Christie, 1976).

The aim of the paper is to explore the influence of communicative context (video-mediated and face-to-face) and cognitive load on speech production in spoken dialogue.

4. Studies of word duration and referential form in video-mediated and face-to-face dialogues

An implication of the Dual Process Model (Bard et al., 2000) is that the formulation of referring expressions does not appear to be based on the same criteria as the control of word articulation. The model suggests that if time is short, there may be insufficient time to access and respond to information relating to the communicative context. Thus one might expect the referring expressions to be more sensitive to communicative context under time pressure than word articulation. In order to answer these questions we conducted two studies which examined word duration and referential form in video-mediated and face-to-face dialogues.

4.1 Study One: word duration in video-mediated dialogues

4.1.1 Method

Design

In order to address the question of whether video-mediated communication would function in the same way as face-to-face communication, we conducted an initial study to examine the

duration of words forming referring expressions in video-mediated dialogues. If the articulation of repeated mentions of words is unaffected communicative context, we would expect the same pattern of results observed in face-to-face communication. Namely, that repeated mentions of word tokens would be shorter in duration than introductory mentions regardless of which speaker introduced the entity into the discourse (Bard et al., 2000). To do this, we compared first and second mentions of words forming the names of landmarks on a map in same- and different-speaker-repetition conditions. In the *same-speaker* condition, both first and second mentions were uttered by the same speaker. In the *different-speaker* condition, first and second mentions were uttered by different speakers.

Materials

The materials for this study were drawn from 48 dialogues recorded as part of a previous study (Anderson et al., 1999) in which two participants performed a collaborative problem-solving task (The Map Task; (Brown, Anderson, Yule, & Shillcock, 1984) in a video-mediated communicative context. The 48 participants were students of the University of Glasgow and the University of Nottingham. They performed two versions of the Map Task, changing roles as Instruction Giver and Instruction Follower. The participants were based at Nottingham and Glasgow and the connection between the two cities was made via the SuperJANET ATM network.

Procedure

Speaker-per-channel recordings of the dialogues were made onto a computer at a sampling rate of 16 kHz. The duration of 649 word pairs words forming part of the same name of landmarks on the map was measured in milliseconds using speech analysis software. The beginning and end of words was determined by examining waveform and spectral representations of the words using the *Syntrillium* waveform editor *Cool Edit*

4.1.2 Results

A two-way ANOVA was carried out on the data with Mention treated as a within subjects variable and Speaker Repetition treated as a between subjects variable. Table 1 shows mean word duration for repeated mentions of word tokens uttered by the same and different speakers.

Table 1: Mean word duration (in milliseconds) for introductory and repeated mentions of lexical items uttered by the same and different speakers

Speaker repetition	**Mention**	
	1st mention	2nd mention
Same	378	342
Different	364	330

There was a main effect of mention [$F1(1,78) = 22.98$, $p < 0.01$; $F2(1,647) = 88.942$, < 0.01] with no effect of speaker-repetition and no interaction. Second mentions of word tokens were articulated more quickly than first mentions regardless of whether repeated mentions of words tokens were uttered by the same speaker or by a different speaker.

4.1.3 Discussion

The pattern of results observed in study one was similar to that observed by Bard et al., (2000). This suggests that, in terms of articulatory quality at least, video-mediated communication functions in the same way as face-to-face communication despite any sense of novelty or social distance that might be associated with this medium.

4.2 Study 2: Word duration and referential form in video-mediated and face-to-face dialogues

4.2.1 Method

Design

The purpose of Study Two was to explore the effect of cognitive load and communicative context on word duration and referential form through a comparison of face-to-face and video-mediated dialogues. Cognitive load was manipulated by imposing a time pressure. Thus a timed condition, where a three-minute time limit was imposed on the task, was compared with an un-timed condition, in which

participants performed the task in their own time. The design was counterbalanced for task order and map, but task role was randomly assigned.

If a three-minute time limit was sufficient to put participants under pressure, timed dialogues should be shorter in duration, contain fewer words and lead to less accurate task performance. Hence, dialogue length (in terms of words and duration) and task accuracy were also examined.

The Dual Process Model predicts that the formulation of referring expressions should be more sensitive to demands on the speaker's time and attention than word articulation. It was expected, therefore, that increased cognitive load would have a greater influence on referential form than word duration. Factors associated with video-mediated communication, such as social distance and novelty, may restrict communication and this may place extra demands on the speaker. Consequently, any effect of cognitive load would be expected to be more pronounced in video-mediated dialogues than in face-to-face dialogues.

Participants

The participants in this study were all undergraduate students of the University of Glasgow and were native speakers of English.

Task

Two electronic versions of the Map Task (Brown et al., 1984) were created using Adobe Photoshop. Printed versions of the maps were used in the face-to-face conditions.

Procedure

Participants were assigned to either a face-to-face or a video-mediated setting and performed the Map Task in timed and un-timed conditions. In the timed condition, participants were interrupted at minute intervals and told that; there were two minutes to go, one minute to go, and the time was up. In the un-timed condition participants completed the task in their own time. Tape recordings of the dialogues were made and, of the 82 dialogues recorded, 64 were retained for analysis. These were transcribed and recorded onto a computer at a sampling rate of 16kHz.

Technical set-up

In the video-mediated condition, participants were located in different rooms and communicated via a desktop videoconferencing system on which the Map Task and a video-window of the other participant were displayed. The audio and visual signals were delivered in synchronisation. Full duplex audio was used and the video image was refreshed at a rate of 25 frames per second. In the face-to-face condition, audio recordings of the dialogues were made using the same equipment to record the video-mediated dialogues. Participants sat opposite each other and were unable to see each other's maps.

4.2.2 Dependent variables

Dialogue length and task accuracy

A measure of task accuracy was obtained by comparing the original route on the map with that reproduced by the Instruction Follower. The area between the two routes provides a measure of the degree to which the Instruction Followers deviate from the route. The duration of the dialogues was measured in seconds and the number of words (including interjections) were counted.

Referential Form

First and second references to landmarks were coded according to the scheme in Table 2 where "0" indicates no shortening from the name printed on the map and "2" indicates the greatest degree of shortening.

Table 2: Shortening scale for referring expressions

Code	Referential form	Example
0	full landmark name	*Do you have a popular tourist spot?*
1	truncated name	*I have a tourist spot, yes*
2	pronoun	*Ok go right the way round it*

There were 774 first-references to landmarks and 631 second-references to landmarks but these were analysed separately. Although functionally distinct, the coding scheme employed here is conceptually similar to that used by Bard and Aylett (2001).

Durational measurements

Duration measurements (in milliseconds) of first and second mentions of words forming the names of landmarks on the map were made from waveform and spectral displays using the *Syntrillium* waveform editor, *Cool Edit.*

4.2.3 Results

Cognitive load response

Two-way mixed ANOVAs were carried out on the dialogue data (N = 32). As expected, timed dialogues were significantly shorter in duration [$F(1,30) = 29.25$, $p < 0.001$] and contained fewer words [$F(1,30) = 26.62$, $p < 0.001$] than un-timed dialogues. Furthermore, route deviation was significantly greater in the timed dialogues than in the un-timed dialogues [$F(1,30) = 10.53$, $p < 0.01$]. These results indicated that subjects did, in fact, respond to increased cognitive load. Under time pressure, dialogues were found to be significantly shorter (in words and duration) and produce less accurate performance.

Referential Form

A two-way, independent measures ANOVA was carried out on second-reference referential forms (N = 631). There was no effect of cognitive load or communicative context, but there was an interaction approaching significance [$F(1,627) = 3.18$, $p = 0.07$]. Analysis of simple effects showed an effect of cognitive load in the video-mediated group [$F(1,627) = 5.47$, $p = 0.02$] but no effect of cognitive load in the face-to-face group ($F < 1$). Referential coding scores were significantly higher in the timed condition than the un-timed condition but only for the video-mediated group. These results suggest that increased cognitive load influenced the formulation of referring expressions in video-mediated communication but not face-to-face communication. Time pressure lead to an increased tendency to use more shortened referring expressions (such as pronouns), but only when the task was video-mediated.

A separate, two-way ANOVA of first-reference referential forms (N = 774) showed an overall effect of cognitive load [$F(1,770) = 12.94$, $p < 0.001$], no main effect of communicative context and no significant interaction. Analysis of simple effects revealed that referential shortening scores were significantly higher in the video-mediated, timed condition than in the video-mediated, un-timed condition [$F(1,770) = 11.24$, $p < 0.01$]. There was no significant difference, however, between timed and un-timed referential coding scores in the face-to-face dialogues. Thus, as with second references to landmark names, time pressure lead to an increased tendency to use more shortened referring expressions (such as pronouns) but only when the task was video-mediated.

Word duration

A three-way ANOVA was carried out on the data with Communicative Context as a between-subjects variable, and Cognitive Load and Mention as within-subjects variables. There were 24 speakers in total with 12 speakers contributing to each cell of the design. As expected, there was a main effect of mention [$F(1,22) = 20.47$, $p < 0.001$]. Second mentions of word tokens were shorter in duration than first mentions. Interestingly, there was a main effect of communicative context [$F(1,22) = 6.11$, $p = 0.02$]. Word tokens in the video-mediated group were articulated more slowly than word tokens in the face-to-face group. There was no effect of cognitive load. Nor were there any significant interactions.

4.2.3 Discussion

Although the speakers articulated words more slowly overall when communicating in a video-mediated context, the communicative context did not lead speakers to resist the pressure to articulate second mentions of words more quickly than first mentions.

5 Conclusion

The purpose of this paper was to explore the effect of communicative context and cognitive load on two aspects of speech production in spoken dialogue, namely, word duration and referential form. Taken together, the results reported here suggest that communicative context does influence speech production in spoken dialogue, but that word duration and referential form are affected in different ways. First, video-mediated communication seems to function in the same way as face-to-face

communication, insofar as speakers reliably reduce the duration of second mentions of word tokens relative to first, in spite of any sense of social distance or novelty that might be associated with this medium. Second, cognitive load, as illustrated by time pressure, did not offset the pressure to reduce repeated mentions of words in either a face-to-face or video-mediated context.

Spoken language did differ between contexts, however, in terms of the influence of cognitive load on referential form and in terms of word articulation in general. Overall, speakers articulated words more slowly in a video-mediated context. Second, when under time pressure, there was a tendency to use a greater number of short referring expressions (such as pronouns), rather than use the name that was given on the map. The tendency only occurred when the task was video-mediated. No evidence was found to indicate an effect of cognitive load in face-to-face communication.

The results of this study are broadly consistent with the Dual Process Model (Bard et al., 2000). The shortening of second mentions of word tokens relative to first mentions can be accounted for in terms of priming processes. Since priming processes are fast and automatic they are deemed too fast for higher level factors, such as a response to the communicative context to make their influence felt or to precede every attempt at articulation. The design of referential form, on the other hand, is thought to be governed by more complex processes which are planned in longer somewhat slower cycles. This allows time for the effects of communicative context to take effect.

Why would cognitive load appear to influence referential form in a video-mediated communicative context while no influence of cognitive load was found in a face-to-face communicative setting? Although it is not clear how the slower articulation of words in a video-mediated context would be accounted for within the framework of the Dual Process Model, (Blokland & Anderson, 1998) suggest that if communication is not smooth, then speakers may compensate by hyperarticulation. Following the same line of reasoning, it could be the case that speakers make a conscious decision to articulate words more clearly if they experience some kind of difficulty communicating in a video-mediated context. Possibly, the technology itself imposes some kind of communicative stress. Following the rationale of Bard et al., (2000), this could place extra demands on the speech production system to extent that certain processes involved in naming a landmark may not be run. The overall conclusion of these studies however, provide evidence that word articulation and the formulation of referring expressions are influenced by different aspects of dialogue. The Dual Process Model seems to give a plausible explanation of the data reported here.

6 Acknowledgement

This work was supported by a University of Glasgow Studentship.

7 References

Anderson, A. H., Mullin, J., Katsavras, E., Brundell, P., McEwan, R., Grattan, E., & O' Malley, C. (1999). *Multimediating multiparty interactions.* Paper presented at the Proceedings of INTERACT 99. IFIP.

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

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

Bard, E. G., & Aylett, M. (2001). *Referential form, word duration, and modeling the listener in spoken dialogue.* Paper presented at the Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society.

Blokland, A., & Anderson, A. (1998). Effect of low frame-rate video on intelligibility of speech. *Speech Communication, 26*, 97-103.

Brown, G., Anderson, A. H., Yule, G., & Shillcock, R. (1984). *Teaching Talk.* Cambridge: Cambridge University Press.

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

Chafe, W. (1974). Language and consciousness. *Language, 50*, 111-133.

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

Clark, H. H., & Clark, E. V. (1977). *Psychology and Language: An Introduction to Psycholinguistics.* New York: Harcourt Brace Jovanovich.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the 'model of the listener'. In D. J. Napoli & J. A. Kegel (Eds.), *Bridges between psychology and linguistics.* Hillsdale NJ: Erlbaum.

Doherty-Sneddon, G., Anderson, A. H., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied, 3*(2), 105 - 125.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 28*, 47-56.

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

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62*, 378-391.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

Hunnicutt, S. (1985). Intelligibility versus redundancy - conditions of dependency. *Language and Speech, 28*(1), 47-56.

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

Lieberman, P. (1963). Some effects of the semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172-187.

Robertson, C., & Kirsner, K. (2000). Indirect memory measures in spontaneous discourse in normal and amnesic subjects. *Language and Cognitive Processes, 15*(2), 203-222.

Rossnagel, C. (2000). Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology, 30*, 429 - 445.

Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of Telecommunications*: Wiley.

Sweller, J. (1988). Cognitive load during problem-solving:Effects on learning. *Cognitive Science, 12*, 257-285.

The Use-Mention Distinction and its Importance to HCI*

Michael L. Anderson[1], Yoshi A. Okamoto[3], Darsana Josyula[2] and Don Perlis[1,2]
(1) Institute for Advanced Computer Studies
(2) Department of Computer Science
University of Maryland, College Park, MD 20742
(3) Microsoft Corporation, Redmond, WA 98052
(301) 405-1746
{mikeoda, darsana, perlis}@cs.umd.edu, yoshio@microsoft.com

The Use-Mention Distinction and its Importance to HCI

Abstract

In this paper we contend that the ability to engage in meta-dialog is necessary for free and flexible conversation. Central to the possibility of meta-dialog is the ability to recognize and negotiate the distinction between the use and mention of a word. The paper surveys existing theoretical approaches to the use-mention distinction, and briefly describes some of our ongoing efforts to implement a system which represents the use-mention distinction in the service of simple meta-dialog.

1 Introduction

We use the term *conversational adequacy* to denote the ability to engage in free and flexible conversation. It is our contention that the ability to engage in meta-dialog is necessary for conversational adequacy, and more importantly, that a robust meta-dialogic ability can make up for weaknesses in other areas of linguistic ability (Perlis et al., 1998). For this reason, we think that time spent understanding and implementing meta-dialog in natural language HCI systems will be well rewarded; the ability to engage in even simple meta-dialog can be used to fruitfully enhance the performance of interactive systems, even those having relatively limited speech recognition and language processing abilities.

Central to the possibility of meta-dialog is the ability to recognize and negotiate the distinction between the use and mention of a word. Although this distinction has attracted some attention from philosophers of language, it has not gotten sufficient notice from those most directly involved in the theory and design of natural language HCI. We hope with this paper to begin to correct this lacuna, and stimulate further research in this important area. To this end, we will first try to establish the importance of meta-dialog, and of mastery of the use-mention distinction in particular, to conversational adequacy. We will then survey and evaluate the existing theoretical frameworks for understanding the use-mention distinction. Finally, we will briefly describe some of our efforts to implement a system which represents the use-mention distinction in the service of simple meta-dialog, and suggest some directions for future research.

* This research was supported in part by the AFOSR and ONR

2 The importance of meta-dialog

Conversation is not generally the exchange of fully formed, grammatically correct, and error-free utterances. Indeed, it is unlikely that there could ever be a fully fluent, error-free dialog; even putting aside problems of signal reception, and assuming perfect syntactic processing, the ability to understand a dialog partner involves such complicated and uncertain tasks as modeling their knowledge state and using context to disambiguate reference. In practice, conversation tends to be like this: a series of contributions in which a common context is constantly assessed and maintained with the use of inference, clarifications and confirmations. The main thread of the conversation starts and stops; there are pauses and hedges, questions are asked and sub-dialogs are spawned. Far from representing a break-down or failure of conversation, these processes, which monitor and establish *grounding*, are central to the nature of dialog. As was modeled by (Clark and Schaefer, 1987; Clark and Schaefer, 1989b) and has been confirmed by a great deal of research since (e.g. (Brennan, 1998; Brennan, 2000; Brennan and Hulteen, 1995; Cahn and Brennan, 1999; Clark and Brennan, 1991; Krahmer et al., 1999a; Krahmer et al., 1999b; Paek and Horvitz, 1999)), conversation is composed of two-phase contributions—an utterance is first *presented* to a dialog partner, but it is not until evidence of understanding is accrued that the utterance is *accepted* and becomes part of the common ground of the two speakers.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 21-28.

Insofar as this process involves monitoring not just of the content of the conversation, but also of one's own *understanding* of the content, as well as making inferences about the other's state of understanding based on aspects of the conversation like timing, feedback, etc., this suggests to us the centrality of meta-linguistic skills to conversational ability. Clark (Clark and Schaefer, 1989a) presents empirical evidence for the use of meta-linguistic skills by young children. Among many other things, these skills include:

1. **Monitoring one's ongoing utterance**: An example of this was seen in a 2 year, 6 month child practicing parts of speech (in this case its pronunciation of "berries") on its own: "Back please/berries/*not* barries/barries, barries/*not* barries/berries /ba ba"
2. **Checking the result of an utterance**: Children at least as young as 5 years, 4 months comment on and correct the utterances of others. They also verify that the listener has understood their utterance and attempt a repair otherwise.
3. **Deliberately trying to learn**: For instance, a 4 year old will ask things like: "Mommy, is it AN A-dult or A NUH-dult?"
4. **Predicting the consequences of using inflections, words, phrases or sentences**: This includes judging the politeness of utterances, which is exhibited by children aged four and a half. Children can also correct word order in sentences judged "silly". Clark cites instances of this being done by two-year olds.

This variety of meta-linguistic skills exhibited by very young children further suggests the central importance these skills have for learning and processing language.

It seems fairly clear that these abilities could not exist apart from a facility with meta-reasoning (reasoning about the reasoning process, whether yours or your interlocutors') and meta-dialog, and in particular without a grasp of the distinction between the use and mention of a word. This can be seen most explicitly in item (3), above, but the general ability to take language or a linguistic item as an object of scrutiny—which ability is at the center of the use-mention distinction—is part of all the abilities listed above.

For current purposes, we would like to draw your attention to the following key conversational ability—reasoning about, discussing, and adapting word meanings and references. This can include reasoning about:

- word meaning—knowledge of the word itself, part of speech, lexical semantics, concept picked out, etc.
- specific reference of a term denoting a particular object, including both names and definite, indefinite and deictic expressions, using a concept to pick out an individual.
- the relation of words to the concept(s) or entities referred to

Clearly the exercise of this capacity—which can be understood, for instance, as the ability to understand sentences like: (a) What does 'felicitous' mean? (b) 'Felicitous' means happy, or well-formed. (c) Is 'shop' a verb as well as a noun? (d) What is a 'VCR'?—requires the ability to take words themselves as objects for reasoning and discussion and, more particularly, the ability to negotiate the distinction between the use and mention of a given term.

3 Survey of Theoretical Approaches to the Use-Mention Distinction

In this section, we briefly sketch the various theories of the use-mention distinction, and the related issue of understanding quotation. Extensive citations are provided for the reader who wishes to explore the issues in more detail.

3.1 Classic theories

There are four main theoretical positions which define the early discussion of quotation: the Name theory, the Description theory, the Demonstrative theory and the Identity theory.[1]

In the Name theory, quotations are considered to be names, that is, quoted expressions name their referents. (Tarski, 1933) explores the frontier by elucidating the issue of quotation in general with this approach.

> Quotation-mark names may be treated like single words of a language ... the single constituents of these names ... fulfill the same function as the letters and complexes of successive letters in single words. Hence they can possess no independent meaning. (Tarski, 1933) p. 159

(Quine, 1940) and (Cram, 1978) travel the same theoretical path. The main point seems to be that, as in a name, "[t]he meaning of the whole does not depend upon the meanings of the constituent words."

[1] The discussion is influenced throughout by (Saka, 1998).

(Quine, 1940) p. 26 "Chief Sitting Bull" refers to Sitting Bull, and not, for instance, to a sitting bull, because that linguistic item is a name. However, whether or not this is generally true of names, it does not seem to be a good model for quotation in general. For instance, I can say: "'α' is a Greek letter," and be entirely understood. Had I been thereby told only the *name* for α, I would not fully understand the sentence until I had determined what the name named. But this does not appear to be the case. More generally, quoted expressions and the linguistic items they refer to seem to be systematically related by the shape or nature of the expression itself. This and like considerations give rise to the Description theory.

The Description theory provides more structurally "descriptive power" to the quoted expressions. There are roughly two approaches. The formal approach, outlined in (Tarski, 1933; Quine, 1940; Richard, 1986), is based on formal elements, such as orthographical elements or phonological elements. The other is the lexical approach, exemplified by (Geach, 1957), which analyzes the quoted expression word-by-word. In general, the idea is that the quoted expression describes its referent, thus "It was the best of times ..." = the expression formed by 'It', plus 'was', etc. The descriptive theory thus appears to rely on having names for some base elements of the language, whether orthographic, phonemic, or lexical; insofar as this is so, it falls prey to the same criticisms as the Name theory.

(Davidson, 1969) explores the Demonstrative account of quotation, and this framework has convinced many linguists and philosophers over three decades. Just to cite a few: (Partee, 1973; Christensen, 1967; Goldstein, 1984; Garcia-Carpintero, 1994; Cappelen and Lepore, 1997). The Demonstrative theory holds that quote marks demonstrate, or point to, the quoted material, on one interpretation to its shape. Hence, (e) is understood as (f) in the following:

e. "Cats" is a noun.

f. Cats. That complex of shapes is a noun.

But taken as a formal transformational rule, this cannot account for the possibility of iterated quotation. It appears that (g) would be transformed into (h) and then something like (i):

g. "'Cats' " is a noun phrase.

h. "Cats." That is a noun phrase.

i. Cats. That that is a noun phrase.

Opposed to the demonstrative theory is the Identity theory, according to which quoted expressions corefer to themselves. The Identity theory is as widely discussed as the demonstrative theory. See (Frege, 18921980; Quine, 1940; Wittgenstein, 1953; Tajtelbaum, 1957; Whitely, 1957; Searle, 1969; Washington, 1992; Reimer, 1996).

> Whereas the Demonstrative Theorist regards quote marks (or context) as referential and the quoted material as an inert adjunct, the Identity Theorist conversely regards the quoted material as (self-) referential and the quote marks as semantically empty. Despite these differences, the Identity and Demonstrative accounts can both be called picture theories, for both claim that quotation resembles its referent, the quoted material. (Saka, 1998)

It appears that the Identity theory also falls prey to the recursion problem, for it does not appear to be able to account for the difference between "'sunset"' and "sunset", which have distinct referents.

3.2 Syntactic and Semantic Treatments of Quoted Expressions

Although the title, *The Syntax and Semantics of Quotation* sounds promising, in her article Partee (Partee, 1973) only deals with the direct quotation of the whole sentence, thereby, unfortunately, excluding the cases in which we are most interested. She classifies quotation into *word quotation* (j) Should "pickup" ever be hyphenated?, *sentence quotation*, (k) "I am speaking now" is always true when spoken, and *direct quotation* (l) Tom said, "My grandfather was in Seattle." Focusing her attention on the last, Partee basically supports Davidson's claim (Davidson, 1969) that the quoted sentence is not syntactically or semantically a part of the sentence that contains it. Partee then argues that the quoted sentence represents nothing more than the surface structure, and it is that surface structure (and not the meaning of the quoted sentence) that contributes to the meaning of the sentence that contains it.

> (A)ll the apparent evidence for deeper syntactic and semantic structure is a result of the main sentence speaker's understanding and analyzing the noises he is quoting as a sentence, just as he understands and analyzes a sentence, a string of noises, that comes to him from someone else. It is in this process that the language/metalanguage distinction is crucially involved. (Partee, 1973) p. 418

Cram (Cram, 1978) considers Partee's conclusion "absurd" (p. 43). He disagrees with Partee's selective application of her constraint only to direct quotations. He thinks that the same constraint "would hold equally well for sentences in general" (p. 43). He instead argues that the quoted expression has the status of a lexical item—more specifically of a noun phrase—and the structure is specified at the level of the matrix sentence. Consider the following examples:

m. "They are cooking apples" ist zweideutig
n. Der Satz "They are cooking apples" ist zweideutig
o. The sentence "They are cooking apples" ist zweideutig.

Cram presents the contrast between (m) and (n) (the tagging test (Reichenbach, 1947)) as a supportive case for the first part of his argument that the quoted expression has the status of a noun phrase. Then he contrasts (n) and (o) to argue that the NP tag "belongs both logically and semantically to the level of the matrix sentence" (p. 44). In order for his analysis to work, however, he has to make one assumption:

> The base rules which generate the matrix sentence also generate a quotational tag under a Noun Phrase node, which may occur either in subject position or elsewhere. The quotation tag is always specified at the level of Deep Structure, but may be deleted by an optional transformation. (Cram, 1978) p.44

From the perspective of computational applications, it is not hard to implement the aforementioned rule as a system designer. In fact, treating a quoted expression as an NP can be a way to solve a simple parsing problem. For instance, following (Reichenbach, 1947), a simple lexical insertion rule can allow one to treat a quoted sentence as a syntactically opaque unit.

However, there are cases when expressions are mentioned without overt quote marking. If we were to design a system simply to look for such quote marks, then we would miss such cases. Yet, we do not want to design a parser or a system that automatically treats any syntactically dubious phrases as noun phrases, either.

3.3 Disambiguated Ostention Theory

In our judgment, the Disambiguated Ostention theory (Saka, 1998) is the most satisfactory of the various theoretical accounts which formalize the use-mention distinction. In general, according to Saka, the capacities for both use and mention stem from the same source, namely from the fact that the human mind associates a multiplicity of deferred ostensions with any exhibited token, thus giving rise to pragmatic ambiguity. Among the possible ostentions are:

p. orthographic form: cat
q. phonic form: /kaet/
r. lexical entry: [cat, /kaet/, count noun, CAT]
s. intension: CAT
t. extension: x: x a cat

Saka then explains how these ostensibly deferred items interact in the following way:

5. Exposure to the written label cat (p) or the spoken label /kaet/ (q) evokes the corresponding lexeme (r) in every competent speaker of English, where a lexeme is an arbitrary ordered n-tuple including orthographic form, phonic form, syntactic category, meaning, register, etc.
6. The lexeme [cat, /kaet/, count noun, CAT] specifies the intension CAT (s) according to the pragmatic function.
7. CAT determines the extension x: x a cat (t) according to some function.

Unfortunately, he does not elaborate much on the nature of the "pragmatic function" (6) nor "some function" (7). However, he does provide a useful characterization of the use-mention distinction:

Speaker S **uses** an expression X iff:

i. S exhibits a token of X;
ii. S thereby ostends the multiple items associated with X (including X's extension);
iii. S intends to direct the thoughts of the audience to the extension of X.

Speaker S **mentions** an expression X iff:

iv. S exhibits a token of X;
v. S thereby ostends the multiple items associated with X;
vi. S intends to direct the thoughts of the audience to some item associated with X *other than its extension.*

Saka's approach is significant in that the use-mention distinction is characterized in terms of extension. And, more importantly, it "allows for the existence of mentioning without quote marks" (p. 127). This is definitely an advantage over the rest of the known approaches, because they assume that quote marks are required for an expression to be considered as mentioned. The Disambiguated Ostention theory has been criticized in (Cappelen and Lepore, 1999). However, that criticism seems to center around the question of whether words really have multiple *referents*, as Saka contends, or only one (perhaps complex) referent. In so far as the matter turns on technical issues in the philosophy of language and on the nature of reference, it does not affect the arguments offered here. It is enough for our purposes to hold that a speaker might, by mentioning (or with the use of quotation), mean to draw our attention to items other than the extension of a word, whether or not these other items are legitimately considered to be *referents* of that word.

4 Description of Current Research

The long-term goal we have set for ourselves is the design of a system modeled not on conversation with a fluent colleague, but rather, for example, on a task-oriented interaction with a stranger who doesn't speak much of a common language. In these situations, e.g., buying a train ticket in a foreign country, speakers are often able to communicate to effectively solve a joint task, in spite of problems in word recognition, or use of unfamiliar words or syntactic structures. Despite the difficulties of understanding the language, interactive dialog behaviors and ongoing repairs allow humans to overcome some of these problems.

One major step toward this goal was the design and implementation of a model of action-directive exchanges (task oriented requests) (Traum et al., forthcoming; Traum and Andersen, 1999) based on an evolving-time ("active logic") model of inference (Elgot-Drapkin et al., 1993; Elgot-Drapkin and Perlis, 1990; Purang et al., 1999). Our model works via a step-wise transformation of the literal request made by a user (e.g. "Send the Boston train to New York") into a specific request for an action that can be performed by the system or domain. In the case of "the Boston train," the system we have implemented is able to interpret this as "the train in Boston," and then further disambiguate this into a specific train currently at Boston station, which it will send to New York. Information about each step in the transformation is maintained, to accommodate any repairs that might be required in the case of negative feedback (if for instance, the system picks the wrong train, and the user says "No" in response to the action). (See (Traum et al., forthcoming) for more detailed information.) This implementation represents an advance not just in its ability to reason initially about the user's intention (e.g., by "the Boston train" the user means . . .") but in its ability to respond in a context-sensitive way to post-action user feedback, and use that feedback to aid in the interpretation of the user's original and future intentions. (For instance, if the user says "No, send the Boston train to New York" the system is able to identify its mistake as its choice of referent for "the Boston train," and choose a different train; if there are no other trains in Boston, it tells the user: "Please specify the train by name.") However, although our current system is able to respond in this way to post-action user feedback, the actual transformation of the user's utterance from its literal form into a performable request takes place without user feedback. We believe that a system which could request and utilize user feedback in this process—one, that is, which contained a more robust meta-dialogic component—would represent a significant advance.

Thus, the next step for us is to design and implement a model of Question-Answer exchanges which can be used not just independently to track ongoing interactions of this type, but also (more significantly for our purposes) to supplement the interpretation of task-oriented requests by allowing the system to request and use user feedback during the interpretive process. It is important not just that the dialog agent be able to respond to questions from the user, but also be able to ask questions *of* the user. For instance, consider the user request "Send the Bullet train to Bean Town." Assuming that the system is unable to determine the meaning of "Bean Town" by itself, it should be able to ask the user for specific help. Understanding and implementing ways of representing the use-mention distinction in natural language HCI is vital to making this possible, for should the user respond: "'Bean Town' means Boston" the system will have to recognize this as an instance of mention rather than use, and deal with the utterance appropriately. It is important that the system recognizes this distinction in order to identify the referent of the word "Bean Town" correctly. In the first case, the system has to disambiguate the referent as a city whereas in the latter case, it has to interpret the referent as a word.

Our own general approach to this issue involves implementing a specialized domain, which (with apologies to Austin (Austin, 1962)) knows how to do things with words. Following Saka's definition, if the system determines that a given sentence is meant to draw attention to something other than the extension of a word, processing of the sentence is passed

on to the "words" domain, where words are objects and are dealt with in terms of ostentions (p)-(s), above. In the simplest cases this means doing things like updating the lexicon and correcting spelling; more complex cases include adopting new conventions for word use, and learning new words.

However, such a system depends on being *able* to determine when a word is being mentioned; methods for doing this reliably is one area in which we would especially like to see more research. If the user were to characterize each instance of a mention with explicit quotation marks, then it would be easy for the system to determine when a word is being mentioned rather than being used. However, in spoken natural language HCI, such explicit quote marks are neither feasible nor reliable, since it is not natural to say "Open quote" before and "Close quote" after each mention, and in any case human-human conversation seems to proceed successfully in most cases without such explicit quotation marks.

Context seems to play a key role in determining whether a word is being used or mentioned. If the system comes across a word that it does not know about, then it could ask the user what it is, and generate the expectation that a future user response would be to provide the answer. Keeping track of such expectations can help the system to identify the mention of a word. For instance, in the example, since the system is not able to determine the meaning of "Bean Town," it can create an expectation that the user is going to provide a definition for "Bean Town" and in that definition the word "Bean Town" would be mentioned rather than used. Still, this approach is clearly not going to cover a large percentage of cases, and other methods are needed. One method that we are exploring is to make the system sensitive to special words like "means" and "is", often used to introduce definitions of unknown words. Such intension words can indicate that a word is being mentioned rather than used. For instance, if the user were to say "'Bean Town' means Boston," "means" would signal that it is a mention of the word "Bean Town" rather than its use that the user intends. Lexical category words (like "noun" and "verb"), can also act as cues for recognizing the possibility of a mention, as in "'Cats' is a noun." If the system misidentifies a mention to be a use, the user would engage in meta-dialog to correct the system. The fact that dialog is proceeding at the level of meta-dialog is itself a hint to the system to look for the possibility of a mention.

A forthcoming paper describes in more detail the progress of our research to date, including our proposed representation scheme and overall system architecture. However, as indicated, this paper is intended primarily to draw attention to a problem, finding solutions (or exploring approaches) to which we believe will have a great impact on the future of natural language HCI.

References

J. L. Austin. 1962. *How to Do Things with Words.* Harvard University Press.

Susan E. Brennan and Eric A. Hulteen. 1995. Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems*, 8:143–151.

Susan E. Brennan. 1998. The grounding problem in conversations with and through computers. In S.R. Fussell and R.J. Kreuz, editors, *Social and Cognitive Psychological Approaches to Interpersonal Communication*, pages 201–225. Lawrence Erlbaum.

Susan E. Brennan. 2000. Processes that shape conversation and their implications for computational linguistics. In *Proceedings of the 38th Annual Meeting of the Association for Computational Lingusitics.*

Janet E. Cahn and Susan E. Brennan. 1999. A psychological model of grounding and repair in dialog. In *Proceedings of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, pages 25–33.

Herman Cappelen and Ernest Lepore. 1997. Varieties of quotation. *Mind*, 106:429–50.

Herman Cappelen and Ernest Lepore. 1999. Using, mentioning and quoting: A reply to saka. *Mind*, 108:741–50.

Niels Christensen. 1967. The alleged distinction between use and mention. *Philosophical Review*, 76:358–67.

H.H. Clark and Susan E. Brennan. 1991. Grounding in communication. In J. Levine L.B. Resnik and S.D. Teasley, editors, *Perspectives on Socially Shared Cognition*, pages 127–149.

H.H. Clark and E.F. Schaefer. 1987. Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2:19–41.

Herbert H. Clark and Edward F. Schaefer. 1989a. Contributing to discourse. *Cognitive Science*, 13:259–294. Also appears as Chapter 5 in (Clark, 1992).

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

Herbert H. Clark. 1992. *Arenas of Language Use.* University of Chicago Press.

David F. Cram. 1978. The syntax of direct quotation. *Cahiers de Lexicologie*, 33(2):41–52.

Donald Davidson. 1969. On saying that. *Synthese*, 19:130–46.

J. Elgot-Drapkin and D. Perlis. 1990. Reasoning situated in time I: Basic concepts. *Journal of Experimental and Theoretical Artificial Intelligence*, 2(1):75–98.

J. Elgot-Drapkin, S. Kraus, M. Miller, M. Nirkhe, and D. Perlis. 1993. Active logics: A unified formal approach to episodic reasoning. Technical Report UMIACS TR # 99-65, CS-TR # 4072, Univ of Maryland, UMIACS and CSD.

Gottlob Frege. 1892/1980. On sense and reference. In Peter Geach and Max Black, editors, *The Philosophical Writings of Gottlob Frege*, pages 56–78. Oxford: Basil Blackwell.

Manuel Garcia-Carpintero. 1994. Ostensive signs. *jphil*, 91:253–64.

Peter T. Geach. 1957. On names of expressions. *Mind*, 59:388.

Laurence Goldstein. 1984. Quotation of types and types of quotation. *Analysis*, 44:1–6.

Emiel Krahmer, Marc Swerts, Mariet Theune, and Mieke Weegels. 1999a. Error detection in spoken human-machine interaction. In *Proceedings of Eurospeech'99*, Budapest, Hungary.

Emiel Krahmer, Marc Swerts, Mariet Theune, and Mieke Weegels. 1999b. Problem spotting in human-machine interaction. In *Proceedings of Eurospeech'99*, Budapest, Hungary.

Tim Paek and Eric Horvitz. 1999. Uncertainty, utility and misunderstanding: A decision-theoretic perspective on grounding in conversational systems. In *Proceedings, AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems.*

Barbara Partee. 1973. The syntax and semantics of quotation. In Stephen Anderson and Paul Kiparsky, editors, *Festschrift for Morris Halle*, pages 410–18. New York: Holt, Rinehart, Winston.

D. Perlis, K. Purang, and C. Andersen. 1998. Conversational adequacy: mistakes are the essence. *Int. J. Human-Computer Studies*, 48:553–575.

K. Purang, D. Purushothaman, D. Traum, C. Andersen, D. Traum, and D. Perlis. 1999. Practical reasoning and plan execution with active logic. In *Proceedings of the IJCAI'99 Workshop on Practical Reasoning and Rationality.*

W.V.O. Quine. 1940. *Mathematical Logic.* Harvard University Press, Cambridge, MA.

Hans Reichenbach. 1947. *Elements of Symbolic Logic.* Macmillan, London.

Marga Reimer. 1996. Quotation marks: demonsratives or demonstrations? *Analysis*, 56:131–42.

Mark Richard. 1986. Quotation, grammar and opacity. *Linguistics and Philosophy*, 9:383–403.

Paul Saka. 1998. Quotation and the use-mention distinction. *Mind*, 107:113–35.

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

Al Tajtelbaum. 1957. It is impossible to be told anyone's name. *Analysis*, 17:52.

Alfred Tarski. 1933. The concept of truth in formalized languages. In J. Woodger, editor, *Logic, semantics, mathematics.* Oxford: Oxford University Press.

D. Traum and C. Andersen. 1999. Representations of dialogue state for domain and task independent

meta-dialogue. In *Proceedings of the IJCAI99 workshop: Knowledge And Reasoning in Practical Dialogue Systems*, pages 113–120.

David Traum, Carl Andersen, Yuan Chong, Darsana Josyula, Michael O'Donovan-Anderson, Yoshi Okamoto, Khemdut Purang, and Don Perlis. forthcoming. Representations of dialogue state for domain and task independent meta-dialogue. *Electronic Transactions on Artificial Intelligence.*

Corey Washington. 1992. Quotation. *jphil*, 89:582–605.

C.H. Whitely. 1957. Names and words. *Analysis*, 17:119–20.

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

Towards a psycholinguistics of dialogue: defining reaction time and error rate in a dialogue corpus

Ellen Gurman BARD
HCRC & School of Philosophy, Psychology, & Language Sciences, U. of Edinburgh
George Square
Edinburgh, UK EH8 9LL
ellen@ling.ed.ac.uk

Matthew P. AYLETT
HCRC, U. of Edinburgh & Rhetorical Systems, Ltd.
4 Crichton's Close
Edinburgh, UK EH8 8DT
matthewa@rhetorical.com

Robin J. LICKLEY
HCRC, U. of Edinburgh & Dept of Speech Science, Queen Margaret U. College
Clerwood Terrace
Edinburgh, UK EH12 8TS
rlickley@qmuc.ac.uk

Abstract

This study uses the multi-level coding of a designed corpus of unscripted task-oriented dialogues to demonstrate that time to respond (Inter-Move Interval, IMI) and rate of disfluency behave like psycholinguistic measures, reaction time and error rate, in reflecting the speakers' cognitive burdens. Multiple-regression analyses show that IMI is sensitive to social distance between interlocutors, to the difficulty of the task which the dialogue serves, and to comprehension of the prior utterance and production of the current one. Rate of simple overt disfluency, in contrast, shows social and task effects, with most of the uniquely explained variance associated with planning and producing the current utterance. The results suggest that coded corpora may be useful in developing models of human interlocutors.

Introduction

Traditionally, the study of dialogue and the study of psycholinguistic processes have been separate fields. Recently, however, a growing body of research is creating a psycholinguistics of communicative speech. Pickering and Garrod (*in press*) have claimed that interactive language use gives us much better clues to psycholinguistic processes than we can find by examining decontextualized 'laboratory speech' or text. Even in laboratory studies, there is a trend towards covert observation of natural linguistic behaviours and away from discrete non-linguistic responses. At the same time, the design of computer-assisted dialogues is often supported by ad hoc studies using exactly the sorts of off-line measures which psycholinguists are at pains to avoid. The large samples of natural dialogue behaviour which are now available in the form of speech corpora yield rich observational data but are often thought to be irrelevant to psycholinguistics, because they do not follow balanced experimental designs. Without some experimental controls, it is difficult to tell which natural dependent variables might reflect the cognitive demands of natural language use. Using a carefully designed corpus, however, this paper shows that two classic psychological measures, reaction time and error rate, can be defined in terms which are likely to be available for a coded corpus of spoken dialogues. The results help to open the way to a naturalistic, on-line, psycholinguistics of dialogue, and allow strategic design of dialogue applications.

1 Motivation

There is good reason to believe that both reaction time and error rate measures will function in dialogue. In laboratory paradigms, reaction time, measured from the presentation of a stimulus to the onset of a verbal response, is used to study perception, comprehension, and problem solving (see Bock (1996) for a summary). Though the dialogue tradition is more interested in what might induce a speaker to initiate a contribution to a verbal *pas de deux,* psycholinguistic studies of monologues (Beattie, 1983; Butterworth, 1980; Wheeldon and Lahiri, 1997) reveal a natural reaction time measure: pause lengths between utterances or between larger discourse segments are related to the complexity of the unit being planned. If time taken to begin speaking in dialogue has these same characteristics, then turn-taking time is a kind of reaction time.

Error rate, usually defined as incorrect answers in forced choice tasks, seems to have a

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 29-36.

natural counterpart in spontaneous speech: overt disfluencies, those which include genuine on-line editing of something already said[1]. Fluency is known to be affected by planning and production processes. Disfluencies tend to occur early in utterances, when planning of later stages is incomplete (Clark and Wasow, 1998). Disfluencies are more common in longer utterances (Clark and Wasow, 1998), in more complex constituents (Clark and Wasow, 1998; Oviatt, 1995), and when response choices are complex (Oviatt, 1995). We propose that disfluencies should be viewed as edited errors, because they indicate the speaker's failure to design an utterance portion conforming to his or her current conversational goals before beginning to speak. Since self-monitoring and self-correction appear to be possible before articulation begins (Blackmer and Mitton, 1991; Levelt, Roelofs, and Meyer, 1999; Levelt, 1983; Maclay and Osgood, 1959; Postma and Kolk, 1992), overt disfluencies comprise failures of those earlier processes of quality control.

Although these results suggest dependent variables for psycholinguistic studies of dialogue, they have not been checked for effects of many of the plausible independent variables, the possible sorces of difficulty for the speaker. Disfluency, for example, has been associated with the process of planning and producing an utterance, but it might also be affected by the other cognitive demands on an interlocutor. In addition to producing their own contributions, interlocutors must also interpret one another's contributions, keep some record of the dialogue so far, and provide appropriate replies promptly enough to make it plain that they wish to take the floor. In task-oriented dialogue, at least, they must use the interaction to achieve an extra-linguistic goal. As with any other task, participants may improve over time or tire as they continue. Reaction time and error rate could be affected by any of these processes.

Most simply, the difficulty of conducting a dialogue should be increased by the complexity and the novelty of the *task* which it pursues. If so, reaction times and error rates should as usual be worse for difficult tasks and early in the course of mastering a task.

Second, there should be effects of the *prior utterance*, because preparing a reply depends on processing the interlocutor's contribution on some level. How hard it is to do this may be critical because the time between conversational turns is so short (with a mode around 150 msec for the materials discussed below) that production processes are likely to start before listening stops entirely. Comprehension difficulty should give rise to slower responses, as it does in laboratory studies. Furthermore, if self-monitoring in speech production actually uses the same comprehension processes which we use for comprehending what is said to us, (Levelt, Roelofs, and Meyer, 1999), then difficulty in comprehending an interlocutor's utterance could interfere with early self-monitoring, increasing rates of overt disfluency. Disfluent input could be particularly damaging. Difficulty in perceiving disfluent utterances (Bard and Lickley, 1998) could delay production of replies. Cross-speaker syntactic priming (Branigan, Pickering, and Cleland, 2000) might even encourage disfluent replies to disfluent utterances.

Third, *production effects*, as outlined above, could independently yield slower responses and more disfluencies where the speaker's planning burden rises. Longer, more complex utterances should yield slower, more disfluent production, as should the need to plan discourse segments.

Finally, *interpersonal factors,* like the number of sensory channels carrying the communication or the familiarity of the interlocutors could affect delicate processes of listener-modelling, planning, and feedback (Branigan, Lickley, and McKelvie, 1999; Doherty-Sneddon, Anderson, O'Malley, Langton, et al., 1997), with better knowledge facilitating responding.

To test for these separate sources of concurrent difficulty, we exploit existing multi-level coding of a dialogue corpus and use multiple regression analysis to determine whether each set of processes makes a statistically independent contribution to reaction time and error rate.

2 Method

2.1 Materials

2.1.1 Corpus.

Materials came from the HCRC Map Task Corpus (Anderson et al., 1991), 128 unscripted dialogues in which 32 pairs of Glasgow University undergraduates communicated routes defined by labeled cartoon landmarks on schematic maps of imaginary locations. Instruction Giver's (hereafter 'IG') and Follower's

[1] We exclude mere hesitations, filled pauses, or rapid self-corrections designed for effect or humour.

(IF) maps for any dialogue matched only in alternate landmarks. Participants knew that their maps might differ but not where or how. Interlocutors could not see each other's maps. Familiarity of participants (within subjects) and ability to see the interlocutor's face (between subjects) were counterbalanced. Each participant served as IG for the same route to two different IFs and as IF for two different routes, yielding 8 dialogues for every group of 4 co-participants. Channel-per-speaker digital stereo recordings were orthographically transcribed and word-segmented. Like the coding systems described below, the segmentations and design variables form part of an XML corpus database.

2.1.2 Unit of analysis.

The word-segmented corpus is framed as a series of Conversational Game Moves (Carletta., Isard, Isard, Kowtko, et al., 1997), turns or parts of turns whose purpose in moving the dialogue forward can be determined by their form and context. Moves are stages of Conversational Games, which are themselves usually stages in completing Transactions, sections of the task (here route communication) which the dialogue serves (Isard and Carletta, 1995). We examined only those Moves which were likely to be a replies to the previous speaker's Move: all were produced by a different speaker from the speaker of the prior Move; none began too early to respond to that prior Move (each included Move began less than 1 sec before the end of the prior speaker's Move and more than 350 msec after the beginning of that Move) and none merely resumed an earlier incomplete Move by the same speaker (all began at least 300 msec after the end of a previous Move by the same speaker and none were coded as continuations of the previous Move) (Bard, Aylett, and Bull, 2000; Bull, 1998; Bull and Aylett, 1998). Bull and Aylett (1998) showed that these criteria were conservative relative to human judgments about the relationship between successive Moves.

2.2 Dependent Variables.

2.2.1 Reaction: Inter-move interval (IMI).

IMI is the time in milleseconds, positive or negative, between the offset of one speaker's Move and the onset of the next starting Move if the latter is produced by the other speaker and is a 'real reply' (as defined above) to the former (Bard et al, 2000; Bull, 1998). IMI ranges from −999 msec to over 13 sec (mean = 506 msec, *s.d.*= 795 msec; mode between 100 and 200 msec)..

2.2.2 Error rate: Disfluency rate

Disfluency rate is the number of simplex disfluencies per Move. Disfluency annotation (Levelt, 1983) was performed on the whole corpus using Xwaves/Entropic xlabel to display the speech waveform and spectrograms where necessary. Disfluencies involved an overt editing of uttered words or part words. Each word or part-word in a disfluent Move was assigned to one of the five possible parts of a disfluency. The optional *original utterance* was composed of pre-error words which should stand after correction. The *reparandum,* the words or partial words which constituted an error, was followed by an *interruption* with or without an optional *filler*, then in most cases by a *repair* and *continuation.*

The current paper reports on only simple disfluencies of 4 types. In ***repetitions*** the speaker repeats a string verbatim, with no additions or deletions: e.g. *[we're going] we're going left of the camera shop.* In ***insertions*** the speaker repeats a string and inserts a word or words within the repeated string: e.g. *[go left] go just left of the camera shop.* In ***substitutions*** a word or string is replaced by another with no major syntactic alteration: e.g. *go [left] right of the camera shop.* In ***deletions***, the speaker interrupts an utterance and either restarts without repeating or directly substituting, or simply surrenders the floor to the other speaker: e.g. *[you're away f-] right see the wee bit that's jutting out?* Combinations of these types, complex forms of disfluency, and mere pauses, filled or otherwise, are ignored. There are 0-4 disfluencies per Move, mean =.13, *s.d.* = .40.

2.3 Independent (Predictor) Variables

2.3.1 Interpersonal

Social distance between IG and IF is reflected in *Eye-contact* condition, (the presence or absence of a flimsy barrier blocking the line of sight between interlocutors), *Familiarity* (whether the pair have just met or are friends), and *Sex of Players* (IG-sex and IF-sex). If speakers can be more efficient at framing their messages when they know more about their listeners, then familiar pairs with eye-contact should have fast, fluent responding. We include *Sex of Players* both because men are more likely to be disfluent than women (Bortfeld et al., 2001; Branigan, Lickley, and McKelvie, 1999; Shriberg, 1994)

and because there is a premium on fluency in initial meetings between different sex pairs (McLaughlin and Cody, 1982).

2.3.2 Task.

Deviation score (mismatch in cm^2 between the route printed on IG's map and the version drawn by IF) measured the dyad's success in communicating the printed route. *Drawing* (whether the IMI included an attempt to draw part of the route) indicated the presence of an ongoing physical task. For the disfluency equations, *IMI* was added so that disfluency rate could be assessed for effects of fast responding (see Carletta, Caley, and Isard, 1995)

2.3.3 Order

Two levels are marked, *Conversation* (among the 8 dialogues produced by each set of 4 the speakers) and *Ordinal Position of the Move* in the dialogue. Both could yield practice or fatigue effects.

2.3.3 Prior Move Comprehension.

For any Move IMI, the immediately prior Move was coded for number of referring expressions (classed as *New* introductions of landmarks *Shared* by IF and IG maps, later *Given* mentions of *Shared* items, *New Unshared* or *Given Unshared*), for each of the 4 categories of *Disfluencies*, and for *Length in Words* excluding any reparandum. Prior moves with higher scores on any of these variables could be more difficult to process.

2.3.4 Current Move Planning and Production.

Planning burden is measured by Speaker's Role (because IGs usually determine the structure of the dialogue), and by the *Size of the Unit Begun* by the Move (Transaction, Game, and Game Internal Move). The production burdens are measured in terms of *Referring Expressions* and *Length in Words* (under the system used for prior Moves).

2.4 Procedure

For IMI, we report results for regression equations for the 64 dialogues whose video record made it possible to eliminate IMIs during which IF drew parts of the route (remaining Moves, $N = 5543$). Predictor variables were first checked for collinearity. No high cross-correlations were found. To determine likely interactions, all predictor variables were entered into a single equation for the 'no-drawing' materials and removed individually, so that the effect on the contribution of the other variables could be examined. Where the effect was significant and the result interpretable, individual interactions were coded and all the new variables were tested by backward-stepping regression (F to remove = 4.0). The resulting model was run for the other 64 dialogues ($N = 7444$) and for the corpus as a whole. We report results only for variables whose sign and significance was maintained in all three equations.

For disfluency rate, we made it possible to examine the effects of the drawing activity by using all the response Moves in the 64 video-taped dialogues except for those with complex disfluencies ($N = 6882$). We report results borne out for all the data ($N = 14389$). To allow comparison with IMI, we used the IMI regression equation as a starting point. No meaningful interactions were found. Contributions of the sets of predictors described above were assessed by noting fall in explained variance (Multiple-R^2) as each set was omitted from the full model.

3. Results

The regression equations for IMI (Multiple $R^2 = .09$; $p < .0001$) and disfluency rate (Multiple $R^2 = .14$, $p < .0001$) show significant effects of psychological variables, but different patterns of effects. Results are stated as β-values, standardized regression coefficients, with $p < .01$, unless otherwise noted. They represent the contribution of a predictor variable with the effects of other predictors statistically removed, Because they are standardized, βs may be directly compared. Tables show raw means.

3.1 IMI

In line with decades of commentary on turn-taking, IMI is significantly affected by *interpersonal* factors: as Table 1 shows, the shortest IMIs are in dialogues between different sex pairs who have just met (Familiarity, .08; Familarity by sex-difference, -.05).

Table 1. Interpersonal effects on IMI: interlocutor familiarity and sex match

	Familiarity	
Sex match	Familiar	Unfamiliar
Same	384	435
Different	440	279

Table 2. Order effect on IMI: position of Move in dialogue

Quartile	Ordinal Position	Mean IMI

1	1-50	447
2	51-100	414
3	101-150	378
	151-200	357
4	201-250	274
	251-300	346
	301-350	334
	351-400	251
	401-450	312
	451-586	364

In line with our hypothesis that IMI is a reaction time measure, all the cognitive variable types have effects. There is a typical *order* effect: IMI falls later in dialogues (-.04) (Table 2). There is also a speed x communication-accuracy trade-off: though IMIs are longest in dialogues where the IF draws the least accurate routes (.10), they are longer in dialogues with the most accurate routes than in those of intermediate accuracy (Table 3).

Table 3. Task effect on IMI: deviation score

Deviation score quartile	1	2	3	4
Mean IMI	411	289	370	463

There is also an effect of *comprehension.* IMIs are shorter after references to landmarks already mentioned in the dialogue (for those on only one map, -.10, for those on both, -.04). They are longer after Moves introducing New landmarks only if the New landmark is not on the responder's map and no reference is made to previously mentioned shared items (-.09) in the introducing Move. In other words, this is a long reaction time during which the speaker conducts an extensive unsuccessful search for a landmark which s/he lacks (Table 4a and 4b).

Finally there are effects of *production* and planning in dialogue much like those known in monologue. IMIs are longer before longer utterances (.06) (Table 5) and those which begin larger units of dialogue: Transaction-initial moves follow longer intervals than Game-initial (.03, $p < .05$), which in turn follow longer intervals than Game-internal Moves (.14) (Table 6).

Table 4. Effects of comprehension on Mean IMI: (a) referring expressions in prior Move by status of first referring expression; (b) interaction of references to New Unshared and Given Unshared map landmarks

(a)	Occurrence of referent on players' maps	
Status in dialogue	Shared	Unshared
Given	227	217
New	277	514

(b)	References to Unshared Given	
References to Unshared New	No	Yes
No	378	350
Yes	557	398

Table 5. Production effects on IMI: current Move length in words (excluding reparanda)

Length (words)	% cases	Mean IMI
1	70	319
2	5	338
3	4	347
5-6	4	410
7-8	4	393
9-12	4	478
13-52	5	505

Table 6. Production effect on IMI: size of unit initiated by current Move

Unit initiated	Mean IMI
Game-internal Move	304
Game	535
Transaction	617

3.2 Disfluency.

Disfluency also responds to putative sources of cognitive difficulty, but the effects differ considerably from the pattern we found for IMI. Among *interpersonal* effects, the sex effect appears in higher rates of disfluency for male IGs (.04). Since the same men take the IF role in other dialogues, this effect amounts to a sex x role interaction, with males more disfluent only when running dialogues. In contrast to IMI, disfluency displays a simple Familiarity effect which does not depend on the sexes of the dyad: speech to unfamiliar interlocutors is more disfluent overall than speech to familiar (.174 vs .140 dislfuencies/Move; -.03, $p < .05$).

The *task* effects now are restricted to a local order effect: though IMI falls across a dialogue, disfluency rates rise (.03, $p < .05$). (Table 7). The increase does not appear to be associated with any other measures of task difficulty.

Table 7. Order effect on disfluency rate: position of Move in dialogue

Quartile	Ordinal Position	Disfluencies / Move
1	1-55	.147
2	56-110	.135
3	111-185	.169
4	> 185	.167

In contrast to IMI, also, disfluency rates are insensitive to *comprehension* variables.

Table 8. Production effect on IMI: size of unit initiated by current Move

Unit initiated	Disfluencies / Move
Game-internal Move	.116
Game	.219
Transaction	.294

Disfluencies do show the predicted robust effects of *production and planning*. IGs, who usually take responsibility for directing the dialogue, are more disfluent than IFs (.181 v .087: $\beta = -.07$) Just as there were longer IMIs at the onsets of larger units of dialogue, there are more disfluencies (.02, $p < .05$) (Table 8). Disfluencies are affected by the complexity of the current Move: longer Moves attract more disfluencies (.28) (Table 9), as do Moves including more referring expressions (.06 for references to Given Shared landmarks, .03, $p < .05$, for New Unshared and .03 for Given Unshared) (Table 10).

Table 9. Production effects on disfluency rate: current Move length in words (excluding reparanda)

Length (words)	Disfluencies / Move
<1	.063
2-5	.099
6-9	.229
10-13	.378
14-17	.495
>17	.709

Table 10. Production effects on disfluency rate: number of referring expressions in current Move (excluding reparanda)

Referring expressions	Disfluencies / Move
0	.09
1	.22
2-5	.47

4. Discussion

These results show that objectively codable characteristics of a dialogue corpus reliably reflect psychological processes. For both IMI and disfluency, multiple regression results confirm the findings of studies exploring single cause-effect relationships. The current work goes beyond earlier studies in testing a range of potential sources of psycholinguistic difficulty against each dependent variable. As Figure 1 shows, the proportions of explained variance attributable to each group of predictors pattern quite differently in the two cases. IMI looks like a general measure of cognitive difficulty with independent contributions made by order, task difficulty, comprehension, production, and interpersonal factors. Disfluency, on the other hand, shows largely the effects of planning for production, with the greater part of the uniquely explained variance attributable to characteristics of the disfluent current Move itself.

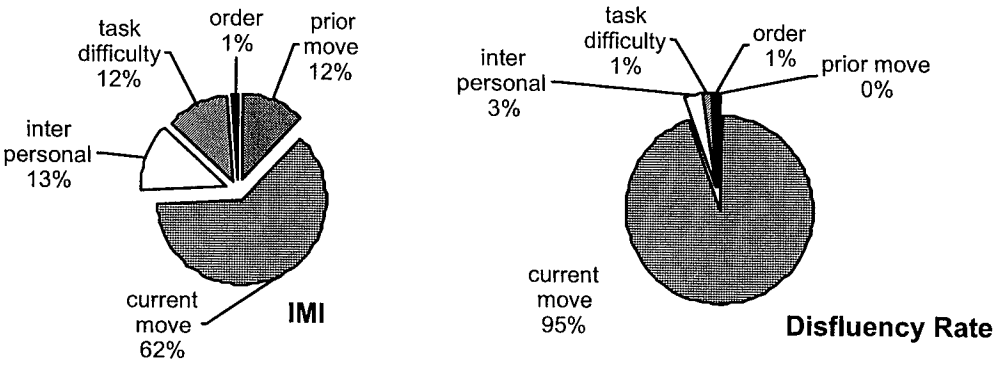


Figure 1. Percentages of uniquely explained variance in IMI and disfluency rate associated with sets of predictor variables

These outcomes have theoretical, methodological and technological implications. Models of speech production, for example, will have to consider why disfluency rate is insensitive to attested difficulties of comprehension: features of a prior utterance which ought to affect comprehension, and do adversely affect IMI, do not increase the rate of disfluency. Nor does it seem that the IMI absorbs the effects of comprehension difficulty, leaving processors clear for language production. If it did, then shorter IMIs, some of which might mark premature onset of the current Move, should make for more disfluency. But IMI is not a significant predictor of variance in disfluency rate and its removal from the equation does not enhance the apparent effects of prior-Move difficulty. Overt self-monitoring appears to be unrelated to factors that are supposed to be related to covert self-monitoring. Either the two kinds of monitoring are unrelated, with only covert monitoring processes interacting with comprehension, or the claim that comprehension processes participate in speech production is overstated.

Instead, the results confirm Clark and Wasow's finding for the effects of sentence length and add pragmatic, discourse and role effects: even when Move length in words is statistically controlled, disfluencies reflect the number of referring expressions being generated, the size of the discourse unit being initiated, or the degree of responsibility which the speaker bears for structuring the dialogue. It is tempting to conclude that all of these facets of speech production belong in a complete psychological model of the process.

It is also of interest that IMI is affected by prior-Move discourse position, length, and referential character, but not by prior move disfluency. Whatever difficulty listeners may have in recognizing words in reparanda (Bard and Lickley, 1998) or whatever advantage they may accrue in interpreting signalled disfluency (Brennan and Schober, 2001), none of this seems to affect the time taken to produce a reply. An explanation awaits further study of the comprehension of real speech in real communicative contexts.

The current techniques should help make such studies more viable. Our multiple regression results confirm several studies examining single cause-effect relationships. If basic results are evident whether or not other sources of variance are partialled out, then large corpora with simpler coding systems might still have something to offer psycholinguists in search of naturalistic on-line data. At the same time, designers of human-computer dialogues can use coded corpora to look for direct effects on delay to response or fluency of response with reasonable confidence that their observations can profit from our wider understanding of how interlocutors work.

Acknowledgements

The authors acknowledge the support of the Economic and Social Research Council, U.K., via HCRC and of the Engineering and Physical Sciences Research Council, U.K. (Speech and Language Technologies grant GR/L50280).

References

Anderson A., Bader M., Bard E. G., Boyle E., et al., (1991) *The H.C.R.C. Map Task Corpus.* Lang. & Speech, 34, pp. 351-366.

Beattie G. (1983) *Talk.* Open University Press, Milton Keynes.

Bard E. G., Aylett M., and Bull M. (2000) *More than a Stately Dance: Dialogue as a Reaction Time Experiment.* Proc. Soc. for Text & Discourse.

Bard E. G. and Lickley R. J. (1998) *Graceful failure in the recognition of running speech.* Proc Cognitive Science Soc., pp. 108-113.

Blackmer E. and Mitton J. (1991) *Theories of monitoring and the timing of repairs in spontaneous speech.* Cognition, 39, pp. 173-194.

Bock K. (1996) *Language production: Methods and methodologies,* Psychon Bull & Rev., 3, pp. 395-421.

Bortfeld H., Leon S., Bloom J., Schober M., and Brennan S. (2001) *Disfluency rates in conversation: effects of age, relationship, topic, role and gender.* Lang. & Speech, 44, pp. 123-147.

Branigan H., Pickering M., and Cleland, A. (2000) *Syntactic coordination in dialogue.* Cognition, 75, pp. 813-825.

Branigan H., Lickley R., and McKelvie D. (1999) *Non-linguistic influences on rates of disfluency in spontaneous speech.* Proc. 14th ICPhS.

Brennan S., and Schober M. (2001) *How listeners compensate for disfluencies in spontaneous speech.* J. Mem. & Lang., 44, pp. 274-296.

Bull M. (1998) *Inter-turn intervals in spontaneous speech.* PhD dissertation, U. of Edinburgh, U.K.

Bull M. and Aylett M. (1998) *An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue.* Proc. ICSLP'98.

Butterworth B. (1980) "Language production: Vol 1., Speech and Talk". Academic Press.

Carletta J., Caley R., and Isard S. (1995) *Simulating time-constrained language production.* Lang. & Cognitive Proc., 10, pp. 357-361.

Carletta J., Isard A., Isard S., Kowtko J., et al. (1997) *The reliability of a dialogue structure coding scheme.* Computational Linguistics, 23, pp. 13-31.

Clark H. H., and Wasow T. (1998) *Repeating words in spontaneous speech.* Cognitive Psych., 37, pp. 201-242.

Doherty-Sneddon G., Anderson A. H., O'Malley C., Langton S., et al., (1997) *Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance.* J Exp. Psych.: Applied, 3, pp. 105-125.

Isard A., and Carletta J. (1995) *Transaction and Action Coding in the Map Task Corpus.* Research paper HCRC/RP-65. HCRC, U. of Edinburgh, U.K.

Levelt W.J.M., Roelofs A., and Meyer, A. S. (1999) *A theory of lexical access in speech production.* Brain & Behav.Sci., *22*, pp. 1-45.

Levelt W.J.M. (1983) *Monitoring and self-repair in speech.* Cognition, 14, pp. 14-104.

Lickley R. J. (1994) *Detecting disfluency in spontaneous speech.* PhD Thesis, U. of Edinburgh, U.K.

Lickley R.J. (1998) *HCRC Disfluency Coding Manual.* Technical Report HCRC/TR-100, HCRC, U. of Edinburgh, U.K.

Lickley R. (2001) *Dialogue moves and disfluency rates.* Proc. of Disfluency in Spontaneous Speech '01, Edinburgh, U.K.

McLaughlin M., and Cody M. (1982) *Awkward silences: Behavioral antecedents and consequences of the conversational lapse.* Human Communication Res, 8, pp. 299-316.

Maclay H. and Osgood S. (1959) *Hesitation phenomena in spontaneous English speech.* Word, 15, pp. 19-44.

Oviatt S. (1995) *Predicting disfluencies during human-computer interaction.* Computer Speech & Lang., 9, pp. 19-35.

Pickering M. and Garrod S. *(in press) Toward a mechanistic psychology of dialogue.* Brain & Behav. Sci.

Postma A. and Kolk H. (1992) *The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs.* J. Speech & Hearing Res., 35, pp. 537-544.

Shriberg E. (1994) *Preliminaries to a theory of speech disfluencies.* PhD Thesis, U. of California at Berkeley.

Wheeldon L., and Lahiri A. (1997) *Prosodic units in speech production.* J. Mem. & Lang, 37, pp. 356-81.

Recoverability Optimality Theory: Discourse Anaphora in a Bidirectional framework

Adam Buchwald, Oren Schwartz, Amanda Seidl, and **Paul Smolensky***
The Johns Hopkins University
Department of Cognitive Science
243 Krieger Hall
3400 N. Charles St. Baltimore, MD 21218-2685, U.S.A

Abstract

In this work we explore the possibility of formalizing explanations based in the notion "deletion under recoverability": information should be omitted in expressions if that information can be recovered without being overtly expressed – e.g., from information already encoded in the state of the discourse. This style of explanation is difficult to formalize for many reasons, among them that the contradictory factors involved are not resolved in a simple, consistent way. Optimality Theory permits the conflicts to be formally adjudicated, and allows the conflicts to be resolved in different ways in different languages. We will present what to our knowledge is the first formal theory of the cross-linguistics typology of discourse anaphora systems.

1 Introduction to the framework

This work builds very directly on much previous research on discourse anaphora and bidirectional Optimality Theory (Beaver 2002, Blutner 2000, de Hoop and de Swart 2000, Hendriks and de Hoop 2001, Smolensky 1996a) and on early pragmatic work which discusses the crucial role of the speaker's notion of both interpretation and expression (e.g., Grice's *Logic and Conversation*, 1975). These connections are explicitly identified and discussed in later sections of the paper. But first, we begin with an intuitive introduction of how discourse is conceived in the framework we propose, called ROT for *Recoverability Optimality Theory*. The concepts introduced in this introductory section will be formally defined in the next section.

Consider the following discourse fragments. Each begins with (1a) and (1b), but sentence (1c) varies from t1 to t3.[1]

(1) a. Alice gave Betsy a bloody nose.
b. She$_A$ reminded her$_B$ that her right hook was devastating.
c. t1. She$_A$ hurt her$_B$.
t2. Betsy hit her$_A$ back.
t3. Alice hurt her$_B$.
d. She$_A$ felt awful.

The questions of interest to us are:

Q1 What principles determine the referent of each pronoun?

Q2 What are the intermediate and final states of the discourse?

Q3 Given a communicative intention, why does a speaker choose to reduce some NPs (to overt pronouns or null elements) but not others?

We take the native speaker intuitions concerning the pronoun referents — our data — to be those indicated by the subscripts in (1c), with A/B obviously connoting Alice/Betsy. Consider the case of t1: why is the interpretation not "She$_B$ hurt her$_A$"? In plain English, ROTs answer is this: In the state of the discourse when t1 is uttered, Alice is more salient than Betsy. The Subject is the most prominent argument and so gets its referent from the most salient discourse entity.

In ROT terminology, the state of the discourse when t1 is uttered is encoded by the *Salience List* SaL_{i-1} = [A(lice) B(etsy)]. In this simple declarative clause, the Subject occupies the first argument position of the predicate in the *Logical Form* LF_i = hurt(A, B) = V(A, B).[2] The alignment of the prominent Subject with the most salient entity favors the pairing of the first (left-most) element of SaL_{i-1} with the first argument of LF_i. This preference is asserted by the ROT constraint LINEARITY(LF_i, SaL_{i-1}); when formalized in 2, it will be clear that this constraint is indeed the standard LINEARITY constraint of OT's Correspondence Theory McCarthy and Prince (1995), applied to a new kind of correspondence relation.

* We would like to thank the anonymous reviewers for their valuable insights and helpful comments.

[1] Our intention in designing examples is that semantics/pragmatics not bias interpretations, leaving interpretations free to be governed by discourse syntax. For some readers, achieving this may require altering the particular verbs or arguments in the examples we give.

[2] We assume throughout that LF preserves information about the grammatical roles of its arguments, and that the overt NPs in our sentences are case-marked so there is no ambiguity about their argument positions, i.e., thematic roles.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 37-44.

With t2 in place of t1, we see that now it is the Direct Object that refers to A; now LF_i = V(B, A) even though SaL_{i-1} must be the same as before: [A B]. When the Subject is a pronoun "p" = "she" as in t1, the most salient discourse entity binds it; when the Subject is the Full NP "B" as in t2, the overt information overrides the salience-preference of A from the discourse, and the Subject is interpreted as B. This leaves the Object pronoun "p" needing a referent, which it can get from the most salient discourse entity, A.

ROT conceives this as follows. All features in the interpretation LF= V(x, y) must be licensed, or checked, by a matching feature which resides either in the overt expression uttered – the *Phonetic Form*, PF_i – or in the previously established discourse entities comprising SaL_{i-1}. As we have seen, constraints favor checking features with PF_i if possible; resort to SaL occurs only when PF is insufficient. (Sensible enough given that the overt expression is rather more unambiguously constraining.) So for t2, in V(x, y) it is preferable to check the features of x with the Subject of PF_i, which is possible iff x = B. A feature that is checked with PF_i will by convention be denoted by a Greek letter, one checked with SaL_{i-1}, a Latin letter. The preference for PF-checking is expressed in ROT simply as SΦ = Don't check a feature with SaL.

The distinction between features available in PFand those unavailable is formalized in ROT as follows. Nominal features will be divided into two groups: those realized in an overt pronoun "p" (e.g., [gender = F]), p*-features, and those realized in a Full NP but not in a "p" (e.g., [name = Betsy]), $f_\star$-features. We denote the former features by subscripts and the latter by superscripts; thus 'x = B' means something like $x = X_{[gender=F]}^{[name=Betsy]}$. So if B is expressed with a pronoun, the subscript features are licensed by PF and only the superscript feature need be licensed by SaL: this violates a constraint $S\Phi^k$. If B were expressed with a null element, $S\Phi_k$ would again be violated and in addition another constraint $S\Phi_k$ would be violated: now even the subscript features must be licensed by a discourse entity. When B is expressed by a full NP, none of the SΦ constraints are violated since all features are licensed by PF.

Consider now (1d). Why is the sole pronoun "p" = "she" interpreted as V(A) and not V(B)? Recall that in the interpretation V(x), x must have its superscript features licensed in SaL since the pronoun does not provide them in PF. The primary force of the term 'salience' is precisely that when features need to be licensed by the discourse, the features of the most salient entity in SaL are preferred; or equivalently, the features of less salient entities are more costly. In standard OT fashion this will be formulated as the universal fixed ranking: [SΦ2 = Don't license features by the second element of SaL] ≫ [SΦ1 = Don't license features by the first element of SaL].

Turning to Q2, we ask, why does the speaker not give the hearer a break, sparing her SΦ violations altogether, by always using full NPs? In ROT, the constraint opposing full overt feature expression is simply the all-encompassing economy constraint of OT: *STRUC = There is no structure. The ranking of *STRUC will determine how insistent the speaker is on economizing overt expression by reduction from full NPs to overt pronouns "p" or to null pronouns "∅". In standard OT fashion, the three-way distinction in quantity of structure – Full NP > "p" > "∅" – is encoded in *STRUC via the universal hierarchy *FULL-NP ≫ *PRON(or $*STRUC^f$ ≫ $*STRUC_{p-}$: a full NP violates both, "p" only $*STRUC^f$, and "∅" neither). We will compactly notate the PF Betsy hit her back as "BVp" – quotes included; the order is Subj V Obj, "A/B" denote the full NPs (Alice and Betsy), and "p" denotes an overt pronoun. (Null elements will be denoted "∅".)

Given *STRUCTURE's pervasive pressure to reduce, why are all NPs not totally reduced? In ROT, the answer is: for the Speaker the intended LF must be recoverable from the uttered PF, and if PF is overly reduced, it will be interpreted with the wrong LF. This is formalized via bidirectional optimization; to a first approximation, it works like this. The expressive optimization Exp is the standard OT optimization in which different PFexpressions compete to express a given LF interpretation. In interpretive optimization, Int, different LFs compete as interpretations of a given PF. The Speaker uses Exp to select the optimal PF to express their communicative intention, but all candidate expressions must pass a "pre-screening" Int that eliminates all PFs that will fail to yield the intended interpretation.[3] For Exp, what varies across candidates is PF_i; what is fixed, given — called the index of the optimization — is the discourse state after the previous utterance, SaL_{i-1}, and an intended interpretation LF_i, and an intended new SaL_i: as we see next, different expressions of the same LF will leave the discourse in different states; the optimal expression depends in part on the intended new SaL. For Int, what varies is the new interpretation LF_i and the new discourse state SaL_i; the given index is the discourse context SaL_{i-1} and the expression PF_i.

Finally, we take up Q3 and consider t3. What is the final state of the discourse? As a diagnostic, we can imagine adding a sentence after t3, e.g., (1d). Who does "p" refer to now? B. This tells us that after t3 the salience list is SaL= [B A], because from

[3] Pre-screening in a logical sense only: we make no claims about the time course of processing.

our discussion of (1d) above we know that a single pronoun must be interpreted as the most salient discourse entity (matching the pronoun's features). After t3 B is most salient, even though the Subject of t3 is A and even though A was more salient before t3. This is the power of pronominalization to shift salience. Because the pronoun must take its superscript features from SaL, it is optimal to place B at the head of SaL, where checking features is least costly. Note that the LF of both t1 and t3 is V(A, B), but they affect saliency differently: the resulting SaLs are [A B] and [B A], respectively.

Formalizing the intuitive accounts of this section is a major challenge. The next section systematically lays out the proposed solution, the principles of ROT.

2 The formal architecture of ROT

Before presenting our analysis of English and our typological results, a few of the concepts laid out in 1 require some elaboration. After the necessary discussions, the formalization of the system of constraints used in ROT is presented.

2.1 Recoverability and Bidirectionality

ROT is a theory of grammar. It defines the set of utterances available to discourse participants and the set of discourse conditions associated with those utterances.[4] ROT employs standard OT. In standard OT Prince and Smolensky (1993), Exp models production (performed by a speaker). The architecture of ROT is also derived by considering the point of view of the speaker. The *input* of Exp is an underlying form, and the *output* is the surface phonetic form (Smolensky 1996b). 161996bSmolensky) proposes the optimization be run in the opposite direction (Int), holding the surface form fixed as the index so that the set of candidates varies in their underlying form. This was intended as a model of comprehension (performed by a hearer). Importantly, the set of constraints is presumed to be identical in both directions of optimization.[5]

As described in 1, ROT uses the two directions of optimization to model a speaker's desire to ensure that his intended interpretation is recoverable from the surface form produced. Because ROT takes the perspective of the speaker, Int is essentially the speaker's model of the listener's interpretiveprocess. [6] As such, the speaker has access to the results and may examine the most harmonic candidate of Int, evaluated over the candidate phonological forms under consideration in Exp, to ensure that the most optimal form (evaluated in Exp), whose meaning is recoverable (evaluated in Int) is chosen to convey the intended meaning.

Following Wilson (2002), we employ a version of bidirectional OT that crucially relies on the interaction between directions of optimization. This interaction is motivated the desire of the speaker to ensure that the intended utterance is recoverable from the surface form produced.[7] The architecture can be formally restated as:

(2) In producing an utterance for an intended interpretation $I = \langle \mathrm{LF}_i, \mathrm{SaL}_i \rangle$, a speaker will use the form $F = \mathrm{PF}_i$ of the most most harmonic candidate $o = \langle I, F \rangle$, evaluated by $\mathsf{Exp}(I)$, such that there is no candidate $o' = \langle I', F \rangle \succ o$ (as evaluated by $\mathsf{Int}(F)$).

This formulation of bidirectional OT falls in between the "strong" and "weak" versions considered by Blutner (2000). In Blutner's terms, "strong" bidirection admits pairs $\langle I, F \rangle$ iff $F \in \mathsf{Exp}(I)$ (strong Q-principle) and $I \in \mathsf{Int}(F)$ (strong I-principle). "Weak" bidirection uses recursive versions of these principles:

(3) Q-principle: $\langle I, F \rangle$ satisfies Q iff $\langle I, F \rangle$ satisfies I and $F \in \mathsf{Exp}(I)$.
I-principle: $\langle I, F \rangle$ satisfies I iff $\langle I, F \rangle$ satisfies Q and $I \in \mathsf{Int}(F)$.

In order for "strong" and "weak" versions to be distinct requires an additional stipulation of the architecture: there must be a one-to-one correspondence between form and meaning. This causes an interpretation I_1 not to compete in $\mathsf{Int}(F_2)$ if there is a pair $\langle I_1, F_1 \rangle \succ \langle I_1, F_2 \rangle$, which allows the less harmonic pair $\langle I_2, F_2 \rangle$ will to be admitted into the language. This generates the phenomenon that "marked forms have marked meanings," (often referred to as *partial blocking*) The most important difference between our conception of bidirection and Blutner's is that we do not stipulate this one-to-one correspondence between form and meaning as a part of the architecture. ROT uses the weak Q-principle, but the strong I-principle, which allows for the possibility of

[4] Whether ROT is a viable model of language production is left as an open question for future work.

[5] Recent work in OT-Semantics Hendriks and de Hoop (2001), Zeevat (2000), inter alia, has taken a contrary position on this.

[6] The authors subscribe to the (possibly controversial) view that a speaker's utterances are influenced by her evaluation of whether or not she will be understood to have said what she meant. However, we do not claim that the speaker's model of the hearer's interpretation of a phonological form (which is the version of Intemployed by ROT) is necessarily consistent with the hearer's model of the same.

[7] Note that in the following definition, we omit SaL_{i-1} from the notation, understanding that as the prior context for a given utterance, SaL_{i-1} is fixed and is present in the indices of both directions of optimization for that utterance; note additionally that the shorthand $\langle a, b \rangle \in \mathsf{Int}(b)$ means that there is no pair $\langle a', b \rangle$ s.t. $\langle a', b \rangle$ is more harmonic than ($\succ$) $\langle a, b \rangle$, as evaluated by Int(b) i.e. that $\langle a, b \rangle$ does not lose the interpretive optimization over b.

ambiguous interpretations. Although adding the requirement that there be a one to one correspondance between form and meaning gives an account of partial blocking Blutner (2000), the same stipulation rules out neutralization phenomena, in which different underlying forms *do* have identical surface forms (in production).[8] It seems to us that the subtleties captured by imposing a one to one correspondence could be left to the mechanics of the constraints, and that an architecture should not rule out by definition any linguistic phenomena. Although at present we do not provide an account of partial blocking, our architecture does not seem to *a priori* rule this out.

2.2 Features, Topic, & Salience: from Centering Theory to ROT

Much of the inspiration for ROT derives from Centering Theory (CT: Grosz, Joshi, and Weinstein (1995)). In CT, the discourse context consists of a list of discourse entities (Centers), whose default ordering may vary from language to language. CT analyzes a discourse in terms of a sequence of local "discourse transitions" from one utterance to the next. The type of discourse transition between any two utterances is defined in terms of the changes in the values of (hidden) variables (backward- and forward-looking Centers), and organized in terms of the "local coherence" of discourse segments. These variables are themselves constructed concepts within CT, and their values are not always readily apparent. Given the fact that CT presupposes a good deal of syntactic and semantic analysis, we hypothesized that a careful, pragmatic treatment of the concept of Recoverability at the interface between syntax, semantics, and discourse would derive the same core set of results as Centering Theory.

Beaver (2002)'s important work on redefining CT in an OT framework (COT), replaces the Centers of CT with a system of constraints built around the concept of "Topic," an often used but still controversial notion.[9] We agree with Beaver's intuition that "topichood" is not always easy to unambiguously define. He suggests an OT analysis that selects a topic based on the more well founded concept of "salience." ROT takes another step in this direction, by dispensing of the notion of "topic" altogether, and allowing the whole system emerge from the interactions between constraints that do not depend on special-status elements. The motivation for our set of SΦ constraints is essentially the same as Beaver's salience-based analysis of topic. In future work, we hope to augment ROT with a richer set of features (such as empathy, -wa marking, cue validity, accessibility, etc.) and corresponding sets of constraints that belong to the SΦ family.

The specific types of features ROT presently employs are mostly left unspecified, given the restriction of the present analysis to this limited domain. At present, they are simply meant to correspond to features that may be present or absent from the particular referring expressions under consideration (such as gender, number, or name). Their ability to restrict the set of possible referents of the anaphoric expression is currently the relevant factor.

2.3 Features, Interpretation & Expression

As discussed in 1, in order for an utterance U_i to be interpreted, the entities in the argument positions of LF_i must have each of their features licenced by corresponding features in SaL_{i-1} or the overt form PF_i. This deserves some elaboration. SaL_{i-1} is meant to correspond to the speaker's internal model of the *common ground*. We assume (for now) that all of the entities in SaL_{i-1} appear there with all of their features.As such, in Int, it will always be possible to license any feature of an anaphoric entity that does not appear in PF_i, albeit incurring a cost via SΦ constraint violations. This makes sense in the (speaker's model of a hearer's) interpretive process, as it seems natural to require full feature set specification (or at least enough to uniquely identify the referents) for interpretation,[10] and for the speaker, such features are presumably available.

ROT also requires each feature of LF_i to be licenced in Exp, although this has a slightly different meaning. In Exp, which corresponds to production, the licensing requirement is more naturally thought of as checking each feature to ensure that the entire LF_i is actually expressed. Accordingly, the licensing of features in Exp takes place between LF_i, PF_i, and SaL_i. The identity of elements in SaL_i and their relative ordering is specified in the index of Exp, but since features are only required to be there there if they are not in PF_i (which varies among the set of candidates), SaL_i is not assumed to have features where they are not necessary. Rather, the entities in SaL_i receive the rest of their features as a consequence of becoming a part of the speaker's model of the *common ground* for the next utterance U_{i+1}.

2.4 Constraints

The constraints in ROT come from three families, and are formally defined below. As mentioned, all constraints are assumed to be present in every optimization. As in OT, each possible re-ranking of constraints (i.e., those that do not violate universal subhierarchies) generates a potential grammar of a natural language.

[8] For example, German requires devoicing of syllable-final consonants, which leads to the same relationship, e.g., /rat/ → /rat/ and /rad/ → /rat/

[9] See Jennifer Arnold's thesis (Arnold 1998) for a comprehensive review of different notions of Topic in the literature, and the relationships between Topic, Focus and salience.

[10] Leaving aside for the moment the complexities of underspecification in semantic or discourse theories.

Recoverability

In ROT, the SΦ constraints are often called "recoverability" constraints. SΦ constraints penalize the use of less salient positions for feature licensing. For the present feature set, this yields:[11]

(4) $S\Phi^i$: do not license a p*-feature with one at position i in SaL.
$S\Phi_i$: do not license a f_*-feature with one at position i in SaL.

with the universal rankings: $S\Phi^m \gg S\Phi^n$ iff $m > n$, and $S\Phi_m \gg S\Phi_n$ iff $m > n$.

Linearity

We use two constraints from the LINEARITY family of Faithfulness constraints (McCarthy and Prince 1995) that penalize candidates when the order of features (according to the hierarchy of grammatical role) that are in LF_i (but not PF_i) does not align linearly with the order of elements in SaL_{i-1} or SaL_i.

(5) $LIN\Phi_{i-1}$: For a pair of features x,y in both Sf and SaL_{i-1}, x precedes y in SaL_{i-1} iff x precedes y in Sf.
$LIN\Phi_i$: For a pair of features x,y in both Sf and SaL_i, x precedes y in SaL_i iff x precedes y in Sf.

In ROT, these constraints serve a double role: they introduce the relationship between salience and grammatical role;[12] and they ensure (indirectly and fallably) the coherence of successive utterances (their ultimate preference is for the orders of entities in SaL_{i-1} and SaL_i to align with each other, (and with obliqueness)).

Economy

Finally, we also use two economy constraints from the *STRUC constraint family that penalize phonological material in the candidate:

(6) *FULL: No Full-NPs
*PRON: No pronouns

There is a universal ranking among these two constraints, such that *FULL≫ *PRON.

3 The Mechanics

In this section, we illustrate the ROT account of discourse anaphora using examples from English. We assume throughout that a speaker has a fixed SaL_{i-1}, intended interpretation LF_i, and target SaL_i for any U_i. ROT defines four transitions over these dimensions:

t1: SaL_{i-1} = SaL_i AND LF_i is aligned with SaL_{i-1}
t2: SaL_{i-1} = SaL_i AND LF_i is NOT aligned with SaL_{i-1}
t3: $SaL_{i-1} \neq_{SaL}$[i] AND LF_i is aligned with SaL_{i-1}
t4: $SaL_{i-1} \neq SaL_i$ AND LF_i is NOT aligned with SaL_{i-1}

The following discourse segment from English illustrates these transitions:

(7) a. Alice gave Betsy a bloody nose.
b. She_A reminded her_B that her right hook was devastating.
c. **t1**. She_A hurt her_B.
t2. Betsy hurt her_A.
t3. Alice hurt her_B.
t4. Betsy hurt her_A.

ROT represents the transitive sentences in (7c) as the four-tuples in (3):[13]

(8) $\langle SaL_{i-1}, SaL_i, LF_i, "PF_i"\rangle$
a. **t1**. $\langle [A,B],[A,B],V(A,B),"pVp"\rangle$
b. **t2**. $\langle [A,B],[A,B],V(B,A),"BVp"\rangle$
c. **t3**. $\langle [A,B],[B,A],V(A,B),"AVp"\rangle$
d. **t4**. $\langle [A,B],[B,A],V(B,A),"BVp"\rangle$

To illustrate properties of ROT discussed above, we will demonstrate how the PFs for **t1** and **t2** are selected for English-type languages.

3.1 Int

As mentioned, ROT contains a formal mechanism to ensure that candidate expressions yield the intended interpretation: the evaluation in Int. Tableau 1 shows Int for **t1**, with an index of "she hurt her" (or "pvp") and SaL_{i-1} of [A, B]. The constraint ranking in Tableau 1 is the ranking that generates English. The quivered arrow symbol points to the most harmonic candidates in Int, representing the optimal interpretation.

In Tableau 1, the most harmonic candidate is (a), corresponding to the four-tuple **t1**. Candidates (b-d) each incur fatal violations of LINEARITY constraints: for candidates (c) and (d) the order of the features in LF that are not accessible from PF differs from the order in SaL_{i-1}; for candidates (b) and (c) they differ from the order in SaL_i . These candidates are *harmonically bound* by candidate (a), and will not be optimal under any constraint ranking. According to principles of OT that are employed in ROT,

[11] More generally, for each feature ϕ in the system, there is a corresponding set of $S\Phi\phi, k$ constraints (where k indexes the position in SaL used to license feature ϕ), with the universal ranking $S\Phi\phi, m \gg S\Phi\phi, n$ iff $m > n$.

[12] *Ceteris paribus*, the central cross-linguistic factor that determines the ordering of the Cf listts of CT (Grosz et al. 1995); also similar to the effects of the SALIENTARG constraint of COT Beaver (2002).

[13] ROT has two transitions (t2 and t4) that have the same PF. This occurs as there can be multiple candidates in Int for which no candidate that is more harmonic in the evaluation. Int evaluates the SΦ constraints over SaL_{i-1}, so (in some cases) two candidates that differ only in SaL_i will be equally harmonic.

this result will hold cross-linguistically; that is, ROT predicts that no language uses two pronouns to represent a change from SaL_{i-1} to SaL_i, or when the LF does not align linearly with SaL_{i-1}. The explanation afforded by ROT is that these interpretations are not recoverable from the PF.

	SaL_{i-1}: [A^1,B^2] PF_i: "p_α Vp_β" "She$_\alpha$ hurt her$_\beta$"	LIN ϕ_{i-1}	LINϕ $_i$	SF_2	SF_1	*FU LL	*PR ON	SF^2	SF^1
➽	a. LF$_i$: $V(A_\alpha{}^1,B_\beta{}^2)$ hurt (Alice$_\alpha{}^1$, Betsy$_\beta{}^2$) SaL_i: [A^1,B^2]						**	*	*
	b. LF$_i$: $V(A_\alpha{}^1,B_\beta{}^1)$ hurt (Alice$_\alpha{}^1$, Betsy$_\beta{}^1$) SaL_i: [B^2,A^1]		*!				**	*	*
	c. LF$_i$: $V(B_\alpha{}^2,A_\beta{}^1)$ hurt (Betsy$_\alpha{}^2$, Alice$_\beta{}^1$) SaL_i: [A^1,B^2]	*!	*				**	*	*
	d. LF$_i$: $V(B_\alpha{}^2,A_\beta{}^1)$ hurt (Betsy$_\alpha{}^2$, Alice$_\beta{}^1$) SaL_i: [B^2,A^1]	*!					**	*	*

Tableau 1: Int(SaL_{i-1}=[A B]; PF="pVp")

3.2 Exp

Recall that Exp evaluates PFs (much like standard OT), holding everything else in the candidate fixed. Tableau 2 shows Exp with an index SaL_{i-1} and SaL_i of [A,B], and LF_i of V(A,B). Although LF remains constant for each candidate PF, the features in LF are licensed differently; to facilitate the reader, the licensing for features in LF_i is specified within each candidate in Tableau 2, and the pattern of violations can be read off from the numerical sub- and superscripts of the candidate LFs. For example, candidate (a) contains full NPs for both arguments (and no numerical indices), so no SΦ constraints are violated, whereas both f*- and $p_\star$-features of candidate (e) are licensed by SaL_i (as seen in the indices), so this candidate violates each SΦ constraint.[14] Candidate (b) (the four-tuple from t1) is the optimal output in Exp under the constraint ranking of English. The crucial ranking ($S\Phi^f \gg$ PRON $\gg S\Phi_p$) reflects the fact that, in English, pronominalizing is preferred to full NPs, but having p*-features licensed by an element in SaL_i (as with ∅-forms) is worse than using a pronoun.

	SaL_{i-1}=[A^1,B^2]; SaL_i=[A,B]; LF=V(A,B) Hurt (Alice, Betsy)	SF_2	SF_1	*FU	*PR	SF^2	SF^1
	a. PF$_i$: "$A_\alpha{}^\alpha VB_\beta{}^\gamma$"; LF$_i$: $V(A_\alpha{}^\alpha,B_\beta{}^\gamma)$; SaL_i: [A,B]			*!*			
➽	b. PF$_i$: "$p_\alpha Vp_\beta$"; LF$_i$: $V(A_\alpha{}^1,B_\beta{}^1)$; SaL_i: [A^1,B^2]				**	*	*
	c. PF$_i$: "$p_\alpha VH_\beta{}^{\gamma}$" LF$_i$: $V(A_\alpha{}^1,B_\beta{}^\gamma)$; SaL_i: [A^1,B]			*!	*		*
	d. PF$_i$: "$A_\alpha{}^\alpha Vp_\beta$" LF$_i$: $V(A_\alpha{}^\alpha,B_\beta{}^2)$; SaL_i: [A,B^2]			*!	*	*	
	e. PF$_i$: "ØVØ"; LF$_i$: $V(A_\alpha{}^\alpha,B_\beta{}^\gamma)$; SaL_i:[$A_1{}^1$,$B_2{}^2$]	*!	*			*	*
	f. PF$_i$: "ØV$B_\beta{}^{\gamma}$"; LF$_i$: $V(A_\alpha{}^\alpha,B_\beta{}^\gamma)$; SaL_i: [$A_1{}^1$,B]		*!	*			*
	g. PF$_i$: "$A_\alpha{}^\alpha$VØ"; LF$_i$: $V(A_\alpha{}^\alpha,B_\beta{}^\gamma)$; SaL_i:[A,$B_2{}^2$]	*!		*		*	
	h. PF$_i$: "ØVp_β"; LF$_i$: $V(A_\alpha{}^\alpha,B_\beta{}^\gamma)$; SaL_i:[$A_1{}^1$,B_2]	*!			*	*	*
	i. PF$_i$: "p_αVØ"; LF$_i$: $V(A_\alpha{}^\alpha,B_\beta{}^\gamma)$; SaL_i:[A_1,$B_2{}^2$]		*!		*	*	*

Tableau 2: Exp(SaL_{i-1}=[A,B];SaL_i=[A,B];LF=V(A,B))

3.3 The Interaction of Int and Exp

As discussed earlier, English speakers who want to change the LF and reorder the SaLs use utterances that conform to the four-tuple identified in t2. Tableau 3 shows Exp for this situation.

[14] As each candidate violates LINΦ_{i-1} and satisfies LINΦ_i, these constraints do not figure in the outcome of this optimization, and are omitted. The gray candidates in Tableau 2 are those that do not yield the intended interpretation (see Appendix for the full description of Int).

	SaL_{i-1}=[S^1,B^2]; SaL_i=[B,S]; LF=BSV	SF_2	SF_1	*FU	*PR	SF^2	SF^1
	a. $V(B_\alpha{}^\alpha,S_\beta{}^\gamma)$; "$B_\alpha{}^\alpha VS_\beta{}^\gamma$"; SaL_i=[B,S]			**!			
	b. $V(B_\alpha{}^2,S_\beta{}^1)$; "$p_\alpha Vp_\beta$"; SaL_i=[B^2,S^1]				**	*	*
	c. $V(B_\alpha{}^2,S_\beta{}^1)$; "$p_\alpha VS_\beta{}^\gamma$"; SaL_i=[B^2,S)			*	*		*
➽	d. $V(B_\alpha{}^\alpha,S_\beta{}^1)$; "$B_\alpha{}^\alpha Vp_\beta$"; SaL_i=[B,S^1]			*	*	*	
	e. $V(B_2{}^2,S_1{}^1)$; "ØVØ"; SaL_i=[$B_1{}^2$,$S_2{}^1$]	*	*			*	*
	f. $V(B_2{}^2,S_\beta{}^1)$; "ØV$S_\beta{}^\gamma$"; SaL_i=[$B_2{}^2$,S]	*		*			
	g. $V(B_\alpha{}^\alpha,S_1{}^1)$; "$B_\alpha{}^\alpha$VØ"; SaL_i=[B,$S_1{}^1$]		*!	*			*
	h. $V(B_2{}^2,S_\beta{}^1)$; "ØVp_β"; SaL_i=[$B_2{}^2$,S^1]		*		*	*	*
	i. $V(B_\alpha{}^2,S_1{}^1)$; "p_αVØ"; SaL_i=[B^2,$S_1{}^1$]	*			*	*	*

Tableau 3: Exp(SaL_{i-1}=[A,B];SaL_i=[B,A];LF=V(B,A))

Of the candidates in Tableau 3 that are not neutralized in Int, the constraint alignment of English favors candidate (d), the four-tuple t2. The need to license p-features in PF rules out candidate (g), and candidate (a) incurs two violations of *FULL-NP, the second being fatal.

Tableaux 1 and 3 demonstrate the import of the interaction between Int and Exp. As stated, all candidates that compete in Exp must yield the target interpretation. Although candidate (b) would be the most harmonic candidate in Tableau 3 with this constraint ranking, it does not compete in Exp as it was ruled out in Tableau 1 (candidate (d)). This demonstrates the "bidirectionality" of the ROT architecture.

In this section, we have used a basic analysis of the use of English anaphora to demonstrate how the ROT architecture generates a grammar. Crucially, the interaction of Int with Exp *under the same constraint ranking* rules out certain candidates that do not yield the target interpretation. In the next section, we will see how properties of constraint interaction inherent in OT provide a framework for thinking about the cross-linguistic variation in the use of anaphora.

4 Typology in ROT and other considerations

Inherent in any OT analysis are predictions about cross-linguistic typology. Constraints in OT can be re-ranked, and each possible ranking is a *potential* natural language. To assess the typological predictions of ROT, we examine the results of the full set of Int and Exp for each possible constraint ranking.[15] ROT generates several well-attested language types, listed in Table 3.1. L1 requires Full NPs for any of the four transitions. We refer to this language as children's storybook English. Note that there is no ambiguity in this language, since any PF of the utterance will crucially depend only on LF. English, as discussed in the previous section, fits type L2, although it may also fit type L3. Japanese and Turkish seem to fit the pattern outlined in L4, in which zero pronouns are used for salient discourse referents (Walker, Iida, and Cote 1994, Kameyama 1998, Turan 1995). If we consider pro to be a zero, Italian

[15] This diverges from other prominent treatments of semantics using bidirectional OT (e.g., Beaver 2000, Hendriks and de Hoop 2000, Blutner 2000).

and Greek fit the description of L7 (and possibly L6; Di Eugenio (1990, 1998), Dimitriadis (1996)).[16]
Table 1: Typology

L	t1	t2	t3	t4	Exemplar
L1	"AVB"	"BVA"	"AVB"	"BVA"	Storybook
L2	"pVp"	"BVp"	"AVp"	"BVp"	English(1)
L3	"pVB"	"BVp"	"AVp"	"BVp"	English(2)
L4	"∅V∅"	"BV∅"	"∅VB"	"BV∅"	Japanese
L5	"∅VB"	"BVp"	"AVp"	"BVp"	
L6	"∅Vp"	"BV∅"	"AVp"	"BV∅"	
L7	"∅Vp"	"BVp"	"AVp"	"BVp"	Italian

Each of these languages represents a particular ranking of constraints such that L1's output is realized by a ranking of all SΦ constraints over *STRUC constraints ($S\Phi^{1,2} \gg^*$ STRUC) and L4 and L7 are generated by $S\Phi^2 \gg^*$ FULL $\gg^*$ PRON $\gg S\Phi^1$ and *FULL $\gg^*$ PRON $\gg S\Phi^2$ respectively.

The predictions of ROT are necessarily limited by the distinctions among features and expressions (e.g., Full NP, pronoun, null). As mentioned in section 2, ROT is designed to be augmented. For example, future work may include expressions "in between" pronoun and zero, such as subject agreement on a verb, or clitics. Alternatively, identifying further classes of features would allow us to use finer-grained SΦ constraints. Changes such as these are necessary to broaden the predicted typology, allowing ROT to become more precise in its cross-linguistic account of anaphora. At present, ROT seems promising as a framework for the study of linguistics at the interfaces of several of its subfields.

References

Arnold, Jennifer. 1998. Reference form and discourse patterns. Doctoral Dissertation, Stanford University.

Beaver, David. 2002. The optimization of discourse anaphora. *Linguistics and Philosophy* To appear.

Blutner, Reinhard. 2000. Some aspects of optimality in natural language interpretation. In de Hoop and de Swart (2000), 1–21.

Clark, H. H., and Susan E. Brennan. 1991. Grounding in communication. In *Perspectives on Socially Shared Cognition*, ed. L. B. Resnick, J. M. Levine, and S. D. Teasley, 127–149. APA Books.

Di Eugenio, Barbara. 1990. Centering theory and the Italian pronominal system. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING '90)*, 270–275.

Di Eugenio, Barbara. 1998. Centering in Italian. In Walker, Joshi, and Prince (1998).

Dimitriadis, Alexis. 1996. When pro-drop languages don't: Overt pronominal subjects and pragmatic inference. In *CLS 32: The Main Session*, ed. Lise M. Dobrin, Kora Singer, and Lisa McNair, 33–47. Chicago Linguistic Society.

Grice, H. P. 1975. Logic and conversation. In *Syntax and Semantics*, ed. P. Cole and J. L. Morgan, 41–58. Academic Press.

Grosz, Barbara A., Aravind Joshi, and Scott Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics* 21:203–226.

Hendriks, P., and Helen de Hoop. 2001. Optimality theoretic semantics. *Linguistics and Philosophy* 21:1–32.

de Hoop, Helen, and Henriette de Swart, ed. 2000. *Papers on Optimality Theoretic Semantics*. Utrecht Institute of Linguistics OTS.

Kameyama, Megumi. 1998. Intrasentential centering: A case study. In Walker et al. (1998).

McCarthy, John J., and Alan S. Prince. 1995. Faithfulness and reduplicative identity. Ms., U. Mass. and Rutgers, July 1995.

Prince, Alan, and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical report, Rutgers University, New Brunswick, NJ and University of Colorado at Boulder.

Smolensky, Paul. 1996a. Generalizing optimization in ot: A competence theory of grammar 'use'. Paper presented at the Stanford Workshop on Optimality Theory, Stanford University.

Smolensky, Paul. 1996b. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27:720–731.

Turan, Ümit Deniz. 1995. Null vs. overt subjects in Turkish discourse: A centering analysis. Doctoral Dissertation, University of Pennsylvania.

Walker, Marilyn, Masayo Iida, and Sharon Cote. 1994. Japanese discourse and the process of centering. *Computational Linguistics* 21:193–232.

Walker, Marilyn, Aravind Joshi, and Ellen Prince, ed. 1998. *Centering theory in discourse*. New York: Oxford University Press.

Wilson, Colin. 2002. Bidirectional optimization and the theory of anaphora. In *Optimality-theoretic Syntax*, ed. Géraldine Legendre, Jane Grimshaw, and Sten Vikner. Cambridge: MIT press.

Zeevat, Hank. 2000. The assymetry of optimality theoretic syntax and semantics. *Journal of Semantics* 17:243–262.

[16] Miltsakaki (2001) notes that there are three possible reduced realizations for entities in Greek (as in many other languages) pro, weak pronouns, and strong pronouns, our typology does not yet handle this degree of variation. Nonetheless, we hope we have made clear that such variation is easily dealt with in ROT.

5 Appendix

I1	SaL_{i-1}= [A^1,B^2]; "AVB"	LI (I-1)	LIN (I)	*F U	*P R	*X^2	*X^1
➽	a. [A,B]; V(A_α^μ,B_β^ν)			**			
➽	b. [B,A]; V(A_α^μ,B_β^ν)			**			

I2	SaL_{i-1} [A^1,B^2]; "pVB"	LN (I-1)	LN (I)	*F U	*P R	* X^2	* X^1
➽	a. [A^1,B]; V(A_α^1,B_β^ν)			*	*		*
➽	b. [B,A^1]; V(A_α^1,B_β^ν)			*	*		*
	c. [B]; V(B_α^2,B_β^ν)	*		*	*	*	

I3	$SaL_{(i-1)}$= [A^1,B^2]; "AVp"	LIN (I-1)	LIN (I)	*F U	*P R	* X^2	* X^1
➽	a. [A,B^2]; V(A_α^μ,B_β^2)			*	*	*	
➽	b. [B^2,A]; V(A_α^μ,B_β^2)			*	*	*	
	c. [A]; V(A_α^μ,A_β^1)	*		*	*		*

I4	$SaL_{(i-1)}$= [A^1,B^2]; "BVp"	LIN (I-1)	LIN (I)	*F U	*P R	* X^2	* X^1
➽	a. [B,A^1]; V(B_α^μ,A_β^1)	*		*	*		*
➽	b. [A^1,B]; V(B_α^μ,A_β^1)	*		*	*		*
	c. [B]; V(B_α^μ,B_β^2)	*		*	*	*	

I5	$SaL_{(i-1)}$= [A^1,B^2]; "pAV"	LIN (I-1)	LIN (I)	*F U	*P R	* X^2	* X^1
	a. [B^2,A]; V(B_α^2,A_β^ν)	*		*	*	*	
	b. [A,B^2]; V(B_α^2,A_β^ν)	*		*	*	*	
➽	c. [A]; V(A_α^1,A_β^ν)	*		*	*		*

I6	$SaL_{(i-1)}$= [A^1,B^2]; "BVA"	LIN (I-1)	LIN (I)	*F U	*P R	* X^2	* X^1
➽	a. [B,A]; V(B_α^μ,A_β^ν)	*		*	*		
➽	b. [A,B]; V(B_α^μ,A_β^ν)	*		*	*		

I7	$SaL_{(i-1)}$ = [A,B]; "pV∅"	LIN I-1	LIN (I)	*P R	* X_2	* X_1	* X^2	*X^1
➽	a. [A,B]; V(A^1,B_2^2)			*	*		*	*
	b. [B,A]; V(A^1,B_2^2)		*	*	*		*	*
➽	c. [B,A]; V(B^2,A_1^1)	*		*		*	*	*
	d. [A,B]; V(B^2,A_1^1)	*	*	*		*	*	*

I8	$SaL_{(i-1)}$ = [A,B]; "AV∅"	LIN I-1	LIN (I)	*F U	*X_2	*X_1	*X^2	*X^1
➽	a. [A,B]; V(A,B_2^2)			*	*		*	
➽	b. [B,A]; V(A,B_2^2)			*	*		*	
	c. [A]; V(A,A_1^1)	*		*		*		*

I 9	$SaL_{(i-1)}$ = [A,B]; "∅Vp"	LIN I-1	LIN (I)	*P R	*X_2	*X_1	*X^2	*X^1
➽	a. [A,B]; V(A_1^1,B^2)			*		*	*	*
	b. [B,A]; V(A_1^1,B^2)		*	*		*	*	*
	c. [B,A]; V(B_2^2,A^1)	*		*	*		*	*
	d. [A,B]; V(B_2^2,A^1)	*	*	*	*		*	*

I 10	$SaL_{(i-1)}$ = [A,B]; "BV∅"	LIN I-1	LIN (I)	*F U	*X_2	*X_1	* X^2	*X^1
➽	a. [A,B]; V(B,A_1^1)	*		*		*		*
➽	b. [B,A]; V(B,A_1^1)	*		*		*		*
	c. [B]; V(B,B_2^2)	*		*	*		*	

I 11	$SaL_{(i-1)}$ = [A,B]; "∅VB"	LIN I-1	LIN (I)	*F U	*X_2	*X_1	*X^2	*X 1
➽	a. [A,B]; V(A_1^1,B)			*		*		*
➽	b. [B,A]; V(A_1^1,B)			*		*		*
	c. [B]; V(B_2^2,B)			*	*		*	

I 12	$SaL_{(i-1)}$ = [A,B]; "∅VA"	LN I-1	LIN (I)	*F U	*X 2	*X_1	*X^2	*X^1
	a. [A,B];V(B_2^2,A)	*		*	*		*	
	b. [B,A];V(B_2^2,A)	*		*	*		*	
➽	c. [A]; V(A_1^1,A)			*		*		*

I 13	$SaL_{(i-1)}$ = [A,B]; "∅V∅"	LN I-1	LIN (I)	*X_2	*X_1	*X^2	*X^1
➽	a. [A,B]; V(A_1^1,B_2^2)			*	*	*	*
	b. [B,A]; V(A_1^1,B_2^2)		*	*	*	*	*
	c. [B,A]; V(B_2^2,A_1^1)	*		*	*	*	*
	d. [A,B]; V(B_2^2,A_1^1)	*	*	*	*	*	*

Exp(t3)

SaL_{i-1} = [A,B]; V(A,B); SaL_i = [B,A]	LIN I-1	LIN (I)	*F U	*P R	* X_2	* X_1	* X^2	* X^1
a. "AVB"; V(A,B)			**					
b. "AVp"; V(A, B_2)			*	*			*	
c. "∅VB", V(A_1^1, B)	*		*			*		*

Using Dependent Record Types in Clarification Ellipsis

Robin Cooper
Dept of Linguistics
Göteborg University
Box 200, 405 30 Göteborg,
Sweden
cooper@ling.gu.se

Jonathan Ginzburg
Dept of Computer Science
King's College, London
The Strand, London WC2R 2LS
UK
ginzburg@dcs.kcl.ac.uk

Abstract

We present a sketch of a formulation of an analysis of clarification ellipsis using dependent record types as they have been developed in Martin-Löf type theory. Record types provide a semantic formalism which at the same time as containing the normal paraphernalia of semantics (functions, binding, quantification) also have strong similarities to typed feature structures as used in HPSG. We argue that this gives us the kind of tools we need to account for the interpretation of clarification ellipses which to a large extent need to be constructed from semantic, syntactic and phonological information about the utterance being clarified. Clarification ellipses thus motivate a radical kind of context dependence over and above compositional semantics. Their meanings need to be constructed by the manipulation of structured representations of previous utterances which include more than just semantic information.

1 Introduction

Ginzburg and Cooper (2001) present a sketch of an account of clarification ellipsis (CE) in HPSG. They discuss examples such as (1).

(1) a. A: Did Bo finagle a raise?
B: (i) Bo?/ (ii) finagle?

b. **Clausal reading**: Are you asking if BO (of all people) finagled a raise/Bo FINAGLED a raise (of all actions)

c. **Constituent reading**: Who is Bo?/What does it mean to finagle?

CE is a good example of a dialogue process that emphasizes the fact that interpretation in dialogue is achieved interactively. Dialogue participants do not automatically agree on the contents of utterances (as might be supposed, for example, from a classical approach to formal semantics) and they often need to engage in dialogue in order to come to a better understanding of what was said. Ginzburg and Cooper (ms) argue in detail that the complexity of the contents involved in cases of CE rules out the viability of syntactically-based operations such as copying/deletion or even more sophisticated methods on logical forms such as higher order unification as general methods for resolution. And yet, CE manifests a high degree of syntactic and in some cases phonological parallelism between source and target. This is the reason why their original account employed HPSG-like signs, entities that encode in parallel phonological, syntactic, semantic, and contextual information.

In this paper we sketch an alternative account using dependent record types as they have been developed in Martin-Löf type theory (Betarte, 1998 and Tasistro, 1997 and also work in progress by Thierry Coquand and Randy Pollock). Among the advantages we see for a formulation in terms of dependent record types are:

- Record types are more suitable for semantic notions such as quantification and functions which have to be encoded in unification-based feature structures (cf Pollard, 2001).
- We are thus able to replace signs with functions from contexts to contents in something like the familiar semantic way while at the same time maintaining the insight of HPSG that phonological, syntactic and semantic information should be incorporated

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 45-52.

in record/feature structures and given an equal and parallel status.

- Dependent record types are inherently dynamic and provide a DRT-like analysis of anaphora.
- Record types provide us with structures on which it is straightforward to define the coercion operations from Ginzburg and Cooper (2001) which are used to generate the interpretations of clarification requests from the interpretations of the utterance being clarified.

2 Record types and typed feature structure

In this section we will attempt to explain why we think it is worthwhile exploring the use of record types as opposed to typed feature structures as a framework in which to cast our analysis of clarification ellipsis. After a brief sketch of the theory of records we have in mind, we will take each of the points mentioned at the end of the previous section in turn.

Records and record types The following informal account of records and record types is taken from Cooper (2000) which in turn follows Betarte and Tasistro (1998), Tasistro (1997) and Betarte (1998). The basic idea is represented informally as follows:

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

$$\begin{bmatrix} l_1 & = & a_1 \\ l_2 & = & a_2 \\ \ldots & & \\ l_n & = & a_n \end{bmatrix}$$

is of type:

$$\begin{bmatrix} l_1 & : & T_1 \\ l_2 & : & T_2(l_1) \\ \ldots & & \\ l_n & : & T_n(l_1, l_2, \ldots, l_{n-1}) \end{bmatrix}$$

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

Consider the record type in (2).

$$(2)\quad \begin{bmatrix} x & : & \text{Ind} \\ c_1 & : & \text{man}(x) \\ y & : & \text{Ind} \\ c_2 & : & \text{donkey}(y) \\ c_3 & : & \text{own}(x,y) \end{bmatrix}$$

Here we have used the labels x and y for fields whose values are required to be of type Ind(ividual) and c_i for what a linguist would think of as a constraint but a type theorist would think of as a proof. In Martin-Löf type theory propositions are themselves types, types of proof. A record which is of the type (2) would be of the form (3).

$$(3)\quad \begin{bmatrix} x & = & a \\ c_1 & = & p_1 \\ y & = & b \\ c_2 & = & p_2 \\ c_3 & = & p_3 \end{bmatrix}$$

where
a, b are of type Ind, individuals
p_1 is a proof of man(a)
p_2 is a proof of donkey(b)
p_3 is a proof of own(a, b)

What constitutes a proof of a particular proposition is, at least in our view, dependent on what kind of theory you want to link the type theoretic account to. In a model theoretic setting it might be a proof of truth of the proposition in a model. In a computational setting it might be the result of a database lookup or the result of running a theorem prover given assumptions about what objects are men, donkeys and who owns what. What makes our type dependent is the fact that the propositions to be proved depend on which choice of individual you make. That is, the propositions, i.e. types of proofs, are dependent types. The order of the fields is important. You have to introduce the objects on which types depend before the types themselves.

$$(4)\quad \left[\, f \;:\; \left(\left[\begin{array}{lcl} x & : & \mathrm{Ind} \\ c_1 & : & \mathrm{man}(x) \end{array}\right]\right) \left[\begin{array}{lcl} y & : & \mathrm{Ind} \\ c_2 & : & \mathrm{donkey}(y) \\ c_3 & : & \mathrm{own}(x,y) \end{array}\right] \,\right]$$

Denoting semantic objects Record types such as (2) are good candidates for semantic objects, we think. There is quite obviously a close relationship between them and discourse representation structures. Like DRSs they can considered as truth-bearing objects. A type can be considered true just in case it is inhabited, that is, if there is at least one object of the type. In the case of (2) this will involve finding at least one individual x who we can show to be a man and an individual y who we can show to be a donkey and also be able to show that x owns y. Thus we get the effect of existential quantification just as in DRT. One possible analysis of *every man owns a donkey* uses function types as in (4). (4) is true just in case we can find a function that will map any individual who is a man to a donkey which he owns (or more precisely records representing this information). The parallel with the DRT '$\Rightarrow$' is exact.

Combining syntactic, phonological and semantic information in a single framework The intruiging thing about dependent record types is that they not only behave like DRSs but they also, given that they consist of fields, look very much like typed feature structures. We have begun work on recasting HPSG grammars directly in terms of dependent record types and while it is still an open question as to whether the dependent record types provide you with everything you need for full HPSG it does look promising. (5a) is a small example of part of an HPSG sign (taken from Ginzburg and Sag, 2000) and (5b) is how its translation into a dependent record type could look.[1]

We give this example to give a taste of how exact the correspondence might be between HPSG and dependent record types without further discussion. While it is still an open question whether all of HPSG could be mimicked it is clear that record types can be used to represent syntactic and phonological information as well as semantic information and thus a central part of the HPSG philosophy, namely that semantic, syntactic and phonological information can be integrated in single structures, can also be realized by using record types. But in that case, why not use HPSG as we did in our previous paper? The attraction of using record types is that we can both follow the HPSG philosophy and have all of the paraphernalia of formal semantics close at hand such as functions, quantification and binding. Typed feature structures, being unification based, are just not really equipped to deal with semantic objects. Often semantics in HPSG involves an encoding of semantic representations in feature structures (faking the bound variables since they do not exist in unification-based systems). The semantics of typed feature structures is oriented towards the description of syntactic objects such as natural language expressions or semantic representations and not towards the semantic objects denoted by these expressions. If a semanticist wishes to express, for example, that meaning is a function from contexts to contents this cannot be expressed directly in HPSG. The best you can do is create some particular kind of feature structure and say that this is used to encode such functions. The typed feature structure semantics itself will not denote such a function. Record types, on the other hand, seem to offer us both semantic tools and feature structures in a single formalism and as such seem very attractive.

Dynamic treatment of anaphora The parallels between type theory and DRT extend to the treatment of anaphora. Thus (6) can be used as the record type corresponding to *every man who owns a donkey beats it.* Discourse anaphora can also be treated in an exactly parallel way to classical DRT. Thus *A man owns a donkey. He beats it.* can be represented by

[1] The new work on record types by Coquand and Pollock provides the possibility of an even stronger kind of equality which would be represented here as

arg-st.0 = π_1 : Synsem

This kind of equality requires exact structural correspondence, i.e. "definitional equality", not simply that the two expressions evaluate to the same value.

(5) a.

$$\begin{bmatrix} \textit{lexeme} \\ \text{PHONOLOGY} & < \text{prove} > \\ \text{SYNSEM} & \begin{bmatrix} \textit{synsem} \\ \text{LOC} & \begin{bmatrix} \textit{local} \\ \text{CAT} & \begin{bmatrix} \textit{cat} \\ \text{HEAD} & \dots \\ \text{SUBJ} & < \boxed{1} > \\ \text{SPR} & < > \\ \text{COMPS} & < \boxed{2} > \end{bmatrix} \\ \text{CONT} & \dots \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & < \text{NP}_{\boxed{1}}, \text{NP}_{\boxed{2}} > \end{bmatrix}$$

b.

We use π_1 to represent the path `synsem.local.cat.subj.0` and π_2 for `synsem.local.cat.comps.0`. (The '0's here represent the first member of a list.)

$$\begin{bmatrix} \text{phon} & : & < \text{prove} > \\ \text{synsem} & : & \begin{bmatrix} \text{local} & : & \begin{bmatrix} \text{cat} & : & \begin{bmatrix} \text{head} & : & \dots \\ \text{subj} & : & < \text{NP} > \\ \text{spr} & : & < > \\ \text{comps} & : & < \text{NP} > \end{bmatrix} \\ \text{cont} & : & \dots \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{arg-st} & : & < \text{NP,NP} > \\ \text{constraint-arg-st-0} & : & \text{eq(Synsem,arg-st.0,}\pi_1\text{)} \\ \text{constraint-arg-st-1} & : & \text{eq(Synsem,arg-st.1,}\pi_2\text{)} \end{bmatrix}$$

$$(6) \quad \begin{bmatrix} f & : & \left(\begin{bmatrix} x & : & \text{Ind} \\ c_1 & : & \text{man}(x) \\ y & : & \text{Ind} \\ c_2 & : & \text{donkey}(y) \\ c_3 & : & \text{own}(x,y) \end{bmatrix} \right) \begin{bmatrix} c_4 & : & \text{beat}(x,y) \end{bmatrix} \end{bmatrix}$$

adding to (2) as in (7).

$$(7) \quad \begin{bmatrix} x & : & \text{Ind} \\ c_1 & : & \text{man}(x) \\ y & : & \text{Ind} \\ c_2 & : & \text{donkey}(y) \\ c_3 & : & \text{own}(x,y) \\ c_4 & : & \text{beat}(x,y) \end{bmatrix}$$

Cooper (2000) shows how static and dynamic meanings can be related to each other. The notion of introducing labels and having types which are dependent on previously introduced labels is central to the theory of dependent types and provides what is needed for a dynamic treatment of anaphora.

Structured objects on which to define transformations between types Record types provide us with the kind of functional paraphernalia we need to do semantics but, unlike classical extension based accounts, such as Montague's, they also provide us with richly structured objects since the various components that are used to construct the semantic objects are provided in labelled fields. This is extremely important for our treatment of clarification ellipsis, where the meaning of the clarification ellipsis, we claim, must be constructed out of

parts of the meaning of the previous utterance by the other dialogue participant. What we are proposing is a radical kind of anaphoric dependence where the meaning of the previous utterance can be cannibalized to construct the meaning of the current utterance.

3 Analysis of CE

In this section we illustrate with respect to an example how an analysis of CE could be formulated using dependent record types.[2]

In order to analyze utterances we will use families of dependent record types, i.e. functions from records to record types. We will use the following notation for these functions: $\lambda r : T_1(T_2)$ represents a function from records of type T_1 to the type T_2 (dependent on r). T_1 is the type of the context required to interpret the utterance. T_2 is the type which constitutes the content of the utterance (resulting from application to the context).

We will represent utterances with vertices between the words (as is common in chart representations) as in (8):

(8) u_1 : $_0$ Did $_1$ Bo $_2$ leave $_3$

We will use $u_{1,0-1}$ to represent the utterance of *did* in u_1.

When we use families of record types to analyze such an utterance we will use an abbreviation:

$\left[\begin{array}{lcl} f_{u_{i,n-m}} & : & T \end{array} \right]$

is to represent

$\left[\begin{array}{lcl} f_{u_{i,n-m}} & : & T \\ pf\text{-}f_{u_{i,n-m}} & : & f(u_{i,n-m}, f_{u_{i,n-m}}) \end{array} \right]$

We give an example in (9). In (9), in order to provide the utterance time we need to introduce an object of type Time, and a proof that this is the utterance time for the object. Since labels are arbitrary they do not of themselves require that anything is true of the object introduced.

In (10) we give an analysis of (8).

The labels for referent, event-time, speaker and hearer refer back to the context r. More properly, we should write as in (11) where each of these labels is explicitly marked as occurring in r.

[2]We defer to the longer version of this paper a reformulation of the coercion rules and the grammatical analysis developed in Ginzburg and Cooper (2001,ms).

However, we will suppress all the extra r's when there is no risk of confusion in order to make things more readable.

The context is required to contain information about phonology, utterance and event time and the relation between them as well as referents, speaker, hearer and syntactic category. Note that, given the definition of typing for records, that the context-record is required to contain at least this information but may contain more fields and still be of this type.

Suppose your context is defective in that you do not have a referent for $_1$ Bo $_2$, i.e., according to our analysis, the record representing your context does not have a subrecord of the type (12).

(12) $\left[\begin{array}{lcl} \mathrm{ref}_{u_{1,1-2}} & : & \mathit{Ind} \\ \mathrm{res}_{u_{1,1-2}} & : & \mathrm{named}(\mathrm{ref}_{u_{1,1-2}}, \text{“Bo”}) \end{array} \right]$

One option you have is to coerce the meaning of (8) to a family of types exactly like (10) except that the fields in (12) have been "lowered" to the content type. Thus the content type of the utterance resulting from such a coercion would be paraphrasable as "There's somebody named Bo and she (the other dialogue participant) is asking whether he left". This is the strategy that in Ginzburg and Cooper (2001) was called existential quantification of deficient parameters. The result of the coercion is shown in (13).

The second strategy identified in Ginzburg and Cooper (2001) is parameter identification. This involves asking a question for the value of the missing parameter as in (14).

(14) u_2 : $_0$ Bo? $_1$

On the reading in question this utterance is paraphrasable as "Who is referred to by your utterance of "Bo"?" Note that what is needed to achieve this is more than any standard compositional semantics for the noun-phrase *Bo*. The interpretation of the clarification has to make reference to a particular part of the previous utterance. It is not sufficient just to refer to the word 'Bo' or to the referent that was introduced by the previous speaker's utterance of *Bo* as in standard anaphoric cases – after all

(9) $\left[\ \text{utt-time}_{u_{1,0-3}} \ : \ \textit{Time}\ \right]$ (*Time* the type of time intervals)

abbreviates

$$\left[\begin{array}{lll} \text{utt-time}_{u_{1,0-3}} & : & \textit{Time} \\ \text{pf-utt-time}_{u_{1,0-3}} & : & \text{utt-time}(u_{1,0-3}, \text{utt-time}_{u_{1,0-3}}) \end{array}\right]$$

(10) $\lambda r : \left[\begin{array}{lll} \text{phon}_{u_{1,0-1}} & : & /\text{dId}/ \\ \text{phon}_{u_{1,1-2}} & : & /\text{bu}/ \\ \text{phon}_{u_{1,2-3}} & : & /\text{liv}/ \\ \text{phon}_{u_{1,0-3}} & : & /\text{dIdbul\'iv}/ \\ \text{utt-time}_{u_{1,0-3}} & : & \textit{Time} \\ \text{ev-time}_{u_{1,0-3}} & : & \textit{Time} \\ \text{tense}_{u_{1,0-3}} & : & \text{ev-time}_{u_{1,0-3}} < \text{utt-time}_{u_{1,0-3}} \\ \text{ref}_{u_{1,1-2}} & : & \textit{Ind} \\ \text{res}_{u_{1,1-2}} & : & \text{named}(\text{ref}_{u_{1,1-2}}, \text{“Bo”}) \\ \text{sp}_{u_{1,0-3}} & : & \textit{Ind} \\ \text{hearer}_{u_{1,0-3}} & : & \textit{Ind} \\ \text{cat}_{u_{1,0-3}} & : & [\text{V}, +\text{fin}] \end{array}\right]$
$\left(\left[\begin{array}{lll} \text{msg}_{u_{1,0-3}} & : & ?\text{leave}(\text{ref}_{u_{1,1-2}}, \text{ev-time}_{u_{1,0-3}}) \\ \text{cont}_{u_{1,0-3}} & : & \text{ask}(\text{sp}_{u_{1,0-3}}, \text{hearer}_{u_{1,0-3}}, \text{msg}_{u_{1,0-3}}) \end{array}\right]\right)$

(11) $\lambda r : [\ \ldots\ ]$
$\left(\left[\begin{array}{lll} \text{msg}_{u_{1,0-3}} & : & ?\text{leave}(r.\text{ref}_{u_{1,1-2}}, r.\text{ev-time}_{u_{1,0-3}}) \\ \text{cont}_{u_{1,0-3}} & : & \text{ask}(r.\text{sp}_{u_{1,0-3}}, r.\text{hearer}_{u_{1,0-3}}, \text{msg}_{u_{1,0-3}}) \end{array}\right]\right)$

(13) $\lambda r : \left[\begin{array}{lll} \text{phon}_{u_{1,0-1}} & : & /\text{dId}/ \\ \text{phon}_{u_{1,1-2}} & : & /\text{bu}/ \\ \text{phon}_{u_{1,2-3}} & : & /\text{liv}/ \\ \text{phon}_{u_{1,0-3}} & : & /\text{dIdbul\'iv}/ \\ \text{utt-time}_{u_{1,0-3}} & : & \textit{Time} \\ \text{ev-time}_{u_{1,0-3}} & : & \textit{Time} \\ \text{tense}_{u_{1,0-3}} & : & \text{ev-time}_{u_{1,0-3}} < \text{utt-time}_{u_{1,0-3}} \\ \text{sp}_{u_{1,0-3}} & : & \textit{Ind} \\ \text{hearer}_{u_{1,0-3}} & : & \textit{Ind} \\ \text{cat}_{u_{1,0-3}} & : & [\text{V}, +\text{fin}] \end{array}\right]$
$\left(\left[\begin{array}{lll} \text{ref}_{u_{1,1-2}} & : & \textit{Ind} \\ \text{res}_{u_{1,1-2}} & : & \text{named}(\text{ref}_{u_{1,1-2}}, \text{“Bo”}) \\ \text{msg}_{u_{1,0-3}} & : & ?\text{leave}(\text{ref}_{u_{1,1-2}}, \text{ev-time}_{u_{1,0-3}}) \\ \text{cont}_{u_{1,0-3}} & : & \text{ask}(\text{sp}_{u_{1,0-3}}, \text{hearer}_{u_{1,0-3}}, \text{msg}_{u_{1,0-3}}) \end{array}\right]\right)$

$$(15)\ \lambda r : \left[\begin{array}{ll} \text{phon}_{u_{1,0-1}} & : /\text{dId}/ \\ \text{phon}_{u_{1,1-2}} & : /\text{bu}/ \\ \text{phon}_{u_{1,2-3}} & : /\text{liv}/ \\ \text{phon}_{u_{1,0-3}} & : /\text{dIdbulív}/ \\ \text{utt-time}_{u_{1,0-3}} & : \textit{Time} \\ \text{ev-time}_{u_{1,0-3}} & : \textit{Time} \\ \text{tense}_{u_{1,0-3}} & : \text{ev-time}_{u_{1,0-3}} < \text{utt-time}_{u_{1,0-3}} \\ \text{sp}_{u_{1,0-3}} & : \textit{Ind} \\ \text{hearer}_{u_{1,0-3}} & : \textit{Ind} \\ \text{cat}_{u_{1,0-3}} & : [\text{V, +fin}] \\ \text{phon}_{u_{2,0-1}} & : /\text{bu L-H}/ \\ \text{utt-time}_{u_{2,0-1}} & : \textit{Time} \\ \text{sp}_{u_{2,0-1}} & : \textit{Ind} \\ \text{hearer}_{u_{2,0-1}} & : \textit{Ind} \end{array}\right] \left(\left[\begin{array}{lll} \text{cont}_{u_{2,0-1}} & : & ?\ \lambda r' : \left[\begin{array}{ll} \text{ref}_{u_{1,1-2}} & : \textit{Ind} \\ \text{res}_{u_{1,1-2}} & : \text{named}(\text{ref}_{u_{1,1-2}}, \text{“Bo”}) \\ & (\text{ref}(u_{1.1-2}, \text{ref}_{u_{1,1-2}})) \end{array}\right] \end{array}\right]\right)$$

the question concerns what the intended referent was. What we are lead to when we consider clarifications is that there is a need for a radical kind of anaphoric or context dependence where various different aspects of particular previous utterances can be needed in order to define the meanings of utterances.

Ginzburg and Cooper's third kind of coercion is called parameter focussing and yields an alternative interpretation of (14) paraphrasable as "Are you asking whether Bo left? (The relevant issue is *who are you asking whether (they) left?*)". It is treated in a similar way by constructing a meaning for it based on (10), given in (16).

4 Conclusion

One important conclusion from this work is that meaning in dialogue is relative to the dialogue participants and the information they have about the context. Note that this is not the same as saying that content is dependent on context, represented in Montague-Kaplan terms by defining meanings as functions from contexts to contents. What we are saying is that context helps determine which function from contexts to contents is used and that context includes richly structured meanings for previous utterances (of the kind which are provided by a record-based approached to grammar). It is perhaps still not widely accepted that meaning is relative to a dialogue participant, i.e. that utterances in a dialogue do not have meaning simpliciter.

The second important conclusion is that meanings for utterances are in part constructed out of parts of structured meanings for previous utterances. (Note again that this concerns meanings, not just contents.) This kind of meaning construction is not included in classical approaches to compositional semantics and anaphora.

A third important conclusion is that meaning is constructed not only in terms of information about semantic objects but also in terms of phonological and syntactic information about speech events.

Acknowledgements

For very useful discussion and comments we would like to thank Pat Healey, Howard Gregory, Shalom Lappin, Peter Ljunglöf, Dimitra Kolliakou, David Milward, Bengt Nordström, Matt Purver and Aarne Ranta. The research described here is funded by grant number R00022269 from the Economic and Social Research Council of the United Kingdom, by INDI (Information Exchange in Dialogue), Riksbankens Jubileumsfond 1997-0134, by grant number GR/R04942/01 from the Engineering and Physical Sciences Research Council of the United Kingdom, the EU project Siridus (Specification, Interaction and Reconfiguration

$$
(16)\ \lambda r : \left[\begin{array}{lcl}
\text{phon}_{u_{1,0-1}} & : & /\text{dId}/ \\
\text{phon}_{u_{1,1-2}} & : & /\text{bu}/ \\
\text{phon}_{u_{1,2-3}} & : & /\text{liv}/ \\
\text{phon}_{u_{1,0-3}} & : & /\text{dIdbul\'iv}/ \\
\text{utt-time}_{u_{1,0-3}} & : & \textit{Time} \\
\text{ev-time}_{u_{1,0-3}} & : & \textit{Time} \\
\text{tense}_{u_{1,0-3}} & : & \text{ev-time}_{u_{1,0-3}} < \text{utt-time}_{u_{1,0-3}} \\
\text{ref}_{u_{1,1-2}} & : & \textit{Ind} \\
\text{res}_{u_{1,1-2}} & : & \text{named}(\text{ref}_{u_{1,1-2}}, \text{“Bo”}) \\
\text{sp}_{u_{1,0-3}} & : & \textit{Ind} \\
\text{hearer}_{u_{1,0-3}} & : & \textit{Ind} \\
\text{cat}_{u_{1,0-3}} & : & [\text{V}, +\text{fin}] \\
\text{phon}_{u_{2,0-1}} & : & /\text{bu L-H}/ \\
\text{utt-time}_{u_{2,0-1}} & : & \textit{Time} \\
\text{sp}_{u_{2,0-1}} & : & \textit{Ind} \\
\text{hearer}_{u_{2,0-1}} & : & \textit{Ind} \\
\text{max-QUD}_{u_{2,0-1}} & : & ?\ \lambda x : \textit{Ind}\ (\text{ask}(\text{sp}_{u_{1,0-3}}, \text{hearer}_{u_{1,0-3}}, ?\text{leave}(x, \text{ev-time}_{u_{1,0-3}}))) \\
([\ \text{cont}_{u_{2,0-1}} & : & ?\text{ask}(\text{sp}_{u_{1,0-3}}, \text{hearer}_{u_{1,0-3}}, ?\text{leave}(\text{ref}_{u_{1,1-2}}, \text{ev-time}_{u_{1,0-3}}))\ ])
\end{array}\right]
$$

in Dialogue Understanding Systems, IST-1999-10516 and the ILT (Interactive Language Technology) project funded by Vinnova, 2001-6340.

References

Betarte, Gustavo (1998) *Dependent Record Types and Algebraic Structures in Type Theory*, Ph.D. thesis, Department of Computing Science, Göteborg University and Chalmers University of Technology.

Betarte, Gustavo and Alvaro Tasistro (1998) Extension of Martin-Löf's type theory with record types and subtyping, in *25 Years of Constructive Type Theory*, ed. by G. Sambin and J. Smith, Oxford University Press.

Cooper, Robin (2000) Information States, Attitudes and Dependent Record Types, in *Logic, Language and Computation, Volume 3)*, ed. by Lawrence Cavedon, Patrick Blackburn, Nick Braisby, and Atsushi Shimojima, CSLI Publications, pp. 85–106.

Ginzburg, Jonathan and Robin Cooper (2001) Resolving Ellipsis in Clarification, *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse*, pp. 236–243

Ginzburg, Jonathan and Robin Cooper (ms) Clarification, Ellipsis, and the Nature of Contextual Updates, http://www.dcs.kcl.ac.uk/staff/ginzburg/papers.html

Ginzburg, Jonathan and Ivan Sag (2000), *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*, CSLI Publications, Stanford.

Pollard, Carl (2001) Hyperintensional Semantics: A Nonextensional Higher-Order Meaning Logic, invited talk at HPSG 2001, Trondheim.

Tasistro, Alvaro (1997) *Substitution, record types and subtyping in type theory, with applications to the theory of programming*, PhD thesis, Department of Computing Science, Göteborg University and Chalmers University of Technology.

Towards a Framework for Rhetorical Argumentation

Floriana Grasso
Department of Computer Science
University of Liverpool
floriana@csc.liv.ac.uk.

Abstract

This paper presents an approach towards the definition of a formal framework for rhetorical argumentation. Before describing the competence theory behind the framework, we introduce the notion of rhetorical argument, and describe the New Rhetoric, a well known theory in the philosophy of argumentation, upon which our work is based.

1 Motivation

The study of argumentation has undergone periods of decadence alternate with periods of renaissance. Currently, we are witnessing one of the latter, as proved by research flowering in many issues and aspects, both in Philosophy and in Artificial Intelligence. The work presented in this paper is concerned in particular with *rhetorical argumentation*, which on the other hand has enjoyed consideration to a somewhat lesser extent. By rhetorical argument we want to denote arguments which are both heavily based on the audience's perception of the world, and concerned more with evaluative judgments than with establishing the truth or otherwise of a proposition, like those in the following dialogue[1]:

A *Do you like cooking?*
B *Not especially. [...] Cooking feels to me like a lot of effort for something (ie. eating) that's over quite quickly. Also, I often feel tired at the end of a day's work and don't want to spend too much time in the kitchen.*
A *You do not cook just because you have to eat! Cooking can be a very relaxing and interesting activity, better than watching TV!*
B *I know you're right but that still doesn't make it easy to do!*

Previous research on argumentation in AI has concentrated either on logic based approaches, typically with a focus on negotiation, or on discourse modelling approaches, with argumentative discourse as a by-product of more general discourse structures. We want to specifically focus on discourse which is rhetorically argumentative, and we do this by explicitly drawing upon concepts in the philosophy of argument. This makes our approach different from those to the study of rhetorical discourse, usually based on linguistic or psycho-linguistic theories. In this paper we focus mainly on the *competence theory* behind our framework (Cohen and Perrault, 1979), and we will therefore ignore any reasoning or planning mechanism. Before presenting our framework, the paper will briefly explain the philosophical aspects of our work and the argumentation theory upon which it is based.

[1]The dialogue is taken from a corpus of e-mail conversations collected at the beginning of our investigation (Grasso et al., 2000).

2 A Theory of Rhetorical Argumentation

We think of rhetorical argumentation, in a tradition going back to Corax and Tisias in 500 BC, as argumentation where the emphasis is put more on the audience than on the argument itself, with the rhetorician concerned with finding what will persuade *in the given circumstances* (Aristotle, 350 BC), by appealing to the audience's set of beliefs in order to achieve persuasiveness to that audience, rather than general acceptability. The richest sort of argument for Aristotle, with premises which are not only deductive and dialectical, but also related to extra-rational factors, it is unfortunate that with time rhetorical arguments have acquired the reputation of being artificious, futile, insincere. So much so that their peculiarities have been much less investigated in the modern philosophy of argument, where a larger emphasis has been placed, instead, on the analysis of the *structure* of good arguments, in an attempt to capture, for argumentation, a formal aspect which could affiliate its study with the study of other, typically formal sciences, such as logic or mathematics (van Eemeren and Grootensdorst, 1992; Toulmin, 1958; Walton, 1989).

In a project aimed at giving back to rhetoric its deserved place in the study of argumentation, philosophers Perelman and Olbrechts-Tyteca published in 1958 a theory which they called the *New Rhetoric*, by referring explicitely to Aristotle's definition. The theory aims at identifying "discursive techniques allowing us to induce or to increase the mind's adherence to the thesis presented for its assent" (Perelman and Olbrechts-Tyteca, 1969, p. 4). With a Fregean approach that goes from examples to generalisation, rather than by establishing a "logic" or argumentation, the work is descriptive: a collection of argument schemata which are successful in

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 53-60.

practice, ordered and classified in terms of an ontological account of the objects of the argumentation, and the types of audience's beliefs the schemata exploit.

The treatise starts by analysing the concept of what can serve as premises in argumentation, that is what the speaker can assume adherence to by the audience in order to build up the argument. Such beliefs, or "objects of adherence", account not only for factual knowledge, but also for less clear cut beliefs and feelings. The theory infact distinguishes between premises *relating to the "real"*, consisting of statements that are believed to a greater ("facts") or lesser ("presumptions") extent by the audience, and of more general conceptions the audience accept ("truths", like scientific or religious theories), and premises *relating to the "preferable"*, expressing the preferences of the audience ("values") and their priorities ("hierarchies").

Dealing with discoursive techniques, Perelman and Olbrechts-Tyteca are not only preoccupied with the agreement and choice of premises, but also with how premises are presented to the audience: the exposition of the argument is sometimes more important than its validity. The New Rhetoric is, in fact, a collection of ways for arranging premises and conclusions, that is *schemata* that are successfully used by people in ordinary discourse. The main objective of each schema is to exploit associations among concepts, either known or new to the audience, in order to pass the audience's acceptance (positive or negative) from one concept to another. The description of the schemata is augmented with both practical examples and potential weaknesses, or counterarguments. A discussion of one of these schemata is given later, as an example.

The theory has inspired our framework in many ways, from the model of argumentation described here, to the tasks of knowledge and belief modelling, and dialogue structuring (Grasso et al., 2000; Grasso, 2000).

3 A Framework for Rhetorical Argumentation

If rhetorical argumentation aims at reaching an evaluation of an object or state of affair, then a rhetorical argument is a way to "pass value" from one topic to another, in the same way as a deductive argument "passes truth" from one proposition to another. So, as a deductive argument expresses something like: "if X is true, and Y derives from X, then Y is true", a rhetorical argument expresses something like "if X has value, and Y *is related to* X, than Y has value" (we will better explain what "related to" means later on). The notion of evaluation is therefore central to our framework, and in our vision it also has to encapsulate a *perspective* from which the evaluation is made. A perspective is a way to express what Perelman and Olbrechts-Tyteca, and indeed classical philosophers, call a *locus* of argumentation, indicating the fact that there are some *fields*, or points of view, that help the process of evaluation of a given concept. Therefore, we would want to say not only that "X has value", but also that "X has value from perspective P", hence contemplating the fact that X might have a different evaluation when seen from another perspective.

We will assume that the rhetorical argumentation can take place on a certain set of objects of discourse, forming an ontology. We have defined elsewhere (Grasso, 2000) more comprehensively the ontology we need, but now it will suffice to say that we need a set of concepts, C_o, of an ontology O, and a set of relationships among these concepts, R_o, again in the ontology O. We think of a relationship, in this context, as any of the possible ways in which two concepts may be linked in the ontology: they not only might have a semantic association between them, but they also might be one the generalisation of the other, or they may both contribute to define a third object, they may be an action and one of its roles, they may have identical values for one of the attributes, and so on and so forth. We express the fact that there exists a relationship r_k among the objects c_i and c_j in an ontology O by writing that:

$$r_k(c_i, c_j)$$

with

$$r_k \in R_o, c_i, c_j \in C_o$$

For the sake of simplicity in the notation, we will consider, without loss of generality, only binary relations, but the definitions below can be naturally extended to consider n-ary relations.

Definition 1 (Evaluation) *We say that there exists an evaluation of a concept c, in the set of concepts C_o of an ontology O, from a certain perspective p, from a set P_o again in the ontology O, if there exists a mapping E of the pair (c, p) into a set V of "values". Assuming that V is a set consisting of two elements (representing "good" and "bad"), we write this as:*

$$E : C_o \times P_o \rightarrow V = \{good, bad\}^2. \quad (1)$$

2

The definition of rhetorical argument is, however, not solely based on the notion of evaluation: to reflect the communicative dimension of rhetorical argumentation, the value passing must also be done by means of a predefined argumentative schema:

Definition 2 (Rhetorical Argument) *We define a rhetorical argument as the act of putting forward the evaluation of a concept, on the basis of a relationship existing between this concept and another concept, and by means of a rhetorical schema. If we call S the set of the available rhetorical schemata, and $E' = C_o \times P_o \times V$ the set of all the evaluations, then we define a rhetorical argument A_R as a function:*

$$A_R : E' \times R_o \times S \rightarrow E' \quad (2)$$

2

In other words, if we have a concept $c_i \in C_o$ and an evaluation of such concept $E(c_i, p_i)$ from a given perspective $p_i \in P_o$, we can put forward a rhetorical argument in favour or against a second concept $c_j \in C_o$, if and only if (i) a relationship exists in the ontology between the two concepts $r_k(c_i, c_j), r_k \in R_o$, and (ii) a schema can be identified that exploits such relation $s_l \in S$.

3.1 Rhetorical Schemata

In our vision of rhetorical argumentation, a rhetorical schema is meant to express when it is admissible to use a given relationship, by capturing the restrictions that, from case to case, have to apply for the relationship to be used legally, or safely. It is beyond the scope of this exercise to establish the exact set of constraints that need to apply to the definition of each schema, and in this sense our approach differs perhaps from what Perelman and Olbrechts-Tyteca intended, and certainly differs from other approaches, such as (Mann and Thompson, 1988) to the definition of discourse schemata. Our sole claim is that it is important that a definition of a rhetorical schema should explicitly represent its admissibility, together with the relationship that the schema is meant to formalise, and we propose here an ontological discussion of what this notion of admissibility refers to, and what it should entail.

Literature in the philosophy of argument, and in particular in informal logic, a branch explicitly concerned with analysing and evaluating natural argumentation (Walton, 1989; Groarke et al., 1997; Johnson and Blair, 1994b), provides good insights as to which characteristics a good argument should possess, characteristics which are well summarised in the so-called "RSA-triangle" (Johnson and Blair, 1994a). We have already mentioned the *Acceptability* criterion: as rhetorical arguments have to be convincing to "the particular audience", one cannot ignore what that audience perceive as true, or valuable, so the arguments have to be explicitly based on the audience's set of beliefs. In addition, the argument proposed must be *Relevant* to the discussion, although this is a criterion which is not simple to define univocally, but depends on the context, and on presumptions on the speaker and hearer knowledge and ability to infer implicatures (Grice, 1975; Walton, 1998). Finally, the argument proposed must be *Sufficient* for the audience to be able to reach an evaluation, by proposing a not biased and balanced point of view (Groarke et al., 1997).

Definition 3 (Rhetorical Schema) *We define a rhetorical argumentation schema as a 6-tuple:*

$$R_S =< N, C, O_c, A_c, R_c, S_c > \quad (3)$$

where:

- N *is the name of the schema,*
- C *is the claim the schema supports,*
- O_c *are the ontological constraints the schema is based on,*
- A_c *are the acceptability constraints,*
- R_c *are the relevance constraints, and*
- S_c *are the sufficiency constraints.* 2

We explicitly differentiate among constraint types not only for the sake of clarity, but also with operational purposes. We assume an ordering among constraints reflecting their importance in the context of the argument, where those in the first slots are more fundamental to the evaluation of the argument than those in the last slots. This also allows us to envisage an environment where these constraints act in fact as a set of desiderata, that can be relaxed if needed: so, relaxing the ontological constraints should "cost more", in terms of the quality of the argument produced, than relaxing the sufficiency ones. Let us examine each constraint type in turn.

Ontological Constraints The ontological constraints represent the characterisation, in the language of the particular knowledge base in use, of the r_k element of the definition of rhetorical arguments (Def. 2), that is they spell out how the two concepts that are used in the rhetorical argument should relate to one another in the ontology. It should be therefore clear that they represent the condition *sine qua non* for the application of the schema: when putting forward the schema, the speaker have to believe that the schema can be actually applied, and the value passing can take place[3]. An argument where violation of the ontological constraints occurs is therefore a deceptive argument from the speaker's side.

Acceptability Constraints The acceptability constraints accounts for the beliefs and perceptions of the audience. As mentioned before, this aspect is crucial in rhetorical argumentation, where the argument is somewhat tailored to the addressees and the circumstances: disregarding the audience's beliefs can lead to perfectly valid and sound arguments that might nevertheless be not persuasive enough. Acceptability is an important aspect, but not as fundamental as the previous one: in fact, in cases, for instance, where the speaker is presenting totally new information, acceptability constraints are likely to be less crucial. In modelling the audience's beliefs, all the considerations typical of agent cognitive modelling apply (Ballim and Wilks, 1991; Rao and Georgeff, 1995; Cohen et al., 1990), but here we want to stress on the distinction made by Perelman and Olbrechts-Tyteca between "facts" and "presumptions". In the hypothesised absence of an objective, omniscient source, we interpret it as the distinction between what the audience has explicitly committed to, and what the arguer can only presume is believed. In checking the A_c constraints, the speaker should therefore give different weight to

[3] We do not assume a situation in which a third, omniscient party can judge the veracity of what the arguing parties say.

their lack of satisfiability depending on how they have been acquired. This accounts for what, in Perelman and Olbrechts-Tyteca's view, the speaker does when assessing tokens of adherence to the objects of discussion: it may be done explicitly (like the dialectical use of question and answer in Socratic dialogues) or may be assumed. We have outlined elsewhere (Grasso et al., 2000) how we have translated this consideration into a categorisation of belief nestings, but, in what follows, we will ignore the problem of expressing various levels of mutual belief, and for the sake of simplicity we will express the acceptability constraints by means of a meta-predicate, β, which stands for "the audience believe", and has as argument an expression of an ontological constraint.

Relevance Constraints The problem of relevance has been much studied in artificial intelligence, in philosophy and linguistics/pragmatics (Grice, 1975; Greiner and Subramanian, 1995; Walton, 1982). We follow (Sperber and Wilson, 1995) in giving to reference an operational interpretation, by capturing the *effort* that the listeners have to put in processing the information, as done in (Reed, 1999; Walker, 1996). We therefore define the relevance constraints as the representation of which premises have to have been communicated for the audience to act towards the acceptance (or rejection) of the conclusion. Paraphrasing (Sperber and Wilson, 1995), we stress that such premises, or beliefs, should be not only possessed by the audience, but also "accessible", that is they can be "retrieved" without great effort. This property has been often referred to as being in the "focus of attention" (Grosz and Sidner, 1986), that is the beliefs should be part of the entities that are salient to what is being said. Again for the sake of simplicity, while acknowledging that a greater sophistication might be desirable (for instance, we do not specify here the rules to manipulate the focus), we represent the relevance constraints as another meta-predicate, which we denote Φ, whose meaning is "the audience have in focus", and whose argument is an expression of an ontological constraint.

Sufficiency Constraints Sufficiency constraints are defined in the informal logic literature as the representation of all the premises that are needed in order for the audience to establish that the conclusion is more likely than not (Groarke et al., 1997). They are meant to measure the speaker's "fairness" in putting forward the argument: in support of the point made, the speaker should have a well balanced list of issues that lead to the proposed conclusion. In non deductive argumentation, it is not always possible or feasible to give precise rules to establish whether this constraint is satisfied, nor a logical characterisation of *all* is needed to satisfy it is in the scope of our investigation. We adopted a circumscriptive approach by encapsulating in the notion of sufficiency Perelman and Olbrechts-Tyteca's list of counter-arguments to each schema, that is the expression of what can "go wrong" when that schema is chosen and used. However, we do this from a positive perspective, by listing instead the beliefs that complete the line of reasoning of the schema, and that can be provided if needed in support to the main point, but that will weaken the argument if they cannot be taken to hold.

The meaning and the scope of the constraints will be best clarified by an example.

4 An Example: Argument from Reciprocity

We will present here, as an example, the characterisation, according to our framework, of the New Rhetoric's *Reciprocity Schema*. In the example we will not commit ourselves to one particular knowledge representation, but we will assume that the system can rely on a knowledge representation including the notion of a Concept, c, for which one can enumerate a certain set of Attributes, $c.a_1 \ldots c.a_n$, and a set of Relations, $r_1 \ldots r_k$ that associate c with other Concepts in the knowledge representation. We indicate the semantics and the value of the attributes respectively as $\mathcal{S}(c.a_j)$ and $\mathcal{V}(c.a_j), \forall j$. In what follows we will use lower case word to denote variables (*x, y, val*), capitalised words to denote constants (*Biscuits, Good, Health*), and we will abbreviate $E(c,p)$ as $E_{c,p}$.

In the New Rhetoric, the argument from reciprocity is based on the concept of symmetry. The aim is to propose that two situations should be given the same evaluation on the grounds of some symmetrical property that they share (Perelman and Olbrechts-Tyteca, 1969, § 53). It is a *quasi-logical* argument in Perelman and Olbrechts-Tyteca's classification. Quasi-logical arguments are those which are presented in a structure that resembles a formal (logical or mathematical) way of reasoning, a rhetorical shift that is aimed at persuading of their soundness, although they not as rigourous as a formal argument can be. In particular, in the argument from reciprocity the symmetry does not necessarily have to be a formal property of relations, but it rather expresses a more or less objective interchangeability of two states of affair. This can happen because of some complementarity between them (e.g. between the acts of accusing and condemning) or because they can be thought of as one the inverse of the other (e.g. who borrows and who lends) and so on. An example or reciprocity argument can be: *You said you like biscuits. But apples are as crunchy as biscuits, so you should like apples too!*[4]. Let us analyse the description of the reciprocity argument in our framework.

Claim. The schema aims at passing good or bad value on to one concept from a certain perspective. In the example this would instantiate to passing good value on to the concept of apples (or eating apples),

[4]The verbalisation does not have necessarily to be as above, and in fact the argument will typically be put forward in a more concise way, as outlined in Sect. 5.

from a certain perspective p to be defined:

$$E_{Apples,p} = Good \quad (4)$$

Ontological constraints. The O_c element of the 6-tuple for the reciprocity schema is based on the analysis of the attributes that two concepts, or two instances of a concept, share. As this is a quasi-logical schema, we almost never will pay attention to the actual meaning of the relationship: it will suffice that a relationship exists with some properties. A relation of reciprocity can be defined to hold between two concepts when for each of them an attribute can be identified such as the two attributes have same semantics and same value. That is, if for two concepts, $c_1, c_2 \in O_c$, we have that $\exists k, j : \mathcal{S}(c_1.a_k) = \mathcal{S}(c_2.a_j)$, and $\mathcal{V}(c_1.a_k) = \mathcal{V}(c_2.a_j)$, then an argument of reciprocity may claim that the two concepts should have the same evaluation on the basis of this symmetry. The specific argument depends of course on the particular structure of the ontology in use. In the example, we can assume we have a knowledge base in which both biscuits and apples have an attribute "texture" (we assume they have the same name, but this is not necessary) which has the same semantics (can assume values in the same domain), and the same value "crunchy":

$$\begin{aligned} O_c \quad : \quad & \mathcal{S}(\text{Apples.Texture}) = \mathcal{S}(\text{Biscuits.Texture}); \\ & \text{Apples.Texture=Crunchy}; \\ & \text{Biscuits.Texture=Crunchy} \end{aligned} \quad (5)$$

Acceptability Constraints. In order for the argument to be successful, the audience should share the view that that the "reciprocal" concept used in the argument has the desired evaluation, from a certain perspective: $\beta(E_{c_2,p} = val)$. In the example we can assume that Biscuits are good from a "taste" perspective, for the audience:

$$A_c \quad : \quad \beta(E_{Biscuits,Taste} = Good) \quad (6)$$

Note that the argument does not ask that the audience should like biscuits *because they are crunchy*: this would entail a level of knowledge of the reality that is not required for quasi-logical argumentation, as opposed to other schema categories. In order to express a quasi-logical argument one only need to create a "skeleton" that aims at making two concepts comparable (Perelman and Olbrechts-Tyteca, 1969, § 45).

Relevance Constraints. The conclusion the argument proposes will be reached with little effort by the audience if all of the elements involved are in the focus of attention. In the example, the crunchiness of apples and biscuits, and the tastiness of biscuits should be in the focus of attention:

$$\begin{aligned} R_c \quad : \quad & \Phi(\text{Apples.Texture} = \text{Crunchy}); \\ & \Phi(\text{Biscuits.Texture} = \text{Crunchy}); \\ & \Phi(E_{Biscuits,Taste} = Good) \end{aligned} \quad (7)$$

If this can be assumed, one can safely use a abbreviated form of the argument (*if you like biscuits, then you should like apples!*), but if it cannot be assumed, the audience might be left bemused if the argument is not spelt out completely. This, however, does not necessarily mean the audience will not be persuaded: as a matter of fact, the relevance constraint might be relaxed maliciously, for instance when the arguer can assume to have a strong influence on the audience (e.g. peer pressure situations).

Sufficiency Constraints. These constraints should enable the audience to reinforce the symmetry between the two concepts, as opposed to it being limited to one attribute only. An exhaustive list of what should be included here is perhaps not feasible, and it definitely depends on the underlying ontology, but we can assume that an appropriate constraint could require that other pairs of attributes of the two concepts in question can be identified sharing the same semantics and the same value. Of course, the more attributes with equal semantics the two concepts have, the more they can be compared to each other, and the more pairs of comparable attributes have equal values, the stronger the reciprocity relation will be. Measures of similarity among concepts can be established in the ontology to have a sounder definition of the reciprocity relation, but for the moment we can assume a sufficiency constraint that asks that, in the example:

$$\begin{aligned} S_c \quad : \quad & \forall j, k \text{Biscuits}.a_j, \text{Apples}.a_k : \\ & \mathcal{S}(\text{Biscuits}.a_j) = \mathcal{S}(\text{Apples}.a_k) \rightarrow \\ & \mathcal{V}(\text{Biscuits}.a_j) = \mathcal{V}(\text{Apples}.a_k) \end{aligned} \quad (8)$$

where the $\forall$ sign should rather be read as "the more the better".

As a remark, we should underline again that, as the schema is quasi-logical, no emphasis is put on the set of attributes, or roles, that are more meaningful to the success of the schema. In the example, an attribute referring to the "edibility" of both biscuits and apples is clearly crucial: a reciprocity argument based on the crunchiness of apples as opposed to, say, gravel will definitely miss the point altogether.

A general schema for reciprocity argumentation can be summarised as in Fig 1.

5 Schemata and Their Use in Generating Rhetorical Arguments

Our main objective in the study of rhetorical argumentation is to draw some insights leading to the generation of

N	RECIPROCITY
C	$E_{concept,perspective} = val$
O_c	$\exists concept_2 \in O, k, j :$ $\mathcal{S}(concept.a_k) = \mathcal{S}(concept_2.a_j)$, $\mathcal{V}(concept.a_k) = \mathcal{V}(concept_2.a_j)$
A_c	$\beta(E_{concept_2,perspective} = val)$
R_c	$\Phi(\mathcal{S}(concept.a_k) = \mathcal{S}(concept_2.a_j))$; $\Phi(\mathcal{V}(concept.a_k) = \mathcal{V}(concept_2.a_j))$; $\Phi(E_{concept_2,perspective} = val))$
S_c	$\forall j, k :$ $\mathcal{S}(concept.a_j) = \mathcal{S}(concept_2.a_k) \rightarrow$ $\mathcal{V}(concept.a_j) = \mathcal{V}(concept_2.a_k)$

Figure 1: Argument from Reciprocity Schema

arguments, rather than their analysis. Our work, therefore, has not as its main focus typical pragmatics issues (Lascarides and Asher, 1993), although we hope our exercise might be useful in this respect as well. We envisage our definition of the rhetorical argument schema, perhaps along the tradition of informal logic and critical thinking teaching (Groarke et al., 1997), as the formalisation of what the speaker has to take into account when producing a "good" argument. So one might see our collection of constraints also as a collection of tokens that need to be added or otherwise to the presentation of the argument in order for it to be successful. For instance, in the reciprocity example shown in the previous section, one might associate with each constraint a piece of discourse to be put forward, that can be eliminated if the constraints can be assumed to hold by the speaker. A complete exposition of the argument may therefore go along the line of:

(i) I think you should like (the taste of) apples (Claim)

(ii) as I think they have the same texture of biscuits (O_c)

(iii) and I believe you like biscuits (A_c)

(iv) and I think apples and biscuits are comparable because... (S_c)

In the exposition, (ii) and (iii) can be eliminated if the speaker can assume that they are not only believed by the hearer, but also in the focus of attention (R_c). Seeing this from another point of view, we can say that the speaker may eliminate, or partially eliminate (ii) and (iii) if it can be assumed that the hearer is able to make the relevant associations with the previous discourse utterances by means of suitable rhetorical relations, so that the current utterance results coherent (Lascarides and Asher, 1999). For instance, if the interlocutor has just mentioned how tasteful crunchy biscuits are, the speaker could simply say *Then you should like apples, they are crunchy too!*, with the second part of the sentence that can be further dropped if the crunchiness of apples can be assumed as a belief the interlocutor adheres to (and has in focus). Finally, the speaker can choose whether to add (iv) to the argument from the start, or wait until challenged by the hearer.

6 Discussion

As mentioned above, we have been mainly interested in the rhetorical structure of arguments, and as a consequence, in the structure of rhetorical argumentative discourse. From this point of view, not many researchers have been undertaking works that are relevant to the one presented here. They can be divided into those mainly concentrating on the representation of multiple viewpoints, perhaps ranked in order of their importance to the audience, such as (Fox and Parsons, 1997; Karacapilidis, 1996; Lowe, 1985; Zukerman et al., 1996), and those focusing more on the generation of argumentative discourse, such as (Elhadad, 1995; Hovy, 1990; Maybury, 1993; Reed, 1999).

The work presented here is perhaps more akin to the latter, although research in discourse modelling usually appeals to theories from linguistics rather than philosophy, most often the Rhetorical Structure Theory (RST) (Mann and Thompson, 1988). It should hopefully be evident that our formalisation does not compete, and indeed can be comfortably used in a system for generating argumentative discourse based on planning, such as (Hovy, 1990; Maybury, 1993), where our schemata can be seen as planning operators. While not suggesting an approach to discourse generation different from one based on RST (which, despite criticisms (Moore and Pollack, 1992) is still most widely used), this work sets itself apart from the issue "RST or not RST", in order to advocate for a more principled collection of the knowledge necessary to make any discourse strategy, hence also RST, work better. For example, RST has a perspective on discourse based on relations among text spans, which allows one to build a tree to represent the whole text, a concept which is very powerful, and suited to formalisation in computational terms (Marcu, 2000). The semantics of the relations, however, is not very formal in RST: relations are defined in terms of constraints, or assumptions, that can be made on the goals of the speaker, which are probably more useful to the analysis of a piece of text than to its generation. Using RST, or any other discourse structure theory, for the generation of text, requires precisely a better definition of the grounds from which the text should be generated, and therefore the objectives of the speakers and the intended effect on the hearer need to be better formalised. Moreover, researcher in discourse processing typically concentrate on discourse in general, which happen to be argumentative in some cases. For such cases standard argumentation schemata might be embedded in the system (Dalianis and Johannesson, 1998; Reed, 1999), but there are very few discourse processing systems that base themselves on one of the frameworks among the plethora that philosophi-

cal theories on rhetoric provide (a notable exception is (Elhadad, 1995)). This work constitutes an attempt in that direction, and is in line with a series of recent efforts to put together two communities, the AI/NLP and philosophy of argument ones, that we believe have much to gain from cross-fertilisation (Reed and Norman, 2000; Carenini et al., 2002).

7 Conclusion

In his position paper presented at the Symposium on Argument and Computation, Crosswhite pictures three possible avenues for research should an AI scholar wish to undertake the task of creating a computational model of rhetorical argumentation (Crosswhite, 2000). The first is the exploitation of the argumentative schemata, of which literature in rhetoric provides a rich repository. The second is the exploitation of the figures of speech, and the ways they influence argumentation. The third is the explicit representation of the audience. In this paper we have shown how we have addressed the first issue. Researchers in computational pragmatics and belief modelling address, from different perspectives, the third one. As to the second issue, we believe this relates, speaking in natural language generation terms, to the surface representation of the argumentative text (Reiter and Dale, 1997). We leave this issue outside the scope of our work, although it definitely constitutes its natural extension.

References

Aristotle. 350 B.C. *Rhetorica (tr. J.H. Freese, 1926).* Loeb, London.

A. Ballim and Y. Wilks. 1991. *Artificial Believers.* Hillsdale, NJ: Lawrence Erlbaum Associates.

G. Carenini, F. Grasso, and C. Reed, editors. 2002. *Proceedings of the ECAI 2002 workshop on Computational Models of Natural Argument.*

P.R. Cohen and R. Perrault. 1979. Elements of a Plan-Based Theory of Speech-Acts. *Cognitive Science,* 3:177–212.

P. Cohen, J. Morgan, and M. Pollack, editors. 1990. *Intentions in Communication.* Cambridge, Mass: MIT Press.

J. Crosswhite. 2000. Rhetoric and Computation. In Reed and Norman (Reed and Norman, 2000).

H. Dalianis and P. Johannesson. 1998. Explaining Conceptual Models - Using Toulmin's argumentation model and RST. In *Proceedings of The Third International workshop on the Language Action Perspective on Communication Modelling (LAP98),* pages 131–140.

M. Elhadad. 1995. Using Argumentation in Text Generation. *Journal of Pragmatics,* 24:189–220.

J. Fox and S. Parsons. 1997. On using Arguments for Reasoning about Actions and Values. In *Proceedings of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning,* Stanford.

F. Grasso, A. Cawsey, and R. Jones. 2000. Dialectical Argumentation to Solve Conflicts in Advice Giving: a case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies,* 53(6):1077–1115.

F. Grasso. 2000. Using an Ontology Conceptualisation Method to Capture an Advice Giving System Knowledge. In W. Horn, editor, *ECAI 2000. Proceedings of the 14th European Conference on Artificial Intelligence,* pages 214–218, Amsterdam. IOS Press.

R. Greiner and D. Subramanian, editors. 1995. *Proceedings of the AAAI Fall Symposium on Relevance,* Menlo Park, CA. AAAI. Tech. Rept. FS-94-02.

H.P. Grice. 1975. Logic and Conversation. In P. Cole and J.L. Morgan, editors, *Speech Acts,* volume 3 of *Syntax and Semantics,* pages 41–58. Seminar Press.

L. Groarke, C. Tindale, and L. Fisher. 1997. *Good Reasoning Matters!* Oxford University Press, Toronto.

B.J. Grosz and C.L. Sidner. 1986. Attention, Intentions and the Structure of Discourse. *Computational Linguistics,* 12(3):175–204, July-September.

E. Hovy. 1990. Pragmatics and Natural Language Generation. *Artificial Intelligence,* 43:153–197.

R. H. Johnson and J. A. Blair. 1994a. *Logical self-defense.* McGraw-Hill, Toronto, 3rd edition.

R. H. Johnson and J. A. Blair, editors. 1994b. *New Essays in Informal Logic.* Informal Logic, Windsor, Ontario.

K. Jokinen, M. Maybury, M. Zock, and I. Zukerman, editors. 1996. *Proceedings of the ECAI-96 Workshop on: Gaps and Bridges: New directions in Planning and NLG.*

N. Karacapilidis. 1996. An Argumentation Based Framework for Defeasible and Qualitative Reasoning. In Jokinen et al. (Jokinen et al., 1996), pages 37–42.

A. Lascarides and N. Asher. 1993. Temporal Interpretation, Discourse Relations and Common Sense Entailment. *Linguistics and Philosophy,* 16(5):437–493.

A. Lascarides and N. Asher. 1999. Cognitive States, Discourse Structure and the Content of Dialogue. In *Amstelogue: 3rd Workshop on the Semantics and Pragmatics of Dialogue.*

D. Lowe. 1985. Co-operative Structuring of Information: the Representation of Reasoning and Debate. *International Journal of Man-Machine Studies,* 23:97–111.

W.C. Mann and S.A. Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text,* 8(3):243–281.

D. Marcu. 2000. *The Theory and Practice of Discourse Parsing and Summarization.* MIT Press.

M. Maybury. 1993. Communicative Acts for Generating Natural Language Arguments. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI93),* pages 357–364. AAAI, AAAI Press / The MIT Press.

J. Moore and M. Pollack. 1992. A Problem for RST: The Need for Multi-level Discourse Analysis. *Computational Linguistics*, 18(4):537–544.

C. Perelman and L. Olbrechts-Tyteca. 1969. *The New Rhetoric: a treatise on argumentation.* University of Notre Dame Press, Notre Dame, Indiana. (Transl. of *La Nouvelle Rhétorique : Traité de l'argumentation* Presses Universitaires de France, Paris, 1958).

A.S. Rao and M. Georgeff. 1995. BDI Agents: from Theory to Practice. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, June.

C.A. Reed and T. Norman, editors. 2000. *Symposium on Argument and Computation: position papers.* `http://www.csd.abdn.ac.uk/~tnorman/sac/`.

C.A. Reed. 1999. The Role of Saliency in Generating Natural Language Arguments. In T. Dean, editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, pages 876–881. Morgan Kaufmann Publishers.

E. Reiter and R. Dale. 1997. Building Applied Natural Language Generation Systems. *Natural Language Engineering*, 3(1):57–87.

D. Sperber and D. Wilson. 1995. *Relevance: communication and cognition.* Blackwell, Oxford, 2 edition.

S. Toulmin. 1958. *The Uses of Argument.* Cambridge University Press.

F.H. van Eemeren and R. Grootensdorst. 1992. *Argumentation, Communication, and Fallacies: A Pragma-Dialectical Perspective.* Hillsdale, NJ: Lawrence Erlbaum Associates.

M.A. Walker. 1996. The Effect of Resource Limits and Task Complexity on Collaborative Planning in Dialogue. *Artificial Intelligence*, 85(1-2):181–243.

D.N. Walton. 1982. *Topical Relevance in Argumentation.* Cambridge University Press, Cambridge.

D.N. Walton. 1989. *Informal Logic: a Handbook for Critical Argumentation.* Cambridge University Press.

D.N. Walton. 1998. *The New Dialectic: Conversational Contexts of Argument.* Toronto Studies in Philosophy. University of Toronto Press.

I. Zukerman, K. Korb, and R. McConachy. 1996. Perambulations on the Way to an Architecture for a Nice Argument Generator. In Jokinen et al. (Jokinen et al., 1996), pages 32–36.

A DRT-based framework for presuppositions in dialogue management

Samson de Jager, Alistair Knott and Ian Bayard
Dept of Computer Science, University of Otago, New Zealand
{sdejager/alik/ibayard}@cs.otago.ac.nz

Abstract

In this paper we describe a dialogue system which makes use of the notion of presuppositions to provide a unifying perspective on the function of a dialogue manager. We extend van der Sandt's (1992) treatment of presuppositions in DRT to allow presuppositions to support question-answering, the generation of targeted follow-up questions, the identification of dialogue acts and the semantic disambiguation of questions.

1 Introduction

A key component of any system which engages in natural language interaction with a user is its **dialogue manager**. At a high level of abstraction, what this module does is to determine the dialogue act associated with a user's incoming utterance and to decide on a suitable dialogue act to produce in response. However, this high-level description often hides a disparate collection of specific strategies, which are closely tied to the semantics of the incoming utterance and of the response to be generated. For instance, the answer to a question relates in a very specific way to the semantics of the question, while a followup question produced in response to a partially-understood utterance relates in a very specific (but different) way to the semantics of the utterance. As a consequence of these linkages, 'dialogue manager' often ends up being a dignified word for a heterogeneous set of if-then statements.

In this paper we describe a system in which the linguistic notion of presuppositions is extended to provide a unifying perspective on the job of a dialogue manager. Te Kaitito ('the improviser/interpreter') is a web-based bilingual dialogue system, supporting the authoring and querying of a database of facts in English and Māori. The system is eventually to be embedded in an educational application, to support the learning of Māori as a second language, but at present there is no high-level 'task' underlying the dialogue between the system and the user: the user simply tells the system about states and events or asks questions about them. We are therefore mainly interested in the mechanisms that underlie 'exhaustive' answers to questions (in the sense of Ginzburg, 1996) and 'direct' answers to questions (as opposed to the indirect answers which commonly occur in task-oriented dialogue, c.f. for instance Asher and Lascarides, 1998). The system is accessible over the web, at `http://tutoko.otago.ac.nz:8080/teKaitito/` (though not every feature mentioned in this paper is fully integrated in the publically available version). An example dialogue with the system is given in Figure 1; we will use this figure to illustrate the features of the system as they are introduced. The sentence-processing mechanisms used in the system are briefly described in Section 2, and the dialogue manager is introduced in Section 3. In Section 4 we describe a range of novel ways to use the concept of presuppositions in dialogue management. In Section 5, we compare the system's model of presupposition and dialogue to a number of other existing models.

2 Syntax and semantics for Te Kaitito

Sentence interpretation and generation in Te Kaitito is performed by the LKB system

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 61-68.

Speaker	Utterance	
User	a blue dog barked	(a)
System	Ok	
User	a red dog walked	
System	Ok	
User	which dog walked	(b)
System	the red dog it was the red dog nä te kurï whero	(c)
User	Ok	(d)
User	the blue dog chased a green cat	(e)
System	Ok	
User	a weasel barked	
System	Ok	
User	a big cat chased the weasel	
System	Ok	
User	the cat barked	
System	Ok	
User	the red dog chased the weasel	
System	Ok	

User	the blue dog walked	
System	Ok	
User	the dog ate the weasel	(f)
System	which dog ate the weasel	(g)
User	the red dog	
System	Ok	
System	Ok	
User	the blue dog chased a slow monkey	(h)
System	Ok	
User	a loud monkey ate the blue dog	
System	Ok	
User	which monkey did the blue dog chase	
System	the slow monkey it was the slow monkey nā te makimaki pōturi	
User	Ok	
User	what chased the monkey	(i)
System	the blue dog it was the blue dog nā te kurī kikorangi	

Figure 1: An extended dialogue session with Te Kaitito

(Copestake *et al.*, 2000), a grammar development environment for HPSG-like unification-based grammar formalisms. LKB is a bidirectional system: the parser and the generator both consult the same declarative grammar, so that any utterance the system can understand, it can also in principle generate. The grammar we have designed is bilingual: all rules and lexical items are associated with a LANGUAGE feature, and agreement between these features is enforced throughout a parse tree. (Any words or rules which do service in both languages can naturally remain undefined for LANGUAGE.) LKB's grammar provides a mapping between sentences and semantic representations given in a formalism called Minimal Recursion Semantics or **MRS** (Copestake *et al.*, 2000).[1] The details of MRS are not important in the present context. For more about the architecture of Te Kaitito and its motivations, see Knott *et al.* (2002).

[1]Sentences which translate or paraphrase each other are associated with the same MRS representation, which means that Te Kaitito can also function as a simple sentence translation and paraphrase system.

3 Overview of the dialogue manager

The dialogue system which embeds LKB needs to consider the role of sentences in context; we therefore decided to use DRT (Kamp and Reyle, 1993) as its main representational tool. MRS structures are converted to a modified form of DRS: a sentence's assertional content is stored in a conventional DRS, and a subsidiary DRS is created to hold any of the sentence's presuppositions, along the lines proposed by van der Sandt (1992). The DRS computed for utterance (a) in the sample dialogue is given in Figure 2.

DRSref57 e50 e51 e52 e53
h55: DOG-REL(e50 DRSref57)
h55: BLUE-RELN(e50 DRSref57)
h56: BARK-REL(e53 DRSref57)
h56: RE-S-RELN(e51 e52 e53)
h56: ASSERTION-RELN(e53)

Figure 2: DRS for *a blue dog barked*

Note that event arguments are introduced by several syntactic elements, including verbs, nouns and adjectives. The event

argument introduced by a sentence's main verb also serves as the argument for a set of predicates relating to the tense and aspectual type of the sentence (for instance, RE-S-RELN encodes a particular combination of Reichenbach's speech, event and reference times), and to its syntactic mode (in our case, either assertive or interrogative).

The system's database is also stored in the form of a DRS, which is termed the **global DRS**. (We do not turn the DRS into a first-order logical representation, because the system does not currently support theorem-proving.) The user can add facts to this database by entering assertive sentences, and can query the database by asking *yes/no* or *wh*-questions. Entities in the database are termed **DRS referents**, while entities in MRSs and in sentence DRSs are termed **variables**. Figure 3 gives a snapshot of the global DRS at the point just before the first question of the sample dialogue (sentence (b)) of Figure 1) is asked.

e70 e69 e68 e67 e66 DRSref74 e61 e60 e59 e58 DRSref65
h73: RE-S-RELN(e68 e69 e70) h73: WALK-REL(e70 DRSref74) h72: DOG-REL(e67 DRSref74) h72: RED-REL(e66 DRSref74) h64: RE-S-RELN(e59 e60 e61) h64: BARK-REL(e61 DRSref65) h63: DOG-REL(e58 DRSref65)

Figure 3: Database DRS after two statements by the user

The system's dialogue model uses a hierarchical taxonomy of dialogue acts, the top levels of which are given in Figure 4. This hierarchy is, roughly speaking, a subtree of the hierarchy of dialogue acts given in Alexandersson *et al.* (1997), but it bears a resemblence to several other such taxonomies (see e.g. Allen and Core, 1997). There are two kinds of STATEMENT, both of which provide new information to the interlocutor: an ASSERTION is a new statement 'apropos of nothing', while an ANSWER is a statement which answers a question. Likewise, there are two kinds of QUERY, both of which pose a question: a QUESTION is a new query apropos of nothing, while a FOLLOW-UP QUESTION is a question about the proper interpretation of a previous utterance.

A simple model of dialogue structure is supported: a QUERY should be followed by an ANSWER; a STATEMENT of any kind should be followed by an ACKNOWLEDGE; a SENTENCE of any kind can be followed by a FOLLOW-UP QUESTION; and (at a meta-level) an UTTERANCE of any kind which doesn't follow these rules is followed by an appropriate ERROR report and ignored. FOLLOW-UP QUESTIONS allow nesting of dialogue moves in the usual way, so a familiar kind of **dialogue stack** is maintained (c.f. for instance Ginzburg's (1996) partially-ordered set of 'questions under discussion', Lemon *et al.*'s (2001) 'issues raised' stack). An example of an embedded structure is shown after utterance (f) in the sample dialogue. (Note that we currently require all ACKNOWLEDGEMENTS to be given explicitly, so at the end of a nested dialogue game, two successive ACKNOWLEDGEMENTS are given.)

An incoming utterance is interpreted by LKB, which delivers an MRS to the dialogue manager. The dialogue manager first computes the dialogue act associated with the utterance, partly from its MRS and partly from dialogue expectations (see Section 4.4 for an example of this). It also converts the MRS into a DRS, and attempts to resolve the presuppositions associated with this DRS. What happens next depends in a complex way on the success of this operation, and on the DRS, and on the dialogue act. It is at this stage that we intend a treatment of presuppositions to provide a unifying perspective.

4 Presuppositions for dialogue management

Correct resolution of presuppositional elements of a sentence requires a reasonably complex computational mechanism, even within a simple database query domain like that of Te Kaitito. We have found that this

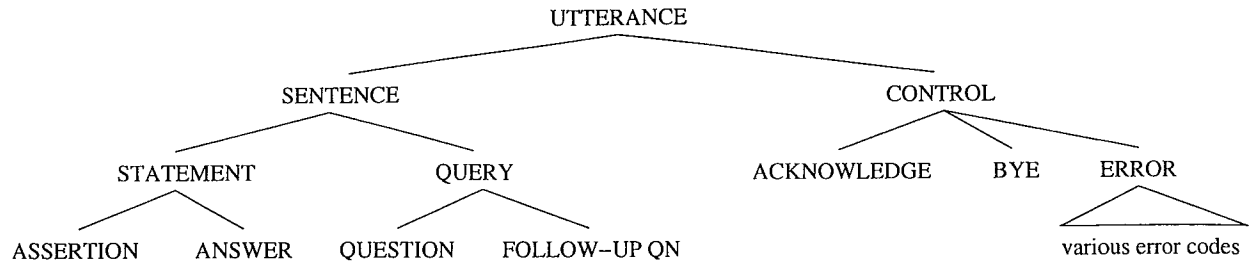


Figure 4: Dialogue act types in the dialogue model

mechanism is very similar to that required for various other tasks in dialogue management (notably question answering and classifying some dialogue acts). From a software engineering perspective it makes sense to enquire whether these similarities are close enough to allow the same software module to be used for the different tasks. This approach has been successful with only minor changes to the representation of simple presuppositions. It should be emphasised, however, that we are not claiming that all uses of this mechanism actually are instances of presupposition resolution. We only wish to draw attention to the fact that a single mechanism can be used to solve what appear to be quite distinct problems in construction of a dialogue system.

4.1 Presuppositions for computing reference

It is conventional to treat certain referring expressions (in particular definite descriptions) as presuppositional; for instance, utterance (e) (*The blue dog chased a green cat*) presupposes the existence of a uniquely identifiable blue dog. The DRS for this utterance is shown in Figure 5, with a presupposition box for *the blue dog* in the style of van der Sandt (1992). This box explicitly represents the fact that to match the variable $x4$ to a DRS referent in the global DRS we will need to find some referent for which the predicates in the box are true. Implicitly it requires uniqueness of that referent, but this is enforced by the code, not the representation. (Section 4.5 will show a case where we wish to relax this restriction.)

In cases where several DRS referents

e69 e68 e67 e66 DRSref79 x4 e70	
h73: ASSERTION-RELN(e70)	
h75: CAT-REL(e69 DRSref79)	
h75: GREEN-REL(e68 DRSref79)	
h73: RE-S-RELN(e66 e67 e70)	
h73: CHASE-REL(e70 x4 DRSref79)	
Presup-box:	**e76 x4 e77** h78: DOG-REL(e76 x4) h78: BLUE-REL(e77 x4)

Figure 5: DRS for *The blue dog chased a green cat.* The marker variable for the presupposition box is shown in bold.

match the variable in a presupposition box, we use a simple representation of the saliency of referents to disambiguate. The system maintains a **saliency list**—at present, simply a list of DRS referents in order of their most recent appearance. In resolving the presuppositions of an utterance we use the saliency list to limit the number of possible distractors. The most salient individual matching the referring expression is first found. If there is another matching individual in the list within a certain **saliency range**, the presupposition is taken to be ambiguous. (The saliency range increases in relation to the depth in the list of the most salient matching individual.)

If an incoming utterance's presuppositions can be fully resolved, then if it is an assertion its DRS is merged with the global DRS, if it is a question it is answered (see Section 4.3), and if it is an answer it is acknowledged. If they cannot be resolved, a follow-up question is generated (see Section 4.2). The

presupposition-handling routine thus in large part determines what happens next.

4.2 Multiple presupposition boxes for follow-up questions

Utterance (f) is a statement whose presuppositions cannot be resolved, even using the saliency list. To resolve the ambiguity the system asks a follow-up question. One innovation in our system is that when building our DRS representations we create a separate presupposition box for each presuppositional referring expression. The marker variable in the presupposition box shows which variable it is intended to disambiguate. This makes it easy to construct a specific follow-up question for the variable whose referent cannot be uniquely identified. In the example above, it is *the dog* which is ambiguous, not *the weasel.* Since these are kept in separate presup boxes, only one is queried in the follow-up question.

4.3 Presuppositions in the representation of questions and answers

Our grammar accepts a range of sentence structures specific to answers, such as *It was the dog* or *The dog.* These sentences clearly require context in order to be interpreted. There are a great number of proposals about how to model this context-dependence; Ginzburg and Sag (2000) is one particularly well worked out method. Our treatment is relatively simple: we have chosen to account for such presuppositions by modelling a question as something which introduces a meta-level predicate about the questioner's (lack of) knowledge about a domain object. For example, the DRSs for the *wh*-question *Which red cat barked* and its answer *The big cat* are shown in Figure 6. We take the *wh*-question to comprise an assertional and a presuppositional component, just like a declarative statement. Its presupposition is that some red cat $x4$ barked. Its assertional content, which is added to the discourse in the usual way, is that $x4$ is unknown—the UNBOUND-RELN predicate conveys this. Naturally, this predicate is of a different type from predicates like RED-REL, CAT-REL etc; it conveys information

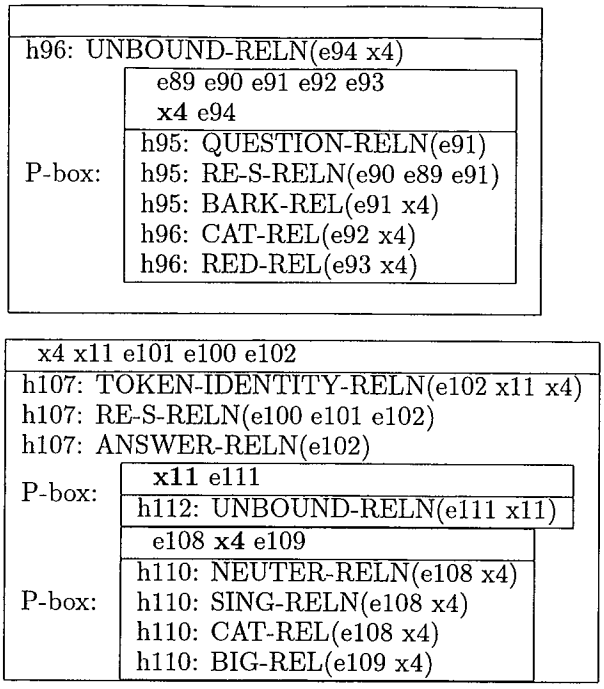


Figure 6: DRS for a question (*Which red cat barked*) and its answer (*The big cat*)

about the questioner's epistemic state rather than about the world. However, this liberty allows us to deal with the presuppositions of the answer sentence in a very simple way. As the right-hand DRS in Figure 6 shows, *The big cat* contains two separate presuppositional elements, which we store in separate boxes (as discussed in Section 4.2). The first is the regular presupposition associated with the definite NP *the big cat.* The second is simply that an unknown object is present in the discourse. (For a more elaborate answer sentence such as *It was the big cat which barked* we would be able to say more about this unknown object.) The main assertional content of the answer sentence is provided by the TOKEN-IDENTITY-RELN predicate, which states that the unknown object is identical to some other variable introduced by the sentence—in our case, the variable associated with the presupposed big cat.

4.4 Follow-up questions and the stack DRS

As already mentioned, the system maintains a stack of dialogue acts to support nested follow-up questions. This raises the question of storage of the DRSs associated with the sentences associated with the dialogue acts sitting in the stack. These DRSs cannot always be merged with the global DRS, because they may (in the case of a statement producing a follow-up question) have variables that have not yet been bound to DRS referents. Our solution is to maintain an intermediary **stack DRS**, that contains a mixture of DRS referents and variables. At the point where all follow-up questions have been answered, all variables have been bound to global DRS referents (this is the function of presupposition boxes) so it is safe to merge the stack DRS into the database DRS.

The stack DRS has two other useful functions. Most straightforwardly, it provides a way of limiting the domain of 'epistemic' predicates like UNBOUND-RELN and TOKEN-IDENTITY-RELN. Occurences of the former predicate can simply be removed when the stack DRS is merged with the global DRS, and occurrences of the latter predicate can be removed by variable substitutions. A second use for the stack DRS is in determing the dialogue act associated with an incoming utterance in certain specific cases. As already mentioned, the system allows for two kinds of question, QUESTION (apropos of nothing) and FOLLOW-UP QUESTION. However, there are currently no syntactic means to distinguish between these two dialogue act types (we have not yet implemented echo questions). Fortunately, the system can use the stack DRS to disambiguate. A follow-up question can only make reference to the the entities in the utterance it is querying, or those lying higher in a chain of follow-up questions. For instance, the query *Which dog chased the cat* can be a follow-up question to the statement *The dog chased the cat* but not to the statement *The aardvark chased the cat.* We specify that an incoming question is a FOLLOW-UP QUESTION if and only if its presuppositions can be resolved using only the variables and DRS referents in the stack DRS. In a few cases this strategy will wrongly treat a QUESTION as a FOLLOW-UP QUESTION; however in these cases the answer will be the same, and the error will be overcome as soon as a question involving other entities or a statement is received.

4.5 Question answering as presupposition resolution

The use of presuppositions in representing the semantics of questions, as described in Section 4.3, also has some beneficial procedural consequences: the process of generating answers to questions occurs largely as a side-effect of the regular process of presupposition resolution. The presupposition resolution algorithm returns a list of possible lists of bindings from variables to DRS referents—these are in fact exactly what we need in order to generate the answers to questions. Multiple bindings might represent a list answer (such as *The blue cat and the red cat both chased the dog*). In the case of a yes/no question involving the indefinite article (*Did a cat chase the dog*) no distinction is made between a single binding or multiple bindings: either represents a *yes.*

Reducing the process of question-answering to the process of presupposition-resolution has the effect of simplifying the operation of the dialogue manager, which is one of the intentions in our system. However, this simplification also results in some interesting and subtle question-answering behaviour. In the interpretation of a question such as *What chased the monkey* (line (i) in the dialogue), only those monkeys which were chased will be considered as possible referents for the referring expression *the monkey.* In the example dialogue, at the point where *What chased the monkey* is asked, there are two possible referents for *the monkey.* If presupposition resolution and question-answering are treated separately, *the monkey* would therefore be interpreted as ambiguous. However in our treatment we can determine that the monkey we want is the one that was chased. This is because the DRS for the question has two presupposition

boxes: one for the definite DP *the monkey* and one for the (unknown) object that chased the monkey. The same set of bindings has to satisfy both presuppositions; this means in turn that the monkey was involved in a chasing event, which removes the other monkey as a possible distractor. Presupposition resolution functions naturally here both to control the procedural component of dialogue management (the generation of an answer) and to control the semantic content of the system's responses in relation to the content of incoming utterances.

5 Comparison with existing dialogue systems and theories

This final section situates Te Kaitito in relation to some other dialogue systems and models.

5.1 Task-based dialogue

As Te Kaitito is currently being developed as a database interface, the interaction is entirely user-driven. There is no provision in the architecture for system initiative, except in asking follow-up questions. Lemon *et al.* (2001), Larsson *et al.* (2000) and Xu and Rudnicky (2000) have developed systems that drive the interaction through some form of dialogue plan on the part of the system. We hope that the approaches to dialogue we have developed with Te Kaitito will allow the addition of a module containing more complex dialogue plans, allowing the system to make *motivated* dialogue moves, rather than simply responding to user demands. This module will be essential for the successful development of the language teaching application which is the eventual aim for the project.

5.2 The dialogue stack

Lemon *et al.* (2001) suggest that the use of stacks to represent dialogue is overly restrictive, because in a task-based dialogue the system may need to shift back and forth between topics, rather than either pushing a new topic onto the stack or popping the current topic to reveal an older one. Again, the database query domain allows this simplification; however for a more general dialogue system their system of dialogue move trees may need to be adopted.

5.3 Semantics of questions and answers

We have used an extremely simple semantics for questions and answers, which is sufficient for the database applications (and perhaps for language teaching also) but does not provide a general account. Particularly, there is no provision in Te Kaitito for partial or indirect answers to questions. Asher and Lascarides (1998) provide an account which explicitly represents the status of questions as steps in a plan to gain information. In this account, a statement is considered an indirect answer to a question if it allows the questioner to complete their plan (to gain the required knowledge) only after applying some deductive or active steps. A partial answer, on the other hand, leaves the possibility of providing more information at a later date. Neither of these types of answers is necessary for the database query domain, so we have chosen not to represent them.

5.4 Public and private information states and the stack DRS

It is common in dialogue systems to distinguish between the information assumed to be shared between the participants in the dialogue (the propositions that have been stated and accepted during the dialogue) and information that has been stated but has not yet been acknowledged, so that the agent has no guarantee that the information has been understood (for example Larsson *et al.* (2000) distinguish between the 'public' and 'private' sections of the agent's information state, and include in the private section information that has been stated but not yet accepted). In Te Kaitito, this distinction corresponds quite closely to the distinction between the context DRS and the dialogue stack DRS. We believe that the use of this distinction to provide dialogue move characterisations of sentences is innovative (at least in dialogue implementations).

References

Alexandersson, J., Buschbeck-Wolf, B., Fujinami, T., Maier, E., Reithinger, N., Schmitz, B., and Siegel, M. (1997). Dialogue acts in Verbmobil-2. Technical Report 204, DFKI Saarbrücken.

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

Asher, N. and Lascarides, A. (1998). Questions in dialogue. *Linguistics and Philosophy*, **23**(2), 237–309.

Copestake, A. (2000). The (new) LKB system. CSLI, Stanford University.

Ginzburg, J. (1996). Interrogatives: Questions, facts and dialogue. In S. Lappin, editor, *The handbook of contemporary semantic theory*, pages 385–422. Blackwell, Oxford.

Ginzburg, J. and Sag, I., editors (2000). *Interrogative investigations*. CSLI Publications, Stanford.

Kamp, H. and Reyle, U. (1993). *From discourse to logic*. Kluwer Academic Publishers, Dordrecht.

Knott, A., Bayard, I., de Jager, S., and Wright, N. (2002). An architecture for bilingual and bidirectional nlp. Manuscript, Dept of Computer Science, University of Otago.

Larsson, S., Ljunglof, P., Cooper, R., and Engdahl, E. (2000). GoDiS—an accommodating dialogue system. In *ANLP/NAACL 2000 Workshop on Conversational Systems*.

Lemon, O., Bracy, A., Gruenstein, A., and Peters, S. (2001). Information states in a multi-modal dialogue system for human-robot conversation. In *Proceedings of the 5th Workshop on Formal Semantics and Pragmatics of Dialogue (Bi-Dialog)*, pages 57–67.

Van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, **9**, 333–377.

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

Dialogue as Collaborative Tree Growth

Ruth Kempson and **Masayuki Otsuka**
ruth.kempson@kcl.ac.uk, masayuki.otsuka@kcl.ac.uk
King's College London

Abstract

This paper applies Dynamic Syntax (Kempson et al., 2001) to dialogue modelling, and provides a characterisation of production that is relative to a converse process of tree growth which constitutes a parse check. As evidence for this approach, we place it in a psycho-linguistic perspective, using it to model (i) the parsing/production of elliptical expressions, (ii) dialogue properties such as a speaker-feedback mechanism, speaker/hearer alignments of structure, and speaker/hearer role reversal in shared utterances (Pickering and Garrod, forthcoming).

Very generally in the current research perspective, production and parsing are taken to be independent applications of a neutral centrally available grammar formalism. In abandoning this assumption and articulating a parsing-directed grammar formalism, the Dynamic Syntax framework (Kempson et al., 2001) faces the challenge of articulating the relationship between them. In this paper, we explore the formal mechanisms needed to model generation (as an idealised form of production) given the parsing-directed assumptions of Dynamic Syntax. We then show how the closeness of correlation posited between parsing and production mechanisms allows a natural extension to dialogue modelling that meets the challenge set by (Pickering and Garrod, forthcoming) that grammar formalisms be evaluated relative to their success in capturing key properties of dialogue.

1 Background Assumptions

Dynamic Syntax is a model of NL understanding in which parsing is defined as the progressive projection of a decorated tree structure following the left-right sequence of words in the string. Logical forms are represented as decorated trees, whose topnode is decorated with a formula $Fo(\alpha)$ of type t, and whose dominated nodes are decorated with subterms of the formula α. The central concept is that of goal-directed tree growth, defined using the modal tree logic LOFT (Blackburn and Meyer-Viol, 1994) with basic operators $\langle\downarrow\rangle$, $\langle\uparrow\rangle$ for mother and daughter relations respectively. Tree growth is defined over partial trees, each sequence of such trees starting from the requirement at a rootnode $?Ty(t)$ constituting the overall requirement to establish a logical form of type t at $Tn(0)$ (Tn for treenode). To this node, additional requirements such as $?\langle\downarrow\rangle Ty(e)$, $?\langle\downarrow\rangle Ty(e \rightarrow t)$, are added as subgoals, and lead to the introduction of daughter nodes with requirements $?Ty(e)$, $?Ty(e \rightarrow t)$. Tree nodes are thus created from the root downwards with imposed *requirements* which are subsequently met by tree growth actions dictated by incoming lexical items as they are parsed in a left-right sequence. Processes such as function application are defined in tandem with type-deduction to decorate nonterminal nodes as pairs of terminal nodes are successfully decorated. (At each stage, there is one itemised node under development, indicated by a pointer, $\Diamond$.) A complete tree is one which projects a logical form (with a formula of type t decorating its top node). ALL update processes are monotonic, progressively developing a tree structure meeting the requirements which are imposed on nodes as they are introduced.

In addition to the concept of requirement are other concepts of structural underspecification. Formula values may be underspecified,

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 69-76.

eg for anaphoric expressions, which project formulae of the form $Fo(\mathbf{U})$, $\mathbf{U}$ a metavariable. Underspecified tree relations also may be introduced (for left-dislocated expressions), with a treenode identified as dominated by some treenode a without at that point in the treegrowth process any fixed extension (its treenode described as $\langle\uparrow_*\rangle Tn(a)$ with a requirement for a fixed extension, $?\exists x.Tn(x)$). These various forms of underspecification interact in the process of progressive satisfaction of all imposed requirements through computational, lexical and pragmatic actions,[1] each of which constitutes a monotonic step of tree growth ($Ty(e)$ is a development of $?Ty(e)$, $Fo(Mary)$ is a development of $Fo(\mathbf{U})$, $\langle\uparrow\rangle Tn(a)$ is an update of $\langle\uparrow_*\rangle Tn(a)$, and so on). Wellformedness is defined in terms of the result of such actions: a sentence is wellformed if and only if at least one completed tree structure can be derived from a sequence of words, with no requirements outstanding. In using a tree description language, the system is like other parsing formalisms using descriptions of partial trees (Duchier and Gardent, 2001), (Joshi and Kallmeyer, forthcoming), (Koller et al., 2000), or descriptions of trees (Sturt and Crocker, 1996). But unlike them, partial trees are the basis for the grammar formalism, and not merely for semantic characterisations or parsing algorithms relative to an independently defined syntax. The entire concept of syntax is founded in this concept of growth of semantic representations along a left to right dimension without any intermediate and independent syntactic level of representation (Kempson et al., 2001).

Quantification in this system is expressed using the epsilon calculus (Meyer-Viol, 1995) with variable-binding term operators of type e in the style of arbitrary names of natural deduction systems for predicate logic. A sentence such as *A man smokes* is taken to project a logical form $Smoke(\epsilon, x, Man(x))$, derived like other aspects of the interpretation process through incremental construction from lexical specifications which only partially determine the resulting formula. Like [Copestake *et al*, 1999], scope is defined through an incrementally collected set of scope constraints allowing lexical variation; and these scope constraints jointly determine the evaluation of the constructed logical form. The form $Fo(Smoke(\epsilon, x, Man(x)))$, for example, is evaluated relative to a scope statement $S < x$ (S a variable representing the index of evaluation) as $Fo(S : Man(a) \wedge Smoke(a))$, where $a = (\epsilon x, Man(x) \wedge Smoke(x))$. Underspecification of scope determination is thus not expressed through underspecified tree-relations (as in (Erk et al., forthcoming),(Joshi and Kallmeyer, forthcoming)) but through metavariables in scope statements projected, for example, by indefinites. Like all other aspects of underspecification, these must be resolved during the construction process.

2 Dynamic Syntax and Production

Assuming this model of parsing, we wish to articulate the relation between parsing and an idealised production model.[2] What we aim to provide is a "tactical" generation system which takes a source tree as input and incrementally "checks" off nodes in this tree as a progressively enriched partial tree can be successfully induced by some selected word. The pointer in this partial parse tree picks out the node whose analogue in the source tree is being "checked", an action which is matched by the selection of some word from the lexicon and "writing" it at the right-hand edge of a sequence of already established words. Such checking action is licensed if and only if the word selected projects a compound parse action which lead to an update of the parse tree from that defined over the words already decided upon, mapping that partial tree onto an update reflecting the annotations on the node being checked. To see the general dynamics, consider a simple tree representing the content of *John saw a woman* with a formula $Fo(S_{PAST} : See(\epsilon, x, Woman(x))(John))$ decorating its rootnode, with accompanying scope statement

[1] A pragmatic process of substitution is presumed to provide the update for the meta-variables lexically projected by anaphoric expressions. Defined as an architecture within which steps of parsing can be articulated, the framework has nothing to say about the pragmatic constraints that determine actual choices.

[2] This paper does not address the problem of phonological/phonetic levels of realisation and their relationship to parsability of the output.

$S_{PAST} < x$.[3] We assume initially that the starting point for any production task is a full tree as source representing interpretation of the string, and a parse tree made up of a single rootnode with pointer and requirement $?Ty(t)$ (though see section 5 where both inputs to production and parsing are generalised to arbitrary partial trees). Generation steps are then licensed relative to some associated parsing step. Within the source tree, the generating system can, for example, "check off" the subject node, and "generate" the first word in the string, because there is a parsing routine whereby a combination of node-introducing rules operating on the rootnode together with the lexical actions of the name *John* can lead to the successful annotation of the subject node as $Fo(John), Ty(e)$. The search, then, through the lexicon is to find the word which provides these actions. In this paper, this aspect of the search is trivialised, by making the Fodorian assumption that words and concepts correspond one-to-one. However the selection of a pronoun is nontrivial, and is licensed as long as the underspecified input provided is sufficient for the hearer to identify the term intended, given the context provided.[4] Having checked off the node associated with the subject in the source tree in virtue of a successful step in the parse tree annotating the subject node in that tree, the next production step again follows the move of the parsing pointer which is back to the rootnode of the parse tree. From there a following computational step allows the introduction of a predicate-requiring node. Accordingly, the generator attempts to check the content of its predicate node in the source tree. With verbs defined as being of the form $IF\ ?Ty(e \rightarrow t), THEN...$, the lexicon is scanned for a word which will lead to the introduction and decoration of a node with $Fo(see), Ty(e \rightarrow (e \rightarrow t)$. The word *saw* is duly selected, which in addition enables the information about the tense construal also to be checked; and it is written at the right edge of the sequence, yielding *John saw*. Given that the actions of the word *saw* leave the pointer in the parse tree at the object node, that node alone remains to be realised by some sequence of words. The term $Fo(\epsilon, x, Man(x))$ can then be checked off because there is a sub-sequence of actions provided by the determiner and noun that introduce a subtree decorated with $Fo(\lambda P.\epsilon, P)$ and $Fo(x, Man(x))$ for some fresh variable x combining together to yield $Fo(\epsilon, x, Man(x))$. Once the hearer can be presumed to have a tree with all terminal nodes annotated, the nonterminal nodes then have to be taken as checked, given the automatic parse steps of functional-application/type-deduction that would apply to decorate nonterminal nodes in the parsing routine. The pattern is quite general. Generation involves three trees: one a fully annotated tree which forms the input to the process; second a tree with a subset of its nodes checked; and third, a tree which reflects the corresponding parse tree commensurate with establishing the node currently being checked. The provided sequence of actions is by no means unique. Any sequence of actions conforming to the pattern of pairing source tree incrementally with emergent parse tree following the sequence of words is licensed: eg left-peripheral placement of any "dislocated" NP presuming on a parse-sequence projecting an initially unfixed node (see section 4 below, and (Kempson et al., 2001)).

3 Ellipsis as Tree Abstraction

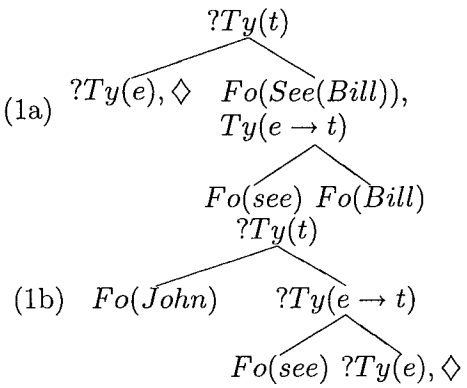
Without independent motivation, such a checking process would be nothing more than stipulation, and indication of the need for a separate production "grammar". The challenge is to reduce this checking system to some operation independently needed for natural language processing. In the interpretation of ellipsis, we find the mechanism we need, this being a tree-abstraction process enabling partial structures to be re-used. For example, consider the parsing of the elliptical fragment in (1), which can be interpreted either as "Harry saw Bill" or as "John saw Harry", the choice being free and determined pragmatically:

[3] We take S_{PAST} as a shorthand representation of tense, S_{PAST} the representation of the index relative to which the logical form $Fo(See(\epsilon, x, Woman(x))(John))$ is evaluated.

[4] In a fuller account of proper names, these too would require nontrivial identification of the individual being talked about with a lexical characterisation involving a meta-variable requiring substitution. However, we leave all such issues aside.

(1) John saw Bill. Harry too.

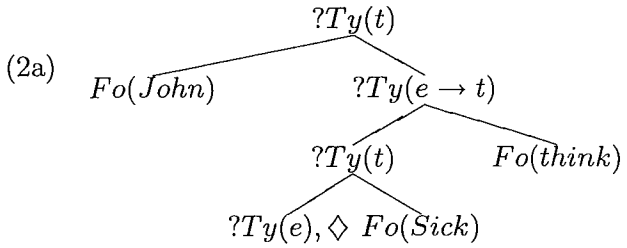
For this fragment, a subtree needs to be constructed from the tree established from the first sentence as the basis for interpreting the fragment *Harry*. This subtree is (1a) or (1b):



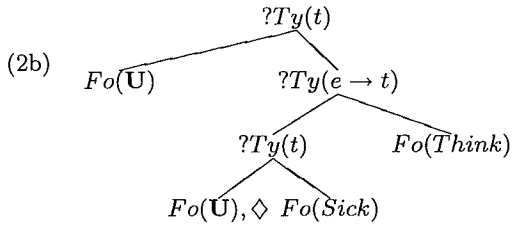
This process takes some decorated tree as input, and replaces some annotated node(s) of that tree with requirements, "abstracting" out established values and in so doing providing a "newly incomplete" tree with which to establish some different completion (notice how such reversal along the process of tree growth at a terminal node automatically applies also to nodes dominating that terminal node).

More generally, we allow a process of tree abstraction back along ANY dimension of tree growth defined over pointed partial trees, returning annotations to requirements ($Ty(e)$ to $?Ty(e)$), Formula values to meta-variables ($Fo(Mary)$ to $Fo(\mathbf{U})$), and so on, to create a partial tree needing completion.[5] We can then define ellipsis construal as the parsing of the expression relative to an arbitrary partial tree such that the result yields a propositional formula as value. This process of abstraction for the processing of a fragment yields (like (Dalrymple et al., 1991)) the parallelism of construal between antecedent structure and interpretation of the fragment characteristic of ellipsis. Sloppy/strict interpretations for ellipsis are expressed as different forms of abstraction or anaphoric construal.[6] For example, the interpretation of the elliptical element as 'John thinks Bill is sick' in (2) requires abstraction of decorations at one node:

(2) John thinks he is sick. Bill too.



Replacing one annotated term of type e with $?Ty(e), \Diamond$, will allow the parsing of the fragment to provide a new object value and so lead to $Fo(Think(Sick)(Bill))(John)$ decorating a newly annotated root node. The sloppy interpretation of (2) can be characterised as an abstraction that returns *Formula* values to metavariable values, yielding a tree with $Fo(\mathbf{U})$ decorating both matrix and subordinate subject nodes, an abstraction which upon parsing the fragment will enable $Fo(Think(Sick(Bill))(Bill))$ to be derived:



The abstraction process, being free, will include abstraction over only non-terminal nodes. This provides a basis for modelling fragments as in

(3) John read a book. Mary's. Which he enjoyed.

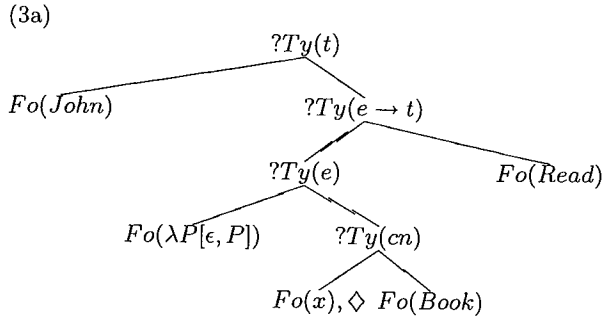
Abstraction of the nonterminal nodes dominating the variable in (3a) allows further extension of the information to be compiled at the dominating nominal node by the building of adjunct structures:[7]

[5] This recovery of a partial tree by abstraction is the structural analogue of the recovery of a suitable relation in the HOU account of ellipsis (Dalrymple et al., 1991). See (Kempson et al., 1998) for arguments that ellipsis construal involves a structural representation of content.

[6] Strict interpretations of the VP pronominal form *do so* involve substitution of the pronominal predicate by the provided antecedent predicate.

[7] In (Kempson et al., 2001), relative clauses and genitives are analysed as projecting quasi-independent linked structures with propositional formulae con-

(3a)



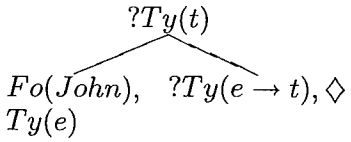
The availability of such abstraction allows elliptical fragments whose construal extends the contextually provided tree, a type of ellipsis that is more problematic for a purely semantic account.

4 The Ellipsis Generation Parallelism

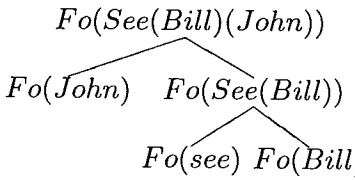
The significance of such abstraction for generation, and the primary focus of this paper, is that it provides a basis for the checking process. Notice that the partial trees defined by this abstraction process need not be those constructed in the course of parsing. (1a), (2a), or (2b) are not intermediate structures in parsing an English string. However, such trees notably correspond to intermediate steps of generation if we model the generation process of "checking" a node as a process of successively abstracting out information from the source tree. On this view, checking of nodes is an emptying process, reversing the growth process by replacing annotations on nodes with weaker decorations, such as the replacement of a formula by $?Ty(X)$, for suitable type. For example (1a) is the tree established when the subject term is the first constituent to be "checked" in uttering *John saw Bill.*

Exploring this further, let us define the checking process as an abstraction from some partial tree replacing it with some partial tree further along a process of tree growth away from a complete tree, eg replacing annotations with requirements. The steps of abstraction licensed in generation are constrained as before to be those for which there is a corresponding parse step from the partial tree associated with a previous generation step to that resulting from the sequence of actions associated with the word selected for linearisation. But now the effect of the constraint is that at each intermediate step, the pair of progressively abstracted partial tree and progressively enriched parse tree unify to yield back the source tree. For example, the linearisation of the word *John* in uttering (1) and the abstraction from source tree to (1a) is licensed, because there is a well-defined parse step from a tree with single node $?Ty(t)$ to the tree (1c), (1a) and (1c) unifying to provide all that is needed to establish the source tree (1d):

(1c)

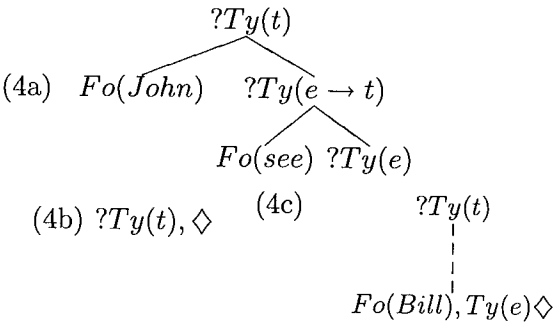


(1d)



Alternative word orders follow the same pattern, the generation of left dislocation sequences also being incremental. Thus the non-subject term in the source tree for (4) is licensed to be linearised first and abstracted from the source tree to yield (4a), in virtue of a corresponding parse step from the start state (4b) to an update (4c) introducing an unfixed node decorated with a non-subject, (4a) and (4c) merging to yield the source tree:

(4) Bill, John saw



Pronouns, equally, are licensed by this generation process. A node in a tree can be abstracted over and replaced by a requirement, with presentation in the linear sequence of words of a pronoun, as long as the hearer can

structed from them that are used to extend the restrictor decorating the $Ty(cn)$ node.

be presumed to have access not only to the lexical actions given by the pronoun (which project a meta-variable as their formula decoration), but also a contextually provided occurrence of some appropriate term with which to substitute that meta-variable.[8]

One advantage of this assumption of the correspondence of the checking process intrinsic to generation and tree abstraction is that checking of nonterminal nodes is now expected. Abstraction of a terminal node that removes a Fo value automatically induces removal of any annotations on dominating nodes which contain that formula as a sub-term.

A second major advantage of this approach is the characterisation it provides of the parsing-production relation while nevertheless maintaining their distinctiveness.[9] Parsing is the development of tree growth from start state ($?Ty(t)$) across intermediate pointed partial trees (T_i) to some completed goal (using Computational, Lexical and Pragmatic actions (C, L, P) involving a string $w_1, ...w_n$), with all subgoals emptied. Generation is the reverse shift from some completed tree as the source tree, with progressive emptying (tree abstraction) as each word $w_1, ...w_n$ is selected until a tree skeleton is reached essentially equivalent to the start state, with all nodes decorated only with requirements. However, generation is not simply a sequence of parsing actions in reverse. Rather there is an interdependence between them – the success of abstraction relies on the parsability of the linearised output. As Figure 1 indicates, the difference between parsing and generation lies in the metalevel aspect to generation. Each step in generating a string with interpretation $Fo(\phi)$ represented as a tree structure T involves correlating some proposed abstracted tree (T_i) and some corresponding parse tree (T_i') which must unify to yield back the source tree T. The tightness of this parsing-generation correspondence is due to there being no intermediate system of constraints: the parsing architecture is the grammar formalism, syntactic structure is no more than progressive construction of semantic representation, and production is defined relative to this tree-growth process.

[8]There are many similarities between this account and the system of SPUD (Stone and Doran, 1997), both providing a syntactic and semantic characterisation in which pragmatic steps are integrated. However, SPUD, in using LTAG, is head-driven and assumes a family of elementary trees for each head, a tree for a left-dislocation string being an input variant of a tree reflecting the canonically ordered string. In the DS account, generation of left-dislocation strings is fully incremental.

[9]The lack of simple reversability of parsing and generation actions is wellknown. See (Levelt, 1989).

5 Dynamic Syntax as a basis for dialogue

While this articulation of the correspondence between production and parsing remains no more than a blueprint for future research, it notably satisfies desiderata for successful dialogue modelling set out in (Pickering and Garrod, forthcoming), who list as properties of dialogue: widespread alignment between speakers of syntactic, lexical and semantic properties, a large proportion of ellipsis and repetition in dialogue, common occurrence of utterances in which speakers finish each others sentences, and the need in any model for a parsing feedback mechanism. First, Dynamic Syntax assumptions lead one to anticipate maximal alignment of syntactic/semantic/situational levels by eliminating levels other than that of semantic representation. Despite no separate syntactic representational level, even alignment of double object constructions is expected:[10]

(5) R: "I bought Eliot a scooter."
H: "I bought him a skateboard."

Lexical specifications involve projection of concept plus a sequence of tree-update actions; and repetition of a word ensures identity of both concept and sequence of actions employed in uttering/parsing the first occurrence. Verbs such as *give*, which have two discrete sequences of actions, require different lexical specifications to reflect these strategies. In consequence, repetition will ensure recovery of the concept and actions used in the previous

[10]This distinguishing of discrete sub-specifications of such lexical entries according to the discrete actions they induce will, equally, provide a basis for explaining the oddity of switching from one to another in ellipsis, as in (i):
(i) I gave Eliot a scooter. ??And a skateboard to Bill.

occurrence, whichever that is. Hence 'syntactic' as well as 'semantic' alignment. Nonalignment would involve shifting from one lexical entry to another.

The account automatically yields an incremental self-monotoring device, as the checking of each generation step against the grammar formalism is a self-monitoring parsing step. We expect immediate production breakdown with parsing feedback delay (contra (Levelt, 1989): (Mackay, 1987)).

Dynamic Syntax assumptions also lead one to expect ellipsis to play a large role in dialogue, with generation and construal of ellipsis both defined in terms of a partial tree as context.[11] By the proposed account of generation, utterance of elliptical fragments is licensed because the fragment may be all that is required to enable a hearer to construct a second propositional formula, given the availability of tree abstraction on some preceding tree to provide a partial tree as context which unifies with the partial tree which the fragment provides:

(6) A: I bought a scooter for Bill. A skateboard too.

For example, in (6), relative to the context provided by the tree resulting from processing *I bought a scooter for Bill*, all that is required as input to interpreting the fragment *a skateboard* is to abstract the decoration $Fo(\epsilon x, Scooter(x))$ from the tree provided by the first sentence, replacing it by the requirement $?Ty(e)$ and pointer at the object node, this inducing a shift back at the predicate and root nodes to the requirements $?Ty(e \rightarrow t)$ and $?Ty(t)$. In consequence all that is required to communicate such a structure is to utter the fragment *a skateboard*, for with this utterance, a second propositional formula can be compiled by the hearer. The processing of ellipsis, that is, requires a partial tree as input to the processing task. The generation of ellipsis can accordingly presume on the recovery of such a partial tree by the hearer. Hence its economy to both speaker and hearer.

It is notable that in all cases of parsing, whether complete or incomplete, each step is defined relative to some partial structure as context. Ellipsis differs only in that the input to the interpretation process is generalised from the initial start state to an arbitrary partial tree. Given the ellipsis-generation parallelism, this allows a corresponding generalisation of the generation task. The source tree too can be from an arbitrary partial tree. This provides an appropriate basis for characterising what it means to convey an incomplete thought, thereby incorporating the phenomenon of speaking without a complete structure in mind within the general account of production without special stipulation.

As a final testcase, we show how the account can model role-reversals in collaborative utterances, presented by Pickering and Garrod as a major challenge to current orthodox grammar formalisms: Consider (7) in the context of a discussion of what to buy R's new grandson (H being the father):

(7) R: What shall I give to ...
H: Eliot?

In linearising the partial string, R's generation task starts from a source tree with a metavariable WH as the object argument whose value is to be requested and some constant denoting the person to be bought a present. R initiates the string with the *wh* expression on the grounds that H has a parse strategy for introducing an unfixed node, decorating it with the lexically projected WH metavariable, and merging it later with some appropriate fixed position. H, parsing, constructs such a partial tree, updating it by merging the unfixed node with the object node introduced in parsing the verb *give*. This leaves one further node in the parse tree needing completion to meet the three-argument requirement of *give*. At the juncture of uttering/parsing the indirect object expression, this partial tree is shared by R and H. With R signalling name loss, H provides it from independent knowledge, taking over the generation task, completing the pair of emptied tree (associated with the linearised string) and the completed tree (as constructed interpretation). R, now parsing, already has the partial tree constructed as a check on her own incomplete utterance; so all she has to do is to process the name *Eliot* to complete the partial parse tree. The construction of a tree in parsing some string, and the construction

[11] The account requires anaphora/ellipsis resolution at the level of tactical generation (contra (Dale, 1992)).

of the same tree in producing that string as a constraint on the generation steps licensed ensures the naturalness of the shift from one activity to the other.

Overall, the characterisation of NL syntax in terms of left-to-right growth of a structural representation of content and the characterisation of NL production as a metalevel activity making essential reference to such processing steps, jointly provide the basis for a natural account of dialogue with no stipulation of actions other than those independently required in NL processing.

References

P. Blackburn and W. Meyer-Viol. 1994. Linguistics, logic, and finite trees. *Bulletin of Interest Group of Pure and Applied Logics*, 2:2–39.

R. Dale. 1992. *Generating Referring Expressions.* MIT Pres, Cambridge MA.

M. Dalrymple, S. Shieber, and F. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14:399–452.

D. Duchier and C. Gardent. 2001. Tree descriptions, constraints and incrementality. In H. Bunt et al, editor, *Computing Meaning.* Reidel, Dordrecht.

K. Erk, A. Koller, and J. Niehren. forthcoming. Processing underspecified semantic representations in the constraint language for lambda structures. *Research on Language and Computation.*

A. Joshi and L. Kallmeyer. forthcoming. Factoring predicate argument and scope semantics: Underspecified semantics with ltag. *Research on Language and Computation.*

R. Kempson, W. Meyer-Viol, and D. Gabbay. 1998. Vp ellipsis: towards a dynamic, structural account. In S. Lappin and S. Benmamoun, editors, *Fragments: Studies in Ellipsis and Gapping*, pages 227–290. Oxford University Press, Oxford.

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

A. Koller, J. Niehren, and K. Striegnitz. 2000. Relaxing underspecified semantic representations for re-interpretation. volume 3.

W. Levelt. 1989. *Speaking: From Intention to Articulation.* MIT Press, Cambridge, MA.

D. Mackay. 1987. *The Organization of Perception and actions: A theory for Language and Other cognitive Skills.* Springer, New York.

W. Meyer-Viol. 1995. *Instantial Logic.* Ph.D. thesis, University of Utrecht.

M. Pickering and S. Garrod. forthcoming. Toward a mechanistic psychology of dialogue. *Brain and Behavioral Science.*

M. Stone and C. Doran. 1997. Sentence planning as description using *TAG. ACL 1997*, pages 198–207.

P. Sturt and M. Crocker. 1996. Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11:449–94.

Parsing

$$\frac{\begin{array}{ccccc} START(?Ty(t)) & \longmapsto^{\{C,L,P\}} & T_i & \longmapsto^{\{C,L,P\}} & GOAL(Fo(\psi), Ty(t)) \\ String(w_1) & & \langle w_1 \dots w_i \rangle & & \langle w_1 \dots w_n \rangle \end{array}}{}$$

Generation As Tree Abstraction Plus a Parsing Control

$$\begin{array}{lccccc} Abstraction: & Source(Fo(\phi), Ty(t)) & \longmapsto^{\{ABSTR\}} & T_i & \longmapsto^{\{ABSTR\}} & START(?Ty(t)) \\ \hline Parsing: & START(?Ty(t)) & \longmapsto^{\{C,L,P\}} & T_i' & \longmapsto^{\{C,L,P\}} & Source(Fo(\phi), Ty(t)) \\ String: & 0 & & \langle \dots w_i \rangle & & \langle \dots w_i \dots w_n \rangle \end{array}$$

Figure 1: **Parsing and Generation**

An algorithm for processing referential definite descriptions in dialogue based on abductive inference

Peter Krause

1 Introduction

The following example of Kripke (1979) illustrates the phenomenon of referential uses for definite (mis-)descriptions pointed out by Donnellan (1966).

A and **B** are talking at a party. They are looking at a scene on the other side of the room. **B**, but not **A**, has seen that a woman has poured mineral water in her glass. In this situation **A** says to **B**:

(1) The woman drinking champagne over there is happy tonight.

The puzzle is to explain how **A** can use the definite description to refer to someone who does not meet its descriptive content. For Kripke's example, it is characteristic that the intended referent of the definite description is perceptually available to both the speaker and the hearer. In some of Donnellan's original examples, this is not the case. For instance, the sentence

(2) Smith's murderer is insane.

may be used in a situation in which a person believed to be Smith's murderer by the speaker (but not necessarily by the addressee) is known to the conversation participants from a previous encounter but not present in the utterance situation.

In this paper, an algorithm is proposed which models this understanding process. The algorithm uses (i) revising abductive presupposition justification, (ii) defeasible inferences to extend immediately given perceptual information, (iii) a pragmatics of acceptance based on belief revision, (iv) abductive belief explanation by assumptions about information sources.

I will present the algorithm in action by exemplifying it with a trace for the paradigmatic case of (1)[1].

2 The algorithm

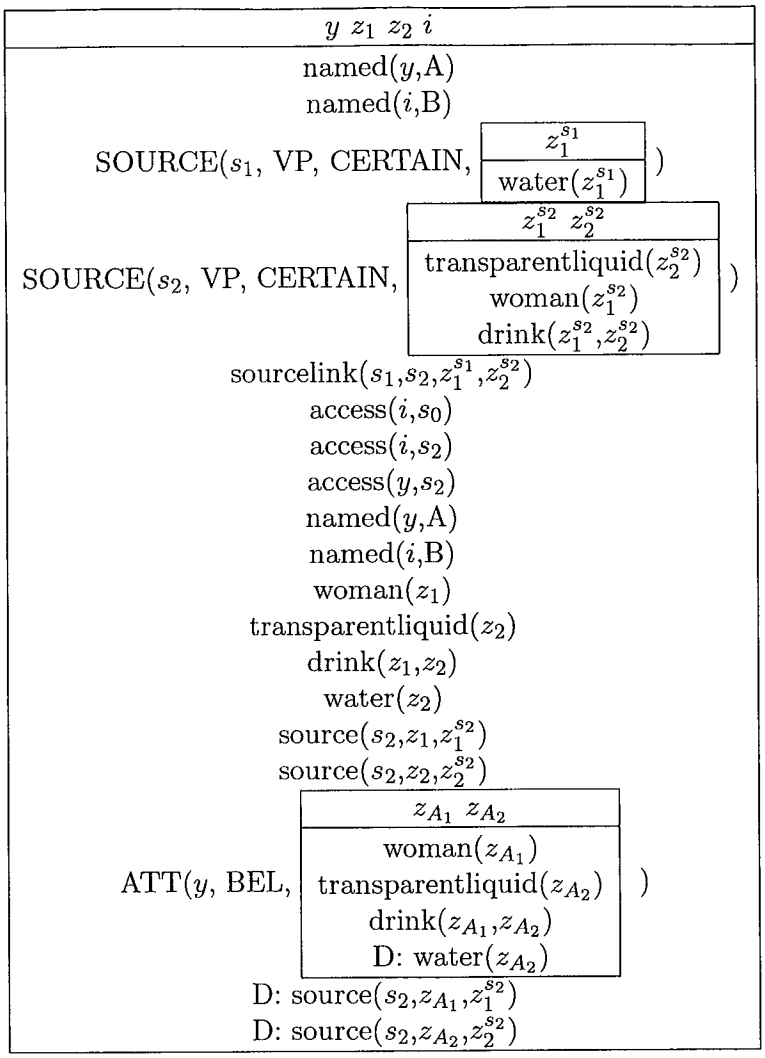
The two interlocutors are named **A** and **B**. The algorithm traces the development of **B**'s assumptions. **B** will therefore be represented using the indexical discourse referent i.

The starting point

The following DRS represents the initial communication situation.

[1]The underlying theory and formalism will appear in an article.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 77-84.



The setting involving common perceptual access to the scene on the other side of the room is modelled abstractly by the two source inscriptions on top of the representation. The first source represents the information that The source mode indicator **VP** qualifies immediate visual perception. The source assumptions anchoring the speaker belief to the information source s_2 [source(s_2,$z_{A_{1/2}}$,$z_{s_{1/2}}$)] express the hearer's belief that the speaker's discourse referents z_{A_1} and z_{A_2} are based on information derived from the referents declared in the source z_1^{s2} and z_2^{s2}. They are labelled as defeasible because it is uncertain whether the speaker has used this information source, and because the belief ascription is subject to revisions during the algorithm. The source link expresses the hearer's belief or knowledge that two referents from different information sources derive from the same object. In the case at hand, this hearer belief will derive from continuous observation of the scene.

The non-defeasible part of the belief ascription to **A** is the result of applying information transfer axioms for the hearer and for his interlocutor to the information source s_2. Information transfer axioms are conditionals. The antecedent consists of an information source declaration such as the one in the representation above and the assumption that an agent has access to the source. The hearer belief D: water(z_2) and the speaker belief ascription D: water(z_{A_2}) are the result of a default extension: transparent liquids may turn out to be water.

The sentence is formalized as happy($\delta(x_1, \overline{x_2}; \text{champagne}(x_2); \text{drink}(x_1, x_2))$). Here δ is the definite description operator of Krause (2001). The hearer must justify the presuppositions with respect to the context ascribed to the speaker. This gives rise to the following abductive inference problem.

Background Theory	Explanandum
z_{A_1} z_{A_2} woman(z_{A_1}) transparentliquid(z_{A_1}) drink(z_{A_1}) D: water(z_{A_2})	*defined*($\delta(x_1, \text{woman}(x_1); \overline{x_2}; \text{champagne}(x_2); \text{drink}(x_1, x_2))$)

There is no consistent justification without accommodation, because being water is incompatible with being champagne. A justification involving the accommodation of a new person discourse referent is ruled out by the assumption that the participants see the situation clearly. Accommodating the assumption that a woman is drinking champagne about whose drink (if any) **B** has no perceptual information yet would be another option, which leads to a different interpretation of the sentence. Now, we are interested in the interpretation in which a property ascription has to be revised. There is no consistent justification, but one revising one:

z_{A_1} z_{A_2}
woman(z_{A_1})
champagne(z_{A_2})
drink(z_{A_1},z_{A_2})
$z_{A_1} \leftarrow x_1$

The result of revising justification explains the definedness of the presupposition trigger. The corresponding result context is

z_{A_1} z_{A_2}
woman(z_{A_1})
champagne(z_{A_2})
woman(z_{A_1})
drink(z_{A_1},z_{A_2})
$z_{A_1} \leftarrow x_1$
happy($\delta(x_1, \text{woman}(x_1); \overline{x_2}; \text{transparentliquid}(x_2); \text{drinks}(x_1, x_2))$)

The satisfied presupposition can be rewritten using the α-operator:

z_{A_1} z_{A_2}
woman(z_{A_1})
transparentliquid(z_{A_2})
drink(z_{A_1},z_{A_2})
champagne(z_{A_2})
$z_{A_1} \leftarrow x_1$
α(x_1 x_2 | woman(x_1), transparentliquid(x_2), drink(x_1,x_2))
happy(x_1)

Once the presuppositions have been justified, they are semantically inactive and are removed by the algorithm. In the same step, the presuppositional information is marked as such by the label *Pre*. This has no semantic effect, but is useful for constructing the input to the following belief explanation step. Furthermore, the resolution equation is simplified away by substitution:

z_{A_2}
woman(z_{A_1})
champagne(z_{A_2})
drink(z_{A_1},z_{A_2})
happy(z_{A_1})

The justification results are turned into beliefs ascribed to the speaker (the first 13 conditions of the "starting point" DRS have been abbreviated to $\mathrm{K}^{hearerbelief}$):

$y\ z_1\ z_2\ i$	
$\mathrm{K}^{hearerbelief}$	
ATT(y, BEL,	$z_{A_1}\ z_{A_2}$
	Pre: woman(z_{A_1})
	transparentliquid(z_{A_2})
	Pre: champagne(z_{A_2})
	Pre: drink(z_{A_1},z_{A_2})
	Ass: happy(z_{A_1})
)	

The source assumptions for z_{A_1} and z_{A_2} had to be removed because the content of the belief ascription has changed.

The presupposition and the assertion have been marked as such by the labels **Pre** and **Ass**. This labelling of the source of the information is useful in the next step, where only the presuppositional part is put into the explanandum of the abductive inference step.

Now the hearer is confronted with an incongruency between the speaker belief he can predict on the basis of the commonly perceived situation encoded in the information source s_2 and the belief ascription derived from the presupposition (and the assertion) of the utterance. How is it possible that the speaker believes of an object he is talking about that it is champagne although the only likely candidate is in fact water ? Again, an abductive inference problem has to be solved. This time, the explanation is the fact that the speaker appears to have a certain belief. I will assume that the hearer limits himself to an attempt of explaining the presuppositional part of the speaker belief, because he can not hope to find an explanation for the assertion which is typically new and surprising information[2].

Background theory:

[2]In the example under discussion, the happiness of the woman in question may actually be perceptible as well, but in the general situation, this will not be the case.

The information transfer axioms (K^{modax}) together with the lexical knowledge (K^{mp}) and the hearer belief without the belief ascription to the speaker (*hearerbelief*(K_{heacon})):

$K^{modax}; K^{mp}$

$y\ z_1\ z_2\ i$
named(y,A)
named(i,B)
SOURCE(s_1, VP, CERTAIN, [$z_1^{s_1}$ | water($z_1^{s_1}$)])
SOURCE(s_2, VP, CERTAIN, [$z_1^{s_2}\ z_2^{s_2}$ | woman($z_1^{s_2}$), transparentliquid($z_2^{s_2}$), drink($z_1^{s_2}$,$z_2^{s_2}$)])
sourcelink(s_1,s_2,$z_1^{s_1}$,$z_2^{s_2}$)
access(i,s_2)
access(y,s_2)
named(y,A)
named(i,B)
woman(z_1)
transparentliquid(z_1)
drink(z_1,z_2)
D: water(z_2)

Explanandum:

The part of the belief ascribed to **A** based on the presupposition of the sentence is the explanandum of the belief explanation step.

ATT(y, BEL, [$z_{A_1}\ z_{A_2}$ \| woman(z_{A_1}), champagne(z_{A_2}), drink(z_{A_1},z_{A_2})])

The belief explanation problem is solved by the *explanation*

source(s_2,$z_1^{s_2}$);source(s_2,z_A)

The two discourse referents have the same specific source s_2. This is indeed an explanation: applying the information transfer axiom once results in the belief ascription

ATT(y, BEL, [$z_{A_1}\ z_{A_2}$ \| woman(z_{A_1}), transparentliquid(z_{A_2}), drink(z_{A_1},z_{A_2})])

But then we also have a (weak) explanation for champagne(z_{A_2}): this formula is in the default extension of transparentliquid(z_A). The following belief ascription is thus also explained.

$$\text{ATT}(y, \text{BEL}, \boxed{\begin{array}{|c|}\hline z_{A_1}\ z_{A_2} \\ \hline \text{woman}(z_{A_1}) \\ \text{transparentliquid}(z_{A_2}) \\ \text{D: champagne}(z_{A_2}) \\ \text{drink}(z_{A_1},z_{A_2}) \\ \hline \end{array}})$$

This is already the explanandum. The assumption champagne(z_{A_2}) has been marked as defeasible because according to the explanation, this speaker belief results from a defeault inference. If the hearer considered alternative explanations such as *the speaker saw the woman pour the liquid in her glass from a bottle of champagne* or even more convincingly *the speaker has tasted the liquid in the woman's glass*, more plausible, then the defeasibility label would be absent.

Update with the abductive explanation

The explanatory source assumptions are added to the hearer belief and marked as defeasible. The first part of the DRS has again been abbreviated.

$$\boxed{\begin{array}{|c|}\hline y\ z_1\ z_2\ i \\ \hline \text{K}^{hearerbelief} \\ \text{ATT}(y, \text{BEL}, \boxed{\begin{array}{|c|}\hline z_{A_1}\ z_{A_2} \\ \hline \text{Pre: woman}(z_{A_1}) \\ \text{transparentliquid}(z_{A_2}) \\ \text{D,Pre: champagne}(z_{A_2}) \\ \text{Pre: drink}(z_{A_1},z_{A_2}) \\ \text{Ass: happy}(z_{A_1}) \\ \hline \end{array}}) \\ \text{D: source}(s_2,z_{A_1},z_1^{s_2}) \\ \text{D: source}(s_2,z_{A_2},z_2^{s_2}) \\ \hline \end{array}}$$

In this case, the hearer did not have to revise his own beliefs in order to explain the belief manifested by the speaker.

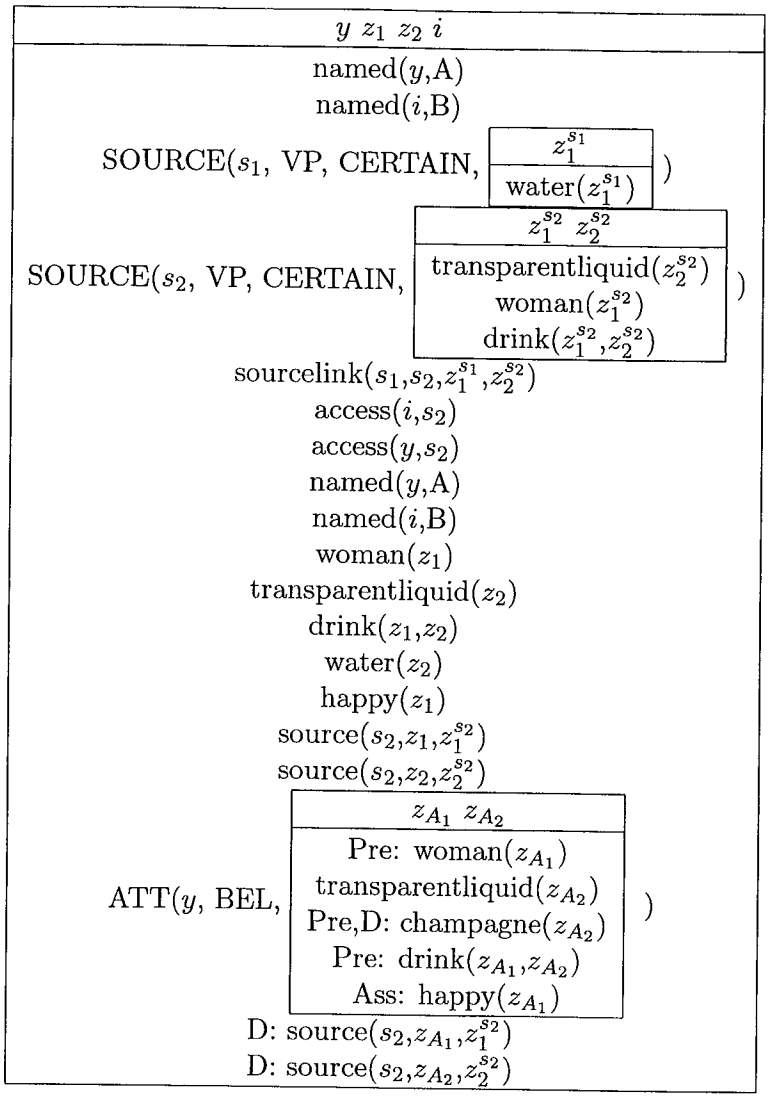
Hearer contextualization

In order to be able to integrate (parts of) it into his own context, the hearer has to adapt it to his own discourse referent naming conventions. The hearer contextualization of the belief ascribed to the speaker with respect to the hearer's belief context is the result of substituting z_{A_1} and z_{A_2} by the corresponding (according to the two source assumptions) referents z_1 and z_2:

$$\boxed{\begin{array}{|c|}\hline z_1\ z_2 \\ \hline \text{Pre: woman}(z_1) \\ \text{D, Pre: champagne}(z_2) \\ \text{Pre: drink}(z_1,z_2) \\ \text{Ass: happy}(z_1) \\ \hline \end{array}}$$

Hearer belief update

Finally, the hearer has to resolve the question which parts of the hearer contextualization he should accept as his own beliefs. **Consolidating** the merge of the previous hearer belief and the hearer contextualization results in the following DRS.



Now the context-dependent proposition happy(z_1) has been communicated to the hearer. Note that the condition of z_{A_2} being champagne has not been transferred. The communicated content has been made independent of the way in which the referent was described, and the effect of communicating a singular proposition has been achieved.

3 Conclusion

Abductive pragmatics of natural language can be fruitfully applied to referential definite descriptions. It plays a double role: firstly, abductive presupposition justification with respect to the belief ascribed to the speaker paves the way, then abductive belief explanation accounts for the introduction of source assumptions which are the basis for extracting the communicated context-dependent proposition.

References

Donnellan, K.: 1966, Reference and definite descriptions, *Philosophical Review 75* pp. 281–304.

Krause, P.: 2001, *Topics in Presupposition Theory*, PhD thesis, Universität Stuttgart.

Kripke, S.: 1979, Speaker's reference and semantic reference, *in* P. French (ed.), *Contemporary Perspectives in the Philosophy of Language*, University of Minnesota Press, Minneapolis, pp. 6–27.

From Dialogue Acts to Dialogue Act Offers: Building Discourse Structure as an Argumentative Process

Jörn Kreutel
SemanticEdge
Berlin
joern.kreutel@semanticedge.com

Colin Matheson
Language Technology Group
University of Edinburgh
colin.matheson@ed.ac.uk

Abstract

Based on the Theory of Communicative Action developed by Jürgen Habermas we propose to analyse the actions of participants in a dialogue in terms of DIALOGUE ACT OFFERS that raise VALIDITY CLAIMS to be evaluated by the addressee of a move. We argue that this approach allows us to model disputes and misunderstandings in conversations in a way that minimises revisions of discourse structure and reconstructs the process of reaching agreement in a more natural way than current formal dialogue theories.

1 Introduction

In the 'Theory of Communicative Action' (Habermas, 1981) Jürgen Habermas provides an account of speech acts that both elaborates and revises John Searle's proposals in Searle (1969). On the one hand, Habermas strengthens the intersubjectivist view of Searle in emphasising that the illocutionary force of an utterance can primarily be determined in a conventionalised way on the basis of what has been said in a given discourse context and does not require reasoning over the intentions of a particular speaker.[1] On the other hand, he introduces the concept of VALIDITY CLAIMS as an alternative to Searle's notion of acceptability conditions for speech acts. Habermas argues that acceptability conditions alone lack an intersubjective, 'pragmatic', dimension and therefore are not sufficient to reconstruct the orientation towards mutual agreement as the primary and fundamental purpose of language use. A successful interaction essentially requires that the participants implicitly or explicitly accept or reject each other's contributions by making judgements about whether the conditions for their successful performance are satisfied or violated. What drives the dialogue forward is exactly this process of a speaker raising the satisfaction of acceptability conditions as an issue pending evaluation by the addressee, and it is the notion of validity claims that incorporates this analysis in the definition of a speech act itself: performance of a speech act means raising validity claims that are 'open to criticism and designed for intersubjective recognition' (see Habermas (1991)).

Habermas assumes that each speech act raises the four universal claims of UNDERSTANDABILITY, propositional TRUTH, normative RIGHTNESS and expressive TRUTHFULNESS,[2] and proposes to classify speech acts depending on which is the major claim. The main debates about his assumptions have focussed on this typology, the universality claim (see for example Goldkuhl (2000)), and on the question of whether intention-based approaches to rational interaction or Habermas' radically intersubjectivist theory prove to be more powerful and explanatorily adequate (Skjei (1985), Harnad (1991), Dietz and Widdershoven (1991)).

In this paper we want to focus on one aspect of Habermas' analysis that tends to be neglected as it can be considered a minor issue or a matter of linguistic expression.[3] Throughout his work, Habermas tends to use the term SPEECH ACT OFFER ('Sprechaktangebot') rather than speech act, thus underlining the fact that the contributions of dialogue participants (DPs) should be seen as 'offers' which may then be accepted, (partially) rejected, replied to with a counteroffer, and so on. Dialogue thus is analysed as a genuinely argumentative process, where argumentation may take place about the propositional content of an utterance as well as about the illocutionary act it is intended to perform. We think that this view of dialogue can be very useful in resolving some problems which current formal dialogue theories have in analysing all kinds of disputes, or cases of misunderstanding. We therefore propose to analyse dialogue structure, both at the level of common ground (see Clark and Wilkes-Gibbs (1986), Poesio and Traum (1998)) *and* at the level of the individual participants' point of view, as something that emerges out of the argumentative process in which DPs *a priori* engage. According to this

[1] From this it follows that Habermas must consider indirect speech acts as non-standard uses of language.

[2] We do not address the claim for understandability (of linguistic content) in this paper. The other three claims are discussed in more detail in section 3.1 below.

[3] See, for example, the argument against Habermas' use of validity claims in Skjei (1985).

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 85-92.

view, the interpretations of the DPs' contributions will always be seen as preliminary, as 'speech act offers', unless the respective addressees have made their evaluation evident.

We first discuss the problems existing theories have in analysing conflictive situations in dialogue. We then outline the basics of what a formalisation of the interpretation process in our model would look like, and illustrate our proposals with an example analysis.

2 Disagreements in Dialogue

In this section we will address some of the problems that current formal dialogue theories face in dealing with situations where there is a mismatch of the DPs' opinions with respect to the content of their actions. In particular, we will refer to the way these cases are dealt with in Segmented Discourse Representation Theory (SDRT), developed by Asher and Lascarides.[4] SDRT is a framework for analysing discourse that has proven very successful in the analysis of the interdependencies between linguistic content and discourse structure, for which the theory employs a modularised approach which keeps reasoning over information content logically distinct from reasoning over discourse structure. However, in the change of emphasis from monologue to dialogue (Asher and Lascarides (1998), Asher and Lascarides (2001)), the theory exhibits some shortcomings which, we argue, are manifested in the way actions which are characterised by an explicit or implicit disagreement between the DPs' points of view are dealt with. We will attempt to illustrate how these problems result from the theory's conception of the interpreter's position and from its construction principles for discourse structure.

2.1 Modelling the Interpreter

In their 1998 paper on questions in dialogue, Asher and Lascarides analyse examples such as (1) below, in which a response to a question does not immediately result in a resolution of the latter (Asher and Lascarides, 1998):

(1) A[1]: How can I find the treasure?.
B[2]: It's at the secret valley.
A[3]: But I don't know how to get there.

Here, the RHETORICAL RELATION between A's question A[1] and B's answer B[2] is immediately interpreted as `nei` ('not enough information') by the SDRT interpretation rules. As discourse interpretation at this step not only has access to the overt content of what has been said in the given discourse context, which includes the manifestation of the DPs' mental states, but also to the mental states of the DPs themselves, this interpretation is plausible if one assumes that A does not know where the secret valley is located. However, we argue that this interpretation strategy not only neglects B's assumptions about B[2] for the given dialogue step, but also if we generalise it for the whole dialogue it corresponds to a conception of the interpreter as an omniscient observer. This approach is not problematic if one sees the purpose of a dialogue theory as providing a consistent interpretation of the DPs' actions in terms of truth conditional semantics. However, if in addition the theory is meant to reconstruct the point of view of an actual third party observer (not to mention the DPs' views), interpretation rules such as those proposed by Asher and Lascarides seem to be inappropriate.

As far as alternative analyses of examples like (1) are concerned, Kreutel and Matheson, who use INFORMATION STATES[5] as developed on the TRINDI[6] project, would interpret B's move B[2] as `answer`, irrespective of A's belief state, and would keep this interpretation in the common ground as the dialogue proceeds, even after A's actual response in A[3]. This account corresponds to a rather lax notion of 'answerhood', which is merely seen as 'a DP's attempt to provide a felicitous answer'. It therefore cannot provide a fine-grained semantic analysis of answers in dialogue, which is a clear strength of the SDRT model.

2.2 Building Discourse Structure

In recent, as yet unpublished, proposals for analysing disputes in dialogue (Asher, 2002), Asher and Lascarides have partially revised their ideas about the interpreter's position, assimilating it to that of a third party observer such as the reader of a textual representation of a dialogue. However, with respect to the way discourse structure is generated and maintained, the theory still exhibits an arguable lack of naturalness which is mainly due to the fact that assignment of rhetorical relations takes place immediately on the level of what is effectively the DPs' common ground. If the DPs' actions exhibit a conflict between their private beliefs, as in discussion scenarios, interpretation thus has to revise the common ground itself and the discourse structure represented therein. We can illustrate this problem with an adaption of the kind of example used by Asher and Lascarides:

(2) A[1]: John went to jail.
A[2]: He was caught embezzling funds.
B[3]: No.
B[4]: He went to jail because he was convicted of tax evasion.

[4] See Lascarides and Asher (1993). Further references can be found at `http://www.cogsci.ed.ac.uk/alex/papers.html`

[5] See Matheson et al. (2000) and Kreutel and Matheson (2000) for the background to this approach

[6] Telematics Applications Programme, Language Engineering Project LE4-8314.

The main point to be emphasised here is that only the rhetorical relation connecting A[1] and A[2], which Asher and Lascarides interpret as `explanation`, is disputed by B[3] and B[4]; the contents of the assertions themselves are accepted. However, as the interpretation of the rhetorical relation is added to the common ground immediately after A[2] is processed, it is necessary to change the representation of the dialogue structure as the analysis progresses. Specifically, it is necessary to replace `explanation` with `Dis(explanation)`, where the `Dis` predicate indicates that the relation is disputed. In this case, then, the process of building dialogue representations is non-monotonic.

We feel that the truth-conditional complexities which arise in such cases can be avoided if one assumes that after the performance of A[2] the content of the move, including the rhetorical relation between A[1] and A[2], has to be acknowledged, either explicitly or implicitly, before it can be considered part of the common ground. Seeing the actions of DPs thus as principally pending for acknowledgement, interpretation of the whole dialogue can proceed monotonically even if no or only partial acknowledgement takes place, as in (2).

The following section outlines such an alternative approach to the examples discussed above using Habermas' notion of speech act offer to analyse how the interpretations of speech acts, including rhetorical relations,[7] are managed by the DPs during an interaction. Our theory can be formulated in a way that is compatible with most of SDRT's assumptions about how speech act assignment and interpretation of linguistic content takes place.

3 The Interpretation Process

In this section we sketch an interpretation process based on Habermas' notion of speech act offer. Our objective is to model the process of the introduction and evaluation of speech act offers independently from other aspects of interpretation, such as speech act assignment and interpretation of linguistic content. In thus keeping the interpretation algorithm as modularised as possible, our proposals are compatible with both SDRT and dialogue models which use a TRINDI-style approach based on the notion of information state.

The basic notions we employ adhere to the proposals in Poesio and Traum (1998) and decompose speech acts into DIALOGUE ACTS of different types. We use CORE SPEECH ACTS such as `ask` or `assert`, on the one hand, and ARGUMENTATION ACTS like `answer`, `request-evidence`, and `correct`, on the other. Argumentation acts can be seen as expressing the context dependent aspects of a core speech act, and SDRT's rhetorical relations will be dealt with as argumentation acts. Following Habermas' terminology we will use the term DIALOGUE ACT OFFER (DAO) rather than 'dialogue act' in the remainder of this paper.

[7] See Asher and Lascarides (2001) for a discussion of the interpretion of rhetorical relations as speech acts.

3.1 Validity Claims

Our primary goal in this paper is to analyse the way in which DAOs are managed in the course of a conversation, and we therefore do not give an exhaustive description of how the validity claims raised by a DAO can be represented. Briefly, we propose to identify the content of the truth claim with the propositional content π of the DAO and its presuppositions. We further think one can determine the contents of the claims of rightness and truthfulness on the basis of π and the type of DAO that has been performed. As for truthfulness, we argue that its content can be seen as the intention that is conventionally associated with a DAO of a certain type, for example the intention that some content π become *shared-belief* and be *resolved* with `assert` and `ask` acts, respectively. From intentions that express interactive goals of this kind, we can then infer the SINCERITY CONDITIONS of dialogue acts in Searle's terms (Searle, 1969), i.e. that the speaker believes π and wants to know π, respectively.

As for the claim of rightness, there is an ongoing discussion of how to generalise Habermas' standard example for rightness claims as given in his analysis of requests (see Goldkuhl (2000)): When requesting π, speakers claim that there is a social norm that entitles them to make the request, hence the rightness claim is related to the acceptability of an action in a certain social context. Given this generalisation, the rightness claim for assertions and questions could, first of all, be modelled in terms of conditions that require that the respective contribution be informative. However, in addition to this, the rightness claim will also comprise the appropriateness of the topic that has been raised in the given social context, so in the following example, B's response can be seen as a rejection of the rightness claim of A's question:

(3) A[1]: Is it true that John went to jail?
B[2]: I don't want to talk about this now.

In his discussion of Habermas' universality assumption for the validity claims, Goldkuhl remarks that the rightness claim has become a sort of 'residual category' in academic discussion which covers all aspects of social interaction (Goldkuhl, 2000). As it is evident that social interaction cannot fully be reduced to issues of concordance to norms, the question is whether one should extend the repertoire of validity claims as proposed by Goldkuhl or rather relax the definition of the rightness claim as adherence to actual social norms. Given this unresolvedness,

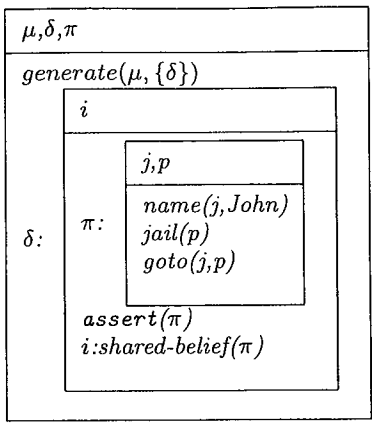


Figure 1: Interpretation of *John went to jail.*

we will currently not provide an explicit representation of the contents of the rightness claim given a DAO of some type, and only represent the claims of truth and truthfulness, on which there is relative agreement in the research community.

Depending on the theoretical framework, the assignment of validity claims to some DAO can be expressed by the discourse interpretation algorithm either in terms of SDRT-style axioms for dialogue act types or alternatively as update rules in the TRINDI framework similar to the ones in Kreutel and Matheson (to appear). In Figure 1 we have sketched an example of what the full analysis of an assertion in a zero context would be after interpretation rules have applied, where μ, δ and π denote the move that has been made by A, the dialogue act offer made in μ and the propositional content of μ, respectively. i denotes the intention conventionally associated with an assertion of content π, which we have identified as expressing the truthfulness claim. Following proposals by Poesio and Traum (Poesio and Traum, 1997) we further assume a *generate* relation between a move that is made by a DP and the set of dialogue act offers that is performed therewith.[8]

Making a dialogue act offer thus is interpreted as first of all producing some propositional content with some illocutionary force that may be associated with an intention related to the former. As for the formal semantic interpretation of DRSs such as that in Figure 1, we propose to interpret the DRS K_δ that represents the content of a DAO δ as being in the scope of some modal operator. In a first approximation this operator could be seen as expressing the beliefs of the relevant speaker, however this would mean modelling agents in dialogue as genuinely sincere. Our model could not then account for a strategic agent who manipulatively *exploits* the way validity claims and their acceptance influence the intersubjectively shared assumptions of DPs.

Rather than associating DAOs with beliefs that apply to the content of some action itself, we think a DAO should be seen as expressing an agent's assumptions about the impact the presentation of this content may have on the addressee of the action. This way, it will be possible to model agents who act insincerely, for instance by asserting something which they do not believe to be true, but which for strategic reasons they wish to convince their addressees of. We therefore assume a more complex modality for DAOs, which can, preliminarily, be described as in (4) (note that the formalisations in this paper are relatively shallow and do not cover the temporal semantics which would be needed for a complete model of the processes described here):

(4) If a dialogue act offer δ is made by some agent A towards an addressee B, then $\mathbf{Bel}_A(\mathtt{accept}(B,\delta))$.

The performance of a DAO with some content K_δ is primarily seen as expressing an agent's belief that K_δ may become part of the common ground of the DPs via the acceptance of the DAO by the addressee. This semantics leaves open whether the agent actually believes K_δ itself or not, and can thus account for both cooperative and strategic behaviour of discourse participants.

However, in order to have the semantics reflect Habermas' critique of the expressive power of Searle's acceptability conditions it is necessary to extend this analysis and to augment the modal operator described in (4) with an interactive dimension, which is not provided by the simple association of DAOs with the beliefs of a speaker about a future action of an addressee. Only this way can we account for the fact that a *claim* is made by the speaker and directed towards an addressee. An intuitive way to model this is to assume that a dialogue act offer introduces a DISCOURSE OBLIGATION[9] on the addressee to take a position with respect to it. The intersubjective nature of DAOs is thus accounted for by having DAOs enforce a reaction on the part of the addressee, be it an acceptance or a rejection. In the definition below, the action, which an addressee is obliged to perform, is classified as `address`, where addressing can be realised explicitly or tacitly:

(5) **Semantics of Dialogue Act Offers**
An agent A makes a dialogue act offer δ towards an addressee B iff

1. $\mathbf{Bel}_A(\mathtt{accept}(B,\delta))$ and
2. $\mathbf{obliged}_B(\mathtt{address}(B,\delta))$.

[8] Poesio and Traum assume *generate* as holding between locutionary acts, i.e. a speaker uttering something, and the performance of dialogue acts, which in their framework are analysed as CONVERSATIONAL EVENTS.

[9] See Traum and Allen (1994) for the basis of the use of discourse obligations in dialogue modelling.

Note that the obligation to address a dialogue act offer that is introduced by the performance of a DAO is distinct from the obligation that may be introduced once the addressee has accepted the offer, for example the obligation to answer a question one has been asked.[10] Whereas the latter obligations are specific to the type of DAO, the obligation to address a DAO is generic and does not depend on the type of dialogue act that was meant to be performed.[11] As the following section will outline, it is the fact that an addressee acts according to these generic obligations that may result in the DPs' shared assumption that a dialogue act of some type actually *has* been performed.

3.2 Evaluating Dialogue Act Offers

As our discussion in the previous section has shown, we think it is a rather natural and intuitive approach to conceive of the evaluation of dialogue act offers by the addressee of a move as a process of GROUNDING. In this respect, our analysis can be seen as an elaboration of some proposals made by members of the TRINDI consortium, in particular the work of Poesio and Traum (Poesio and Traum (1997), Matheson et al. (2000)).

In contrast to earlier proposals in Traum (1994), Poesio and Traum (1997) assume that the domain of the grounding process is a DISCOURSE UNIT that comprises the full interpretation of a dialogue move. Our model adds a further level of granularity to this analysis and proposes to have grounding apply to the single DAOs into which a move can be broken down, and to determine for each DAO whether all of its content can be grounded or whether it contains information that must be exempt from the grounding process and be made subject to argumentation.

Our analysis is thus closer, on the one hand, to the ideas in Traum (1994), which describe grounding processes on the basis of an abstract COMMUNICATION RECIPE. Here, dialogue is analysed on the conceptual level as a number of acts of presenting and acknowledging content – in a generic sense – at a granularity finer than the actions of discourse participants. On the other hand, our model retains the advantages of the highly structured account of the overt realisation of dialogue as provided by a theory of discourse which is based on the notions of dialogue acts and validity claims.

[10] See Kreutel and Matheson (2000) for how these obligations can be employed for modelling the actions of DPs.

[11] In a similar fashion, Poesio and Traum (1997) assume that all core speech acts introduce an obligation on the addressee to perform an understanding act expressing acknowledgement. In addition, they propose the introduction of an address obligation for some classes of core speech acts. In the context of Habermas' theory of validity claims, this analysis can be extended to all classes of core speech acts and argumentation acts.

Consider again as an example the dialogue below, introduced in (1) above:

(6) A[1]: How can I find the treasure?.
B[2]: It's at the secret valley.
A[3]: But I don't know how to get there.

We assume that B's answer B[2] to A's question A[1] introduces two DAOs of type assert (δ_1) and answer (δ_2). A's response A[3] can then be interpreted as fully grounding δ_1, in particular its propositional content π, but only as partially grounding δ_2. While A does not reject the truthfulness of B's answer, and thus grounds B's intention that A's question be resolved,[12] A does reject the claim that in asserting π B provides an answer that actually *does* resolve the question. Depending on what types of dialogue acts one assumes, this rejection could be seen as A interpreting the relation between A[1] and B[2] as nei (as in Asher and Lascarides (1998)), thus introducing a counteroffer δ_2' to B's claim in δ_2 that the relation is actually answer. Now, B has basically two options: one is to revise the belief that A knows a way to the secret valley, accept the rejection, and come up with an alternative answer. However, if there is reason for B to assume that A is not sincere in A[3], B may reject this claim and provide arguments for the view that B[2] provides an acceptable answer to A[1].

Our assumptions about the introduction and evaluation of the DAOs that underlie this analysis can be summarised in the following way:

(7) **Discourse Interpretation Algorithm**

1. For each move μ determine the set Δ_μ of dialogue act offers generated by μ applying common interpretation rules such as those in SDRT or in TRINDI models.
   (a) If μ performs a mere acceptance or correction δ and if there is a subsequent move $\mu+1$ by the speaker of μ then $\Delta_\mu = \{\delta\} \cup \Delta_{\mu+1}$.
2. If there is a move ν by the addressee of μ that precedes μ and if Δ_ν contains ungrounded dialogue act offers, then
   (a) Check for each ungrounded $\delta_i \in \Delta_\nu$ whether μ grounds δ_i or whether there is some $\delta_j \in \Delta_\mu$ that exempts parts of the content of δ_i from being grounded and/or provides a counteroffer δ_i' to δ_i.
   (b) If a counteroffer is provided, add δ_i' to Δ_μ, otherwise merge δ_i with the common ground.

[12] We think that this intention could be assigned – and grounded – also in an interpretation that analysed B's response as a hint rather than as an answer. On the other hand, if B was considered to be insincere, just trying to tease A, the intention would be assigned to the DAO but then A's response could be interpreted as a rejection of the truthfulness claim.

In general the merge of a dialogue offer δ with the common ground mentioned in the algorithm can – in a DRT-based formalisation – be thought of as joining the sets of discourse referents and conditions of the DRS K_δ that contains the content of δ with the corresponding sets of that DRS that contains the grounded information.[13] In the particular framework of SDRT the site where the merge takes place will be determined by the way dialogue act offers attach to the previous dialogue context. Assuming a model that integrates our notion of dialogue act offers with Lascarides and Asher's algorithms for discourse interpretation, we propose to merge the content of a DAO with the content of that SDRS that contains the condition $\delta : K_\delta$.

3.3 Example Analysis

We will now give a more elaborate analysis of example (2), repeated here as (8):

(8) A[1]: John went to jail.
A[2]: He was caught embezzling funds.
B[3]: No.
B[4]: He went to jail because he was convicted of tax evasion.

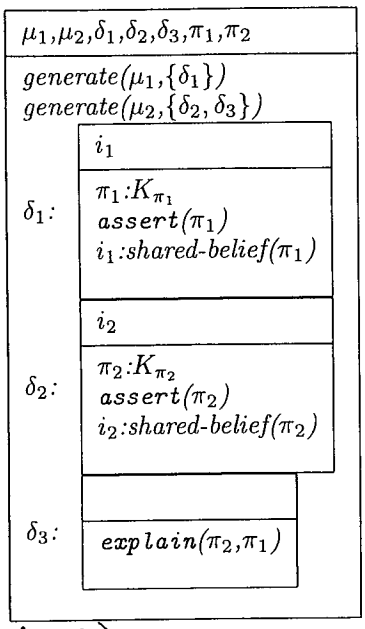


Figure 2: Dialogue State after A[2]

Figure 2 shows the result of applying step 1 of the algorithm to A's moves A[1] and A[2], which are represented by the discourse referents μ_1 and μ_2. We see that three DAOs, assigned the types `assert` (δ_1, δ_2) and `explain` (δ_3), are introduced. Note that whereas δ_1 and δ_2 comprise propositional content (π_1,π_2), δ_3, being an argumentation act, simply adds illocutionary force to μ_2 claiming that π_2 actually provides an explanation for π_1.[14] Our interpretation currently does not comprise an intention associated with δ_3. However, for an argumentation act like `explain`, we can extend interpretation rules of the kind described in Kreutel and Matheson (to appear), which conventionally associate intentions with dialogue acts in the following way:

(9) $\texttt{explain}_{DP}(\pi_2, \pi_1)$
$\vdash i_{DP} : \textit{shared-belief}(\textit{explanation}(\pi_1, \pi_2))$

According to this rule, if some DP makes a DAO of type `explain`, we can infer that they have the intention that the fact that *explanation* holds between two propositional contents π_1 and π_2 become shared belief.

Apart from the mere fact that two moves have been made, raising DAOs δ_1-δ_3 with contents π_1 and π_2, Figure 2 shows that all the actual content generated by μ_1 and μ_2 is embedded under the modality associated with DAOs. Step 2 of the algorithm does not apply here as we are at a discourse initial state and there is no preceding move by B. The result of grounding parts of A's contribution through B's performance of B[3] and B[4], represented by μ_3 and μ_4, is then shown in Figure 3. Here, we see, first of all, that in addition to issuing a DAO δ_4 typed as `correct`, two further `assert` DAOs are introduced (δ_5 and δ_6). Furthermore, δ_7 claims that `explain` holds between the contents of the assertions.

As B's *No* is a simple correction which is followed by another move by B, step 1a of the algorithm applies in this case, and μ_1 and μ_2 are evaluated against $\{\delta_4\} \cup \{\delta_5, \delta_6, \delta_7\}$ according to step 2. Note further that we interpret `correct` acts – as well as `accept` acts – as anaphoric, applying to some set of DAOs Δ, the reference of which is not resolved before the application of step 2.

Considering the propositional content of δ_5 and δ_6, we then can assume that μ_3 and μ_4, first of all, fully ground δ_1, the content of which is echoed in δ_5.[15] In addition, the proposed interpretation assumes that δ_2 is accepted as well, where the DRS shows the result of merging the contents of δ_1 and δ_2 with the

[13] see Poesio and Traum (1997) for a formalisation of the merge operation in terms of Compositional DRT.

[14] One could, nevertheless, associate a propositional content with an argumentation act such as explain(π_2,π_1), assuming that it expresses the fact that *explanation*(π_1,π_2) holds. Note that in a two-level approach to discourse interpretation, *explanation* will be a predicate in the language of information content whereas explain will belong to the language used for reasoning over discourse structure. As noted, Asher and Lascarides do not distinguish the level of the actions of a DP from the level of rhetorical relations, whereas our model can derive the latter in a secondary step as information content given the interpretation that some act has been performed, which, on its part, will be achieved by applying the logics of dicourse structure.

[15] δ_5 and δ_6 are not fully represented in the DRS. The $\rightsquigarrow$ symbol should be read as introduction of the indicated discourse referents.

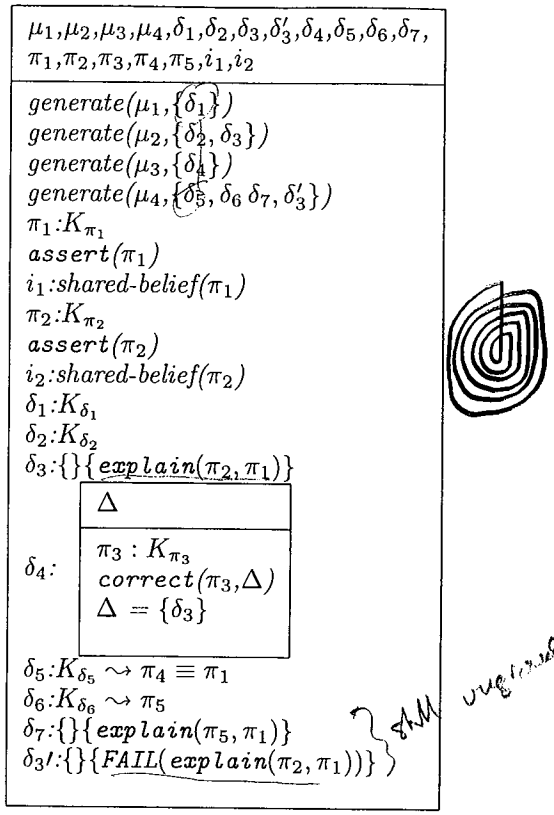


Figure 3: Dialogue State after B[4]

root DRS. Given the correct and explain DAOs δ_4 and δ_7, we can further assume that what is rejected by B is actually δ_3, which allows us to resolve the anaphor Δ to $\{\delta_3\}$.[16] In addition, we propose to have δ_4-δ_7 introduce a counteroffer δ_3' to δ_3, which claims that π_2 does not actually provide an explanation for π_1.

As Figure 3 demonstrates, our analysis is very close to that in SDRT as far as the successfully grounded information is concerned.[17] However, in contrast to the SDRT analysis of (8), our interpretation builds discourse structure monotonically; in particular it maintains the information in δ_3 while at the same time expressing the fact that it is disputed by the presence of the counteroffer δ_3'.[18] As in the previous example (6), given the state represented in Figure 3, A now has the options of either acknowledging B's claims, including δ_3', or starting a discussion to provide arguments in support of δ_3.

Given these example analyses, we think that an algorithm of the kind described in (7) is powerful enough to account not only for an intuitive analysis of cases like (6) and (8), but also for situations such as the one exemplified in (10) below, in which an utterance is misinterpreted by the addressee:

(10) A[1]: Where's the water tap?
B[2]: You shouldn't drink the water here. There's mineral water in the fridge.
A[3]: I actually just wanted to water the plants.
B[4]: Oh, I see. It's downstairs in the cellar.
A[5]: Thanks.

Here, B's response B[2] grounds the overt content of A[1] and, in addition, introduces a DAO containing B's assumptions about A's motives underlying A[1]. In the same way as in the examples discussed above, A[3] then provides a counteroffer to this interpretation, which is finally grounded and responded to in B[4].

4 Conclusion

In this paper we propose an approach to dialogue modelling that is based on Habermas' notion of speech act offer, incorporated in our theory as dialogue act offer. Leaving aside how all the validity claims raised by a DAO will actually be represented for the different types of DAOs, we have concentrated our argument on the general issue of how a discourse interpretation algorithm can be formulated on top of these basic assumptions.

As the discussion of the example dialogues has made clear, an algorithm of the kind proposed in (7) is powerful enough to analyse various situations in which DPs have contrary assumptions about the interaction in a way that does not require any revision of the discourse structure that was built before the conflict became evident. On the other hand, it is able to reconstruct the third party perspective of an interpreter who does not have to be seen as omniscient, but rather as an interpreter for whom the representation of the dialogue emerges out of the *observable* actions of the discourse participants and the interpretation rules that are conventionally associated with these actions.

References

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

[16] An alternative interpretation of μ_3 and μ_4 would be to assume that only δ_1 is grounded and to resolve Δ to $\{\delta_2,\delta_3\}$.

[17] In particular with respect to the example dialogue (8) in Asher (2002). What is not represented here, however, is the fact that μ_1 and μ_2 are grouped together and trigger a package of information which some rhetorical relation could refer to as a whole.

[18] Whereas Lascarides and Asher's Dis operator marks some rhetorical relation as being under discussion, the application of the FAIL operator to some dialogue act α indicates that the speaker considers the performance of the act as not felicitous. The fact that some dialogue act offer of action type α is under discussion is thus reflected by having the dialogue state comprise both α and FAIL(α) embedded under the belief operator for DAOs.

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

Nicholas Asher. 2002. An analysis of corrections. Presentation to ICCS/HCRC seminar series, University of Edinburgh, April.

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

Jan L. G. Dietz and G. A. M. Widdershoven. 1991. Speech acts or communicative action? In L. Bannon, M. Robinson, and K. Schmidt, editors, *Second European Conference on Computer-Supported Cooperative Work.*

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.

Jürgen Habermas. 1991. Comments on John Searle: 'Meaning, Communication and Representation'. In *John Searle and his Critics.* Cambridge/MA.

Stevan Harnad. 1991. Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1:43–54.

Jörn Kreutel and Colin Matheson. 2000. Obligations, intentions, and the notion of conversational games. In *Götalog 2000, the 4th Workshop on the Semantics and Pragmatics of Dialogue.* University of Gothenburg.

Jörn Kreutel and Colin Matheson. to appear. Incremental information state updates in an obligation-driven dialogue model. *Language & Computation.*

Alex Lascarides and Nicholas Asher. 1993. Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16(5).

Colin Matheson, Massimo Poesio, and David Traum. 2000. Modelling grounding and discourse obligations using update rules. In *NAACL.*

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: An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge.

Erling Skjei. 1985. A comment on performative, subject and proposition in habermas's theory of communication. *Inquiry*, 28.

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 R. Traum. 1994. *A computational theory of grounding in natural language conversation.* Ph.D. thesis, Computer Science, University of Rochester, New York, December.

Enhancing collaboration with conditional responses in information-seeking dialogues*

Ivana Kruijff-Korbayová and **Elena Karagjosova**
University of the Saarland, Saarbrücken, Germany
{korbay,elka}@coli.uni-sb.de

Staffan Larsson
Göteborg University, Sweden
sl@ling.gu.se

Abstract

We aim at improving the collaborative character of the responses in the GoDiS dialogue system participating in information-seeking dialogues. In this paper we describe how the system currently deals with successful and unsuccessful database search and show the need for generating more collaborative responses. We consider conditional yes/no responses as one kind of collaborative responses and discuss issues related to their implementation.

1 Introduction

Our goal is to improve the collaborative character of the responses of a dialogue system participating in information-seeking dialogues (ISDs). We employ the information-state approach to dialogue developed in the TRINDI and SIRIDUS projects (Cooper et al., 1999; Lewin et al., 2000). Within this framework, we propose an implementation of collaborative conditional responses as answers to yes/no-questions. Such responses are illustrated in (2).

(1) Not if you want to fly economy class.

(2) Yes, if you can fly business class.

In (1), the negative response is contingent on flying economy class, whereas in (2) the positive response is contingent on flying business class.

It has been argued that task-oriented dialogues are natural and efficient when they are collaborative (Chu-Carroll and Brown, 1997; Rich et al., 2000). Collaboration means that both the system and the user are contributing to solving the task at hand. ISDs are a particular kind of task-oriented dialogues. For instance in the travel domain, the task is to determine a set of parameters of a possible journey with respect to a database to which only the system has access. However, a typical ISD system does not enable any collaboration in determining the journey parameters: The system does not provide the user with any indications of what journeys are (still) available in the database. The user has to "blindly" specify her desires, and equally blindly revise them, refining (if they are under-constraining) or relaxing them (if they are over-constraining).

In the travel domain, it is not a viable option for the system to guide the user in the initial specification of journey parameters. It is impossible to enumerate all available options for individual parameters since the number of potential journeys in the database is typically large. On the other hand, once the search space is restricted by setting some initial set of parameters, the system could be collaborative in the subsequent phases. Our proposal mainly concerns collaboration by providing responses that help the user to revise or refine the initial parameters.

We show that one useful way of accomplishing collaboration in the parameter revision and refinement phase is by providing a conditional response (CR): a positive or negative response clarifying the condition(s) under which this response holds. It is the making of this condition explicit that makes the response collaborative. For example, a negative CR can be collaborative by mentioning such parameter(s) in the condition, whose relaxation could result in a positive response instead (*economy class* in (1)). A positive CR can be collaborative by mentioning such parameter(s) that are necessary for preserving the positive response (*business class* in (2)).

The paper is structured as follows. In Section 2 we describe the dialogues GoDiS can participate in, and show the need for collaborative responses after both successful and unsuccessful database search. Section 3 discusses the meaning of CRs also taking into account the context in which they occur. Section 4 concerns the implementation. We briefly describe the information-state update approach in

* Work on this paper was supported by SIRIDUS (Specification, Interaction and Reconfiguration in Dialogue Understanding Systems), EC Project IST-1999-10516.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 93-100.

GoDiS and then discuss the implementation of CRs in terms of the conditions under which they can be selected and the effects they have on the information state. We also describe the underlying processing of the database search results. In Section 5 we review some related work on generating cooperative responses in over-constrained situations. Finally, in Section 6 we indicate some issues for future work.

2 Information-seeking dialogue in GoDiS

Our work is set in GoDiS (Gothenburg Dialogue System). We use the information state (IS) approach to dialogue (Larsson and Traum, 2000; Larsson, 2002a) (cf. Section 4.1 for details).

GoDiS has been initially designed to handle information-seeking dialogues in the travel domain, such as (3). It collects from the user a set of parameters constraining a journey, and performs a database search with these constraints. If the search succeeds, GoDiS returns the price of the journey (3e). Otherwise, it indicates the search has failed (3f).

(3) a. S: Welcome to the travel agency!
b. U: The price for a flight from Malmö to Paris on the first of April please.
c. S: What class did you have in mind?
d. U: Economy.
e. S: The price is 7654 crowns.
f. S': Sorry, there is nothing matching your request about price.

Normally the system returns the value of one parameter, e.g., *price*. But when the user does not specify all the search parameters expected by the system, as in (4c), multiple answers and the corresponding parameter values are returned, as in (4d).

(4) a. U: I want to fly from Malmö to Paris on the first of April.
b. S: What class did you have in mind?
c. U: It doesn't matter.
d. S: The price is 7654 crowns. Business class. The price is 456 crowns. Economy class.

The result of a database search is currently one of the following: (i) the proposition *fail(q)* indicating that no answer to the question was found (realized as in (3f)), (ii) a proposition specifying a unique value of a requested parameter (realized as in (3e)), (iii) a set of alternative answers (realized as in (4d)).

However, a dialogue like (3) often does not stop at the result of the database search, be it negative or positive. The user may revise or refine some parameter(s) and initiate a new search. For example, after (3f), the user may continue by changing the departure day (5a). Then, after (5b), the user may continue by trying to further constrain the search by specifying an additional parameter, for example the airline as in (5c). However, such continuations can become dull as the user is trying to find out which combinations of parameters succeed, as in (5c)-(5f).

(5) a. U: Can I fly on the second?
b. S: Yes.
c. U: Can I fly with Ryanair?
d. S: Sorry, there is nothing matching your request.
e. U: What about Lufthansa?
f. S: Sorry, there is nothing matching your request.

We propose that the system should be collaborative and help the user find a satisfiable set of parameters. This involves the system being able to indicate a parameter to relax upon failed database search, or to indicate a parameter to keep in cases where some hits are found but they are too many to enumerate, and thus the search criteria need to be refined.

There are of course various ways to realize this collaborative system behavior. In this paper, we concentrate on the realization by generating conditional responses, such as those in (6b), (6c) and (7b).

(6) a. U: Can I fly on the second?
b. S: Not if you want to fly economy class.
c. S': Yes, if you can fly business class.

(7) a. U: Can I fly on the second?
b. S: Yes, if you can fly with SAS.

In (6a) (as a continuation of (3f)), the user changes the departure day to April 2^{nd}. In (6b) the system not only gives a negative answer, but also indicates that the failure of the database search with the changed parameter is conditional on the parameter *economy class* which the user has specified earlier. In an alternative response (6c) in this context, the system suggests *economy class* as an alternative for which the database query would be successful.

The result of the database search, and therefore the answer to a user's question, can also be contingent on a parameter the user has not specified. In

Figure 1: Patterns of conditional responses

	Negative CR	*Positive* CR
Question	$?p$	$?p$
Response	Not if c	Yes if c
Assertion	If c, then not-p	If c, then p
Implicature	If not-c, then p	If not-c, then not-p

this case, too, it makes sense to indicate this contingency to the user: in (7b) the system gives a positive answer to the question and also indicates that the database search is successful (and thus the answer to the question is positive) as long as an additional parameter, namely the SAS airline, is assumed.

CRs are briefly described in the next section.

3 Conditional responses

The CRs we currently consider have the form *Not if c / Yes if c*. As an answer to a yes/no question $?p$, a CR not only answers the question (positively or negatively), but it also implicates that if c does not hold, the answer would have the opposite polarity.[1] For example, (6b) suggests that the negative answer concerns only the case where the parameter *class* is set to *economy* whereas for a different value of this parameter the answer may be positive. Figure 1 summarizes the patterns of CRs.

Conditional responses are discussed in (Green and Carberry, 1999), and characterized in terms of the speaker's motivation to provide information "about conditions that could affect the veracity of the response". However, they consider only the cases in which the speaker does not know whether the condition holds, while utterances in which the condition is already specified are left unnoticed.

In (Karagjosova and Kruijff-Korbayová, 2002b; Karagjosova and Kruijff-Korbayová, 2002a), we proposed an analysis distinguishing two types of CRs with respect to the contextual status the parameter in the condition, on which the CR is contingent:[2] (i) CRs contingent on a contextually-determined parameter (**CDCRs**), as in (6b, 6c). (ii) CRs contingent on a contextually non-determined parameter (**NDCRs**), as in (7b). We also argued that this distinction is important for the *dialogue move* that such responses perform: A CDCR makes a proposal for revising the contextually determined parameter, and a NDCR raises the question whether the parameter holds. In fact, CRs perform multiple dialogue acts: they are a response (backward-looking function) and a question or a proposal (forward-looking function) at the same time. Both these functions of CRs are reflected in the implementation we are developing.

[1] Also corpus examples reported in (Karagjosova and Kruijff-Korbayová, 2002b; Karagjosova and Kruijff-Korbayová, 2002a) justify this treatment which distinguish berween the asssertion and implicature of the CR.

[2] This distinction is also supported by data from the Verbmobil appointment-scheduling corpus and the SRI American Express (AMEX) travel agency data (cf. (Karagjosova and Kruijff-Korbayová, 2002a)).

4 Generating CRs in GoDiS

In this section we describe our implementation of CRs as *answer moves* in GoDiS. Dialogue moves in GoDiS are modelled in terms of information state *update rules*. Specifying such rules involves defining preconditions for selecting a particular move (*move selection*) and the effects of the move on the information state (*move integration*). In § 4.1 we describe the GoDiS information state. In § 4.2 and § 4.3 we discuss the selection and integration of a CR as an answer move, respectively.

4.1 GoDiS information state

GoDiS is an experimental dialogue system built using the TrindiKit (Larsson et al., 2000) and using the Information State (IS) approach to dialogue modeling (Larsson and Traum, 2000; Larsson, 2002a). The type of record assumed for the GoDiS IS is a version of Ginzburg's Dialogue Game Board, (Ginzburg, 1996), and is shown in Figure 2.

The IS is divided into a PRIVATE and a SHARED part, the latter containing information that the system assumes to be shared by the user. The SHARED part contains information about the latest utterance (speaker and move(s)), the shared commitments (a set of propositions) and the stack of questions under discussion (QUD, (Ginzburg, 1996)).

In the PRIVATE part, the plan contains the system's long-term goals (a list of actions), whereas the agenda contains more immediate actions. For example, the action to ask the user where she wants to go, is represented on the agenda as $findout(\lambda x.dest(x))$, which will then be converted into the move $ask(\lambda x.dest(x))$.

Current versions of GoDiS use keyword and keyphrase spotting, and a simplified semantics. A user's utterance *I'd like to go to London* will be recognized as a move giving a destination through the word 'to' followed by a city, and its contents will be represented in the shared commitments as the

Figure 2: The GoDiS Information State

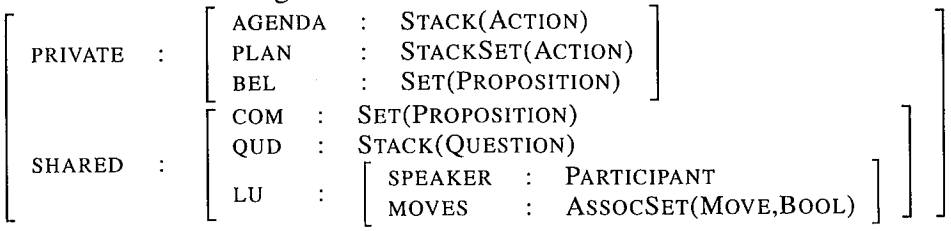


Figure 3: CR selection algorithm

Given a set of search parameters $P = \{p_1, \ldots, p_n\}$, where each $p_i = \alpha_i(v_i)$ is a proposition specifying the value v_i of an attribute α_i.
Given a (possibly empty) set of database search solutions $\text{SOL} = \{S_1, \ldots, S_m\}$ *s.t.* $\forall S_i \in \text{SOL} : P \subseteq S_i$.

If responding to a yes/no question q which specifies search parameters $Q = \{q_1, \ldots, q_k\}$

if SOL $= \emptyset$ (database search failed)

then

if $\exists p_i \in P/Q$ s.t. $\exists$SOL' $= \{S_1, \ldots, S_k\}$ *s.t.* $\forall S_i \in$ SOL' : $P/p_i \subset S_i\}$ (some parameter identified as responsible for search failure; if this parameter is relaxed, search succeeds)

then answer q with a conditional negative response with condition p_i

else answer q with an unconditional negative response

else (database search succeeded)

if $\forall S_i \in$ SOL : $S_i/P \neq \emptyset$ (there are other user-unspecified parameters in search result)

then

if the size of SOL is less than Max (the results are enumerable)

then answer q by enumerating SOL

else (the results cannot be enumerated)

if $\exists p_j$ *s.t.* $\forall S_i \in$ SOL : $p_j \in S_i/P$ (all results share some parameter p_j)

then answer q with a conditional positive response with condition p_j

else answer q with a positive response

predicate-logic proposition $dest(london)$. The corresponding question, where does the user want to go, will be represented on the QUD as $?\lambda x.dest(x)$.

4.2 CR Selection

The selection rules specify the conditions for selecting particular dialogue moves by the system. The conditions relevant for selecting a CR are specified in the selection algorithm in Figure 3.

The first step we make towards enhancing collaboration with CRs is to enable them under certain circumstances as answers to yes/no questions. This requires extending the treatment of user's questions in the GoDiS system for the travel agency domain beyond dealing only with questions concerning flight prices or visa requirements of various destination countries. We need to add new domain plans for dealing with other questions. Since questions like (5a) concern the availability of flights, we deal with them by defining a domain plan which consists of searching the database for flights satisfying a set of parameters. The parameters used for the search consist of those provided in the question plus other parameters provided ealier (if any), retrieved from the shared-commitments part of the information state.[3]

If the question revises a parameter that has already been specified earlier, the parameters provided in the question itself take precedence; for example in answering (8e) the search needs to be for a flight on April 2^{nd} even though the user earlier specified the departure day as April 1^{st}.

(8) a. S: Welcome to the travel agency!
b. U: The price for an economy flight from Frankfurt to Paris on April first please.
c. S: Sorry, there is nothing matching your request about price.
d. S′: The price is 200 Euro.
e. U: Can I fly on the second?
f. S: Not if you want to fly economy class.
g. U: Can I fly from Luxembourg?

[3] In a more cautious version of the plan, we should also ensure that the search only be triggered if some minimal amount of information about the flight has already been provided by the user (e.g., the departure or destination city).

In GoDiS, the parameters specified in a follow-up question replace the earlier specified parameters. (This is implemented as downdate of the previous parameter and an immediate update with the new one.) For example, before (8e), the IS contains $dep\_day(first)$, and after them it contains $dep\_day(second)$.

It is an open question whether such immediate revision of the parameter is warranted. (8f) is an example where the immediate revision could lead to trouble: Intutively, what the user is asking about in (8g) are economy flights from Luxembourg to Paris on April 1^{st}. However, the immediate revision strategy would lead to a search for economy flights from Luxembourg to Paris on April 2^{nd}. This arises as follows: In (8b), the user specifies the parameters as $dept\_city(frankfurt)$ & $dest\_city(paris)$ & $month(april)$ & $dept\_day(first)$. No matter whether the search result is negative (8c) or positive (8d), the user may want to explore other possibilities by asking (8e). If the system uses the immediate revision strategy, the departure day parameter will now be set to $dept\_day(first)$, and remain that also when (8g) is being interpreted, which is wrong.

The revision should therefore be at least conditioned upon success of the database search and possibly also upon the user's subsequent acknowledgement. Another solution is to treat such follow-up questions as introducing alternatives. Presently we leave this is an issue for future work.

If the system is indeed responding to a yes/no question, the next decision is based on the result of the database search, and on the subsequent processing of the results. Our version of the database search retrieves a (possibly empty) set of records that satisfy the search criteria. This differs from the previous versions of GoDiS, where the search retrieves the value of one parameter, e.g., *price*.

Currently we are dealing with two of the possible four cases of CRs: (i) a negative CDCR instead of a plain negative response, as a collaborative recovery from a failed database search; (ii) a positive NDCR instead of a plain positive response, as a collaborative continuation after a successful database search. We discuss the respective portions of the selection algorithm below.

Negative CDCR. When the database search with a set of user-specified parameters fails, the system attempts to collaborate by suggesting which (if any) parameter the user might relax to get a successful search instead. To find out, the system performs additional database searches with the parameters relaxed one at a time. If the relaxation of some parameter leads to a successful search the system produces a negative NDCR contingent on this parameter.

Figure 4: A toy database

dep	**dest**	**month**	**dep_day**	**class**	**airline**
London	Paris	April	2^{nd}	busin	BA
London	Paris	April	3^{rd}	econ	BA
Malmö	Paris	April	2^{nd}	econ	SAS
Malmö	Paris	April	2^{nd}	busin	SAS

For illustration, consider (6a) and the database records in figure 4. A search with the parameter set (9) fails. However, a search with the modified parameter set (10), where the *class* parameter is relaxed, succeeds, so (6b) can be generated, indicating that the negative answer depends on *class*.

(9) $\{dept\_city(malmoe), dest\_city(paris), month(april), dept\_day(second), class(econ)\}$

(10) $\{dept\_city(malmoe), dest\_city(paris), month(april), dept\_day(second)\}$

At the moment, we leave the order in which the parameters are tested arbitrary. Moreover, we currently only model the relaxation of one parameter p_j, but it is conceivable to look for a combination of parameters. More sophisticated techiques for conflict resolution are proposed for example in (Qu and Beale, 1999). It is therefore one possible strand of future work for us to incorporate such techniques.

We do not allow a NDCR contingent on parameters mentioned in the question, because of the oddity of answering a question such as (11a) with (11b).[4]

(11) a. S: Can I fly from Malmö to Paris on April 1^{st}?

b. U: Not if you want to fly on April 1^{st}.

Positive NDCR. Our system is collaborative also when the database search with the user-specified parameters succeeds, but there are more results than can be sensibly conveyed to the user at once.[5] In this case, the system attempts to use a CR to indicate when the success of the search depends on any

[4] As an anonymous reviewer pointed out, other follow-up responses are possible, such as: *You cannot fly on April* 1^{st}. *But April* 2^{nd} *would be possible.* We are aware of this, but leave the use of other forms for future work.

[5] The maximum number of results conveyed at once depends on the system's output modality: a graphical interface enables to output more than spoken mode or text mode on a small display. GoDiS uses either spoken or textual mode.

Figure 5: Effect of CRs on information state

	-CDCR *Not if c*	*+CDCR* *Yes if c'*	*-NDCR* *Not if c*	*+NDCR* *Yes if c'*
IS before:				
QUD	$?p$	$?p$	$?p$	$?p$
Shared	c	c'		
IS after:				
Shared	$\{c \wedge \neg p,$ $c' \wedge p\}$	$\{c' \wedge p,$ $c \wedge \neg p\}$	$\{c \wedge \neg p,$ $c' \wedge p\}$	$\{c' \wedge p,$ $c \wedge \neg p\}$
QUD	$?c$	$?c'$	$?c$	$?c'$

paramater as yet unspecified by the user. This helps to avoid future failed search as in (5c)-(5f).

This part of the algorithm relies on the database search returning all the records that satisfy the search criteria. We implement a simple subsequent processing which determines whether all results have any other parameter(s) in common. When this is the case, a positive NDCR can be generated.

For illustration, consider (7a) again. Given the database in Figure 4, a search with the parameter set (12) returns more than one hit. They all share the parameter $airline(sas)$, so (7b) can be generated, indicating this parameter as one on which the positive answer is contingent.

(12) $\{dept\_city(london), dest\_city(paris),$ $month(april)\}$

4.3 CR integration

The integration rules specify the effects of particular dialogue moves on the IS. The way we model the various effects of a CR on the information state is shown schematically in Figure 5.[6]

In GoDiS, the question $?p$ that a CR can be used to answer is represented as the question under discussion (QUD, (Ginzburg, 1996)). The answer provided by a CR is added to the shared commitments. The propositional content of a CR is a conditional. Since GoDiS does not support implication, the proper encoding of CRs as conditionals requires an extension of the semantic representation used in GoDiS. Presently, we use an approximation instead: We define the propositional content of a CR as a combination of its assertion and its implicature, and encode it as a set of two alternatives (e.g. $\{c \wedge \neg p, \neg c \wedge p\}$ for negative CDCR). This approximation is made possible by the fact that we only alow the generation of a CR in situations where the implicature is known to hold, and thus the conditional can be turned into a bi-conditional.

In addition to answering the question $?p$, the effect of a CR is that it puts a question corresponding to the condition on QUD. The effects of CDCRs and NDCRs on the IS differ, because they occur in different contexts: A NDCR contingent upon c as an answer to question $?p$ has the effect of raising the question whether this condition c, which had not been previously mentioned, should hold. In contrast, asserting a CDCR contingent upon c as an answer to question $?p$ cannot simply raise the question whether the condition c should hold, because c has been already determined and is part of the shared commitments. A CDCR therefore has the effect of *re*raising the question whether c should hold (Larsson, 2002a). This is consistent with the claim that CDCRs and NDCRs perform different dialogue moves (Section 3).

One issue that needs to be considered in relation to this is whether a CDCR involves removing the previously established proposition c (downdate). We believe that the examples (13c) and (13d) as alternative continuations of (13a) (as a continuation of (3f) above) show that the previously established proposition should not be downdated as a direct result of the CR, because the proposed revision can still be either accepted or rejected by the user.

(13) a. U: Can I fly on the second?
b. S: Not if you want to fly economy class.
c. U: Ok, I'll fly business class.
d. U': What about the third?

In (13a), the user asks about economy flights from Malmö to Paris on April 1^{st}, and the system gives a negative CDCR suggesting the parameter $class(economy)$ as responsible for the failed database search. In (13c), the user accepts the alternative of flying in business class. In this case, the proposition $class(economy)$ can be deleted from the shared commitments, and replaced by $class(business)$ as a result of (13c).

(13d), on the other hand, should be interpreted as asking about economy flights from Malmö to Paris on April 3^{rd}, indicating (in an indirect way) that the user does not want to revise the parameter $class(economy)$, but rather tries the parameter $dept\_day(third)$ instead of $dept\_day(second)$. In this case it would have been wrong to remove the proposition $class(economy)$ from the shared commitments as a result of the CDCR.

[6] The signs '-' and '+' abbreviate "positive" and "negative". c and c' are contextual alternatives.

Therefore the established proposition c should be kept in the shared commitments, while the alternative c' proposed in the CDCR is being considered. This is a modification of the way reraising is handled in GoDiS that we have to consider in more detail in future work.

5 Related work

Some of the problems we are dealing with in providing alternative suggestions to the user in information-seeking dialogues in over-constrained sitations, have been addressed in work concerned with conflict resolution (Qu and Beale, 1999), (Chu-Carroll and Carberry, 1994) and others (see (Qu and Beale, 1999) for further references).[7]

(Qu and Beale, 1999) propose a constraint-based model for cooperative response generation aiming at detecting and resolving situations in which the user's information needs have been over-constrained. In contrast to earlier work using heuristics for identifying relaxation candidates (e.g., based on constraint weights, (Abella et al., 1996; Pieraccini et al., 1997)), (Qu and Beale, 1999) employ AI techniques like constraint satisfaction, solution synthesis and constraint hierarchy.

(Chu-Carroll and Carberry, 1994)'s work on detecting invalid beliefs or plans and suggesting alternative solutions by relaxing over-constrained queries and proposing relaxation modification is also related to the work presented in this paper.

In comparison, our technique to identify a relaxation candidate is much simpler. We are currently exploring the possibility of employing the more elaborate techniques proposed in (Qu and Beale, 1999) also in the GoDiS system.

In contrast to the works cited above, we also deal with under-constrained situations where a positive CR indicates additional parameters in order to prevent failed future search. The only other work addressing also this issue we are aware of is described in (Hochberg et al., 2002), but their paper not provide much details in this respect.

6 Conclusions and future work

We proposed to improve the collaborative character of the responses of the GoDiS dialogue system participating in information-seeking dialogues. We described the way the system currently deals with successful and unsucessful database search and argued for the need to generate more collaborative responses that indicate sensible further search options to the user. We showed that CRs are suitable for this purpose, because they provide a positive or negative response contingent on a particular parameter. We then defined CRs in terms of the conditions on selecting them and their effect on the dialogue context.

We are currently implementing our proposal in the GoDiS system. We have preliminary versions of the update rules for CRs which are subsequently tested for improvement. We have implemented parts of the selection algorithm and a preliminary version of the database search and the subsequent processing of the results. Besides completing the implementation, we are presently considering several future extensions of our work.

First of all, we intend to implement a more sophisticated version of the further processing of the database search results by employing smarter techniques for identifying relaxation candidates and additional parameters.

Another extension concerns the integration rule for CDCRs. Currently, we are implementing the effect of a CDCR on the IS as reraising a question. However, reraising the question whether a proposition should hold can be seen as opening a negotiation whether the already specified answer to that question should be preserved or revised (because it has already once been determined). It would therefore be natural to provide an account of CDCRs as proposals opening negotiation. We plan to do this using the issue-based account of collaborative negotiative dialogues proposed in (Larsson, 2002b), which is based on (Sidner, 1994).

Another possible extension is to consider uses of CRs in other cases than as answers to yes/no-questions. That is, CRs can be used as answers to wh-questions, as in (14), or as grounding confirmations, as in (15).[8]

(14) a. U: Which day can I fly?
b. S: Monday, if you take business class.

(15) a. S: There is a flight from Philadelphia to Frankfurt on Sunday.
b. U: So I would arrive on Monday.

[7] An anonymous reviewer noted that the general issue of negotiation strategies after failed database search is presently being addressed by various teams developing commercial dialogue systems, e.g., the Soliloquy system. Another such system we have recently seen demonstrated is developed at IBM (Hochberg et al., 2002).

[8] In (Karagjosova and Kruijff-Korbayová, 2002a) we provide a corpus example illustrating this use of CRs.

c. S: Not if you leave in the morning.

d. S': Yes, if you leave in the evening.

Furthermore, CRs can also be realized by surface forms other than the ones we have considered so far (for example with *unless*). We plan to integrate different forms into our analysis and implementation.

Finally, collaborative responses in over- and under-constrained situation can be achieved in other ways than by using CRs, and it would be intereseting to investigate how the different ways relate and differ, in order to be able to capture systematic choices among them.

References

A. Abella, M.K. Brown, and B. Buntschuh. 1996. Development principles for dialogue-based interfaces. In *Proceedings of ECAI'96 Workshop on Dialogue Processing in Spoken Language Systems*, pages 1–7.

Jennifer Chu-Carroll and Michael K. Brown. 1997. An evidential model for tracking initiative in collaborative dialogue interactions. *Meeting of the Association for Computational Linguistics*, pages 262–270.

Jennifer Chu-Carroll and Sandra Carberry. 1994. A plan-based model for response generation in collaborative task-oriented dialogues. In *Proceedings of AAAI-94*, pages 799–805.

Robin Cooper, Staffan Larsson, Colin Matheson, Massimo Poesio, and David Traum. 1999. Coding Instructional Dialogue for Information States. http://www.ling.gu.se/projekt/trindi/.

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

Nancy Green and Sandra Carberry. 1999. A Computational Mechanism for Initiative in Answer Generation. *User Modeling and User-Adapted Interaction*, 9(1/2):93–132.

Judith Hochberg, Nanda Kambhatla, and Salim Roukos. 2002. A flexible framework for developing mixed-initiative dialogue systems. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, Philadelphia, PA, USA. Universtity of Pennsylvania.

Elena Karagjosova and Ivana Kruijff-Korbayová. 2002a. An Analysis of Conditional Responses in Dialogue. In *Proceedings of the 5th International Conference on TEXT, SPEECH and DIALOGUE*, Brno, Czech Republic.

Elena Karagjosova and Ivana Kruijff-Korbayová. 2002b. Conditional Responses in Information Seeking Dialogues. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, Philadelphia, PA, USA. University of Pennsylvania.

Staffan Larsson and R. Traum, David. 2000. Information state and dialogue management in the trindi dialogue move engine toolkit. http://www.ling.gu.se/publications/GPCL.html.

Staffan Larsson, Alexander Berman, Johan Bos, Leif Grönqvist, Peter Ljunglöf, and David Traum. 2000. Trindikit 2.0 manual. Technical Report Deliverable D5.3 - Manual, Trindi.

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

Staffan Larsson. 2002b. Issues under negotiation. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, Philadelphia, PA, USA. Universtity of Pennsylvania.

Ian Lewin, C.J.Rupp, Jim Hieronymus, David Milward, Staffan Larsson, and Alexander Berman. 2000. Siridus System Architecture and Interface Report (Baseline). http://www.ling.gu.se/projekt/siridus/.

R. Pieraccini, E. Levin, and W. Eckert. 1997. AMICA: The AT&T mixed initiative conversational arcitecture. In *Proceedings of EUROSPEECH*.

Yan Qu and Steve Beale. 1999. A constraint-based model for cooperative response generation in information dialogues. In *Proceedings of AAAI-99*, Orlando, FL.

Charles Rich, Candance L. Sidner, and Neal Lesh. 2000. COLLAGEN: Applying collaborative discourse theory to human-computer interaction. Technical Report TR-2000-38, Mitsubishi Electric Research Laboratories.

Candance Sidner. 1994. Negotiation in collaborative activity: A discourse analysis. *Knoweldge-based systems*, 7(4):265–267.

Making Sense of Partial*

Markus Löckelt and **Tilman Becker** and **Norbert Pfleger** and **Jan Alexandersson**
DFKI GmbH,
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
{loeckelt|becker|pfleger|janal}@dfki.de

Abstract

We address the challenging task of processing partial multimodal utterances in a mixed-initiative dialogue system for multiple applications such as information seeking, device control etc. Based on four different types of expectations computed by our action planner (aka dialog manager), we distinguish between expected utterances, various degrees of unexpected but plausible utterances, and uninterpretable utterances. We include a description of the backbone architecture and the representation formalisms used.

1 Introduction

An important characteristic of mixed-initiative dialogue are partial utterances that can only be interpreted in the context of the previous dialogue. The setting in which we investigate the interpretation of partial utterances is the multimodal, mixed-initiative dialogue system SMARTKOM. However, it is our larger goal to develop a core dialog back-bone. In fact, some of the modules described here have already been used in other projects. In SMARTKOM, communicative actions include spoken utterances and two-dimensional gestures (pointing, encircling etc.) by the user and also by an animated presentation agent. Since the analyses of both modalities are integrated into a common representation, our processing mechanisms are actually modality-independent. In the following we will use *utterance* to include communicative actions in all modalities.

There are a number of phenomena that occur naturally in such dialogues and systems, including the need for robustness against fairly common recognition errors. In this paper, we focus on partial utterances: Those in the context of user-initiative, usually additions to or changes of the current discourse context and those in the context of user-response, usually a reaction to a system request.

This paper first presents the dialog system SMARTKOM, its discourse and domain representation language and the data flow of processing. Section 2.2 presents the discourse modeler and 2.3 goes into more detail about planning a system action and supplying expectations for the next user utterance. The central sections are 3 and 4 which describe the details of interpretation as the integration of partial utterances into a coherent dialog context. Before we conclude our paper we discuss and compare our approach in the light of QUD in section 5.

* The research within SMARTKOM presented here is funded by the German Ministry of Research and Technology under grant 01 IL 905.

2 Architecture

Our back-bone is divided into different modules (see figure 1), each specialized on a different task. A multi-blackboard architecture (see, e.g., (Wahlster, 2000)) is used for communication where different modules publish or subscribe to indicate read or write permission for so-called data pools. All modules within the back-bone exchange information using a common representation: the domain model (see below). For the analysis part of the back-bone, either a single *application object* or one or more *sub-objects* are wrapped into a *hypothesis*, containing additional information, like syntactic information, scores, dialogue acts etc. A sequence of hypotheses forms a *hypotheses sequence* and finally several (alternate) hypotheses sequences form a *hypothesis lattice.*

2.1 Domain Model

The processing of discourse information is not independent of the representation formalism. Thus we will briefly characterize the approach taken in SMARTKOM, see (Gurevych et al., 2002) and also section 4. The representation is similar to frame-based approaches (Minsky, 1975). It is based on an ontology of application objects and subobjects. Application objects, e.g., a movie_ticket_reservation event, contain subobjects, e.g., movie_theater, show_time, and movie_information. Subobjects are recursively composed of other subobjects, e.g., the movie_theater subobject contains contact information which includes an address which includes a street_name etc. The list of top-level objects, i.e., application objects is of course determined by the set of applications. In SMARTKOM, there are currently about 10 different application objects. In ad-

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 101-107.

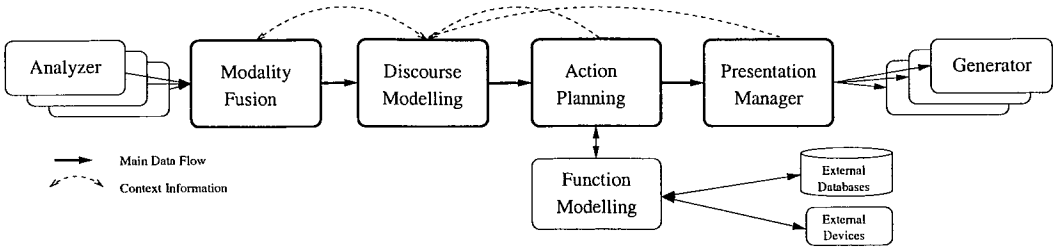


Figure 1: Architecture of the back-bone

dition to an application object, an analyzed utterance will contain a *goal*, e. g., *info* or *reserve* for a movie. However, the objects cannot be combined freely, as in logical frameworks such as those used for natural language semantics. Instead, the entire ontology is predefined for the given applications.

We call the position of a subobject in an application object a *path*. Some paths are called *slots*. These are paths within an application object that are meaningful for the action planner, e. g., the path leading to a (begin) time expression in a movie application object.

Furthermore, the application objects and subobjects are represented as *typed feature structures* with an inheritance *tree*[1] for the types. The type hierarchy is used to express relations – especially shared substructures – between objects.

2.2 Discourse Modelling

In a system where the user communicates with, e. g., spoken language and gestures, the recognition modules produce many hypotheses in addition to the semantic ambiguities arising while analyzing. We have chosen to tackle this challenge by letting all analysis components compute a score for each hypothesis. These scores form the decision base for the selection of a most probable hypothesis sequence.

Modality hypotheses are brought together by the modality fusion module into an intention hypothesis lattice (see (Johnston et al., 1997) for a similar approach) and are passed on to the discourse modeller. Before selecting a hypotheses sequence[2] the discourse modeller performs two tasks: First, it scores (see (Pfleger et al., 2002)) each hypothesis based on how well it fits into the current discourse context (*validation*). Second, appropriate contextual information (see section 3) is added to the hypothesis (*enrichment*). Important here is that although the input hypothesis may contain subobjects representing partial analyses, the output from the discourse modeller should contain only an application object (see also section 2.3). The enrichment is ensured using one simple and robust processing mechanism. In the case of a hypothesis containing an application object, the application object is compared with the discourse context using a non-monotonic operation we call OVERLAY (Alexandersson and Becker, 2001; Pfleger et al., 2002). The overlay operation is based on unification and works on two structures we denote *covering* and *background* where the former represents new and the latter old information. In case of conflicting information, i. e., unification would fail, overlay overwrites the conflicting information in the background. Important is that in case of a type clash, the background is first *assimilated* to the type of the covering thus making it possible to inherit parts of the background despite a type clash. The assimilation is computed via the least upper bound of the covering and the background[3]. If the hypothesis contains one or more subobjects, in a first step the most probable meaning of the subobject(s) is determined. Here, the expectations from the action planner give the expected/possible/filled paths which represent possible interpretations of the subobject(s). These paths are used to construct a new instance of the appropriate application object, an operation we call *bridging*. We call the result of the bridging operation the *interpretation* of the subobject(s). In the next step, the interpretation is overlayed over the appropriate background. During the overlay operation, a score is computed which represents how well the hypothesis fits the context (Pfleger et al., 2002). Finally, the hypothesis with the best overall score is selected and passed to the action planner which plans and triggers actions like questioning a data base or operating on an external device (see section 2.3) and/or appropriate presentations.

In case the subobjects could not be integrated into

[1] Although the implementation in SMARTKOM is based on an inheritance tree, our approach is general enough to handle inheritance lattices.

[2] SMARTKOM features two additional modules for intention selection and, e. g., incorporating default information to the hypotheses. The former may also take other factors into account; however, it is not an integral part of our core backbone.

[3] ... which might result in no inherited information. THis is the case if the least upper bound is top

an application object, fallback processing is responsible for resolving the meaning of this particular intention.

2.3 Action Planning

As an example consider the following user utterance:

(1) User: I want to record a film on channel two.

It would result in an intention hypothesis containing a goal to be reached in the VCR application, that of recording a film on the VCR, and provide an application object identifying the channel to be recorded.

If a hypothesis sequence arrives, the action planner processes the hypotheses in turn and devises a sequence of steps to establish the goals of the user. In the example task, the system needs some additional information. Minimally, this would be start time and end time. To reach the goal, a sequence of steps – a plan – is devised to accomplish completion of the task. This may involve actions on external functionalities like databases or abstract devices as well as seizing the dialogue initiative to request additional information from the user.

Figure 2 schematically depicts the plan for the goal VCR_record. The boxes indicate plan operators with their pre- and postconditions, as well as the actions and presentations generated during execution. The arrows indicate the (partial) plan execution ordering. To reach the goal, several "tracks" must be followed, each one ultimately establishing conditions needed by the goal. Since the channel was already provided in the utterance, no VCR_getChannel operation needs to be part of this plan.

To execute the plan, its tracks can be followed in arbitrary order. One possibility is to start with the plan operator VCR_getStartTime. The action planner seizes the dialogue initiative and triggers a presentation asking for the start time as specified in the plan operator, then halts since the postcondition does not hold (VCR_startTime is not known). It then waits for a message providing needed information – in this case, new user input in response to the presented request – then tries to proceed with plan execution.

In the example, two pieces of information of type "time specification" are needed to complete the overall task, a start and an end time. If it can be inferred from the form of the user utterance alone whether it is about start or end time, the data can be integrated at the correct position in the application object unambiguously. This is not always the case, though. In the example, an answer giving a time could be about a start or end time, but an end time is not really plausible because the system just asked for a start time. The discourse modeller will try to use discourse context to resolve ambiguities and insert subobjects at the correct positions. The action planner tries to help this resolution by publishing its expectations regarding user input.

2.4 Expectations

An *expectation* is a data structure providing predictions about anticipated user input. When the system has seized the initiative and the user is expected to react, an expectation is published to support the scoring of different interpretation alternatives for the reaction by specifying context information. Most importantly it divides the slots into four classes, depending on the current context:

1. A list of *expected slots*: Typically, application subobjects that the user has just been asked to provide.
2. A list of *possible slots*: application subobjects that are not filled and not focused on by the dialog at the moment, but of which it is in principle possible that they may be filled even if not asked for, e. g., the user may specify a recording time when asked to insert a medium.
3. A list of *filled slots*: subobjects that already have been assigned a value in the course of the dialogue. If the user utterance is about retracting or changing already present information, it should be interpreted with regard to filled slots.
4. All remaining slots implicitly belong to the class *other slots* which cannot be interpreted in the current context and are either misinterpretations or must be considered un-cooperative user behavior.

The expectation contains some additional information like the *type of the application object* that is currently active and/or talked about.

3 Integration of Partial

As discussed in section 2.2, the major tasks of the discourse modeler are validation and enrichment of the user utterance. For partial utterances which are analysed as subobjects, this includes the integration into the preceding context to achieve a coherent discourse.

The general procedure for this, see also section 2.2, is to integrate the subobjects into a new application object of the same type as the object of the focused application. Enrichment is achieved by applying the OVERLAY-operation (see (Alexandersson and Becker, 2001)) to the new application object and the application object in focus. We consider three cases of partial utterances with respect to their interpretability:

1. the system has the initiative and the user responds elliptically to a system request

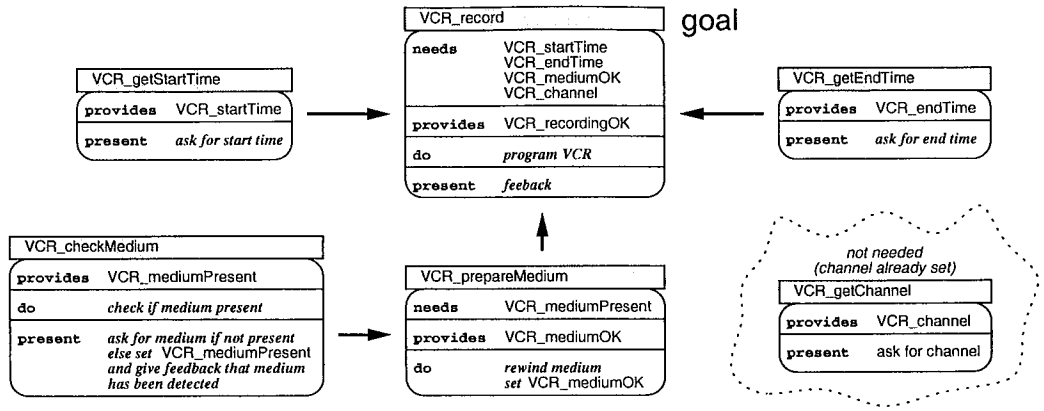


Figure 2: A schematic plan for *"I want to record a film on channel two"*

2. an elliptical user utterance does not correspond immediately to a request but can be interpreted in context
3. an elliptical user utterance cannot be interpreted, caused either by misinterpretation by the analysis module or non-cooperative user behavior.

In case 1 the answer uttered by the user contains the requested subobject (i. e. the user skips the constituents that are already mentioned in the question). The identification of an elliptical answer to the system request is based on the presence of an *expected slot* of the same type as the incoming subobject. Additionally, it is possible that the user is providing other relevant information, as in case 2. This can be interpreted via *possible* or *filled slots*. Section 3.2 describes the processing of user responses and an example is given in section 4.1.

In case 2 the user might provide unasked-for information that still fits into the discourse history, e. g. a restriction of "only channel two" upon presentation of an overview over the current TV program. Another possibility is that the user intends to change the value of an already set slot of an application object by uttering only the changed subobject, e. g., "and tonight?" in the same context.

The correct interpretation of such a subobject is a replication of the request from context including the new or changed subobject. This requires the presence of *possible* or *filled slots*. Either there are no *expected slots* or the subobject is incompatible with all of them. Section 3.3 describes the identification and processing of such partial utterances and in section 4.2 an example is given.

Case 3 is met when the discourse modeller is not able to integrate the subobject into an application object; this might occur in case of non-cooperative user behaviour or recognition errors. In such cases, supplemental recovery strategies are used.

Given these three classes of partial utterances and the four classes of slots as defined for our expectations, we can summarize the possible combinations (of which more than one can occur in the same utterance) in the table 1.

	User Initiative	System Initiative
Expected slot	NA	expected, unify, plan continues
Possible slot	plausible, unify, replanning	plausible, unify, replanning
Filled slot	plausible, overlay, replanning	plausible, overlay, replanning
Other	implausible – recover strategies	

Table 1: Possible combinations of partial utterances and different expectations given user or system initiative.

Note that if the user retains the initiative, no expected slots are available as marked by NA in the upper left cell. If the initiative remains with the system and the user supplies an expected answer (upper right cell), the action plan continues as in a simple system-initiative dialog system.

The processing of possible and filled slots does not depend on the status of initiative. Also, both are handled by a single mechanism, using the OVERLAY-operation. However, in the case of possible slots, OVERLAYreduces to pure unification. The special properties of overwriting are only used for already filled slots. These differences are reflected in different scores, see (Pfleger et al., 2002). The (re-)setting of filled and possible slots can give rise to the need to adapt the current plan, reflecting the mixed-initiative situation.

To illustrate the process of integrating a partial

user utterance in the context preceding it, we now take a brief look at the underlying structures and operations of the discourse modeller.

3.1 Context Representation

Our approach to context representation (Pfleger, 2002) is based on a three-tiered context representation (LuperFoy, 1991) and mental representation as in (Salmon-Alt, 2000).

For the work presented here, we focus on the three layers of the context representation. In the discourse layer, a *discourse object* (DO) represents a concept which can be referred to. It is introduced into the discourse either by means of language (user and system) or graphics (system). A DO keeps information about each mentioning depending on the manner of presentation. Additionally, each DO has access to its corresponding application object or subobject. Access to DOs in the discourse layer is restricted by a local focus stack. In the *Modality Layer* there are three types of objects: *Linguistic Objects* (LO), *Gesture Objects* (GO), and *Visual Objects* (VO). The third layer consists of *domain objects* are instances of our domain model, including, e. g., the application objects. Figure 3 shows how the three-tiered context representation and the local focus stack are connected. For each turn there exists exactly one application object or at least one subobject representing the current user or system intention. These objects are stored in order of occurrence in the domain layer and provide the domain information for the lower layers.

3.2 Processing User Responses

In the case of a user response to a system request, the integration of subobjects is guided by the list of *expected slots* which determine the paths of the expected slot fillers within the current application object. For the integration of the subobjects, first, a new application object of the same type as the current application is constructed. For each subobject in the hypotheses its type is compared with the types of the *expected slots*. In the case of a match, the subobject is integrated into the new application object using the path information of the matching slot. Then, the newly created and extended application object is OVERLAY-ed over the focused application object for enrichment with previous discourse information. If some subobjects do not match with any of the expected slots, the integration continues as described in the following section.

3.3 Processing Plausible Partial Utterances

In general, *possible slots* are processed just like *expected slots*. If the user tries to change a *filled slot*, this is implemented by searching for a discourse object of the same type in the local focus stack. Such a matching discourse object will always be found since the slot had been filled at some point in history. The scoring, however, might be low, e. g., in case it was mentioned a long time ago. A new application object of the same type as the current application is constructed. The subobject is integrated into this application object using the path information of the matching discourse object. Finally, the new application object is OVERLAY-ed over the application object of the previous user turn.

4 Two Examples

The first, shorter example shows the distinctions between expected and possible slots. The more elaborate second example, dealing with a filled slot, gives more details of the context representation.

4.1 Programming a VCR

On the basis of the following example we demonstrate the process of integrating a subobject in the case of system-initiative with a matching *expected slot*:

(2) User: I want to record a film.

(3) System: When should I start recording?

(4) User: From 1:30pm on channel two.

Analysis of (2) produces an intention hypothesis containing the setting of a goal VCR_record and an empty application object for the VCR application. The action planner accepts VCR_record as the current goal and constructs a plan to acquire the necessary data for a recording.

In this case, it decides that the time slot, VCR_startTime, should be filled first, and the plan operator that provides this slot takes the discourse initiative by triggering presentation (3) that asks for it. An expectation is published that contains VCR as the type of the current application object, VCR_startTime as the (only) expected slot, no filled slots (save maybe for possible application defaults), and, some possible slots, e. g., VCR_endTime and VCR_channel.

When the user answers (3) by (4), an intention containing two subobjects is passed from the modality fusion to the discourse modeller, which then uses the expectation to determine how to integrate the new information into a single application object. In (4), the expected slot is filled by *"1:30pm,"* but also one of the possible slots by *"channel two."* The two subobjects are integrated into a new application object which is afterwards OVERLAY-ed over the current application object stored from the previous turn. The resulting application object is passed on to the action planner, which will store the slot settings and continue the dialogue to reach the goal of programming the VCR.

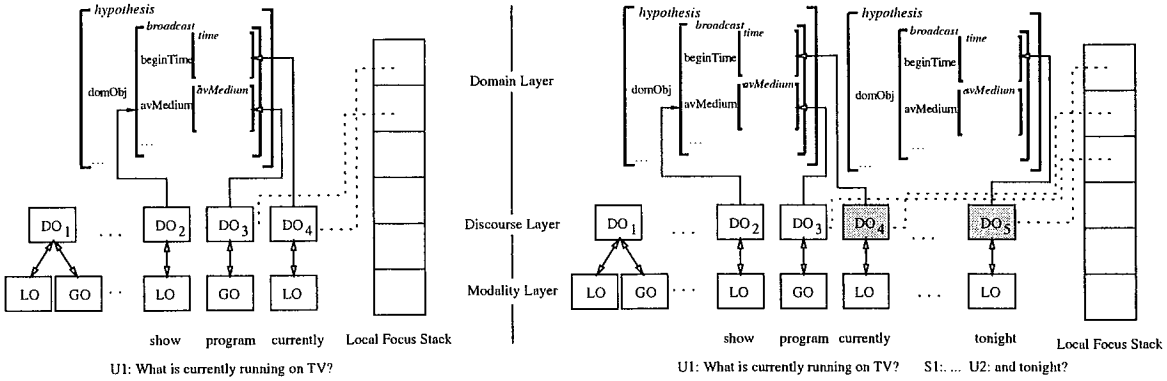


Figure 3: Two sample configurations of the context representation

4.2 Watching TV

With this example we demonstrate in some more detail the process of integrating a subobject in the case of user-initiative with no matching *expected-slot*:

(5) User: What is currently running on TV?

(6) System: The following (SMARTAKUS points at a list of programs) programs are currently running.

(7) User: And tonight?

The left part of figure 3 shows an excerpt of the configuration of the context representation after (5) has been processed. For example, the discourse object DO_4, introduced by the user's uttering of *"currently,"* points to the slot TV_beginTime in the current application object stored in the domain layer. It is also accessible through the focus stack.

The right part of figure 3 shows the situation after (7) has been analyzed. DO_5 is introduced by the user's utterance of *"tonight."* There is no *expected slot* and DO_5 does not match any of the *possible slots*. This triggers a search in the local focus stack for a discourse object corresponding to a *filled slot*. During this search, DO_4 is found. Then, a new application object containing the subobject of the user utterance (7) under the path for the filled slot (TV_beginTime) is created and enriched with the remaining information of the application object of (5), using the OVERLAY-operation. Finally, the resulting application object is passed on.

5 Discussion

We characterize our approach in the light of (Ginzburg, 1996). In (Ginzburg, 1996), his *Dialogue Score Board* (DSB) keeps track of:

- The set of currently accepted FACTS.
- The syntax and semantics of the LATEST-MOVE.
- QUD: A partially ordered repository that specifies the currently discussable questions.

Although Ginzburg's work on QUD aims for a more general framework than ours, there are some similarities between our approaches. Our notion of filled slots seems to correspond the the facts in the DSB. There is a close relation between the QUD and our exptected and maybe also our possible slots.

Amongst the differences between our approaches we have meta-talk. Instead of allowing for meta-talk we specialize on narrow but multiple domains and tasks, such as information search, device control, etc. In these scenarios, however, we consider our entire context for resolution of partial utterances. Another difference is the absence of traditional semantics in our approach. Instead we use an ontology-based domain model.

The salient feature of our frame work comes out clear in table 1. There we show a fine grained characterization of possible short utterances (which we call partial utterances) in different situations and their corresponding consequences. For all different cases, we use one robust and simple mechanism which treats all different expectations (expected, possible and filled slots): OVERLAY. Our scoring function distinguishes the different exptectations in such a way that unifiability is rewarded whereas overwriting is not (see (Pfleger et al., 2002; Pfleger, 2002) for more details).

6 Conclusion

We have shown how partial (multimodal) utterances can be interpreted by integrating them into an appropriate dialog context. While our approach hinges on a frame-like representation of dialog context and corresponding processing operations, some of the details have proven very helpful but are not essential to the core approach, e. g., the use a three-tiered discourse history, typed feature structures and the OVERLAY operation.

We have shown how in such an approach action planning can provide expectations to make sense of partial utterances in various situations. Especially for the interpretation unexpected (partial) utterances, a complex discourse history and focus stack are necessary to find the most plausible context for their integration. The approach is general enough to handle system expectations with the same processing mechanism.

References

Jan Alexandersson and Tilman Becker. 2001. Overlay as the Basic Operation for Discourse Processing in a Multimodal Dialogue System. In *Workshop Notes of the IJCAI-01 Workshop on "Knowledge and Reasoning in Practical Dialogue Systems"*, Seattle, Washington, August. `http://www.ida.liu.se/labs/nlplab/ijcai-ws-01/alex.pdf`.

Christian Ebert, Shalom Lappin, Howard Gregory, and Nicolas Nicolov. 2001. Generating full paraphrases of fragments in a dialogue interpretation system. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, Aalborg, Denmark, September, 1-2.

Jonathan Ginzburg. 1996. Dynamics and the semantics of dialogue. In J. Seligman, editor, *Language, Logic and Computation*, volume 1. CSLI Lecture NOTES, CSLI, Stanford.

Iryna Gurevych, Robert Porzel, and Michael Strube. 2002. Annotating the Semantic Consistency of Speech Recognition Hypotheses. In *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*, pages 46–49, Philadelphia, PA, July. Association for Computational Linguistics.

Michael Johnston, Philip. R. Cohen, David McGee, Sharon L. Oviatt, James A. Pittman, and Ira Smith. 1997. Unification based multimodal integration. In *Proceedings of the 35th ACL*, pages 281–288, Madrid, Spain.

Susann LuperFoy. 1991. *Discourse Pegs: A Computational Analysis of Context-Dependent Referring Expressions.* Ph.D. thesis, University of Texas at Austin, December.

Marvin Minsky. 1975. A framework for representing knowledge. In Patrick Winston, editor, *The Psychology of Computer Vision*, pages 211–277. McGraw Hill, New York.

Norbert Pfleger, Jan Alexandersson, and Tilman Becker. 2002. Scoring functions for overlay and their application in discourse processing. In *KONVENS-02*, Saarbrücken, September – October.

Norbert Pfleger. 2002. Discourse processing in multimodal dialogues. Master's thesis, Unversität des Saarlandes. Forthcoming.

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.

Dialogue Analysis for Diverse Situations

William Mann

SIL International

Abstract

Research on human language must cope with diverse and abundant complexity. One common approach is to reduce the scope of study to single languages, single cultures, single situations and homogeneous sources of data. This succeeds, at the cost of becoming unaware of the broader complexities of the phenomena studied, and also at the cost of missing the broad applicability of narrow, successful results.

In the case of studies of human dialogue, it has been common to restrict or ignore the variability and complexity of the subject, so that in multiple studies of "conversation," for example, it is treated as if the subject matter were held in common.

This paper explores diverse applications of a particular descriptive theory of dialogue structure. It reports application of the theory to an opportunistic diversity of dialogues in English, seeking to sketch the generality and limits of the theory. One use of such explorations is to guide work seeking to overcome the limits that are found.

There have been studies of ***task oriented dialogue*** for decades, but little exploration of what other kinds of dialogue the results of such studies might cover. This paper finds that methods that work for task oriented dialogue sometimes work beyond that boundary as well.

This paper reports an exploration of the nature of dialogue coherence. There are many papers that describe contributions to coherence without identifying the nature of coherence. Here we are concerned with the nature of coherence as a whole.

There are many views of coherence in language, and of dialogue coherence in particular. To appreciate them, a distinction must be made between coherence and cohesion. Cohesion represents connectedness of text that arises from the use of particular linguistic devices such as pronouns, anaphoric reference, patterns of topic or focus and many more. In detail it is language dependent. The classic exploration of cohesion is. (Halliday and Hasan 1976)

In contrast, coherence is a kind of impression that arises (or not) in a person who attempts to understand particular language use. In general it is not language dependent, in the sense that a translation of a coherent text is coherent, even if the cohesive devices of the source text are absent in the target language. Although not all of the views of coherence make this distinction, it is essential for comparisons.

One view says that coherence is simply the collective effect of use of cohesive devices, a view that is more often assumed than stated. Another view sees coherence as arising from propositional consistency, plus relevance in Grice's sense. (Goldberg 1983) Another group of views sees coherence as arising

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 109-116.

from genre or common recurrent patterns of expression. (Farrell 1983) Using Grice's Cooperative Principle and maxims, contributions to coherence are found in signaling creative violation of maxims (Mura 1983). Some scholars regard coherence as identified with continuity of topic, theme or focus. (Crow 1983) Others seek (or criticize) a more formulaic scheme, as for example well formedness of sequences of acts. (Jacobs and Jackson 1983) Others, notably Searle, seek for an extension of the concept of speech act to account for the structure of "conversation"; the extension is finally seen as implausible. (Searle 1991)

The view being explored here rejects each of those views as insufficient.

Coherence represents integrity of intentions that, in the processes of interpretation, are imputed to the text creator – here the dialogue participants. This integrity includes consistency, sequential continuity and apparent completeness of the imputed intentions and their representation in language.

Coherence as a technical focus shapes the exploration in ways that differ from the effects of an interpretation focus, a generation focus or more generally a participant or agent focus. It is simply an alternate technical focus, expected to be informative in its own way. It requires an uncommonly broad exploration of natural dialogue data.

The intentions exhibited in dialogue depend very strongly on the dialogue situation. In a hostage negotiation or a medical interview, intentions may arise that are never seen in tutoring transcripts or laboratory elicitations. Yet the notion of coherence seems equally applicable. So, for any particular view of coherence, it is worthwhile to examine situations in which it produces a credible account, other situations where it does not, and eventually how these collections differ.

Situation here is an informal term. We do not yet have enough evidence to select a technical view of situation that helps in modeling coherence. The models and the relevant notion of situation can be expected to develop together.

Dialogue Macrogame Theory

There is an approach to intention-based dialogue annotation that is useful in this exploration. It is called Dialogue Macrogame Theory (DMT). (Mann 2002) Because it is widely unknown, it is sketched here.

The terminology of DMT is intended to avoid confusion. There are various uses of "dialogue game" and closely related terms. (Mann, Carlisle et al. 1975; Levin and Moore 1977; Mann, Carlisle et al. 1977; Mann 1979; Carlson 1983; Mann 1988; Kowtow, Isard et al. 1993; Traum 1994; Carletta, Isard et al. 1996; Poesio and Traum 1998; Traum 1999; Stolcke, Reis et al. 2000; Traum 2000; Kreutel and Matheson 2001; Mann 2001a) DMT differs from these precedents enough that to simply call it "dialogue game theory" would be inappropriate.

DMT is a framework in which analysts can describe dialogues in a systematic way. It does not explain why such analysis is possible, but rather it enables the identification of certain observable regularities and patterns in dialogue. The key analytic facility of DMT is the capability of analysts to understand particular kinds of human interactions.

DMT has two major types of abstract constructs, *Dialogue Macrogames* and *Unilaterals.*

Dialogue Macrogames

Dialogue Macrogames are used to describe intervals of dialogue interaction. For convenience these can be thought of as sets of adjacent turns, although more finely divided elements are used. These intervals can be as short as an utterance and a response: "Ready?" "Yup." but in general they are indefinitely long. The intervals are distinguished by the use of a joint intention. (Clark 1996; Tuomela 2000)Clark, Tuomela. Each such interval is thereby collaborative.

In DMT it is possible for each participant to have a continuously definite notion of the intentions governing the course (but not the intellectual or other content) of the dialogue. Also, both participants can continually control the course of the dialogue, in the sense that each proposed course of talk can be vetoed.

These possibilities are expressed in DMT using a negotiation metaphor. A participant proposes using a particular dialogue macrogame by a ***bid***, and the bid is ***accepted*** or ***rejected***. If the bid is accepted, then the macrogame is ***pursued***. Either participant can ***bid termination*** of the macrogame, and the other participant can ***accept termination*** or ***reject termination***. If termination is rejected, pursuit of the game continues. As we will see below, these actions are typically implicit.

The novelty of this negotiation scheme, in comparison to other "game" notions applied to dialogue, is mostly in the way of explicitly representing possibilities of rejection, of choosing not to enter into a particular course of discussion, or to leave one. These possibilities have long been known, but have not always been given theoretical status. For example, in designing the Maptask representation, this was recognized:

> "It is also theoretically possible at any point in the dialogue to refuse to take on the proposed goal... .Often refusal takes the form of ignoring the initiation ... However, it is also possible to make such refusals explicit... . These cases were sufficiently rare ...that it was impractical to include a category for them.", (Carletta, Isard et al. 1997) p. 21,22.

In the hope of broad coverage the possibilities of rejection were made explicit in DMT.[1]

A dialogue macrogame, (henceforth a ***game***), consists entirely of a set of three goals (intentions):[2]

1. A goal of the ***initiator***
2. A goal of the ***responder***
3. A joint goal

A game is a convention, loosely comparable to a lexical item or a grammatical pattern.

DMT expects that people will generally hold hierarchic goal structures, one use of them being to allow embedded uses of games. However, the three goals of a game definition are not in hierarchic configuration. When a game bid has been accepted, the goal of the initiator and the joint goal will be in the memory of the initiator as commitments. (DMT does not constrain the relationship of these two.) Similarly, the goal of the responder and the joint goal will be in the memory of the responder as commitments. In each memory, the two

---

[1] The relationship between DMT and the Maptask work is discussed in a paper presented at the July 2002 SIGdial workshop and available at http://www-rcf.usc.edu/~billmann/dialogue. The two schemes are superficially quite different but very comparable.

[2] There is no reason why these could not be three sets of goals, but in DMT development so far this flexibility has not been required.

goals are committed and dismissed (un-committed) simultaneously, and at the same time each person's knowledge of the other's commitments is adjusted. This, with the negotiation acts, makes it possible for each participant to know the committed intentions of the other, because by acceptance of the bids of the other they simultaneously adjust their own intention sets and those they impute to the other participant.

These three goals are not unrelated. The initiator's goal and the responder's goal are compatible, not in conflict. They both support achievement of the joint goal. Since in dialogue the participants need to know what they are doing in order to do it, and they generally want to know why they are doing it, we can expect that ordinary dialogue that seeks cooperative action will be clear enough about what goals each party holds so that the parties can proceed.

The analysts' understanding of dialogues is generally somewhat less than that of the participants, but comparable. Neither the participants nor the analyst need supernatural knowledge of other persons' thoughts.

One of the games is named the ***Information Offering game***. Like all of the other games, a single occurrence of this game can be used to account for an indefinitely long interval of interaction. Currently in DMT there are about 17 defined games.

The definition of the Information Offering game is:

1. goal of the initiator:	to provide particular information to recipient
2. goal of the responder:	to identify and receive the particular information offered
3. joint goal:	the responder comes to possess the particular information

These, of course, are not fully specified goals. Rather, they (and all of the goals of the framework) have places for unspecified arguments, such as the particular information above. DMT does not assume that the dialogue is task oriented. That is, it does not assume that every dialogue involves pursuit of joint intentions known to both participants. In task oriented dialogues, the goals of a particular use of a game may be task goals, and a particular goal can appear in more than one position in a game use.

All of the goals that are used in DMT games have the following attributes in the (Mann 2001b)framework of the attributes of intentions in language theories: partialness, priorness, tacitness, immediacy , interaction-configuring, intended to be recognized, structuredness, complementarity, conventionality. The joint goals have the additional attribute of jointness, and they lack complementarity.

Game definitions exist for about 17 games exemplified in actual natural cases. General annotation instructions exist for identifying bids and acceptances of both games and terminations.

Unilaterals

In natural dialogue there is a lot of activity that is not collaborative. To represent dialogue intervals where there is no joint goal being pursued, DMT has a separate set of categories, the Unilaterals. Up to now, they have been documented far less than the games. Unilaterals are particularly useful for establishing and ending communication, for managing the media (such as conversation using a noisy radio channel), for various kinds of acknowledgment and expressions of politeness. Unilaterals are also used to

represent individual direct action during an interaction, and to represent announcements or other non collaborative giving of information.

A DMT Example

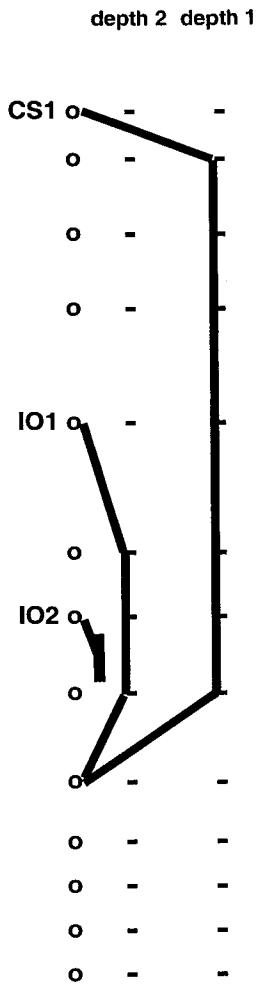
The example below is very simple, but it illustrates several of the basic mechanisms of DMT. It is an administrative phone call, quoted in (Clark 1996)from a corpus by Quirk and Svartik. It is a call from an academic secretary to someone outside of the department. The dialogue transcript and analysis are shown in the figure below.

Since this is a complete dialogue rather than an excerpt, it is appropriate to note the placing of the call as a way of seeking a conversation, by bidding the Conversation Seeking game (bg CS). For uniformity, like all other game uses in the annotation scheme, it is labeled with letters and a number, in this case CS1.

In turn 2 Blake accepts the bid. Since turns 2 through 4 establish the identities of the participants, and they do not otherwise change the goals held by either participant, they are labeled as Media Management (MM).

In turn 5 Allen bids the Information Offering game, in this case by beginning to convey the offered information. Turn 6, "right", accepts the movement to engage in information offering. Since the offering

Turn	Who	Speech by Allen	Speech by Blake	Game acts and Uni laterals	Game name
1	Allen	**[rings Blake]**		bg CS	Conversation Seek
2	Blake		**hello Michael Blake,**	ag, MM	
3	Allen	**um, this is Professor Worth's secretary, from Pan-American College,**		MM	
4	Blake		**oh, yes**	MM	
5	Allen	**um Professor Worth suggested contacting a Mr. L. S. Worenam, W O R E N A M, -he's in the department of English, at Royal Paramount College**		bg IO	Info Offer
6	Blake		**right**	ag, ACK, bt IO1	
7	Allen	**do you know the address ?**		rt IO1, bg IO2	Info Offer
8	Blake		**um oh I I'll find that, um yes indeed, well, thank you very much**	rg IO2; bt IO1; bt CS1 POL	
9	Allen	**right**		at IO1, at CS1, POL	
10	Blake		**thank you**	POL	
11	Allen	**OK**		POL	
12	Blake		**goodbye,**	POL	
13	Allen	**bye**		POL	



seems to be complete, and thus the goals of IO1 satisfied, turn 6 also bids termination of IO1, (bt IO1).

In turn 7 Allen offers the address of Mr. Worenam. The offer is refused in turn 8. This is really a somewhat rare occurrence, a definite refusal that is clearly so, not simply an ignoring of what was said.

Given that refusal, the information offering episode and the entire conversation are, from a goal pursuit point of view, over. This is acknowledged in turn 9, which accepts termination of both IO1 and CS1. However there are still things to attend to: the medium and politeness. Beginning with "well, thank you very much" in turn 8, there is a six-element closing sequence of polite acts that involve gratitude, acknowledgement and termination, here all labeled POL, which represents the undivided politeness category.

Recent Use of DMT

DMT has been applied opportunistically to the following sorts of situations: Apollo 13 emergency radio conversations, medical interviews, airline cockpit pre-crash radio, travel agents' conversations, computer mediated computer assistance, administrative phone calls, courtroom questioning of witnesses, laboratory conversational tasks, hostage negotiation, mathematics tutoring, electronics tutoring.

Examples (8 or more) are on http://www-rcf.usc.edu/~billmann/dialogue . They include tutoring, hostage negotiation, spacecraft emergency, travel agent work and laboratory map following.

The set of games and unilaterals has continued to grow slowly, especially when new situations are encountered.

Coherence and Situation

How can judgments of coherence depend on situation? A possible path runs as follows. The appearance of coherence of a dialogue involves attributing intentions to the language user. The intentions of the language user depend on what that user hopes to accomplish. In dialogue, this is specifically what that user hopes to accomplish in the present place and time by interacting with the present other participant. Situation strongly limits what can be accomplished, and may also constrain what can be said. Interactions in courtroom actions are strongly constrained by tradition and by law, for example. More generally, plausibility of intentions, and of their use in particular structures, can be strongly affected by situation.

Recent Findings

- Rejection of bids is unexpectedly frequent in general, but rare or absent in certain situations including some involving social dominance or extreme cooperativeness.
- Task oriented dialogue can often be accounted for using general purpose structures.
- For interpreting particular dialogue elements, the non equality of participants is important. Medical interviews illustrate this.
- What counts as a coherent next element of dialogue is often dependent on the situation.
- Dialogue resumption is a basic structural feature. Natural dialogues often function as resumptions of previous dialogues.
- It is easy to find natural dialogues that DMT cannot represent. Most commonly

such dialogues do not involve joint intentions known to both parties. [3]

References

Beach, Wayne A. (1996). *Conversations About Illness: Family Preoccupations with Bulimia.* Mahwah, NJ: Lawrence Erlbaum Associates.

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

Carletta, J., A. Isard, S. Isard, J. C.. Kowtko, G. Doherty-Sneddon and A. Anderson (1997). The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics* **23**(1): 13-32.

Carlson, Lauri (1983). *Dialogue Games: An approach to discourse analysis.* Boston: D. Reidel.

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

Crow, Bryan K. (1983). Topic Shifts in Couples' Conversations In R. T. Craig and K. Tracy (eds,). *Conversational Coherence: Form, Structure and Strategy* Beverly Hills: Sage Publications, 136-156.

Du Bois, John W. (2000). *The Santa Barbara Corpus of Spoken American English*: .

Farrell, Thomas B. (1983). Aspects of Coherence in Conversation and Rhetoric In R. T. Craig and K. Tracy (eds,). *Conversational Coherence* 2: Beverly Hills: Sage Publications, 259-284.

Goldberg, Julia A. (1983). A Move Toward Describing Conversational Coherence In R. T. Craig and K. Tracy (eds,). *Conversational Coherence: Form, Structure and Strategy* Beverly Hills: Sage Publications, pp. 25-45.

Halliday, M. A. K. and R. Hasan (1976). *Cohesion in English.* London: Longman.

Jacobs, Scott and Sally Jackson (1983). Speech Act Structure in Conversation In R. T. Craig and K. Tracy (eds,). *Conversational Coherence* 2: Beverly Hills: Sage Publications, 47-66.

Kowtow, Jacqueline C., Stephen D. Isard and Gwyneth M. Doherty, (1993). *Conversational Games Within Dialogue*, HCRC: Human Communication Research Centre, Edinburgh, 12. Previous version 1991: Espirit Workshop on Discourse Coherence.

Kreutel, J. and C Matheson (2001). *Context-Dependent Interpretation and Implicit Dialogue Acts.* 5th Workshop on Formal Semantics and Pragmatics of Dialogue (Bi-Dialog 2001). P. Kuhnlein, H. Rieser and H. Zeevat. Bielefeld: , 68-78.

Levin, James A. and James A. Moore (1977). Dialogue Games: Meta-Communication Structures for Natural Language Interaction. *Cognitive Science* **1**(4).

[3] There are several interesting, accessible sources of such dialogues. A particularly interesting and significant one is the centerpiece of the book *Conversations about Illness* (Beach 1996). The Santa Barbara Corpus of Spoken American English (Du Bois 2000)contains several others, generally with more than two parties. With young children, the CHILDES Database, a resource under ongoing development, contains many more (http://childes.psy.cmu.edu/). See also the rich index at http://devoted.to/corpora.

Mann, William C., (1979). *Dialogue Games*, USC Information Sciences Institute, Marina del Rey, CA, ISI/RR-79-77,

Mann, William C. (1988). Dialogue Games: Conventions of Human Interaction. *Argumentation* **2**(1988): 511-532.

Mann, William C. (2001a). *Extension of Dialogue Game Theory*. Workship on Coordination and Action (on occasion of ESSLLI XIII). P. Kuhnlein, A. Newlands and H. Rieser. Helsinki, organized by G. Sandu and A. Pietarinen (Dept. of Philosophy, Univ. Helsinki): .

Mann, William C. (2001b). *Models of Intentions in Language*. 5th Workshop on Formal Semantics and Pragmatics of Dialogue. P. Kuhnlein, H. Rieser and H. Zeevat. Bielefeld: .

Mann, William C. (2002). *Dialogue Macrogame Theory*. SIGdial Workshop 2002. Philadelphia: .

Mann, William C., James H. Carlisle, James A. Levin and James A. Moore, (1975). *Observation Methods for Human Dialogue*, USC Information Sciences Institute (ISI), Marina del Rey, CA, ISI/RR-75-33; U.S.Govt. ERIC ED112871,

Mann, William C., James H. Carlisle, James A. Levin and James A. Moore, (1977). *An Assessment of Reliability of Dialogue-Annotation Instructions*, USC Information Sciences Institute (ISI), Marina del Rey, CA, ISI/RR-77-54; U.S.Govt. ERIC ED136779,

Mura, Susan S. (1983). Licensing Violations: Legitimate Violations of Grice's Conversational Principle In R. T. Craig and K. Tracy (eds,). *Conversational Coherence* 2: Beverly Hills: Sage Publications, 101-115.

Poesio, M. and D. R. Traum (1998). *Towards an axiomatisation of dialogue acts*. Twente Workshop on the Formal Semantics and Pragmatics of Dialogues. J. Hulstijn and A. Nijholt. Enschede, Universiteit Twente, Faculteit Informatica: , 207-222.

Searle, John R. (1991). *(On) Searle on Conversation*. Amsterdam: John Benjamins.

Stolcke, Andreas, Klaus Reis, Noah Coccaro, Elisabeth Shriberg, Rebecca Bates, Van Ess-Dykema and Marie Meteer (2000). Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics* **26**(3): 339-374.

Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Department of Computer Science. Rochester, N.Y., University of Rochester: .

Traum, David R. (1999). Speech Acts for Dialogue Agents In M. Wooldridge and A. Rao (eds,). *Foundations of Rational Agency* 14: Dordrecht: Kluwer, 169-202.

Traum, D. R. (2000). 20 Questions for Dialogue Act Taxonomies. *Journal of Semantics* **17**(1): 7-30.

Tuomela, Raimo (2000). *Cooperation: A Philosophical Study*. Dordrecht: Kluwer Academic Publishers.

A simulation of small group discussion

Emiliano G. Padilha* and **Jean Carletta**
ICCS – University of Edinburgh
2 Buccleuch Place, EH8 9LW, Edinburgh, UK
[emilianp,jeanc]@cogsci.ed.ac.uk

Abstract

We describe a simulation of turn-taking in small group discussion which includes behaviours from a range of modalities such as speech, gaze, facial expression, gesture, and posture. The simulation is intended to show how these behaviours contribute to the turn-taking process. The agents that participate in the discussion make probabilistic decisions about whether or not to exhibit behaviours such as speaking, making a backchannel, or shifting posture, within a general framework suggested by the existing largely descriptive literature on turn-taking. The group behaviours that characterize turn-taking models, such as turns, simultaneous speech, and competition for the floor, emerge from the individual behaviours of the agents. At this stage in the project the basic model has been designed and prototyped.

1 Introduction

Simulation is a well-established method for investigating computational models of dialogue. There have been a number of two-party simulations (Power, 1979; Houghton and Isard, 1987; Carletta, 1992; Walker, 1994; Guinn, 1996) focused on intentional structure aimed at task-oriented dialogue. Multi-party conversation or *group discussion*, on the other hand, has seldom been modelled, although much is understood about it. The best known simulation (Stasser and Taylor, 1991) reproduces only the order and distribution of turns by choosing the next speaker via two parameters: their relative talkativeness and how many turns have passed since they last spoke.

In group discussion, non-verbal behaviours play an important part. Besides complementing the discourse and providing emotional content, they help coordinate the turn-taking. It is not surprising therefore that they are starting to be applied in conversational agents that interact with human users, as in e.g. Cassell et al. (1994), Beskow et al. (1997), Rickel and Johnson (2000). This improves the communicative efficiency and naturalness of such agents.

This paper describes a simulation of small group discussion, i.e. three to seven equal-status participants engaged in unstructured conversation. Communication in such groups proceeds as interactive dialogue, whereas bigger ones have a character more like serial monologue (Fay et al., 2000).

In this simulation, simplified behaviours on a range of modalities are reproduced: speech, gaze, head movement, facial expression, gesture, and posture. The contents and language processing of the conversation are abstracted away to focus on how these behaviours contribute to the turn-taking process. Within a framework suggested by the existing literature on turn-taking, participants in the discussion are modelled as independent agents with a set of parameterizable attributes that define likelihoods of performing the various behaviours. Group phenomena that characterize turn-taking models, such as normal and overlapped transitions, simultaneous speech, and the competition for the floor, emerge from individual behaviours, such as talk in turn, feedback, and turn-claiming gestures.

At this stage in the project the basic model has been designed, prototyped, and is being tested. We intend to evaluate it against a corpus of group discussions.

* This research was supported by a studentship from CNPq, Brazil's National Research Council.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 117-124.

2 Background

In group discussion, participants take *turns* at talk. During a turn, the speaker monitors whether others want to speak. This could be because he wishes to end the turn or in case someone else has a question or comment to interpose. The speaker's discourse has natural points for others to begin their turns. These points are called *transition relevance places*, or TRPs. According to Sacks et al. (1974), when the speaker selects one addressee, for instance by asking a question of him specifically, then that addressee will take the next turn. When the speaker does not select a particular addressee, leaving a *free* TRP, anyone can self-select to speak. If this does not occur, the speaker may (but need not) continue to talk.

2.1 Speaker transitions

Although these rules capture the basic properties of turn-taking, they are somewhat simplified. More than one participant can self-select to talk at a TRP. The earliest one generally takes the turn, but in simultaneous starts, the loudest usually wins (Meltzer et al., 1971), but not always. For instance, negative feedback can take precedence (Sacks et al., 1974). More specifically, the decision to continue in multiple talk has more to do with eagerness to make a contribution and the involvement in the discussion, that affect the priority given to one's own turn over the others (Oreström, 1983): i.e. some sort of confidence in oneself.

There are also *interruptions*, when someone starts to talk in the middle of the speaker's turn, between TRPs. They can happen because the interrupter has misjudged the location of a TRP, because he has anticipated what the speaker will say and wants to cut him short (Oreström, 1983), or when collaborating in an apparent "free-for-all" joint building of an idea (Edelsky, 1981). Theoretically, it could also be that the interrupter is not really listening anymore. The frequency of such interruptions depends on the size of the group (Fay et al., 2000) and in general three outcomes can be expected: either the interrupter stops in a *false-start*, or the speaker is *cut-off*, or he finishes his turn in an *overlapped transition*, talking simultaneously with the new speaker who takes the subsequent turn.

However, most speaker transitions do occur at TRPs, with a slight gap or overlap, or none at all (Sacks et al., 1974). *Gaps* are the silent intervals between transitions. They are an emergent phenomena of the group, and different from a speaker's *pause* within his turn. Mean intervals between turns in a corpus of goal-oriented dialogue were observed to be around 450–650ms (Bull and Aylett, 1998). In cases of slight gap or overlap, hearers still perceive the transition as being smooth.

Such tight timings are possible because listeners can anticipate, or *project* the TRP, chiefly from the prosodic, syntactic and semantic characteristics of the speech (Schaffer, 1983; Oreström, 1983), and also by various non-verbal behaviours. Speakers tend to look away from their interlocutors in the early stages of planning an utterance, gazing back when done planning what to say so they can monitor uptake (Kendon, 1967). They typically break mutual gaze with their interlocutors shortly after taking the turn, but sometimes maintain it a little longer before gazing away (Novick et al., 1996). Although speakers make gestures in most of their clauses (McNeill, 1992), varying with the individual, situation, and culture, they always stop them before finishing to speak (Duncan, 1972). Changes of posture occur often when speakers initiate a turn and at boundaries of discourse segments (Scheflen, 1972), which generally coincide with TRPs (Cassell et al., 2001). Listeners wanting to talk use these cues together with the speech in their attempts to take the floor.

2.2 Listener activity

While the speaker is talking, listeners provide information to the speaker about how the communication is going. *Feedback* signals at a TRP indicate that the listener does not want to speak. Positive feedback such as nods or backchannel continuers like "uh-huh," "mhm" or "right" may mean continued attention, agreement and possibly various emotional reactions, showing that the speaker can continue. Negative feedback such as a puzzled facial expression, "eh?" or "what?", on the other hand, indicates some problem in the hearing or understanding of the message. Usually, negative feedback induces the speaker to reformulate or further explain his meaning. Feedback in gen-

eral is extremely common in two-party dialogue, but is less frequent in groups (Boden, 1994) and certainly varies across cultures, gender and individuals.

Other than feedback, listeners may also indicate that they wish to speak at the next TRP. These are *turn-claiming* signals (Duncan, 1972): for instance, posture shifts accompanied or not by gestures and feedback between TRPs[1] (Beattie, 1985), or simultaneous talk and false-starts (Oreström, 1983). All of these behaviours would tend to draw the attention of the other participants. Meanwhile, listeners maintain long gazes at the speaker interspersed by short glances away (Argyle and Cook, 1976). Speaker and listeners thus interact actively in the discussion producing a complex pattern of verbal and non-verbal behaviours.

3 The simulation

Since we are not generating actual language, with its timing, prosody, syntax and semantics that allow one to anticipate and identify TRPs, a number of simplifications have to be made. TRPs are currently provided explicitly by the speakers, as well as being announced ahead of time by a "pre-TRP" cue. This is so that listeners can decide whether they want to talk and if so, start behaving as if they are going to take the turn. Also, in order to simulate the various short intervals between turns and to distinguish precedence in multiple starts (when generally the first continues), starting turns are timestamped in fractions of a second before or after the TRP, i.e. negative or positive offsets representing overlaps and gaps, respectively.

The behaviours of the simulation fall on the following modalities:

- speech: **start** a turn (with a timestamp), **talk** in turn, **announce** a TRP (the pre-TRP cue), **arrive** at a TRP possibly **selecting** next-speaker, or make a **positive** or **negative feedback**.
- head/face: **nods** are a positive feedback, and a **puzzled expression** is a negative feedback.
- gesture: participants can **gesture**. The speaker can gesticulate throughout his turn, while listeners can indicate they wish to speak by gesturing before a TRP.
- posture: participants can **shift posture**. Listeners can shift posture when wanting to talk, or as they start a turn. As speaker, he can shift posture again when finishing his turn, a strong indicator that he has indeed finished.
- gaze (visual monitoring): participants can **look at one another** or **at no one**. Currently, listeners all look at the speaker with occasional glances away, and he looks at the previous speaker and then looks away, gazing back when finished planning what to say.
- listening (audio monitoring): participants **pay attention to the speaker** and the speaker to the others except when busy planning what to say, in which case he pays attention **to no one** (Butterworth, 1980).

Compatible feedback from the verbal and non-verbal modalities can occur simultaneously at TRPs[2]. Participants start to speak at a TRP, but occasionally can interrupt in the middle of a turn too. While simultaneous talk persists, as in multiple starts or at these interruptions, some or all of them can decide to stop. If everyone stops, then no one is speaking and everyone can again decide to start to speak.

3.1 Participants

Participants in the discussion are modelled as independent, autonomous agents defined by a set of constants that govern their behaviours probabilistically. Currently, the attributes we defined for each agent are:

talkativeness likelihood of wanting to talk;

transparency likelihood of producing explicit positive and negative feedbacks, and turn-claiming signals;

confidence likelihood of interrupting, and continuing to speak during simultaneous talk;

interactivity the mean length of turn segments between TRPs;

[1] The same signals used for feedback can also be used as *turn-precursors*: e.g. nodding before starting a turn.

[2] Clearly, non-verbal feedback alone is only effective when its producer holds speaker's gaze, but people give it even when they do not: e.g. in telephone conversation.

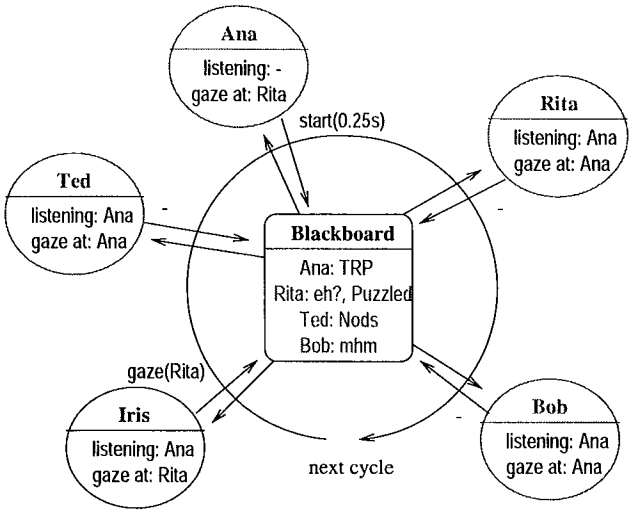


Figure 1: One cycle of conversation.

verbosity likelihood of continuing the turn after a TRP at which no one self-selected.

These values can be generated automatically for a group of agents at the start of the simulation (for example, based on a given mean and standard deviation), or set individually. They are a means of obtaining a variable pattern of behaviour for the group.

3.2 Architecture

In order to simulate independent agents performing sometimes simultaneous behaviours, the discussion is run in a loop with a clock *cycle* of an arbitrary length[3] during which agents read and write behaviours to a blackboard (figure 1). This represents the environment of the conversation: everything that is said and done. At each cycle, agents read the blackboard to *perceive* what has happened in the previous one. All behaviours in a cycle are meant to be "simultaneous."

With this framework, speaking turns and gesturing spread across several cycles. Group phenomena such as gaps and simultaneous talk at speaker transitions, too, can stretch for more than one cycle. Other, more instantaneous behaviours such as feedback, shifts of gaze and posture occur within only one cycle in the simulation.

[3]A good choice for its length is 500ms, or half-a-second. Thus turn starts, which vary in this range from the TRP (positively, or negatively in overlaps), can be contained in one such cycle.

3.3 Processing

At each cycle, an agent reads the behaviours from the blackboard, decides what to do and writes back the new behaviours. Decisions are all probabilistic, based on the likelihoods given by his attributes. The agent has some variables recording the discourse context, so that they know, for instance, when they are in the middle of a turn. His decisions in a cycle are as follows:

If (no one is speaking)
  test **talkativeness** to start to speak here
  if so, start with a random interval
    test **transparency** to shift posture.
If (listening to a single speaker)
  look at him; occasionally, look away
  if (read the pre-TRP)
    test **talkativeness** to decide to start
    if so, test **transparency**
      to make turn-claiming signals now
    mark next cycle as the TRP
  if (at a TRP *and* decided to start)
  or (at a TRP *and* was selected)
    start with a random interval
    test **transparency** to shift posture
  if (at a *free* TRP *and* not going to start)
    test **transparency** to do feedback
  if (anywhere else)
    test **talkativeness** *and* **confidence**
    to start to speak, i.e. to interrupt.
If (started simultaneously at a TRP)
  test **confidence** *and* who started first
    to decide whether to continue.
If (speaking simultaneously, *and* not planning)
  test **confidence** whether to continue.
If (speaking alone in a turn)
  use **interactivity** to set the segment length
  decide when to gesture and gaze-away
  decide when planning stage ends
    gaze back at interlocutor at that point
  if (at the last cycle before the TRP)
    decide whether to select next-speaker
  if (arrived at a TRP *and* no one started)
    test **verbosity** to continue talking.

Some explanation is in order. Turn-claiming signals are gesture, posture shift, nods and/or "mhm." They are decided individually according to the agent's transparency. In multiple starts, the agents test confidence reduced appropriately by their precedence in time with regard to the others. The only other combined

test uses talkativeness and confidence in deciding whether to interrupt.

Regarding the speaker, the segment length from TRP to TRP is based on his interactivity: the higher it is the shorter the segments. Other minor decisions could be naturally based on his verbosity: frequency of selecting next-speaker and the length of the planning stage. Still others, such as gestures and gaze-away, are fixed.

As for feedback, the higher the transparency attribute the more likely the agent is to make explicit signals, and in more than one modality at once. The simulation chooses negative cases in a fixed proportion. Feedback is only performed at a *free* TRP, i.e. when the speaker did not select-next[4]. This selection is indicated at the last cycle before the TRP. Only the selected agent, reading this, will decide to start at the TRP.

3.4 Example

Figure 2 presents a log of three turns in the simulation. For the sake of space, only three participants are shown, each with three columns:

- speech: with turn starts shown only by the timestamp from the TRP, and the pre-TRP cue indicated by pTRP;
- "body" behaviours: GESTures, POSTure shifts, NODS and PUZZled facial expressions;
- gaze: at other participants or at no one.

In this example, Iris begins to talk shifting posture and gesticulating in the middle. She passes a TRP being acknowledged by the others. At the next TRP, Rita slightly overlaps her (by 50ms). It is not possible to know whether Iris was intending to continue after the TRP, but since someone started, she stops. Ana also starts to speak at that TRP but gives in, especially since she began later than Iris. Iris continues with talk in turn, shortly ending at a TRP. Rita gives positive feedback at the TRP but Ana takes the turn perceivably later (more than one cycle away). At Ana's first TRP, Rita complains, causing her to start a new turn to address the problem. The timestamp of that

[4]Negative feedback certainly can occur when someone is selected, but this would mean the speaker has to make another select-next turn with the same addressee. We are leaving these contextual complexities for later.

Iris	(body,gaze)	Ana	(body,gaze)	Rita	(body,gaze)
	Rita		Rita		Ana
+.25s	POST Rita		Rita		Ana
talk	-		Iris		Iris
talk	GEST -		Iris		Iris
talk	GEST -		Rita		Iris
pTRP	GEST Rita		Iris		Iris
talk	Rita		Iris		Iris
TRP	Rita		NODS Iris	mhm	NODS Iris
talk	-		Iris		Iris
talk	-		Iris		Iris
pTRP	-		Iris		Iris
talk	-		Iris		Iris
TRP	Rita	+.30s	Iris	-.05s	GEST Iris
	Rita		Rita	talk	GEST Iris
	Rita		Rita	talk	-
	Rita		Rita	pTRP	-
	Rita		Rita	talk	Iris
mhm	NODS Rita		Rita	TRP	Iris
	Ana	+.25s	POST -		Iris
	Ana	talk	-		Ana
	Ana	pTRP	Rita		Ana
	Ana	talk	Rita		Ana
	Ana	TRP	Rita	eh?	PUZZ Ana
	Ana	+.10s	Rita		Ana
	:	:	:		:

Figure 2: Excerpt of three turns of conversation.

start is in this case relative to the TRP in her own talk. Again it is not possible to know whether Ana would have continued or stopped at that TRP had no negative feedback been produced. And so the simulation continues until it is stopped by the user or reaches an indicated number of turns.

4 Evaluation

The most usual type of formal evaluation requires a corpus against which properties of the simulation can be measured. Although there is some group corpus data available, it is still extremely limited. Corpora like that are extremely expensive to produce. Moreover, this is a new research area, and research concerning any behavioural phenomenon must begin with descriptive analysis, with quantitative corpus work following later. Our simulation synthesizes a substantial body of largely descriptive analysis pertaining to turn-taking in small groups, showing how the different behaviours work together. It is thus a preliminary step compared to work currently being undertaken in more mature areas.

We have access to two sources of corpus data. The first, the corpus used by (Fay et al., 2000), consists of group discussions among undergraduates. There are ten groups each of five and ten participants, each of which used a circumscribed scenario which limited the topics in-

volved and allowed them to determine the relationship among group size, discussion structure, and influence. Discussions were recorded on audiotape using two non-directional microphones. The second consists of two group discussions, one of size five and the other of size eight, using Fay's scenario but recorded to give better information for all modalities. These groups were audio-recorded on individual, synchronized tracks. The participants wore baseball hats marked with contrasting arrows, making their head direction and gestures apparent from a ceiling-mounted video recording.

Evaluation of a model such as the one underlying our simulation is performed by showing that the proposed model fits the real data better than some simpler, baseline model. This is the approach used by (Stasser and Taylor, 1991). In this case, the proposed model might, for instance, predispose the group to dyadic turn patterns by giving a starting advantage to the agent at whom the speaker gazes when approaching a TRP and making it likely for the speaker to gaze at the last speaker. Whether or not this elaboration is necessary would be tested by checking whether a model containing it can be made to fit the data better than one that excludes it. The properties of small group discussion which we feel it would be most useful for our model to explain are:

- the distribution of turns among speakers;
- dyadic patterning in turn sequences;
- turn and turn segment length;
- rate, location, and length of simultaneous speech and interruptions;
- placement and frequency of positive and negative feedback in the various modalities;
- placement and frequency of the explicit signals that a listener wishes to take a turn.

Clearly the amount of data that we have with clear recording of all modalities is limited enough to make evaluation difficult for some properties of the model. However, at this early stage, even measurement and synthesis of a coherent, first-pass model is a useful research aim. Our results could be used simply to inform upon the rate and rough placement of the behaviours we study. For many animators wishing to create naturalistic agents, this may be enough information to improve their results.

5 Conclusions

We described a simulation of small group discussion that attempts to reproduce patterns of turn-taking. Actual language contents and processing that drive a dialogue were ignored in order to focus on how turn-taking behaviours in a range of modalities constrain and shape the conversation, that is, those contents. This does not mean that the turn-taking drives the contents, but that the contents do not directly affect turn-taking behaviour.

This simulation improves upon previous work in a number of aspects. Simulations of group discussion such as (Stasser and Taylor, 1991) traditionally run in a central loop where all decisions are made; participants do not take them individually. In this work, they are independent agents capable of behaving simultaneously. They are modelled after a set of attributes that define complex individual 'personalities.' Turn-taking patterns thus emerge from the independent decisions they make. They perform a range of multi-modal behaviours that result in talk in turn, feedback and turn-claiming signals. They engage in various types of speaker transitions including interruption, overlap and selection-of-next. And finally they analyse the turn in progress for when to speak or to perform feedback, and whether others are doing this too, hence determining who takes the turn at the TRP.

The behaviours presented here are *placeholders* or abstractions for future elaboration of the various complex phenomena that surface in multi-party conversation. So they generalize a range of subtly different cases. In real conversations, an interjection like "oh!" or "ah!", for example, can have different meaning and function (as backchannel, turn-precursor, etc) depending on their context, intonation and facial expression accompanying it. Gestures may vary in type, scope and energy and even postures communicate moods and situations: e.g. slouching, stiff, twisted (Blatner, 2002). All these subtleties show how complex non-verbal communication is and how it contributes to the conversation.

We intend to proceed with some elaborations to improve the simulation. One first possibility is to replace explicit TRP indicators by pauses and changes in intonation, volume and rate of speech, thus making agents analyse the actual cues for turn-taking. This might be further enhanced by generation of simple sentence patterns for the contents of the conversation, containing topics, speech-act types and obligations. On top of this, various types of performance phenomena, like pre-starts, tag questions, hesitations, and self-interruptions, might be relevant once a model of the speaker is employed, such as the one proposed by Levelt (1989). All of this can further affect decisions on whether to speak and listen, whether to interrupt and continue simultaneous talk, to perform feedback, where to gaze, and so on.

This work is basic research. It is intended to inform about multi-party conversation, in particular small group discussion, by emulating multi-modal simultaneous behaviours of multiple independent agents in such setting. It could also inform practical work on, for instance, embodied conversational agents, more natural "barge-in" for spoken dialogue systems and to help coordination of video-conferences. One way in which it can help is by emulating the ways by which people anticipate when new utterances are expected (when someone is going to talk) by their behaviours while listening. Another is by generating appropriate behaviours to coordinate speaker transitions and keep the conversation flowing. Thus we expect to provide a first step in the development of better models of agents that can naturally interact in multi-party conversation.

References

M. Argyle and M. Cook. 1976. *Gaze and mutual gaze.* Cambridge University Press, Cambridge, MA.

G. Beattie. 1985. The threads of discourse and the web of interpersonal involvement. *Bulletin of the British Psychological Society*, 38:169–175.

Jonas Beskow, Kjell Elenius, and Scott McGlashan. 1997. OLGA - A dialogue system with an animated talking agent. In *Proceedings of the Eurospeech'97*, pages 1651–1654, Rhodes, Greece.

A. Blatner. 2002. *About nonverbal communications.* Available at `http://www.blatner.com/adam/level2/nverb1.htm`.

D. Boden. 1994. *The Business of Talk: Organizations in Action.* Blackwell, Oxford, England.

M. Bull and M. Aylett. 1998. An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. In *Proceedings of the ICSLP'98*, volume 4, pages 1175–1178, Sydney, Australia.

B. Butterworth. 1980. Some constraints on models of language production. In *Speech and talk*, volume 1 of *B. Butterworth (ed.) Language production*, London. Academic Press.

J. Carletta. 1992. *Risk-taking and Recovery in Task-oriented Dialogue.* Ph.D. thesis, University of Edinburgh, Edinburgh, Scotland.

J. Cassell, N. Badler C. Pelachaud, B. Achorn M. Steedman, B. Douville T. Becket, and M. Stone S. Prevost. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of the SIGGraph'94 Computer Graphics Annual Conference*, volume 1, pages 413–420, New Orleans.

J. Cassell, C. Sidner Y. Nakano, T. W. Bickmore, and C. Rich. 2001. Non-verbal cues for discourse structure. In *Proceedings of the 41st Annual Meeting of the ACL*, pages 106–115, Toulouse, France.

S. Duncan. 1972. Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, 23(2):283–293.

C. Edelsky. 1981. Who's got the floor? *Language and Society*, 10(3):383–421.

N. Fay, S. Garrod, and J. Carletta. 2000. Group discussion as interactive dialogue or serial monologue: the influence of group size. *Psychological Science*, 11(6):487–492.

C. I. Guinn. 1996. Mechanisms for mixed-initiative human-computer collaborative discourse. In A. Joshi and M. Palmer, editors, *Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics*, pages 278–285, San Francisco, USA.

G. Houghton and S. Isard. 1987. Why to speak, what to say and how to say it: modelling lan-

guage production in discourse. In *P. Morris (ed.) Modelling Cognition*, pages 249–267.

A. Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.

W. J. M. Levelt. 1989. *Speaking: from Intention to Articulation.* MIT Press, Boston.

D. McNeill. 1992. *Hand and Mind: What gestures reveal about thought.* The University of Chicago Press, Chicago.

L. Meltzer, W. Morris, and D. Hayes. 1971. Interruption outcomes and vocal amplitude: Explorations in social psycho-physics. *Journal of Personality and Social Psychology*, 18:392–402.

D. G. Novick, B. Hansen, and K. Ward. 1996. Coordinating turn-taking with gaze. In *Proceedings of the ICSLP'96*, volume 3, pages 1888–1891, Philadelphia.

B. Oreström. 1983. *Turn-taking in English Conversation.* Lund Studies in English. LiberFörlag Lund, Malmö, Sweden.

R. Power. 1979. The organization of purposeful dialogues. *Linguistics*, 17:107–152.

J. Rickel and W. Johnson. 2000. Task-oriented collaboration with embodied agents in virtual worlds. In *Cassell, J., J. Sullivan, S. Prevost et al. (eds.) Virtual Worlds, in Embodied Conversational Agents*, Boston. MIT Press.

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

D. Schaffer. 1983. The role of intonation as a cue to turn-taking in conversation. *Journal of Phonetics*, 11:243–259.

A. E. Scheflen. 1972. *Body language and the social order.* Prentice Hall, Englewood Cliffs, NJ.

G. Stasser and L. A. Taylor. 1991. Speaking turns in face-to-face discussions. *Journal of Personality and Social Psychology*, 60(5):675–684.

M. A. Walker. 1994. Discourse and deliberation: testing a collaborative strategy. In *Proceedings of the 15th International Conference on Computational Linguistics*, pages 1205–1211, Kyoto, Japan.

Semantic Negotiation: Modelling Ambiguity in Dialogue

Alison Pease Simon Colton
Alan Smaill
University of Edinburgh, Division of Informatics,
80 South Bridge, Edinburgh, EH1 1HN, Scotland

John Lee
University of Edinburgh, Division of Informatics,
2 Buccleuch Place, Edinburgh, EH8 9LW, Scotland

email: {alisonp@dai, simonco@dai, smaill@dcs, john@cogsci}.ed.ac.uk

Abstract

We argue that negotiation over the meaning of terms in a statement is part of human discussion and that it can lead to richer theories. We describe our preliminary model of semantic negotiation and discuss theoretical examples which we hope to implement. Finally we consider how semantic negotiation fits into existing work on argumentation.

1 Introduction

There many situations in which participants realise partway through an argument that they are interpreting one of the key terms differently. The meaning of the key term is then called into question, and the focus of the argument switches from the truth or acceptability of a claim or offer to the *meaning* of the term. This is particularly common in subjects such as philosophy, law and politics, in which persuasive reasoning is all important and concept definitions are modified according to the proponents' goals. Yet this phenomenon is rarely seen in AI research. Economic agents may well disagree on the price of a potato, even haggle over it in a reasonably sophisticated way – but they never start arguing about what a potato is.

In this paper, we argue that:
1) people sometimes define concepts in a way which supports their beliefs or goals (§2.1), and
2) subsequent disagreement over the meaning of a concept can result in a richer theory (§2.2).

We are currently modelling this phenomenon, (described in §4), with the aim of (a) elucidating it and (b) using it to extend existing theory formation programs. The fact that although participants in a discussion use a shared language, some of the terms are ambiguous, raises questions like – what sorts of things can be ambiguous? How might ambiguity arise? How can it be resolved? Can it be used to produce richer theories? By modelling *semantic negotiation* – the problem of reaching mutually acceptable definitions (a specialised type of *negotiation*, defined by the agent community as the problem of reaching mutually acceptable agreements) – we hope to address these questions.

We draw on work by Jennings et al. (1998) (described in §3), and in §5 place our ideas in the context of previous work on argumentation.

2 Semantic negotiation and its value

2.1 Ambiguity in human reasoning

The failure at the beginning of the last century of the quest for a perfect language in which neither ambiguities, nor paradoxes or redundancies exist, showed the difficulties involved in writing a formal language which can be used to describe a reasonably large domain. Today it is normally accepted that any non-trivial language is likely to contain ambiguities. Different types of ambiguity include *lexical* (where a word has two different meanings); *syntactic* (a sentence with two syntactically correct derivation trees which indicate different meanings); *semantic* (a sentence with two meanings, only one of which makes sense), and *pragmatic* (the meaning of a word is relative to the speaker). Much work on AI and am-

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 125-132.

biguity is concerned with methods to automatically determine a writer's intended meaning (for example Romacker and Hahn (2001) who focus on representing and managing ambiguity in natural language text understanding). In contrast, we are concerned with ways in which ambiguity may be exploited (or introduced into a previously unambiguous concept), in order to support an argument or set of beliefs.

Examples in which ambiguity is used to support an argument include an insurance company which argues that a house damaged in a hurricane is not covered by the owner's accident policy, as a hurricane is an 'act of God' rather than an accident. Similarly, a lawyer may argue that a client who jumped a red light while rushing his wife to a maternity ward is not guilty of reckless driving. In the 1990's the European Union Food Standards discussed the definition of chocolate. The minimum cocoa content which a substance must contain in order to be called chocolate was debated, as countries which produce it with a higher cocoa content did not want their chocolate to be confused with products with a lower cocoa content. Finally, consider the recent controversy over the meaning of 'prisoners of war'. When challenged that their treatment of the Taliban prisoners violated the Geneva Convention – that all prisoners of war should be treated humanely – the American government argued that the prisoners were not 'prisoners of war', but 'battlefield detainees'[1]. In these examples, the terms *accident*, *reckless driving*, *chocolate* and *prisoners of war* are defined by each party in such a way as to aid their argument.

2.2 Using semantic negotiation to enrich a theory

We discuss three examples of semantic negotiation in mathematics, and show how it has enriched mathematical theories.

The concept 'polyhedron'

Lakatos (1976) shows both *that* ambiguity exists even in mathematics, and – of key importance to AI researchers – he shows *how* it might arise and how it can be used effectively. He presents a dialogue between a group of students and their teacher, in which the history of Euler's conjecture and its proof is enacted as a rational reconstruction. As well as providing a rare insight into the way in which theories evolve (where by theory we mean concepts, counterexamples, conjectures and 'proofs' or arguments), this work is of great value to researchers modelling dialogue. Argumentation moves are set out which Lakatos categorises into various methods.

Euler's conjecture states that for all polyhedra, the number of vertices (V) minus the number of edges (E) plus the number of faces (F) equals 2. Starting from this conjecture, one of the students finds the counterexample of the hollow cube – a cube with a cube-shaped hole in it (figure 1).

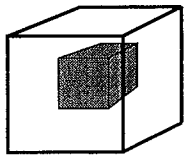


Figure 1: The hollow cube; $V - E + F = 16 - 24 + 12 = 4$

One reaction to this counterexample is to question the meaning of the concept 'polyhedron'. Rather than accept it as a counterexample, it is branded a *monster* since, it is claimed, it is not a polyhedron. Participants in the discussion then argue about the definition of polyhedron. Lakatos (1976) calls this process of arguing that a counterexample is not valid and therefore not a threat to a conjecture, the method of *monster-barring*. Using this method repeatedly participants expand the theory of polyhedra, differentiating between 'a solid whose surface consists of polygonal faces' to 'a surface consisting of a system of polygons' (thus excluding the hollow cube), to 'a system of polygons arranged in such a way that (1) exactly two polygons meet at every edge and (2) it is possible to get from the inside of any polygon to the inside of any other polygon by a route which never crosses any edge at a vertex'. They also discuss the meaning of the concepts polygon, area and edge. As a result definitions are tightened and students gain a greater understanding of the conjecture.

The concept 'number'

In the late 19th century Cantor introduced the mathematical community to the 'number' $\aleph_0$, which is the size of the set of all integers (the first transfinite number). This was a counterexample to the conjecture that if you add a non-zero number to another number, then the second number always changes, since $\aleph_0 + \aleph_0 = \aleph_0$. Similarly it violates the conjecture that any positive number multiplied by 2 is bigger than the number, as $\aleph_0.2 = \aleph_0$. The

[1]The definition of *humane treatment* was also disputed – in particular whether it could ever include interrogation, as the American government felt it important to interrogate the prisoners while not wanting to be open to the charge of inhumane treatment.

law of monotonicity, that for all numbers a, b and c, if $b < c$, then $a + b < a + c$ fails if $a = \aleph_0$ (for any finite b and c). For these and other reasons, initial reaction to Cantor's work was hostile, $\aleph_0$ branded a monster (to use Lakatos' terms), and barred from the concept of number. However Cantor was developing a whole area of mathematics which he considered to be interesting and worthwhile – which included the number $\aleph_0$. Therefore he continued his research and tried to convince the mathematical community (eventually with success) that transfinite numbers *are* a kind of number and a valid area of mathematics. This process of suggesting that a new sort of object *is* a number, initial denial and later acceptance when it proves its worth can be seen repeatedly in number theory. For instance the number zero was initially branded a monster by the Greeks, for various reasons including its violation of the conjecture that if you add a number to another number, then the second number always changes. When zero was accepted this conjecture was modified to that above (*i.e.*, excluding zero). Other examples of ambiguity in the concept of number include initial barring of 1 (barred by the Pythagoreans as it challenged their belief that all numbers increase other numbers by multiplication); $\sqrt{2}$ (it violated the Greek belief that all numbers describe a collection of objects); and $x = \sqrt{-1}$ (violating the law of trichotomy, for any numbers x,y, either $x = y$ or $x < y$ or $x > y$). Now of course 0, 1, irrational and imaginary numbers are accepted without question, and the concept of number has been generalised to complex numbers and beyond (quarternions). Clearly number theory (and other areas of mathematics) have been greatly enhanced by these additions.

The concept 'prime'

Another example in number theory is the definition of prime number. A prime was initially defined to be 'a natural number which is only divisible by itself and 1'. However this definition includes the number 1, which was found to be a counterexample to many theorems and conjectures about primes. For instance, the Fundamental Theorem of Arithmetic (FTA) states that every natural number is either prime or can be expressed uniquely as a product of primes. If 1 is also considered prime then this violates the uniqueness claim, since, for example, $6 = 2*3 = 2*3*1 = 2*3*1*1 = 2*3*1*1*1 = ...$ Rather than explicitly exclude the prime 1 from this (and other) theorems, it is preferable to exclude it from the concept definition, and today the accepted definition of a prime is 'a natural number with exactly two divisors'. This change in the concept definition has enabled many theorems about primes, including the FTA, to be neatly stated.

2.3 Properties of semantic negotiation

By extracting some general properties from the examples above, we can begin to answer the questions we asked in §1, and to see how we might model this phenomenon.

What sorts of things can be ambiguous?

Some ambiguous entities are (sub)concepts within a universe of objects (such as a prime), and some are the main concept, *i.e.*, the universe itself (such as a polyhedron, or number).

How might ambiguity arise?

Some concepts are initially specifically defined (everyone agrees on the definition), and the definition changed (someone argues that a second definition would be more useful). Others are initially vague (it is not known whether some objects are examples of the concept or not) and only when disagreement arises are different concept definitions made explicit.

How can ambiguity be resolved?

Experts evaluate the worth of each of the rival definitions (often with different results). The existing definition is assumed by default, with the onus on proponents of a new definition to convince the other experts of its value. Grounds for accepting the new definition include showing that it produces interesting new theories or results (including preserving the 'truth' of a faulty conjecture). There is usually a period during which it is unclear whether the object in question belongs to a concept or not. It passes through a period of indefinite status with some people accepting its status, others not, others unsure, until it either proves its worth and is generally accepted or fails to convince enough people and gradually disappears.

Lakatos (1976) does not explore reasons for choosing one concept definition over another. Instead the teacher in the dialogue simply asks everyone to accept the strictest, *i.e.*, most limited definition suggested so far (at least for the duration of the discussion). However the only clear end to this process is a tautology (hence a student's sarcastic suggestion that a polyhedron be defined as 'a system of polygons for which the equation $V-E+F = 2$ holds' – (Lakatos, 1976, p. 16). Dunmore (1992) suggests that concept definitions are chosen and developed

according to their use. For instance a concept which is not associated with any interesting conjectures is unlikely to become well known and accepted within the mathematical community.

3 Argumentation-based negotiation

Jennings et al. (1998) emphasise the importance of negotiation in multi-agent research, and outline an informal framework describing its key features. They divide negotiation issues into *protocols* (rules which govern interaction), *objects* (the range of issues over which agreement must be reached) and *Agents' Decision Making Model* (the way in which an agent follows the protocol to achieve its objectives).

Agents move through the space of possible agreements (in our case all the candidate definitions for a concept), defining their own spaces of acceptable points. Negotiation works by agents suggesting points in the space which are acceptable, and evaluating each point suggested. This ranking may change during negotiation, as agents are persuaded that a point is valid. The way in which they rate points may also change. A minimal requirement is that agents be able to propose some part of the agreement space as acceptable, and can respond to other agents' suggestions. A more efficient model would give agents capability to explain *why* they are rejecting/proposing a certain point. This might include rejecting a proposal but stating which aspects were considered good, a *critique*, or making a *counter-proposal* in response to a proposal. Such a model might include *justifications* – in which the agent states its reasons for making a proposal, or *persuasion* – in which an agent tries to change another's agreement space or rating over the space. These arguments help to support an agent's stance.

Jennings et al. (1998) state that an agent capable of argumentation-based negotiation must have a mechanism for:

- communicating proposals and supporting arguments;
- generating proposals;
- assessing proposals and arguments; and
- responding to proposals.

They do not suggest that the meaning of a concept could be a possible object over which agents negotiate, giving as examples issues relating to negotiation over services or products. However, the framework is general enough to include this type of negotiation. For instance, once the disagreement over a concept definition has arisen, an agent might suggest one and justify why it is good (for example it may be used in many of its conjectures) - which may lead another agent to re-evaluate the definition and rate it more highly.

4 A preliminary model of semantic negotiation

We are currently extending the HR system (Colton, 2001) to perform semantic negotiation. In this section we briefly outline the original system, as well as the extended version. We then describe two theoretical examples of semantic negotiation and ideas on their implementation in HR.

4.1 The HR system

HR is an automated theory formation program (Colton, 2001) which is given background information about a domain, including some objects of interest (the 'universe') – such as integers – and some initial concepts – such as multiplication and addition. It forms new concepts by using one of 10 general production rules to transform one (or two) old concepts into a new one. The examples of a concept are used to make conjectures empirically.

HR is able to evaluate concepts and conjectures based on various interestingness criteria, which measure values such as the *novelty* of a concept, the *number of open or true conjectures which the concept is in*, and the *surprisingness* of a conjecture.

4.2 Extensions to HR

In order to model semantic negotiation (and as part of a project to model Lakatos-style reasoning) we have partially implemented an agent architecture in which multiple copies of HR are run which can communicate details of their theory to each other. These consist of 'students' and 'teacher', and all have different weightings of interestingness values, for example one may favour novel concepts while another prefers surprising.

Lakatos (1976) identifies various reactions to a counterexample to a conjecture, and in §2.2 we described one reaction – the method of monster-barring. Another reaction may be to to modify the

conjecture by barring the counterexample. For instance, we may generalise from the hollow cube to 'polyhedra with cavities', and then modify Euler's conjecture to *for all polyhedra without cavities, V - E + F = 2.* This is the method of exception-barring, which we have implemented in the extended version of HR, and expect to use in our model of semantic negotiation.

Using the reflection mechanism available in Java – which enables information about classes and individual objects to be obtained at run-time – we have implemented a cut-down Java interpreter in HR, which can be used to interpret and execute Java code at run-time. Currently, the interpreter can handle the creation of new objects and various constructs, including for loops, if-then-else statements and string manipulations. We intend to extend the functionality of the interpreter to handle more complicated Java code.

Having a run-time interpreter has various advantages, including:

- homogenuity in the code used to control HR at compile time and at run-time;
- easily compiling pieces of run-time code to become compiled parts of HR, which improves efficiency;
- introspective access to the main 'theory' object which HR has built, which enables it to run any methods attached to that object, and to build and execute new functions at run-time.

Various aspects of the way HR forms and presents theories take advantage of the interpreter, including the way it reports theories to the user, and the reaction mechanism, which is a particular heuristic search whereby HR reacts to certain events in the theory formation by taking various additional theory formation steps. The reaction scripts are written in Java and describe which events to react to, and how to react.

In terms of the project discussed here, HR's Java interpreting functionality can be used to avoid an inter-lingua in the communication between agents. That is, the ways in which the various agents instruct, inform and control each other do not have to be prescribed before each run is undertaken. This will enable a more flexible communication between the agents.

4.3 Two theoretical examples

Suppose that *alpha*, *beta* and *gamma* are different versions of HR which have been running independently for a time. We are aiming to generate the sort of dialogue shown here. Note that, as described above, they would communicate in java code rather than natural language – here we paraphrase to aid the example. Superscript numbers indicate places where we discuss ideas for implementing aspects of the discussion.

Example one: proposing to exclude an existing object

alpha: I've noticed that the object o is a counterexample to many of my conjectures about concept C (C may be the universe or a concept)[1]. It would be neater to bar it from the concept rather than my conjectures. What do you all think?[2]. Does anyone have a neat definition which includes exactly those objects in C except for o?
beta: It's a counterexample to many of mine as well. Let's bar it.
gamma: I disagree. It does not violate any of my conjectures. For which conjectures of yours is it a problem?
alpha: For conjectures X, Y and Z.
beta: Also conjectures U, V and W.
gamma: OK these are interesting conjectures. Let's revise our concept definition then[3]. I have a concept C_1 which includes the same objects as C except o. Let's use this from now on[4]

Implementing the dialogue

(1) If an agent has many conjectures of the type $\forall x \neq n, P(x) \rightarrow Q(x)$ in its theory then it may propose a concept definition change. Whether the object in question is to be excluded from the universe or from a concept depends on whether the problem conjectures all involve the *same* concept (in which case exclude the object from this concept), or different concepts (in which case exclude it from the universe).
(2) We now have an ambiguity since it is unclear whether a concept includes a given object or not. In order to decide, we need *(i)* a way for individual agents to determine their position, and *(ii)* a negotiation protocol to determine how agreement is reached. We discuss these below.

(3) any agent can look in its theory to see if it has a concept which covers certain objects.
(4) HR has a mechanism – replace definition – which is currently used to overwrite old definitions if a simpler one is found. The more complex definition is still in the theory (and in all the conjectures prior to the new definition), but is no longer refered to. We can use this mechanism to replace the old definition

with a new one.

(i) Deciding individually whether to exclude an object

Currently the interestingness measures in HR evaluate concepts and conjectures. We can extend these to evaluate how interesting an object is. For instance, we may measure:

- generality of conjectures – the number of conjectures in the theory which involve the concept in question, which the object does *not* violate (the more general a conjecture is, the more valuable);

- counterexample-barred conjectures – the inverse of the number of counterexample-barred conjectures in the theory which involve the concept in question to which the object is a counterexample.

- broken conjectures – the number of conjectures in the theory which are violated by the object.

Of two rival definitions, an agent might prefer the simplest. Colton has already implemented this in HR. It arises when a conjecture that two concepts are equivalent has been proven. HR then evaluates them both to determine the simplest, where simple is defined as the number of production rules used to generate a concept (the fewer the simpler). The user can instruct HR to always keep the simplest definition, with the more complicated definition only appearing in equivalent conjecture statements. An agent might also judge which of two rival concepts is more interesting according to its interestingness criteria already present.

Using these (and other) measures, an agent can evaluate and respond to a proposal to exclude an object – which it may or may not already have in its theory. Clearly, these values will be different for each agent, as they will have a different weighting of interestingness values, so would evaluate the same object or concept differently. Additionally they have different theories – *i.e.*, different conjectures and concepts against which the object or concept is measured. As with other interestingness measures in HR, we anticipate making these flexible (a weighted combination input by the user) and then experimenting to see which combination is the most productive.

(ii) A negotiation protocol

Following Jennings et al. (1998), the space of all possible agreements is all candidate definitions for a concept. Agents define their own space of acceptable points by evaluating their own candidate definitions (those which score the highest according to their interestingness values). Negotiation begins by the agents suggesting points in the space which are acceptable, and evaluating each received definition (which may score more highly than one which they have generated themselves). This constitutes the simple model described in (Jennings et al., 1998). We would then want to enable agents to communicate the reasons they reject or propose a definition, for instance "I reject definition D because it excludes the number 1 and I need this number for all my proofs", or "I propose definition D because it would preserve conjecture C, which is an interesting one". Definitions could then be re-evaluated, for example if the agent receiving the latter proposal also evaluates conjecture C as interesting, it might be persuaded that definition D is more interesting than the one it previously proposed.

The agents re-evaluate their measures when others present reasons for/against accepting a concept definition.

This dialogue demonstrates how we may automate ambiguity in both the universe of objects and (sub)concepts, a period of indefinite status while it is discussed and the sorts of arguments which might be made for changing to a new definition or sticking to an existing one, and methods by which everyone might agree on a new definition and then change to it.

A similar dialogue can be envisaged in which a proposal to accept an object, *i.e.*, to *widen* rather than narrow a concept definition to include rather than exclude it, is made. If agreement is not reached then the teacher could decide, based on a weighting of the results of its own interestingness criteria and those of the students'. It may also put more weight on the existing definition (since it is costly to revise definitions). The students would then use the specified definition, at least for the duration of the discussion.

Example two: object-driven concept formation

Suppose that a conjecture has been proposed, which is true for all numbers between 1 and 100 except 1, 17 and 72. We may wish to find a concept to cover these counterexamples (and possibly others which lie outside the range), *i.e.*, a 'concept-to-exclude'. Therefore the teacher may ask the students to find such a concept. Each would then look in its theory and send a specific (possibly different) concept back. The 'concept-to-exclude' is now ambiguous, and agents rate the definitions and negotiate over which is the best.

This example encompasses:

1) an initially vague concept – it is not known

whether some objects are examples or not;
2) the vague concept developed independently into (possibly) different explicit concepts;
3) each agent suggesting arguments for accepting its concept.

5 The 'fallacy' of ambiguity – where does it fit into argument analysis?

In order to place our work in context, we conclude by briefly describing approaches to argumentation and ambiguity, and noting our position towards these perspectives.

Aristotle (1957, 1955, 1976) identified three motivations behind arguments; *apodiactic*, *dialectic* and *rhetoric*, in which certainty, a general acceptance and convincing an audience is respectively sought. Although many would claim that motivation behind mathematical argument falls into the first category, the second (or even third) is more appropriate to mathematics as Lakatos (1976) describes it. Aristotle treated dialectical argument as a game between a defender and attacker, and suggested guidelines such as forcing the defender to contradict herself, state an untruth or paradox, or a defend a circular argument, for conducting the debate. Certain moves – called fallacies – were disallowed, including the fallacy of ambiguity.

In his work on controversy, Crawshay-Williams (1957) emphasised the need for clarification of concepts prior to discussion. He claimed that if participants in a discussion agree upon the criteria under which a statement will be tested, then agreement regarding its absolute/probable/indeterminate truth will soon be reached. He calls one such criterion *conventional*, to mean the condition that participants agree on the meaning of terms (the others are *logical*, in which inference rules must be agreed, and *empirical*, in which facts and their contextual description should also be agreed).

Naess (1953) also stated that criteria for the verification or falsification of a statement are essential (to the extent that if no such criteria are found then discussion should be abandoned). However he includes agreeing on terms as a stage in the discussion, rather than a pre-requisite to it. The three stages in resolving a discussion, he suggests, are interpretation, clarification and argumentation. For any statement T there is set of possible interpretations of T, and participants must agree on which interpretation they wish to discuss. He claims that *precizating* statements, (being more precise) helps to eliminate misunderstandings, where U is more precise than T if any interpretations of U are also interpretations of T, but there are interpretations of T which are not interpretations of U. This is useful only if the disagreement has occurred through different interpretations, he does not advocate continual precization of statements since discussion would be practically impossible. Disagreements rooted in misunderstandings are termed *verbal*, only if after there is still disagreement after precization is it *real*. In the case of a real disagreement the evidence is weighed up to see which of the two statements is more acceptable.

Carbogim et al. (2000) present a survey of issues which can be handled by automated argumentation systems and suggest directions for future research. They consider the generation and evaluation of arguments, including issues such as drawing conclusions from an incomplete or inconsistent knowledge base, decision making under uncertainty and multi-agent negotiation systems. Argument about the meaning of terms used is not considered either explicitly in the text, nor in any of the examples, although it may turn out that semantic negotiation fits into one of the frameworks outlined.

Our view is that ambiguity plays an important role *within* a discussion. That is, questioning the meaning of terms in a discussion *is* a valid strategy (with restrictions on which words may be questioned). We differ from Crawshay-Williams' approach in that we see debate of terms as an important part of discussion rather than a pre-requisite of it. Participants may not realise initially that they have different interpretations of a word, indeed they may not themselves have a clear interpretation. We also differ from Naess' linear approach, disagreement over terms could arise at any point in a discussion. Of particular interest in Naess' work is the *evidence* used to resolve 'real' disagreements, and we hope by implementing semantic negotiation to elucidate the sort of evidence required. While the survey by Carbogim et al. (2000) is not intended to be exhaustive, it does cover the central issues. The omission of semantic negotiation suggests that it is either irrelevant to argumentation or is a new direction which has received little attention. We hold that it is the latter.

6 Conclusion

We have argued that negotiation over the meaning of terms in a statement is part of human argument

and that it can lead to richer theories. We have also described our preliminary model of semantic negotiation and discussed theoretical examples which we hope to implement. Finally we have considered how it fits into existing work on argumentation. We consider this phenomenon to be a fertile and relevant area of research in the fields of argumentation and agent-based negotiation.

Our goal now is to extend the HR system (Colton, 2001) to model semantic negotiation. Issues which we expect to address in include:

- how one definition is chosen over another. In Lakatos (1976) the narrowest definition is chosen. Other criteria may be the most common definition, or the most interesting;

- how are different interpretations best represented?

- what appropriate argumentation forms are there for this form of negotiation?

- to what extent is it useful to question word meaning?

We also intend to apply the Lakatos-style reasoning enabled by the agency to machine learning problems. In particular, we hope the agency could take over where machine learning programs currently finish. For example, suppose a program has learned a way of predicting whether chemicals are carcinogenetic, based on their molecular structure. If the prediction has a success rate of, say, 85%, then 15% of the chemicals break the prediction rule. If we collect these together as the examples of a new concept, which we define as: "The set of chemicals which break the prediction rule", then we have a suitably ambiguous concept to which the agency can apply their negotiation techniques. If the agency could identify a definition for this concept, then it is possible that the additional information could be incorporated into the prediction rule itself to increase accuracy, or at least provide a better understanding of the domain.

Acknowledgements

We would like to thank the two anonymous reviewers for some very helpful comments on a previous draft. This work was supported by EPSRC grants GR/M45030 and GR/M98012. The second author is also affiliated with the Department of Computer Science, University of York.

References

Aristotle (1955). *On sophistical refutations.* Harvard University Press, USA.

Aristotle (1957). *Aristotle's prior and posterior analytics: a revised text.* Clarendon Press, Oxford.

Aristotle (1976). *Rhetoric.* Berolini.

Carbogim, D., Robertson, D., and Lee, J. (2000). Argument-based applications to knowledge engineering. *The Knowledge Engineering Review.*

Colton, S. (2001). *Automated Theory Formation in Pure Mathematics.* PhD thesis, Dept. of Artificial Intelligence, University of Edinburgh.

Crawshay-Williams, R. (1957). *Methods of Criteria of reasoning: An Inquiry into the Structure of Controversy.* London: Routledge and Kegan Paul.

Dunmore, C. (1992). Meta-level revolutions in mathematics. In Gillies, D., editor, *Revolutions in Mathematics*, pages 209–225. Clarendon Press, Oxford.

Jennings, N. R., Parsons, S., Noriega, P., and Sierra, C. (1998). On argumentation-based negotiation. In *Proceedings of the International Workshop on Multi-Agent Systems*, Boston, USA.

Lakatos, I. (1976). *Proofs and Refutations.* CUP, Cambridge, UK.

Naess, A. (1953). *Interpretation and Preciseness: A Contribution to the theory of Communication.* Oslo: Skrifter utgitt ar der norske videnskaps academie.

Romacker, M. and Hahn, U. (2001). Context-based ambiguity management for natural language processing. In *Proceedings of the Third International Conference on Modelling and Using Context*, pages 184 – 197, Dundee, Scotland.

Modeling the common ground: the relevance of copular questions.

Orin Percus
University of Florence/ University of Tübingen

Abstract. Stalnaker 1978 and much subsequent work argues that we should model the common ground as a set of possible worlds, but there are different ways in which one might imagine doing so. This paper asks which of these ways is most appropriate when it comes to explaining how language works. The paper looks at a certain restriction on the use of copular questions, and suggests that accounting for this restriction is easier given one way of modeling the common ground than given others. Unfortunately, space considerations will force me to omit many details.

1. A question about language.

Copular questions of the kind in (1), where the WH-expression originates in subject position, cannot ask what function John performs.

(1) $\text{Who}_i\ \text{is}_j\ [_{IP}\ t_i\ t_j\ \text{John}\ ]$

(structure before wh-movement and auxiliary raising: $[_{IP}$ **who is John** $]$)

Some evidence for this comes from sentences like those in (2), where -- thanks to the fact that *is* does not move in the embedded clause (cf. Higgins 1973) -- it is easy to see that the WH-expression originates in subject position. To see that (2a) cannot ask what function John performs, note that, on *Scenario I*, I cannot use the question in (2a) to ask you what instrument John plays in the trio. (By contrast, (3a), where the WH-expression originates in *object* position, can be used for this purpose.) What *can* a sentence like (1)/(2a) ask? It can ask which of the people in front of us got introduced as John. On *Scenario I*, this kind of question is infelicitous -- the answer is already common knowledge --but on *Scenario II*, where the answer is not common knowledge, it is clear that the question can have this use.

The question I will ask in this paper is: How should this restriction on the use of (1) be described, and why does it hold?

(2) a. Who (do you think) is John?
b. May I ask [who is John]

▲
structure of the embedded clause
before wh-movement: $[_{IP}$ **who is John** $]$

cannot ask what function John performs

(3) a. Who (do you think) John is ?
b. May I ask [who John is]

▲
structure of the embedded clause
before wh-movement: $[_{IP}$ **John is who** $]$

can ask what function John performs

(4) Scenario I. When we arrive at the piano trio concert, a friend of ours tells us that he is going to introduce us to a couple of the musicians. He brings over two men in tuxedos and introduces one as John (the one on the left), and the other (the one on the right) as Bill. At this point, we know that John is one of the musicians, but we don't know whether he is the cellist, violinist or pianist. Likewise for Bill.

(5) Scenario II: We weren't paying attention, and our backs were turned when the introductions were made. When we turn around, we see those two people in front of us, but we don't know which one got introduced to us as John and which one got introduced to us as Bill.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 133-140.

2. A question about information states.

That was a question about language. Here is a different kind of problem. I want to consider our first problem in the context of this second one.

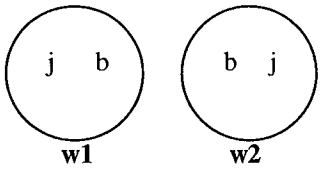
Suppose we assume that the right way to model the common ground is as a set of possible worlds, and assume (pace Lewis) that individuals exist across worlds. Now recall our second scenario, the one in which our backs were turned at the point when introductions were made. The problem is: how should we model the common ground that the scenario leads to? At the point when I ask you the question in (1), what kinds of worlds are in our context set?

Here are two ideas that we might entertain.

2.1. Idea A.

On Idea A, there are two kinds of worlds in our "context set." In some, on the left is a certain individual *j* who exists throughout our possible worlds, and on the right is a certain individual *b* who exists throughout our possible worlds. In others, it's the other way around. On this view, to talk about John is to talk about *j* and to talk about Bill is to talk about *b*.

(6)



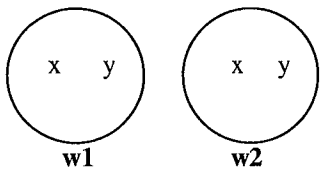
(7) The guy on the left is John.

Suppose we hear a sentence, like (7), that informs us that the person on the left is John. On this view, if we agree to take the sentence to be true, we reduce our context set, leaving in only one kind of world, the kind where the person on the left is *j*. That is, we eliminate worlds like **w2** from our set of candidates for the actual world.

2.2. Idea B.

Idea B assumes that to talk about John, or Bill, is basically to talk about a property – an individual might have the "John" property in one world but not in another. Here, in all worlds in the "context set," the same individuals *x* and *y* are to the left and to the right. The difference is that in some worlds *x* has the "John" property and *y* the "Bill" property, and in others it's the other way around.

(8)



w1: *x* has the "John" property, *y* has the "Bill" property
w2: *x* has the "Bill" property, *y* has the "John" property

On Idea B, when we hear a sentence like (7), we reduce our context set so as to admit only the kind of world in which *x* is the one with the *"John"*-property. We eliminate **w2**-worlds.

The mentality behind Idea B is that names basically encode particularly useful kinds of properties that allow us to maintain a repertoire of those individuals that are important to us. In many cases, the person who has the "John" property in one candidate world will be the person who has the "John" property in others. But cases like our scenario, where we do not know which of two perceptually accessible individuals "is John," will allow for other possibilities.

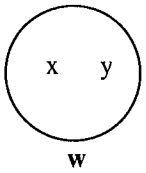
What I will argue now is that the idea we adopt about how to model the common ground will determine how we look at the problem we started out with – the problem of why a question of the form in (1) can convey some things but not others. Interestingly, one idea about how to model the common ground, because it encourages us to look at the problem in a particular way, could make a direction for solution more evident than would another idea. In that case, we would have a reason to favor one idea over the other. I will suggest that this is indeed the case.

(2.3. Digression: An "intermediate idea."

Idea A and Idea B are not the only views that one might entertain about how to model the common ground. There is also a somewhat intermediate idea, which I will sketch now and then neglect, but return to at the end (although in a certain sense it departs slightly from our initial assumptions).

This "intermediate idea" shares with Idea B the view that in all of our possible worlds the same individual – call him x – is on the left, and the same individual -- call him y -- is on the right. And it shares with Idea A the view that names are meant to talk about individuals. But there is an important difference between the intermediate idea and our other ideas. Idea A and Idea B assume that the parties to conversation have decided what individual, or property, it is that a name like John evokes, and therefore that sentences with the name *John* impart some kind of information to them. The intermediate idea assumes instead that the parties to conversation have not resolved what individual to associate with the name "John," and if they were to hear a sentence like *John is happy*, they would not necessarily be able to determine a way in which to reduce the context set.

(9)



A proponent of this intermediate view would look at our scenario as a case in which communication does not proceed quite as smoothly as it could. The idea would be roughly as follows. At the outset, since we do not know that to talk about John is to talk about x, there are a lot of sentences with the name "John" that we could not use to narrow down the worlds in our candidate set – including the phrase that accompanied the introduction, say *I present you John*. The principal effect of a sentence like *The guy on the left is John* ((7)) is then not to eliminate worlds from our candidate set, but just to let us know that there is a certain way of talking about the worlds in our candidate set – that to talk about John is to talk about x. Once we know this, we are then in a position to use the information provided by other phrases, like *I present you John*, and to narrow down our candidates to those worlds in which we have been introduced to x.

In other words, unlike on Idea A and Idea B, on the intermediate idea there is no sense in which the worlds in the context set divide up into two classes according to who is John. And an assertion like *The guy on the left is John* is not designed to get us to eliminate worlds from our context set – in fact, it does not directly cause the elimination of any. Rather, it is designed to allow us to interpret previously uninterpretable messages.

3. Tackling these questions.

How do these ideas of how to model the common ground affect our view of the fact we started out with? Recall the fact: in a situation in which we have been introduced to someone named "John," (1) can ask which of the salient people is the one so introduced, but not which function that person performs. One aspect of this fact happens to be that, in the scenarios we considered, (1) does not elicit responses like (10a) but does elicit responses like (10b).

(10) a. # John is the violinist. b. John is the guy on the left.

We can address this issue via an assumption about questions. I assume that questions are instructions to express one proposition in a certain set that the question[1] makes relevant; a felicitous answer will then be a sentence whose denotation is contextually equivalent to one of these propositions[2]. (Propositions ϕ and ψ are contextually equivalent with respect to a context set C when, for every world w in C, ϕ is true in w iff ψ is true in w.) Bearing this in mind, what do Ideas A and B suggest as far as the set of propositions that the question instructs us to choose among?

Let us start with Idea B, the idea under which there is a "John" property that potentially holds of different people in different worlds. Idea B suggests that the set for (1) is a set of propositions that vary with respect to an individual – a subset of (11).

(11) { λw. n = the person who has the "John" property in w | n an individual }

some propositions in this set:
λw. x = the person who has the "John" property in w
λw. y = the person who has the "John" property in w

Note that on this approach we explain as follows why the question does not elicit responses like (10a) in the scenarios we considered. Idea B suggests an analysis of (10a) along the lines of (12), but on a scenario like ours, where it is not known which of the individuals in front of us plays the violin, (12) will not convey what any of the propositions in (11) conveys. (By contrast, Idea B suggests an analysis of (10b) along the lines of (13), and on our scenarios, (13) *is* contextually equivalent to one of the propositions in (19): the proposition λw. *x* = the person who has the "John" property in w.)

(12) λw. the person who has the "John" property in w = the violinist in w
(13) λw. the person who has the "John" property in w = the guy on the left in w[3]

If instead we adopt Idea A, it seems more promising to say that the question instructs us to choose from a set of propositions that vary with respect to an individual concept – a subset of (14). But there is an important difference between the approach that Idea B suggests and the approach that Idea A suggests. On Idea A, the subset must be restricted in such a way as to exclude concepts like *the violinist* (including concepts like *the guy on the left*). By contrast, on Idea B, where we dealt with a set of propositions that varied with respect to an individual, no individuals needed to be excluded.

(14) { λw. F(w) = j | F a function from worlds to individuals }

some propositions in this set (abbreviated):
λw. the person on the left in w = j
λw. the person on the right in w = j

---

[1] Or, more precisely, the LF of a question. Questions that have different LFs are potentially ambiguous between different kinds of instructions. (See Section 4.)
[2] Or, more precisely, a sentence whose LF has a denotation that is contextually equivalent to one of these propositions.
[3] This is shorthand: to "be on the left" in w is roughly to be standing on the left of where we are in w.

(15) λw. j = the violinist in w

To see that the subset must be restricted, note that, on Idea A, the natural way of analyzing our (10a) is as in (15). If the subset were not restricted, it would include (15), and so we would wrongly predict (10a) to be a possible answer on a scenario like Scenario 1 or Scenario 2.

The tool for deciding between Idea A and Idea B is thus the theory that we adopt of what propositions a question makes relevant. If it is easy for such a theory to account for the constraints on proposition sets that we are forced to once we adopt one of these views, that is a point in favor of the view; if hard, that is a point against. I suggest that it is easier to see what has to be done to account for Idea B's less constrained proposition sets than Idea A's more constrained proposition sets. This argues for Idea B. In the last section of the paper, I assume a theory that links the propositions a question makes relevant to what is in the question's LF, and show that Idea B allows us to see the limits on (1)'s interpretation as posing a syntactic puzzle.

4. A syntactic puzzle.

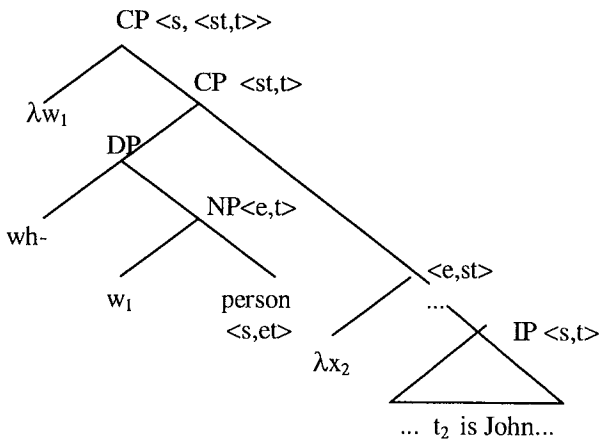
In what follows, I will take for granted a Heim and Kratzer 1998-style theory of how syntactic structures are interpreted. And I will assume Idea B. The point will simply be that, by taking this theory together with Idea B, we can transform the question we started out with into a different kind of puzzle. Maybe we are better equipped to solve that different kind of puzzle.

I assume that what set of propositions a question makes relevant depends on what the question's LF is. In particular, while the denotation of a question's LF is a set of worlds to sets of propositions, it is only licit to use a question when it yields the same proposition set for every world in the context set (cf. Stalnaker 1978) -- so that is how the single set of propositions arises. Given this background, if we assume Idea B, we would like to guarantee that (1) only has LFs with a denotation like (16). (16) will lead to propositions that vary with respect to an individual.

In fact, it is easy to imagine an LF that will yield this result. The simplified structure in (17) gives the general idea.[4] (I assume here that the verb has lowered back into IP, and that *who* actually has a complex structure.) Moreover, it would not be controversial to posit such an LF.

(16) λu. {λw. n = the person who has the "John" property in w | n is a person in u}

(17)



[4] We will need the following denotations: [[**wh**]]^g = $\lambda F_{<e,t>}.\lambda P_{<e,st>}.$ { P(n) | F(n) = 1 } ; [[**person**]]^g = $\lambda w_s.\lambda n_e.$ n is a person in w ; [[[$_{IP}$ **...t_2 is John ...**]]]^g = $\lambda w_s.$ g(2) = the person who has the "John" property in w. Note that **wh** takes a predicate and a property (like the property of "being John,"), and creates a set of propositions each of which says that the property holds of some individual the predicate characterizes.

But at the same time we cannot stop here and say that we have solved our problem. This is because there is no reason to assume that an LF like this is the *only* LF the question has, and other potential LFs might *not* lead to a denotation of the kind we want. What other potential LFs do I have in mind? Some background is in order.

Basically since Engdahl 1986, it has been clear that, once we say that questions like (18) have one LF that makes relevant propositions that vary with respect to an individual ((19)), then we must say they also have a second LF, which makes relevant different kinds of propositions. We must take this step because the questions license answers such as (20) even when such answers are not contextually equivalent to any of the propositions in (19) -- for example, when it isn't known who the candidates are.What kind of propositions does the second LF make relevant? A possible answer is: propositions that differ with respect to an individual concept ((21)). But now the worry should be clear: if (18) admits a second LF that makes relevant propositions that vary with respect to an individual concept, then our sentence (1) potentially might as well. And in that case, we might lose our account of the restriction on (1)'s interpretation.

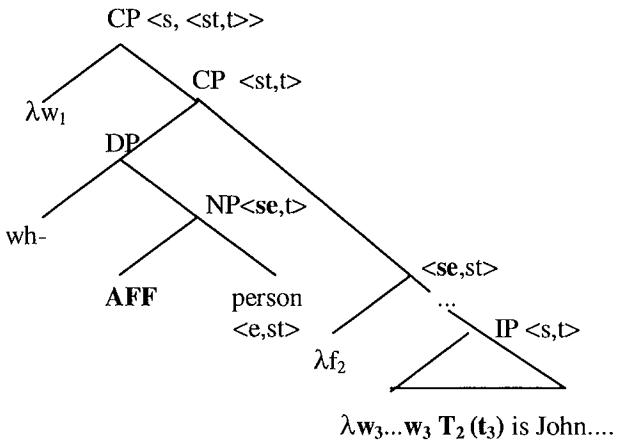
(18) Who will win the next election?
(19) { λw. in w, individual n wins the next election | n an individual }
(20) The candidate with the biggest campaign budget [will be the winner].

(21) { λw. in w, F(w) wins the next election | F a function from worlds to individuals }

(e.g., $\mathbf{F_1}$(w) = the election candidate in w with the biggest campaign budget in w)

To concretize the problem, let us imagine what the potential second LF for (1) might look like. A possibility (along the lines of the functional wh-analyses of Engdahl 1986, Chierchia 1993 and Heim 2001) is given in (22). This LF has the same basic architecture as the earlier one in (17), but differs in two important ways: first, the wh-phrase contains a silent affix; second, movement leaves a complex trace (it consists of an item that functions as a variable over concepts together with an item that functions as a variable over worlds).[5] The net effect of these differences is to give individual concepts the role that individuals play in (17).

(22)



[5] Note the change in the order of *person*'s arguments (the old entry was to give a simplified tree). There will also be a new cross-categorial denotation for *wh*. We will need these denotations, among others: $[[\mathbf{AFF}]]^g(P_{<e,st>}) = \lambda K_{<s,e>}$. for all w in dom(K), P(K(w))(w) = 1 ; $[[\mathbf{wh}]]^g = \lambda F_{<X,t>}.\lambda P_{<X,st>}$. { P(N) | F(N) = 1 } . (The effect of AFF is that NP will be a predicate of individual concepts rather than of individuals. The effect of the big trace and the lambda that binds it will be to create out of the movement remnant too a function that takes individual concepts, rather than individuals, as arguments.)

The LF in (22) will have (roughly) the denotation in (23). This is not the denotation that we would want a tree for (1) to have. If (1) had this LF, we would wrongly predict *The violinist is John* to be a fine answer.

(23) λu. {λw. in w, K(w) has the "John" property | for all v, K(v) yields a person in v}

So if we adopt Idea B, we arrive at a syntactic puzzle. We can account for the restriction on the use of (1) if we say that (1) *admits* the *first* kind of LF – the one in (17), the kind that makes relevant a set of propositions that differ with respect to an individual – but does *not* admit the *second* kind of LF – the one in (22), the one that makes relevant a set of propositions that differ with respect to an individual concept. The puzzle is: why is the second LF excluded? I don't know how tractable this problem is, but at least is clear what the problem is we have to solve. And there are places to start investigating. The second LF, for instance, contains a complex trace. Might it be that complex traces are barred from certain positions?

5. Concluding remarks.

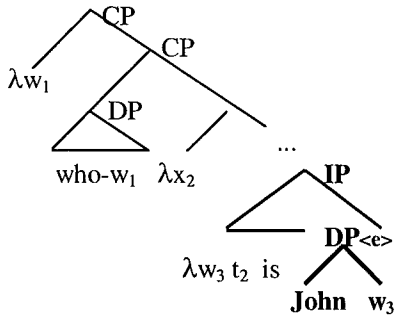
What I hope came out of the discussion is this. When we look at ways of combining assumptions about ontology, on the one hand, with ideas about interpretation, on the other, we find that some ways are more suited to describing natural language than other ways are. Therefore, if we take a stand on one of these things – in this case, a theory of how we interpret syntactic structures – we will be driven to conclusions about the other -- in this case, about what view of possible worlds is the right one for modeling the common ground.

The discussion so far suggested that we should adopt Idea B, on which it makes sense to talk about a "John" property that individuals have in some worlds but not others. This is interesting for a number of reasons. For one thing, it could help us to understand what articles in the copula literature mean when they say that name-like expressions can sometimes behave like "predicates" and sometimes like "referring expressions." On Idea B, there is a natural way of explicating this terminology: a name like *John* behaves like a "predicate" when the "John" property holds of different individuals in different worlds in the context set, and behaves like a "referring expression" when the "John" property holds of the same individual in all worlds in the context set. On Idea A, by contrast, explicating this terminology looks less straightforward.

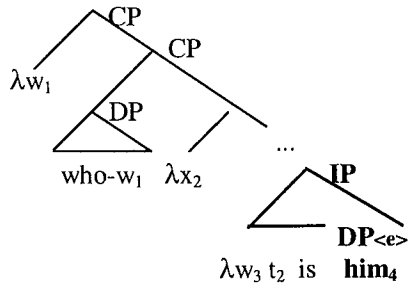
Speculating, here is another advantage that Idea B might have. It might lead to insight as to why, unlike our friend (24a) , parallel sentences with pronouns like (24b) seem bizarre in any context. When we ask how the denotations of questions like (24a) are built out of the parts of the sentence, it is a small step to say that the name *John* denotes an individual concept – the person who has the "John" property. It is another small step to say that in the structure for this sentence, *John* combines with a world variable, and the result occupies a slot reserved for an individual-denoting expression ((25a)). Now, the idea to pursue is that, while names denote properties, pronouns denote individuals. In that case, in the structure for (24b), the pronoun will occupy the position that the name-world variable complex occupies in the structure for (24a). If the pronoun in (24b) is a simple variable, the result will be that, when we compute the denotation for the question in (24b), and we look at the propositions the question will makes relevant, we will find only tautologies or contradictions.

(24) a. Who is [t t John] ? b. ?? Who is [t t him]?

(25) a.



b



$[[(25a)]]^g = \lambda u.\ \{\lambda w.\ n = $ the person who has the "John" property in w $\mid$ n is a person in u$\}$

$[[(25b)]]^g = \lambda u.\ \{\lambda w.\ n = g(4) \mid$ n is a person in u$\}$

(5.1. Final digression.)

Remember the "intermediate idea"? Here are a few words about what consequences that third idea might have for an analysis of our fact about the use of questions like (1) – although in fact it is not completely transparent how to apply this idea.

Recall that on the intermediate view we use names to talk about individuals rather than properties, as on View A. But recall at the same time that the parties to conversation might or might not have decided what individual one uses a name like "John" to talk about. It seems to me that, on the intermediate view, maybe one could maintain what might at first look like a non-starter -- that the question makes relevant a set of propositions that, in terms of the way they are built up, vary with respect to an individual, but that are either a tautology or a contradiction ((26)). However, we must also stipulate that the question requires its answer to take a certain form: to be of the form *X is John*. What will happen then? The idea is roughly as follows. Since one of the propositions is a tautology, the answerer will have the obligation to utter a proposition that is true in all worlds in the context. As soon as he answers a sentence of the form *The guy on the left is John*, the parties to conversation, who know that *x* is the guy on the left, will see that to talk about John is to talk about *x*. This, then, looks like another case where we would like to say that a question makes relevant a set of propositions that vary with respect to an individual. So at first glance at least, it looks as though, since we want to insure variation with respect to an individual, we are going to wind up with the same problems that we have to face with Idea B.

(26) { $\lambda w.\ x = j$,
$\lambda w.\ y = j$,
$\lambda w.\ z = j$,
... }

References.

Chierchia, G.: 1993. "Questions with Quantifiers," *Natural Language Semantics* 1, 181-234.
Engdahl, E.: 1986. *Constituent Questions*. Kluwer, Dordrecht.
Hamblin, C.L.: 1973. "Questions in Montague Grammar," *Foundations of Language* 10, 41-53.
Heim I: 2001. Unpublished lecture notes, MIT.
Heim, I. and A. Kratzer: 1998. *Semantics in Generative Grammar*. Blackwell, Malden.
Higgins, F.R.: 1973. *The Pseudo-cleft Construction in English*. PhD diss., MIT.
Karttunen, L.: 1977. "Syntax and Semantics of Questions," *Linguistics & Philosophy* 1, 3-44. .
Stalnaker, R.: 1978. "Assertion," *Syntax and Semantics* 9, 315-332.

Towards Automated Generation of Scripted Dialogue: Some Time-Honoured Strategies

Paul Piwek and **Kees van Deemter**
ITRI – University of Brighton
Watts Building, Moulsecoomb
Brighton BN2 4GJ
UK
Email: Paul.Piwek@itri.bton.ac.uk Kees.van.Deemter@itri.bton.ac.uk

Abstract

The main aim of this paper is to introduce automated generation of scripted dialogue as a worthwhile topic of investigation. In particular the fact that scripted dialogue involves two layers of communication, i.e., uni-directional communication between the author and the audience of a scripted dialogue and bi-directional pretended communication between the characters featuring in the dialogue, is argued to raise some interesting issues. Our hope is that the combined study of the two layers will forge links between research in text generation and dialogue processing. The paper presents a first attempt at creating such links by studying three types of strategies for the automated generation of scripted dialogue. The strategies are derived from examples of human-authored and naturally occurring dialogue.

1 Introduction

By a scripted dialogue we mean a dialogue which is performed by two or more agents on the basis of a description of that dialogue. This description, i.e., the script, specifies the actions which are performed in the course of the dialogue and their temporal ordering. We assume that the script is created in advance by an author. Automated generation of scripted dialogue involves a computer programme in the role of the author and execution of the script by software agents. André et al. (2000) coin the term 'presentation team' for such a collection of software agents. In their words, presentation teams '[...] rather than addressing the user directly–convey information in the style of performances to be observed by the user' (André et al., 2000:220).

The plan of this paper is to first motivate why the study of scripted dialogues is interesting and useful, whilst also pointing out the limitations and complications of scripted dialogue. We then present a number of strategies for the automated generation of scripted dialogue. Our discussion is illustrated by means of some extracts from mainly scripted dialogues. The paper concludes with a brief overview of the ongoing NECA project in which scripted dialogues are presented by embodied conversational agents.

2 Prospects and Problems

Scripted dialogues have interesting features, both from a theoretical and a practical point of view. Let us start by highlighting A theoretical issue. Scripted dialogues involves two layers of communication. First, in a scripted dialogue there is a layer at which the participants of the dialogue *mimic* communication with each other. The communication is not real because the actions of the participants are based on a script; the participants do not interpret the actions of the other participants in order to determine their own actions. Second, there is a layer of uni-directional communication from the author of the script to the audience of the dialogue. At this level, a scripted dialogue is very much like a monologue. Thus scripted dialogue presents a challenge because it requires simultaneous generation of a layer of real and one of pretended communication, each layer having its own participants with their (real or pretented) goals, beliefs, desires, personalities, etc.

Also from a practical point of view scripted dialogue has something to offer. Before we go into its advantages, let us, however, first discuss a feature of scripted dialogue which might be perceived as one of its limitations. There is a class of applications which requires the generation of dialogue about a subject matter that evolves in real-time. For such applications

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 141-148.

scripted dialogue is not possible. For instance, André et al. (2000)'s dialogues between two reporters about a live transmission of a ROBOCUP soccer event cannot be implemented as scripted dialogue: the verbal reports of the agents are determined to a large extent by events which evolve in real-time. Hence it is impossible to script the verbal reports in advance. Similarly, automatically generated scripted dialogues do not lend themselves well to user involvement in the dialogue: because the script is created in advance there is no scope for reaction to the user's contributions.

These limitations are, however, offset by a large number of benefits. Firstly, it should be noted that for large scale applications it has been argued that a combination of scripted and autonomous behaviours is required. For instance, in the MRE project at USC[1] a combination of autonomous and scripted virtual humans is used to create a realistic training environment. Thus applications do not necessarilly force a strict choice between autonomous and scripted behaviour (more specifically dialogue).

Secondly, staying within the education/training domain it should be noted that in the literature on Intelligent Tutoring Systems (ITS) a case has been made for so-called vicarious learning (e.g., Lee et al., 1998; Cox et al., 1999): learning by watching dialogues of other people being taught or engaged in a learning process. Various studies have been carried out in this area and some positive effects of vicarious learning by overhearing dialogue (as opposed to monologue) have been found (see Scott et al., 2000).

Thirdly, there is a large class of obvious applications for scripted dialogues. The dialogue in film scenes, commercials, plays, product demonstrations, etc. can be treated as scripted dialogue. These are examples of situations in which real-time interaction with the environment or a user are not required.

Finally, there is not only a wide range of potential applications for scripted dialogue, but applying scripted dialogues also has some distinct advantages over relying on dialogue generated by autonomous agents:

1. The generation of a scripted dialogue requires no potentially complicated and error-prone interpretation of the dialogue acts produced by other autonomous agents.
2. More time is available for the generation process, because scripted dialogue is not generated in real time.
3. Not only more time but also more information is available to the generation process of scripted dialogue. Whereas in spontaneous dialogue an individual action can only be constructed using information about the actions which temporally precede it, in scripted dialogue an action can be tailored to both the actions which precede and those which follow it.
4. It is much easier to create dialogues with certain global properties (e.g., a certain pattern of turn-taking), because the dialogue is constructed by a single author. In spontaneous dialogue, such properties emerge out of the autonomous actions of the participants, which makes it difficult to control them directly.

[1] See, e.g., Rickel et al. (2002).

3 Strategies for Scripted Dialogue Generation

We have already pointed out that in one important respect scripted dialogue resembles monologue: information flows from a single author to an audience. Hovy (1988) was one the first to systematically consider how the (communicative) goals of an author can be related to various strategies for communicating information through a monologue. In particular, he describes a natural language generation (NLG) system (PAULINE) which implements a number of such strategies. He thereby abandoned an assumption which was and still is implicit in many NLG systems, namely that a text is generated from a database of facts and that the task is mainly one of mapping these facts onto declarative sentences which express them. Hovy's work on the influence of pragmatic factors on natural language generation is currently followed up by various researchers involved in building embodied conversational agents (for a bibliography of recent work in this newly emerging area of natural language generation see Piwek, 2002).

The new picture that emerges is one where an author uses a text as a device for influencing

the attitudes of his or her audience. Amongst the attitudes which an author might want to influence are the beliefs, intentions (plans for action), goals, desires and opinions (judgements about whether something is good, bad or neutral) of the audience. All of the aforementioned attitudes are *about* something (that is, they are intentional; see, for instance, Searle, 1983). Roughly speaking, one can discern attitudes which are about the subject matter/topic of a text and those which pertain to the context (e.g., the author, the audience and the relation between the two). Attitudes about the context are normally communicated implicitly. Hovy discusses how style (formal, informal, forceful) can be used to do so. Generally speaking, everything that is discussed explicitly in a text is part of its subject matter. Thus, whenever contextual aspects are discussed explicitly (e.g., 'I am your boss, therefore listen to what I have to say'), they become also part of the subject matter. This leaves us with a class of information that is neither discussed explicitly (and therefore not part of the subject matter) and is neither part of the context. For instance, take an opinion about a person which can be expressed explicitly as in 'X is a bad guy', but also implicitly as in 'X killed John'.[2]

Scripted dialogue offers the same communicative opportunities (i.e., for communicating facts and influencing an audience in other ways) as ordinary text, plus a number of other ones in addition. To illustrate some of the issues, let us summarize one of the first dialogues written by the humanist philosopher Erasmus of Rotterdam in 1522 (Erasmus, 1522).[3]

E1
1. A: Where have you been?
2. C: I was off to Jerusalem on pilgrimage.
3. A: Why?
4. C: Why do others go?
5. A: Out of folly if I'm not mistaken.
6. C: That's right; glad I'm not the only one though.
7. A: Was the trip worthwhile?
8. C: No.
9. A: What did you see?
10 C: Pilgrims causing mayhem.
11. A: Were you morally uplifted?
13 C: No, not at all.
14 A: Did you get richer?
15. C: No, quite the contrary.
16 A: Was there nothing good about the trip then?
17 C: Yes, in fact there was. In particular, I can now entertain others with my lies, like other pilgrims do.
18. A: But that's not very decent, is it? [...]
19. C: True. But I may also be able to talk others out of the idea of pilgrimage.
20. A: I wish you had talked *me* out of it.
21. C: What? Have you been as stupid as I?
22. A: I've been to Rome and Santiago de Compostela.
23. C: Why?
24. A: Out of folly I guess [because ...]
25 C: So why did you do it?
26. A: My friends and I vowed to go when we were drunk.
27. C: Surely a decision worth taking when you're drunk [...] Did everyone arrive back home safely?
28. A: All except three: Two died; the third we left but he's probably in heaven now.
29. C: Why? Was he so pious?
30. A: No, he was a scoundrel.
31. C: Then why is he in heaven?
32. A: Because he had plenty of letters of indulgence with him [...]
33. A: Don't get me wrong: I'm not against letters of indulgence but I have more admiration for someone who leads a virtuous life. Incidentally, when do we go to these parties that you mentioned?
34. C: Let's go as soon as we can, and add to other pilgrims' lies.

The central question that we will start addressing in this paper is 'what strategies for influencing the attitudes of the audience are specific to scripted dialogue?' A number of these strate-

[2]Hovy (1988) discusses how reporting that a person is the actor of an action which is generally considered to be bad can be used to implicitly convey the opinion that the person is bad.

[3]Large parts of the dialogue were omitted, reworded, or summarized, since we wil be focusing on a specific set of issues. The (very tentative) translation is our own.

gies will be introduced below. We will return to Erasmus' dialogue at various points in our discussion and highlight parts of the dialogue which illustrate the aforementioned strategies.

3.1 Strategies of information distribution

As we have seen, the main difference between monologue and scripted dialogue is that the latter communicates with the audience *via* the (pretended) communication between the dialogue participants. This has a number of immediate effects:

1. The author can let a participant say something without being directly responsible for the content.
2. Each participant can represent a particular chunk of information, making the combined content more easily digestible.
3. In particular, the participants can represent different points of view on the same subject matter, which may even be inconsistent with each other.
4. One participant may express an opinion concerning something the other participant has raised.

Point 1 was clearly relevant to Erasmus, in whose case direct criticism of the Catholic church could have made him a target for the Inquisition. (A's last utterance seems intended to further milden any criticism.) *Point 2* is relevant because each of the two participants represents a particular journey. Both journeys could have been related in one monologue, but this might have led to confusion and would certainly have been less exciting to read. In fact, it is patently clear from reading the whole dialogue that one participant's questions are used for making more lively what would otherwise have been a story with some rather boring parts.

Point 3 is not directly relevant in the case of the present dialogue but the very fact that both participants agree on all essentials can only reinforce the strengths of Erasmus' implied position. (See also under Strategies of Emphasis.) *Point 4* is relevant again, since both participants frequently express evaluative opinions concerning various elements of the two stories. In many cases, evaluative opinions are expressed in highly indirect fasion, and this brings us to a second class of strategies.

3.2 Strategies of association

One way to influence the attitude of the audience about, for instance, a person is to mention the person in combination with something else to which the audience already has the intended attitude. Hovy (1988) suggests, for example, that we can make somebody look good, bad or neutral by presenting him or her as the actor of an action which is generally (or specifically by the audience) perceived to be good, bad or neutral, respectively: "Mike killed Jim" makes Mike look bad, whereas "Mike rescued Jim" makes him look good.

In fact, this strategy seems to be an instance of a more generally applicable strategy: *To convey that X has property P, one can present X in combination with something which has or implies property P.* Thus, to convey that Mike is a clever guy we might say "Mike managed to solve this partial differential equasion in no time", i.e., Mike is presented as being able to solve a difficult problem quickly, which implies being clever.

Information conveyed in a text is not only presented in the context of other information which can influence its interpretation but also *by* the author (and possibly speaker) of the text. If any properties of this author/speaker are known, these can rub off on the points s/he is trying to make. The appeal to this tendency in an argument is considered to be a fallacy (Argumentum ad Hominem). For instance, one might argue that Bacon's philosophy is untrustworthy because he was removed from his chancellorship for dishonesty (Copi, 1972:72).

Scripted dialogue lends it particularly well to this type of association. The presence of the second layer of communication allows the author to distribute communicative acts over characters which were conceived by the author. These characters can be given certain traits which influence the interpretation by the audience of what they say in the dialogue. These traits can be conveyed by various means. In the case of Erasmus' dialogue, the fact that the protagonists and their pilgrim friends are avid partygoers – evidently something Erasmus didn't approve of – is used to discredit their pilgrimage. If the dialogue is enacted by a

collection of embodied agents, their physical appearance can be used. Alternatively, certain characteristics of the dialogue can also suggest a particular property. For instance, Thomas (1989) discusses various ways in which a speaker can come across as dominant or an authority (interruptions, abrupt changes of topic, marking new stages in the interaction, metadiscoursal comments, etc.). The following is an example of a marking of a new stage in the interaction taken from Thomas (1989:146):

E2 A: Okay that's that part. The next part what I want to deal with is your suitability to remain as a CID officer.

3.3 Strategies of emphasis

For various reasons, an author might want to highlight certain information and suppress other information. In a monologue, repetition of information signals emphasis. For instance, de Rosis and Grasso (2000) analyse an explanation text about drug prescription and point out that certain information is rather redundant, i.e., repeated with identical or equivalent wording such as:

E3 "The good is news is that we do have tablets that are very effective for treating TB", and "but it is something we can do something about".

Here it seems that positive information is repeated intentionally. In dialogue, repetition can be achieved naturally due to the presence of two interlocutors at the second layer of communication. Consider the dialogue fragment below from Twain (1917:11) between a young man and an old man.

E4
1. Y.M. What detail is that?
2. O.M. The impulse which moves a person to do things – the only impulse that ever moves a person to do a thing.
3. Y.M. The *only* one! Is there but one?
4. O.M. That is all. There is only one.
5. Y.M. Well, certainly that is a strange enough doctrine. What is the sole impulse that ever moves a person to do a thing?
6. O.M. The impulse to *content his own spirit*–the *necessity* of contending his own spirit and *winning its approval.*

Here turns 3. and 4., which form a subdialogue, are both about the claim that there is *only one* impulse which moves a person to do things. Now imagine that we have an algorithm which has already distributed the information which it wants to get across to the audience amongst the dialogue participants. Let us call this step I. During step II, the algorithm will determine how this information can be conveyed through a sequence of turns. To generate E4, we might at this stage produce the sequence: 1., 2., 5., 6. Step III would involve the addition of further turns for the purpose of emphasizing information. During step III, such an algorithm could insert subdialogues like the one above (3., 4.), if there are any matters which need particular emphasis.

Note that Erasmus also employs this strategy. For instance, in Erasmus' Dialogue (E1), the turns 4. and 5. form a subdialogue which a system like the one proposed here could insert during step III, after already having created the sequence 1., 2., 3., 6., ...

Note that the sketched approach presupposes that realization (both verbal and non-verbal) is performed only after steps I – III; during steps I – III the algorithm manipulates abstract descriptions of the semantic and pragmatic content of the utterances. The reason for this is that before step III, the algorithm can not yet know whether to realize the beginning of turn 6. as 'That's right' or 'Out of folly', since this depends on whether the subdialogue (4., 5.) is inserted or not.

4 Application in the NECA project

The work reported in this paper is carried out in the context of the NECA project which started in October 2001 and has a duration of 2.5 years.[4] In this project a system is being built that can generate scripted dialogues that are subsequently performed by animated human-like

[4] NECA stands for Net Environment for Embodied Emotional Conversational Agents. The project is funded by the EC. The partners in the project are: DFKI, IPUS (University of the Saarland), ITRI (University of Brighton), ÖFAI, Freeserve and Sysis AG. Further details can be found at http://www.ai.univie.ac.at/NECA/.

characters. One prototype to be delivered by NECA is an electronic showroom (eShowroom).[5] The idea is that a user/customer can select a class of cars and/or attributes (friendly for the environment, luxury, sportiness, etc.) in which s/he is interested. Furthermore, the user can set the personality traits of the characters which are to discuss this car (introverted, extroverted, agreeable, etc.). On the basis of these settings, the system then produces dialogues about specific cars and presents these to the user by means of embodied conversational agents which play out the dialogue. The strategies discussed in the present paper are highly relevant in the context of car sales. For example,

- Information about cars can be complex, so it can be useful to have one or more participants ask clarification questions (somewhat in the style of Conan Doyle's Watson character, who triggers Sherlock Holmes into explanations that benefit us as readers).
- Not all customers are alike and it can be useful to let different types of customers be represented by different animated characters: one who is primarily interested in the performance off the car, one who is interested in chrome and gloss, one who is very aware of environmental and safety-related issues, etc.

Let us describe in more detail how one of the strategies of emphasis will be implemented in the NECA system. The NECA system generates the interaction between two or more characters in a number of steps, where information flows from a Scene Generator to a Multi-modal Natural Language Generator, to a Speech Synthesis component, to a Gesture Assignment component, and finally to a media player.

In the Scene Generator, the basic structure of the dialogue is determined. For this purpose, a top-down planning algorithm is used. The output of this module is a RRL Scene Description (see Piwek et al., 2002). Amongst other things, this Scene Description contains a set of dialogue acts. Individual dialogue acts are specified in terms of the dialogue act type, the speaker, the addressees, the semantic content, the actions which the act is a reaction to, and the and emotions (felt and expressed). The temporal ordering amongst the dialogue acts is represented separately and allows for underspecification.

A Scene Description is constructed stepwise. We might start by constructing the following dialogue fragment:[6]

E5 x_1 B: How fast is this car?
x_2 S: Its top speed is 180mph.
x_2 B: Wow, that's great.

At this point, further elaboration of the dialogue is possible. Assume, for example, that the positive information that the car has a top speed of 180mph is to be emphasized. For this purpose, the strategy of emphasis we discussed in section 3.3 can be employed: a subdialogue can be inserted after x_2 consisting of a question by the buyer ('As much as 180mph?') which provides the seller with the opportunity to repeat a positive piece of information. This procedure yields the following 'enhanced' dialogue:

E6 x_1 B: How fast is this car?
x_2 S: Its top speed is 180mph.
y_1 B: As much as 180mph?
y_2 S: Yes, no less than 180 mph.
x_2 B: Wow, that's great.

Note that the information which required emphasis has been mentioned no less than three times in the dialogue. This has been achieved by exploiting a very natural dialogue phenomenon: the occurrence of confirmation subdialogues.

The thus created abstract representation of the dialogue (the Scene Description) can subsequently be processed further by the Multimodal Natural Language Generator, the Speech Synthesis component, and the Gesture Assignment component. The result is sent to a media player which displays the dialogue to the user by means of a collection of embodied conversational characters.

5 Conclusions

We are not aware of any work which lays out strategies for the automated generation of

[5] This application builds on the work carried out at DFKI and reported in André et al. (2000).

[6] In reality, at this stage in the processing linguistic realization has not yet taken place; thus the texts in E5. should be understood as mere paraphrases of the abstract descriptions of the dialogue acts which are actually passed on.

scripted dialogue. There is a rapidly growing body of work on (Embodied) Conversational Agents (see, e.g., Ball & Breeze, 1998; De Carolis et al., 2001; Loyall & Bates, 1997; Nitta et al., 1997; Prendinger & Ishizuka, 2001; Walker et al., 1996 and Zinn et al., 2002), but to the extent that language generation is discussed there[7], it is from the perspective of the agents who participate in the conversation, rather than from the perspective of an author who produces a script for their interaction. The only exception we came across is André et al. (2000) which reports on implemented systems for both spontaneous and scripted dialogue. Our aim has been to take a step back from the domain-specific implementation which they propose and find out whether it is possible to first identify more general strategies which are valid for the automated generation of scripted dialogue. We have tried to find such strategies on the basis of examples of human-authored and naturally occurring dialogues. We hope that our tentative investigations will encourage further studies into this new topic. A topic which, in our opinion, harbours interesting research questions at the intersection between dialogue processing and text generation, and which also lends itself well for various types of practical applications.

Acknowledgements

We would like to thank Alexandre Direne, Martin Klesen and Neil Tipper for comments on an earlier draft of this paper, and two anonymous reviewers for their comments and encouragement. This research is supported by the EC Project NECA IST-2000-28580. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

References

André, E., T. Rist, S. van Mulken, M. Klesen & S. Baldes (2000). 'The Automated Design of Believable Dialogues for Animated Presentation Teams'. In: J. Cassell, J. Sullivan, S. Prevost and E. Churchill, *Embodied Conversational Agents*, MIT Press, 220-255.

[7]See Piwek (2002) for an overview of the literature in this area, specifically on affective natural language generation.

Ball, G. & J. Breeze (1998). 'Emotion and Personality in a Conversational Character'. In: *Proceedings of the Workshop on Embodied Conversational Characters*, Lake Tahoe, CA, 1998, 83–86.

Cox, R., J. McKendree, R. Tobin, J. Lee & T. Mayes (1999). 'Vicarious learning from dialogue and discourse'. *Instructional Science*, 27, 431–458.

Copi, Irving M. (1972). *Introduction to Logic: Fourth Edition.* Macmillan, New York.

de Carolis, B., V. Carofiglio, C. Pelachaud & I. Poggi (2001). 'Interactive Information Presentation by an Embodied Animated Agent'. In: *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue*, Verona, Italy, 14–15 December 2001, 19–23.

Erasmus, Desiderius (1522). *Colloquia.*

Hovy, Eduard H. (1988). *Generating Natural Language Under Pragmatic Constraints.* Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Lee, J., F. Dineen, J. McKendree (1998). 'Supporting student discussions: it isn't just talk'. *Education and Information Technologies*, 3, 217–229.

Loyall, A.B. & J. Bates (1997). 'Personality-Rich Believable Agents That Use Language'. In: *Proceedings of the first International Conference on Autonomous Agents*, Marina Beach Marriott Hotel, Marina del Rey, California, February 5-8, 1997.

Nitta, K., O. Hasegawa, T. Akiba, T. Kamishima, T. Kurita, S. Hayamizu, K. Itoh, M. Ishizuka, H. Dohi, and M. Okamura (1997). 'An Experimental Multimodal Disputation System'. In: *Proc. of the IJCAI-97 Workshop on Intelligent Multimodal Systems*, Nagoya, 1997.

Piwek, P. (2002). 'An Annotated Bibliography of Affective Natural Language Generation'. *ITRI Technical Report ITRI-02-02*, University of Brighton. (available at: http://www.itri.bton.ac.uk/~Paul.Piwek)

Piwek, P., B. Krenn, M. Schröder, M. Grice, S. Baumann & H. Pirker (2002). 'RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA'. In: Proceedings of the AAMAS workshop "Embodied conversational agents – let's specify and evaluate them!", 16 July 2002, Bologna, Italy.

Prendinger, H. & M. Ishizuka (2001). 'Agents That Talk Back (Sometimes): Filter Programs for Affective Communication'. Contribution to: *Second Workshop on Attitude, Personality and Emotions in User-adapted Interaction* (in conjunction with User Modeling 2001), Sonthofen, Germany, July 13, 2001.

Rickel, J., S. Marsella, J. Gratch, R. Hill, D. Traum & W. Swartout (2002). 'Toward a New Generation of Virtual Humans for Interactive Experiences'. *IEEE Intelligent Systems*, July/August 2002.

de Rosis, F. & F. Grasso (2000). 'Affective Natural Language Generation'. In: A.M. Paiva (Ed.), *Affective Interactions*, Springer Lecture Notes in AI 1814, 204–218.

Scott, D.C., B. Gholson, M. Ventura, A.C. Graesser and the Tutoring Research Group (2000). 'Overhearing Dialogues and Monologues in Virtual Tutoring Sessions: Effects on Questioning and Vicarious Learning'. *International Journal of Artificial Intelligence in Education*, 11, 242–253.

Searle, John R. (1983). *Intentionality.* Cambridge University Press, Cambridge.

Thomas, Jenny A. (1989). 'Discourse control in confrontational interaction'. In. L. Hickey (ed.), *The Pragmatics of Style*, Routledge, London and New York.

Twain, Mark (1917). *What is man? And other essays.* Chatto & Windus, London.

Walker, M.A., J.E. Cahn & S.J. Whittaker (1996). 'Linguistic Style Improvisation for Lifelike Computer Characters'. In: *Proceedings of the AAAI Workshop on AI, Artificial Life and Entertainment*, Portland.

Zinn, Claus, Johanna D. Moore, & Mark G. Core (2002). 'A 3-tier Planning Architecture for Managing Tutorial Dialogue', To appear in: *Proceeding of Intelligent Tutoring Systems, Sixth International Conference* (ITS 2002), Biarritz, France, June 2002.

Knowledge–based Reference Resolution for Dialogue Management in a Home Domain Environment

J. F. Quesada
University of Seville
jquesada@cica.es

J. G. Amores
University of Seville
jgabriel@us.es

Abstract

Several features of spoken dialogue systems require advanced reference resolution strategies. This paper presents a knowledge–based strategy for reference resolution implemented over an agent architecture which differentiates the Knowledge and Action Managers from the Dialogue Manager. The strategy, in conjunction with the ability of the Dialogue Manager to manipulate the dialogue history, has been applied to a home environment domain to solve successfully different phenomena such as quantification, anaphoric expressions and coordination.

1 Introduction

Reference Resolution plays a crucial role in spoken dialogue systems:

- The cognitive structure and content of linguistic utterances incorporates multiple phenomena that require specific strategies to solve their reference. For instance, anaphoric expressions (Eckert & Strube 2001; Palomar & Martinez-Barco 2001), under-specification (Bos 2001), presupposition (Hulstijn 1996; Lemon et al 2001b), or quantification (Cooper 1993).
- In application–oriented dialogue systems (Giachin & McGlashan 1996) there is a representation gap between the logical form manipulated by the semantic interpreter and dialogue manager components, and the interface to the real elements controlled at the application level (Kamp & Reyle 1993; Fraser & Thornton 1995; Larsson & Traum 2000)
- Multimodality adds additional constrains to the reference resolution problem, such as the use of anaphoric (Lemon et al 2001a; Lemon et al 2001b; Lemon et al 2001c), or even deictic anaphoric expressions (Eckert & Strube 1999).

This paper describes the knowledge-based device resolution (k-DevRes) strategy which has been implemented in the Spanish demonstrator developed by the University of Seville within the D'Homme project (DHomme Project 2001). This strategy may be characterised as a *Knowledge–based Device Resolution Algorithm*, given the central role played by the Knowledge Manager module in the resolution process.

In the context of a home environment with a large number of devices, one of the most important tasks to be carried out by the modules in charge of contextual interpretation and dialogue management involves the identification of the device or devices which are the object of a command.

There are two main challenges at this level.

1. First, Device Resolution is a task in which several components or agents are usually involved (Milward 2001; McGlashan & Axling 1996). Even though the semantic analysis module provides the formal representation of the user's instruction, the identification of the device or devices to which the desired command has to be applied requires both static (structure of the house, etc.) and dynamic information (connected devices, their characteristics, state, etc.) which is stored in the Knowledge Manager. On the other hand, taking into account the previous dialogue history, the Dialogue Manager may apply different disambiguation strategies.
2. Second, and related to the previous

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 149-154.

point, Device Resolution must incorporate the resolution of phenomena such as anaphoric expressions, question accommodation, quantification, etc. This, in turn, imposes a higher degree of coordination and integration between the different modules involved.

Section 2 analyses the main agents involved in the resolution of devices by means of a simple example. Plug and Play functionality is achieved as a consequence of the agent architecture and the communication interface designed between them. The main function at this level is situated between the Dialogue Manager and Knowledge Manager agents involved in the resolution of devices, which is studied in section 3.

The strategy, in conjunction with the ability of the Dialogue Manager to manipulate the dialogue history has been applied to a home environment domain to solve successfully different phenomena such as quantification, anaphoric expressions, coordination, etc.

Specifically, the information returned by the Device Resolution function allows the detection of ambiguity and the design of collaborative subdialogues for its resolution. Furthermore, this strategy has been applied to solve disambiguation subdialogues, error repairs, reaccommodation, task accommodation, default actions, and undo operations.

2 Knowledge Specification, Knowledge Management and Action Management

This section describes the main features of the spoken dialogue system (Quesada et al 2001) in which we have implemented the reference resolution strategy:

- It includes a dialogue management system which allows flexible interaction using natural language commands.
- It has been implemented over a distributed agent architecture.
- It allows speech input and output. That is, specific agents connect the dialogue management engine with the modules of speech recognition and synthesis.
- It has been especially designed for the control of devices in a home environment.
- The system supports Plug and Play functionality, both weak and strong. That is, the design of the agent architecture permits the dynamic incorporation of new devices and the automatic reconfiguration of the Action Manager, Knowledge Manager and Dialogue Manager agents.

From the point of view of the device resolution, the agent architecture includes a specific agent for the specification of the domain knowledge (the HomeSetup agent), which permits the design of a semantic structure for a house, or, in general, for the environment of use of the system. The semantic structure is represented by means of a tree, which is internally stored as a graph. Three main nodes are distinguished in the semantic graph: location, devtype and descriptor, which correspond to the Location, DeviceType and Descriptor types of the representation formalism used (Quesada et al 2001).

The GUI included as part of the HomeSetup agent allows the association of real devices installed in the house with their semantic structure. Let us consider, for instance, the following four devices with their corresponding characteristics (location, device type and descriptors):

1. Light01: living_room, light, small
2. Dimmer01: living_room, dimmer, blue
3. Light02: garden, light, yellow, big
4. Light03: bedroom, light

Figure 1 shows the semantic graph represented in the Knowledge Manager, which includes synonyms, and the devices associated with each node in the graph.

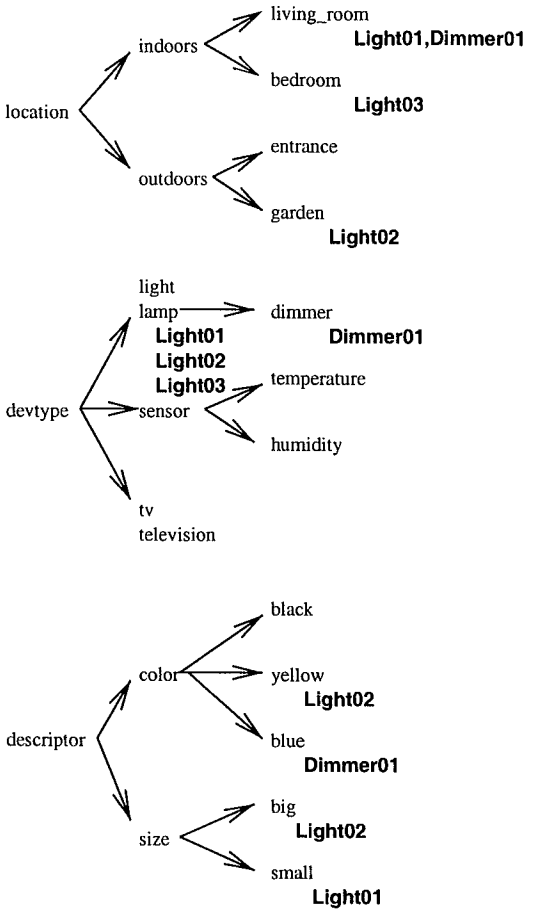


Figure 1: Semantic graph, including nodes, synonyms and devices.

2.1 Reference Resolution

Let us consider that the user prompts the system to turn on all the indoor lights:

- **U(1)**: Could you please turn on all the indoor lights?

The Natural Language Understanding agent (Parser and Semantic Interpreter) will obtain the formal representation (using the DTAC protocol) for the previous utterance (Figure 2).

Upon receiving this structure, the Dialogue Manager detects that it is a command of type CommandOn (turn on) which must be executed over one or more devices (there is a quantification mark).

In order to resolve the device, the Dialogue Manager sends a request of type k-DevRes (knowledge-based Device Resolution) to the Knowledge Manager module, passing the information known about it:

k-DevRes(location:indoor, devtype:light)

The Knowledge Manager traverses the semantic graph (Figure 1) looking for the values for the each of the attributes above. That is, it searches for *indoor* and *light*. Each of these searches obtains the union of the devices hanging from that node, and all its descendents (thus, the search in the light node obtains also the devices of type dimmer, since dimmer has been defined as a daughter or subtype of light).

In a first phase, then, it obtains the following result:

k-DevRes-SolveNode(indoor) =
(Light01,Light03,Dimmer01)
k-DevRes-SolveNode(light) =
(Light01,Light02,Light03,Dimmer01)

In a second phase, it obtains the intersection of the searches performed on each node, and deletes the unplugged devices (in this case, Light01). Thus, the reference resolution of this query will contain Light03 and Dimmer01.

k-DevRes(location:indoor, devtype:light) =
{Light03, Dimmer01}

Finally, the Dialogue Manager will send to the Action Manager the command SwitchOn applied to the devices Light03 and Dimmer01. In turn, the Action Manager will update the state of the devices in the HomeSetup simulation, and will send an instruction to the Device Manager to control the real devices.

So, the new situation is: Light01: unplugged, Light02: off, Light03: on, Dimmer01: on.

2.2 Dialogue Management: Complex Linguistic Phenomena

The previous example has illustrated the use of quantification as part of the specification of devices. Furthermore, the dialogue manager supports a wide range of linguistic phenomena. Next, we illustrate this functionality continuing with the example above in the state obtained in the previous section.

2.2.1 Natural Command Languages, Underspecification and Expectations

- **U(2)**: Turn off.

$$
\left[\begin{array}{lll}
DMOVE & : & specifyCommand \\
TYPE & : & CommandOn \\
ARGS & : & [DeviceSpecifier] \\
CONT & : & \\
 & & \left[\begin{array}{lll}
DMOVE & : & specifyParameter \\
TYPE & : & DeviceSpecifier \\
ARGS & : & [DeviceType, Location] \\
CONT & : & \\
QUANT & : & all \\
DeviceSpecifier & : & DeviceType : \left[\begin{array}{lll} DMOVE & : & specifyParameter \\ TYPE & : & DeviceType \\ ARGS & : & \\ CONT & : & light \end{array}\right] \\
 & & Location : \left[\begin{array}{lll} DMOVE & : & specifyParameter \\ TYPE & : & Location \\ ARGS & : & \\ CONT & : & indoor \end{array}\right]
\end{array}\right]
\end{array}\right]
$$

Figure 2: Formal (DTAC) representation of the utterance: *Could you please turn on all the indoor lights?*

- **S(1)**: What device would you like to turn off?
- **U(3)**: The one in the bedroom.

In this example, the system detects that in order to turn off a device, the user must supply the descriptor of the actual device. This functionality is implemented as an expectation of the dialogue manager ((Quesada, Torre & Amores 2000)).

As a result the system turns off the bedroom light (Light03).

2.2.2 Queries, Anaphora and Multiple Commands

- **U(4)**: Is there any light turned on in the house?
- **S(2)**: Yes, the dimmer light in the living room is on.
- **U(5)**: Turn it off, please, and turn the garden light on.

The dialogue manager allows up to 12 different types of queries. Sentence U(4) illustrates one of these possibilities. Sentence U(5) includes multiple commands and also an anaphoric expression (turn it off) which must be solved taking into account the dialogue history. As a result of this subdialogue, the system should turn off Dimmer01 and turn on Light02.

Other advanced linguistic phenomena covered by the system include error repairs, positive and negative task accommodation, undo operations, conjuntions of device specifiers, default interpretations, plug and play (dynamic incorporation of new devices or change of the characteristics of the devices installed, even change of location).

2.3 Action Management and Device Control

In order to illustrate, analyse and evaluate the whole system in a more natural scenario we have designed, configured and installed a small house prototype. The only difference with respect to a 'real' house is its size, since our prototype had to be small enough for its manipulation and transportation.

The prototype consists of a set of lights installed over a board, in which the floor plan of a house has been drawn.

Figure 3 shows a front photograph of the prototype, with labels (in Spanish) for each 'room' in the house.

3 Conclusion and Future Work: Towards Ambient Intelligence

This paper has concentrated on the description of an agent architecture for the design and implementation of spoken dialogue systems, along

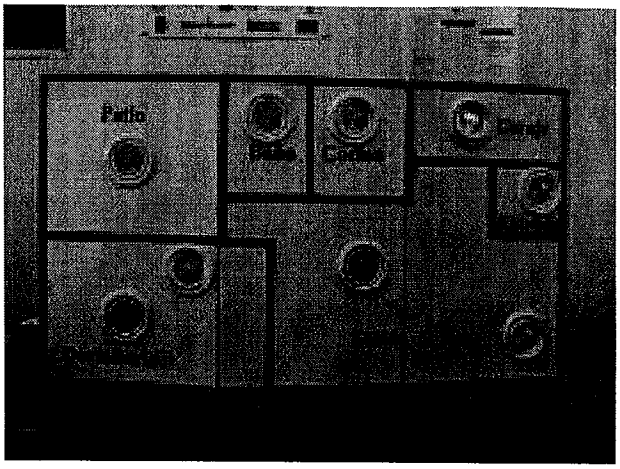


Figure 1: A house prototype with real devices

with its application to a real scenario in the home environment.

The main characteristics of the system are:

- It includes a dialogue management system which allows flexible interaction using natural language commands.
- It has been implemented over a distributed agent architecture.
- It allows speech input and output.
- It has been especially designed for the control of devices in a home environment.
- The system supports Plug and Play functionality.

Although in its current version the system supports a reduced number of control commands over real devices, the design and implementation of agents has been carried out having in mind future demands imposed by more intelligent devices.

Even though the generic goal underlying the goal of Ambient Intelligence (Ducatel et al 2001) was beyond the specific goals to be achieved by the D'Homme project, there exist points of coincidence between both approaches.

Perhaps the most relevant point in the Ambient Intelligence declaration lies in the necessity of achieving the control of devices through natural language and flexible communication with them. Obviously, the incorporation of more intelligence in the devices will necessarily involve more sophistication in the linguistic modules. Nevertheless, a critical study and evaluation of the agent architecture proposed does not foresee a serious *a priori* incompatibility with such linguistic functionality. Rather, the use of a machine translation engine as a lexical, syntactic and semantic analysis module ensures the capability of dealing with complex natural language expressions.

On the other hand, the HomeSetup and Knowledge Manager agents are not limited either by the current functionality of the devices being controlled in the current version. Perhaps the HomeSetup should evolve into an agent with more spatial coverage (something like an AmbientSetup).

Finally, the agent which would require more modifications is the Action Manager, which should support a more sophisticated set of actions. Furthermore, in an Ambient Intelligence environment, the proactive behaviour of devices is a crucial feature.

References

Amores, J. G., Quesada, J. F. 2000. *Dialogue Moves in Natural Command Languages.* Deliverable 1.1. Siridus project.

Amores, J.G., Berman, A., Bos, J., Boye, J., Cooper, R., Ericsson, S., Holt, A., Larsson, S., Milward, D., Quesada, J.F. *Knowledge and Action Management in the Home Device Environment.* Deliverable 4.1. DHomme project. October 2001.

Bos, Johan. DORIS 2001: Underspecification, Resolution and Inference for Discourse Representation Structures. In Patrick Blackburn and Michael Kohlhase, editors, *ICoS-3, Inference in Computational Semantics*, pages 117–124, 2001.

Boye, Johan, Ian Lewin, Colin Matheson, James Thomas, and Johan Bos. 2001. *Standards in Home Automation and Language Processing.* Deliverable D1.1. D'Homme project. November 2001.

Cheyer, Adam; Martin, David. (2001). The open agent architecture. *Journal of Autonomous Agents and Multi-Agent Systems*, **4(1/2)**, 143–148, March 2001.

Cooper, Robin. 1993. Generalised quantifiers and resource situations. In P. Aczel and D. Israel and Y. Katagiri and S. Peters editors, *Situation Theory and its Applications, v.3*, pages

191–212, CSLI and University of Chicago, Stanford, 1993.

http://www.ling.gu.se/projekt/dhomme

Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J., Burgelman, J-C. (compilers). 2001. *Scenarios for Ambient Intelligence in 2010.* Information Society Technologies Advisory Group and Institute for Prospective Technological Studies (Joint Research Center). `http://www.cordis.lu/ist/istag.htm`

Eckert, Miriam; Strube, Michael. 1999. Resolving Discourse Deictic Anaphora in Dialogues. In EACL '99 (pp.37-44).

Eckert, Miriam; Strube, Michael. 2001. Dialogue Acts, Synchronising Units and Anaphora Resolution. Journal of Semantics 17(1). pp 51-59.

Fraser, N.M. and J.H.S. Thornton. Vocalist: A robust, portable spoken language dialogue system for telephone applications. In *"Proc. of Eurospeech '95*, pages 1947–1950, Madrid, 1995.

Giachin, E. and S. McGlashan. 1996. Spoken Language Dialogue Systems. In Corpus-based Methods in Language Processing edited by G. Bloothooft and S. Young, Kluwer, The Netherlands, 1996.

Hulstijn, J. 1996. Presupposition accommodation in a constructive update semantics. In Durieux, G., Daelemans, W., and Gillis, S. (eds.), Proceedings of CLIN VI. University of Antwerp.

Kamp, Hans, and Uwe Reyle. *From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT.* Kluwer, Dordrecht, 1993.

Larsson, S. and D. Traum Information State and dialogue management in the TRINDI Dialogue Move Engine Toolkit. In *Nat.Lang. Engineering*, volume 6, 2000.

Lemon, Oliver, Anne Bracy, Alexander Gruenstein, and Stanley Peters. 2001. Information States in a Multi-modal Dialogue System for Human-Robot Conversation. In Proceedings Bi-Dialog, 5th Workshop on Formal Semantics and Pragmatics of Dialogue, pages 57 - 67, 2001.

Lemon, Oliver, Anne Bracy, Alexander Gruenstein, and Stanley Peters. 2001. "A Multi-Modal Dialogue System for Human-Robot Conversation", In proceedings NAACL 2001.

Lemon, Oliver, Anne Bracy, Alexander Gruenstein, and Stanley Peters. 2001. The WITAS Multi-Modal Dialogue System I", In proceedings EuroSpeech 2001.

McGlashan, S. and T. Axling. 1996. Speech Interfaces to Virtual Reality. In the Proceedings of the International Workshop on Speech and Computers, St. Petersburg, Russia, 1996

Milward, D., Knight, S. (eds). *The Initial D'Homme System Architecture* Deliverable 2.1. D'Homme project. May 2001.

Palomar, M., Martnez-Barco, P. 2001. Computational Approach to Anaphora Resolution in Spanish Dialogues. Journal of Artificial Intelligence Research 15 (2001), pp. 263-287.

Quesada, J. F., Torre, D. & J. G. Amores. 2000. *Design of a Natural Command Language Dialogue System* Deliverable 3.2. Siridus project. December 2000.

Quesada, J. F., J. G. Amores, J. Bos, S. Ericsson, G. Gorrell, S. Knight, I. Lewin, D. Milward, M. Rayner. 2001. *Configuring Linguistic Components in a Plug and Play Environment.* Deliverable 3.1. D'Homme project. October 2001.

Relevance Only

Robert van Rooy*
Institute for Logic, Language and Computation
University of Amsterdam
vanrooy@hum.uva.nl

Abstract

In this paper, a single notion of exhaustivity will be defined in terms of *relevance* that accounts for (i) standard exhaustification; (ii) scalar readings; and the intuition that mention-some answers can sometimes be completely resolving. It will be shown that this notion of exhaustivity leaves something to be done for 'only'. Moreover, an analysis will be given of 'only' in terms of relevance that is stronger than exhaustivity in a subtle way. Finally, it will be suggested how exhaustivity and 'only' could interact with disjunctions.

1 Introduction

Rooth (1985) analyzes *only* as follows (neglecting the assertion/presupposition distinction and assuming that ϕ is of type st):

(1) $[[only\ \phi]] = \{w \in [[\phi]] : \neg\exists p \in [[\phi]]^f : w \in p \text{ and } p \neq [[\phi]]\}$

In this formula, $[[\phi]]^f$ is the focus-semantic value of ϕ, a semantic value that is determined on top of its ordinary semantic value $[[\phi]]$. If ϕ is a sentence like (2a), then its focus-semantic value will be something like (2b).

(2) a. John introduced $[\text{Bill}]_F$ to Sue.

b. {[[John introduced d to Sue]] : $d \in Alt(Bill)$}

For (1) to be an acceptable rule, it is crucial to limit what the alternatives to ϕ are. It has been argued by various authors that the rule makes sense only in case the alternatives to ϕ differ from ϕ only in a constituent of type e, i.e., just for sentences of the form (2a). Only in that case it can be (naturally) assumed that the alternatives are *independent* of one another.[1] In case *groups* of individuals are treated on a par with a 'real' individual like Bill, however, even this limitation is not strong enough anymore. Analyzing sentences of the form

(3) Sue only introduced $[\text{Bill and Mary}]_F$ to Sue.

by Rooth's interpretation rule (1) gives rise to the false prediction that the following elements of the focus-semantic value of (3) are not true.

(4) a. John introduced Bill to Sue.

b. John introduced Mary to Sue.

To get rid of these false predictions, something like the following has been proposed by Zeevat (1994) and used by Butler (2001):

(5) $[[only\ \phi]] = \{w \in [[\phi]] : \neg\exists p \in [[\phi]]^f : w \in p \text{ and } p \subset [[\phi]]\}$

Actually, they are not using alternative semantics, of course, and use more something like a background-focus structure (cf. von Stechow,

* The research of this work is supported by a fellowship from the Royal Netherlands Academy of Arts and Sciences (KNAW), which is gratefully acknowledged. This note originally figured as a discussion paper for the 'only' day organized by Alastair Butler, Paul Dekker, and Henk Zeevat. I would like to thank Henk Zeevat for coming up with a problem to work on; the participants of the 'only' day – especially Maria Aloni, David Atlas, Alastair Butler and Katrin Schulz – for discussion; and the anonymous reviewers of Edilog 2002 for their comments and suggestions.

[1] If all meanings of type $\langle e, \langle e, t\rangle\rangle$ would be alternatives to generalized quantifiers like *a man* and *two men*, Rooth's rule is in deep trouble in case these latter quantifiers would be in focus. See von Stechow (1991) and especially Bonomi & Casalegno (1993) for discussion.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 155-160.

Krifka). Also, their analysis is not an analysis of 'only', but rather one of *exhaustification*. Assume a sentence is of the form $\langle B, F\rangle$, and that F and F' range over groups. Then they treat the exhaustification of $\langle B, F\rangle$, $exh\langle B, F\rangle$, more along the following lines:

(6) $[[exh\langle B, F\rangle]] = \{w \in [[B(F)]] : \neg\exists F'[F' \neq F \wedge B(F')(w) \wedge [[B(F')]] \subseteq [[B(F)]]]\}$

Thus, $exh(B, F)$ is true in world w just in case $B(F)$ is the *most informative* true proposition among its alternatives. Depending on the predicate/question (*Who is sick*, *How far can you jump*, or *In how many seconds can you run the 100 meters*), the group denoted by focus-value F in the answer is predicted to be the *maximal* (first two questions) or *minimal* (last question) value that is true.

Although this is a very pleasing result, it gives, as Zeevat (pc) noted himself, rise to a problem too: if we exhaustify already in case of *free* focus, why would we ever use *only*? The goal of this paper is to help to resolve this question.

2 Scalars and utility

Bonomi & Casalegno (1993) have argued in the last part of their paper that 'only' can also have a 'scalar' reading where it rules out alternatives that are 'higher' on a certain scale. Consider the following sentences:

(7) a. Peter only saw [the secretary of state]$_F$.

b. Peter did not see the special envoy to the Middle East.

Bonomi & Casalegno point out that (7a) can easily mean that the secretary of state was the *highest* official Peter got to see, in which case it would exclude his having seen the president, but not his having seen the special envoy to the Middle East. Notice that these (lack of) implications don't follow from rule (6): That Peter saw the secretary of state is not entailed by him having seen the president, nor does it entail that he saw the special envoy to the Middle East. The scalar reading of 'only' is perhaps most obvious in *games*. Suppose we are playing a card game against each other, and the *goal* is to win, and winning depends exclusively on who has the highest card. The king of diamonds is higher than the jack of hearts. You show me the king of diamonds and ask: 'What do you have?' Although I have three cards in my hands, I say 'I *only* have the jack of hearts', meaning that this is the highest card I have and (thus) the only one that counts.

To account for this scalar reading involving a scale '$>$' between propositions – and notice that this is one that cannot be reduced to entailment – Bonomi & Casalegno propose that 'only' not only has an exhausification meaning, but also a scalar one, and they analyze this as (something like) (8).

(8) $[[only\ \phi]] = \{w \in [[\phi]] : \neg\exists p \in [[\phi]]^f : w \in p \text{ and } p > [[\phi]]\}$

But, of course, not only 'only', but also free focus gives rise to scalar readings. And, as emphasized by Hirschberg (1985), also to ones where the scales cannot be defined in terms of entailment. Assuming that 'exhaustification' simply means 'picking out (one of) the best reading(s)', to account for scalar readings for free focus we have to give an exhaustification-meaning defined in terms of scales too:

(9) $[[exh\langle B, F\rangle]] = \{w \in [[B(F)]] : \neg\exists F'[B(F')(w) \wedge [[B(F')]] > [[B(F)]]]\}$

I propose that to interpret **free focus** – i.e. the use of focus that does not associate with an explicit operator like *only*, *even* or *too* –, the latter rule is the basic one,[2] and thus that we should give up exhaustification rule (6) proposed by Zeevat. But how, then, can we account for the standard exhaustification effect? I propose that the ordering relation $>$ is always one of **relevance**: in w, $[[B(F)]]$ is (one of) the most relevant true proposition(s) among the alternatives. Optimizing relevance simply means optimizing *utility* to be determined in a decision theoretic framework.

3 Decision Theory and Utility of propositions

Decision theory is a theory that tells us which action an agent will or should perform given his

[2] the second conjunct as implicature, perhaps.

beliefs and his desires. It is normally assumed that the beliefs of an agent can be represented by a *probability function*, a function from events to real numbers that add up to one. To represent the preferences, or desires, of the agent, on the other hand, we need not only a probability function, but also an additional function that assigns to each action a payoff given an event, a *utility function*. A *decision situation* can then (in the finite case) typically be represented by a *decision table*, which identifies the conditional gain associated with every possible combination of the acts under consideration and the possible mutually exclusive events with their probabilities of occurrence:

				Actions		
World	Probability	a_1	a_2	a_3	a_4	a_5
e_1	1/9	5	1	4	2	7
e_2	2/9	3	3	5	6	1
e_3	4/9	1	3	0	3	3
e_4	2/9	6	2	3	5	2

Once we have both the probability and the utility function around, the most sensible way to determine which action the agent should perform is by *maximizing the expected value*. The expected value of each action is determined by multiplying the conditional values for each action/event combination by the probability of that event, and summing these products for each act. The expected utility of action a_1, for instance, is $(5 \times 1/9) + (3 \times 2/9) + (1 \times 4/9) + (6 \times 2/9) = 27/9 = 3$. The action which should be chosen is a_4, because it is the action which maximizes the expected utility, $36/9 = 4$.

Assuming that the relevant possible events are indeed exclusive, we can say that the probability functions take *worlds* as arguments. If we assume that the utility of performing action a in world w is $U(w, a)$, we can say that the expected utility of action a, $EU(a)$, with respect to probability function P is

$$EU(a) \quad = \quad \sum_w P(w) \times U(w, a).$$

The decision situation of the agent as described above is also known as a *decision problem*, and can in general be modeled as a triple, $\langle P, U, A \rangle$, containing (i) the agent's probability function, P, (ii) her utility function, U, and (iii) the alternative actions she considers, $\mathcal{A}$. Given such a decision problem, we might denote the utility of the action with maximal expected utility of this decision problem by UV(Choose now):

(10) $UV(\text{Choose now}) \quad = \quad max_a EU(a).$

Before we can determine the utility of new information A, we first have to determine the expected utility of an action conditional on learning A. For each action a, its conditional expected utility with respect to proposition A, $EU(a, A)$ is $\sum_w P(w/A) \times U(a, w)$. Now we can denote the utility of the action which has maximal utility after learning A by UV(Learn A, choose later):

(11) $UV(\text{Learn } A, \text{ choose later}) \quad = \quad max_a EU(a, A).$

The utility of proposition A, $UV(A)$, can be determined as the difference between the expected utility of the action which has maximal expected utility in case you are allowed to choose *after* you learn that A is true, and *before* you learn that A is true:

$$\begin{array}{lcl} UV(A) & = & UV(\text{Learn } A, \text{ choose later}) \\ & & -UV(\text{Choose now}) \\ & = & max_i EU(a_i, A) - max_i EU(a_i) \\ & = & [max_i \sum_w P(w/A) \times U(a_i, w)] \\ & & -[max_i \sum_w P(w) \times U(a_i, w)] \end{array}$$

Notice that this general rule leaves open how utility should be measured, i.e., what does U depend on? For linguistic applications, two sorts of notions of relevance have been proposed. The first notion might be called an *argumentative*, or *goal oriented* notion of relevance. The goal is to make a proposition h common ground, and the relevance of proposition A is measured in terms of in how far it helps to increase h to be common ground. Merin (1999) has proposed to measure goal-oriented relevance in terms of the Peirce/Turing/Good notion of *weight of argument*, $r_h(A)$, and used it to account for several linguistic phenomena.

(12) a. $r_h(A) \quad = \quad log \frac{P(h/A)}{P(\neg h/A)}$

b. $A >_h B$ iff $r_h(A) > r_h(B)$

According to this notion of relevance it can be the case that $A >_h B$ although A and B are propositions that are logically, or informationally, *independent* of each other. In terms of this notion we can define scales that cannot be reduced to entailment. Indeed, in terms of this notion we can account for the scalar readings mentioned above. It is important to note that in case our agent wants to argue for h, we can show (cf. van Rooy, 2002) that $UV(A)$ reduces to $r_h(A)$.[3]

In other cases utility or relevance is thought of in more Gricean and cooperative terms. The relevance of an assertion is measured in terms of in how far it helps to resolve an other agent's question, or decision problem. In case Q is the underlying question/decision problem, Groenendijk & Stokhof (1984) value A as a (strictly) better assertion/answer than B, $A >_Q B$, just in case A eliminates more answers of Q than B does:

(13) $A >_Q B$ iff $\{q \in Q : q \cap A \neq \emptyset\} \subset \{q \in Q : q \cap B \neq \emptyset\}$

Notice that this notion comes down to (one sided) *entailment* in case $Q = W$, where W denotes the finest-grained partition corresponding to the question how the *world* looks like. In van Rooy (2002) it is shown that this comparative notion of relevance is induced by the general $UV(\cdot)$ function, in case (i) only truth is at stake, and (ii) probabilities are irrelevant. Thus, the comparative relevance-relation '$>$' used in (9) and defined in terms of $UV(\cdot)$ reduces either to (12b), corresponding with a *scalar* reading, or to (13), which in special cases comes down to entailment, giving rise to the standard exhaustive reading.[4]

4 Resolving answers and 'only'

Notice, though, that if '$>$' comes down to '$>_Q$' it might be the case that the background question/decision problem Q does *not* give rise to a partition. It might be the case, for instance, that Q corresponds with (14) that most naturally gives rise to a *non-partitional* set of propositions.

(14) Where can I buy an Italian newspaper?

Suppose that you can buy an Italian newspaper at the station in u and w, and at the palace in v and w, then Q can be thought of as non-partitional $\{\{u, w\}, \{v, w\}\}$. If I answer (14) by saying 'At the station', this intuitively does not rule out that I cannot buy one at the palace. And, indeed, that's what comes out according to our rule (9). In both worlds u and w, there is no alternative true answer that would eliminate more elements of Q than answer 'At the station' does. Thus, although optimizing relevance measured in terms of '$>_Q$' gives rise to a *mention all* reading in case Q is a partition, in case Q is not a partition it naturally allows for *mention-some* readings. Thus, according to our analysis of exhaustivity, (9), an exhaustive reading of an answer might be equal to a mention-some reading. I take this to be a desirable feature of the analysis, because in specific circumstances a mention-some reading of the answer 'feels' like being completely *resolving* (cf. Ginzburg, 1995).

Notice that this feature leaves something open to be done for *only*. Indeed, I propose that (9) should *not* be the rule that analyzes this particle. As observed by Butler (pc), it seems that 'only' restores (classical) exhaustivity. Whereas answer 'At the station' to question (14) allows naturally for a mention-some reading, this is not the case anymore for answer 'Only at the station'. How could we account for the extra effect of 'only'? I will propose that this is actually very easy. The meanings of 'exhaustivity' and 'only' are closely related, but the meaning of the latter is somewhat stronger: whereas exhaustivity says that *no better* true answer could be given, 'only' says that there is even *no alternative* true answer that is *equally good*. Slightly differently: whereas saying that F is the exhaustive answer to the question means that there is no alternative that (i) is true, and (ii) *more* relevant than $B(F)$; saying that F associates with 'only' means that there is no alternative different from F such that (i) it is true, and (ii) it is more or *equally* relevant as $B(F)$.

[3] More precisely, in those case $UV(\cdot)$ reduces to a function that is continuously monotone increasing to $r_h(\cdot)$. Thus, to a function that gives rise to the same ordinal scale as $r_h(\cdot)$.

[4] Actually, under special cases also $r_h(\cdot)$ comes down to entailment, see van Rooy (2002).

(15) $[[only\langle B, F\rangle]] = \{w \in [[B(F)]] : \neg\exists F'[F' \neq F \wedge B(F')(w) \wedge [[B(F')]] \geq [[B(F)]]]\}$

Thus, '$only B(F)$' says not just that $B(F)$ is *a* best answer, but it also says that it is *the unique* one. In our model described above it will hold that answer 'Only at the station' will just denote $\{u\}$: world w is ruled out, because there is an alternative answer to 'At the station', nl. 'At the palace', that (i) is true in w, and (ii) is equally relevant as 'At the station'.

Notice that if '$>$' is measured in terms of underlying question Q, and in case $Q = W$, (15) comes out as being equal to (6). I conclude that Zeevat's problem about the effect of 'only' can be met by making a subtle distinction between exhaustification and 'only'.

5 Disjunction and exhaustification

Our analysis seems straightforward, but also problematic for the case of explicitly non-exhaustive answers: in case the answer (focus) is in *disjunctive* form. If our question is of a typical *mention-all* kind, like (16a), the exhaustified reading of disjunctive answer (16b) will be the empty proposition:

(16) a. Who did John invite?

b. Bill or Mary.

The reason is that if we assume that the whole term-answer is in focus, we can conclude from (16b) that the more informative term-answers 'Bill' and 'Mary' must be false, and thus that no world will be in the extension of the exhaustive value of background (16a) applied to focus (16b).

As convincingly argued by Simons (2001), however, an answer like (16b) is appropriate only in case we assume that the disjuncts themselves are appropriate answers. Putting it slightly different, this suggests that a disjunctive answer is only appropriate in case each of its disjuncts by themselves are *exhaustive*, or *resolving*, answers. This suggests that disjunctive answers should be interpreted as follows:

(17) $[[exh(A \vee B)]] = [[exh(A)]] \cup [[exh(B)]]$

In case (16a) gives rise to a mention-all reading this has the effect that answer (16b) is equivalent to (18)

(18) Bill and nobody else or Mary and nobody else.

Notice, however, that this does not, and should not, come out in case (16b) is given as answer to a question like (19) that typically is interpreted as being of the mention-some variety:

(19) Who has got a light?

But meaning (18) can be arrived at in this case as well when we front (16b) with 'only'. This could be accounted for by assuming that 'only' distributes over the disjuncts. In case 'only' does not distribute over the disjuncts, the disjunctive answer fronted with 'only' is now predicted to mean (20):

(20) Only Bill has got a light, only Mary, or only both of them.

6 Conclusion and outlook

In this paper I have shown how a single notion of relevance can account for both the standard concept of 'being a resolving answer' and for scalar readings of answers. In other work I show that a similar analysis can be used to account for the different kinds of readings of *questions* as well. Our relevance-induced notion of exhaustivity leaves something to be done for 'only' and in this paper I show how relevance can help to determine the meaning of this particle as well.

On the other hand, I agree with Zeevat that there is more to 'only' than just the exhaustification effect. According to Zeevat, '*only* ϕ' indicates that ϕ is less than *expected*. I believe that this should rather be 'less than *hoped* for'. According to Merin (1994), 'only' indicates *preference reversal* (John did (only) a little). It seems to me that our notion of relevance can be used to account for this intuition as well. Perhaps this can also explain why 'only nobody' and 'only everybody' are bad answers. A full analysis has to wait for another occasion, however.

Our analysis of 'only' also suggests a similar analysis of another particles like 'just'. Also this particle seems to give rise to both a scalar and an exhaustification reading. But, then, what is the difference between these two particles? According to one of my informants (David Atlas) 'just' more naturally gives rise to a scalar

reading than 'only' does. According to another (Egon Stemle), it is much more natural to say 'just nobody' than to say 'only nobody'. The standard analysis of such particles makes use only of informativity. Our additional use of preferences might help to give a more appropriate treatment of this particles as well. Something similar holds for 'even'. And indeed, Merin (1999) suggests that the scale involved for the analysis of 'even' should be induced by relevance, rather than by informativity. A closer look at these issues, however, must be left for another paper.

References

Bonomi & Casalegno (1993), 'Only: association with focus in event semantics'. *Natural Language Semantics*, **2**: 1-45.

Butler, A. (2001), 'Sentences with exhaustification', In: Bunt et al (eds.), *Proceedings of the 4th International Workshop on Computational Semantics*, Tilburg.

Ginzburg, J. (1995), 'Resolving Questions, I', *Linguistics and Philosophy*, **18**: 459-527.

Groenendijk, J. and M. Stokhof (1984), *Studies in the Semantics of Questions and the Pragmatics of Answers*, Ph.D. thesis, University of Amsterdam.

Hirschberg, (1985) *A theory of scalar implicatures*, Ph.D. dissertation, UPenn.

Merin, A. (1994), 'The algebra of elementary social acts', In: S. Tsohatzidis (ed.), *Foundations of Speech Act Theory*, London, Routledge.

Merin, A. (1999), 'Information, relevance, and social decisionmaking', *Logic, Language, and Computation, Vol. 2*, In L. Moss et al. (eds.), CSLI publications: Stanford.

Rooth (1985), *Association with focus*, PhD dissertation, Amherst.

Rooy, R. van (2002), 'Utility, informativity, and protocols', In Bonanno et al (eds.), *Proceedings of LOFT 5: Logic and the Foundations of the Theory of Games and Decisions*, Torino.

Simons, M. (2001), 'Disjunction and alternativeness', *Linguistics and Philosophy*, **24**.

Stechow, A. von (1991), 'Focusing and backgrounding operators', In: W. Abraham (ed.), *Discourse Particles*, Amsterdam: John Benjamins.

Zeevat, H. (1994), 'Questions and Exhaustivity in Update Semantics', In: Bunt et al. (eds.), *Proceedings of the International Workshop on Computational Semantics*, Institute for Language Technology and Artificial Intelligence, Tilburg.

Resolving Fragments Using Discourse Information

David Schlangen **Alex Lascarides**

ICCS / Division of Informatics
University of Edinburgh
{das|alex}@cogsci.ed.ac.uk

Abstract

This paper describes an extension of RUDI, a dialogue system component for 'Resolving Underspecification with Discourse Information' (Schlangen et al., 2001). The extension handles the resolution of the intended meaning of non-sentential utterances that denote propositions or questions. Some researchers have observed that there are complex syntactic, semantic and pragmatic constraints on the acceptability of such fragments, and have used this to motivate an unmodular architecture for their analysis. In contrast, our implementation is based on a clear separation of the processes of constructing compositional semantics of fragments from those for resolving their meaning in context. This is shown to have certain theoretical and practical advantages.

1 Introduction

Non-sentential utterances that denote a proposition or a question are pervasive in dialogues. This paper describes an extension of RUDI, a dialogue system component for 'Resolving Underspecification with Discourse Information' (Schlangen et al., 2001). The extension handles the resolution of the intended meaning of such fragments in the context of scheduling dialogues. The system models the behaviour of fragments which are used to perform two types of frequently occurring speech acts: (a) question answering, as in (1); and (b) what we call question-elaboration or *Q-Elab* (following SDRT, (Asher and Lascarides, 1998)). *Q-Elab* is a subclass of questioning where all answers to the question elaborate a plan to reach the goal of a prior utterance, illustrated in (2).

(1) A: What time on Tuesday is good for you?
B: 3pm. / B′: #At 3pm. / B″: #The hotel.

(2) A: Let's meet next week.
B: (OK.) Thursday at three pm?

Our claim is that resolving the intended meaning of fragments is a by-product of establishing which speech act was performed, i.e. of establishing the *coherence* of the fragment's contribution to the dialogue. Different speech acts impose different constraints: coherent short answers must meet certain syntactic constraints (which predict B′ in (1) is incoherent) and semantic constraints (which predict B″ is incoherent), while *Q-Elab* imposes constraints on content but not on syntax.

In contrast to (Ginzburg and Sag, 2001), henceforth G&S), who incorporate contextual resolution of fragments into the grammar (see Section 4), we offer a fully compositional analysis, separating grammar from pragmatic processing. This has three advantages. First, the grammatical analysis of fragments is uniform; contextual variation in their meaning is accounted for in the same way as it is for other anaphoric phenomena, via inferences underlying discourse update. This yields the second advantage: resolving fragments is fully integrated with resolving other kinds of underspecification, such as bridging (Clark, 1975). For instance, the system resolves B in (1) to mean "3pm on [the Tuesday A was referring to] is good for B".[1] Third, it enables us to make only those semantic distinctions that are required by the dialogue-purpose. For example, the fragment in (2) will in our system be resolved with a generic *good_time* predicate, abstracting over possible resolutions like "How about we meet on Thursday at ...?" or "Can you make Thursday at ...?".

In the next section, the theoretical basis of the dynamic semantic approach realised in RUDI will be described: in a nutshell, an (underspecified) compositional semantic representation of the current clause is constructed (see Sections 2.1 and 2.2), which is used to update the representation of the discourse context. The co-dependent tasks of computing speech acts and goals and resolving semantic underspecification are tackled as a by-product of computing this update. For this, we use Segmented DRT (SDRT, (Asher and Lascarides, in press)), where *update* computes the pragmatically preferred interpretation (see Section 2.3.1). This is formulated within a precise nonmonotonic logic,

[1] Details on the resolution of bridging relations can be found in (Schlangen et al., 2001); we focus here on the new treatment of fragments.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 161-168.

in which one computes the *rhetorical relation* (or equivalently, the speech act type) which connects the new information to some antecedent utterance. This speech act places constraints on content and possibly even on the form of their arguments, and also on the goals of the utterances; these constraints serve to resolve semantic underspecification (Section 2.3.2). In Section 3 we then describe how this formalisation is implemented in our computer program, and we give a worked example. We conclude with a comparison with related work (Section 4), and with an outlook on further work (Section 5).

2 Theory

2.1 A Compositional Semantics for Fragments

For compositional semantic analysis we use Minimal Recursion Semantics (MRS, (Copestake et al., 1999)), a language in which (sets of) formulae of a logical language (the *base language*) can be *described* by leaving certain semantic distinctions unresolved. This is achieved via a strategy that has become standard in computational semantics (e.g., (Reyle, 1993)): one assigns labels to bits of base language formulae so that statements about their combination can be made in the 'underspecification' language. The (first-order) models of formulae of this latter language then can be seen as standing in a direct relation to formulae of the base language; $M \models \phi$ then means that the base-language formula corresponding to M is described by the MRS ϕ.[2] By way of example, (3) shows an MRS-representation of "Everyone loves someone", where so called elementary predications (EPs) are labelled with *handles* (h_n), with h being the top handle that outscopes all others; '$h_1 =_q h_2$' stands for an 'outscopes' relation between EPs where only quantifiers can be scoped in between h_1 and h_2; *prpstn_rel* signals that the MRS describes a proposition.

(3) $\langle h, e, \{h{:}prpstn\_rel(h_1), h_2{:}love\_v\_rel(e, x_1, x_2),$
$h_6{:}\_every\_rel(x_1, h_8, h_9),$
$h_{10}{:}\_person\_rel(x_1),$
$h_{11}{:}\_some\_rel(x_2, h_{12}, h_{13}),$
$h_{14}{:}\_person\_rel(x_2)\},$
$\{\ h_1 =_q h_2, h_8 =_q h_{10}, h_{12} =_q h_{14}\ \}\rangle$

The compositional semantics of fragments leaves more information unresolved than just semantic scope, however. All we know about the meaning of fragments like those in (1) and (2) independent from their context is: (a) they will resolve to a proposition or a question respectively, of which (b) the main predicate is unknown, but (c) one participant in the main event is specified although its exact role isn't. We represent this with an anaphoric relation *unknown_rel*, and so the NP-fragment "3 pm" (regardless of the context it stands in) is represented as:

(4) $\langle h, e, \{h{:}prpstn\_rel(h_1),$
$h_2{:}unknown\_rel(e, x),$
$h_6{:}def\_rel(x, h_8, h_9),$
$h_{10}{:}numbered\_hour\_rel(x, 15)\},$
$\{\ h_1 =_q h_2,\ h_8 =_q h_{10}\ \}\rangle$

The *unknown_rel* acts as a 'place-holder' for a potentially complex sub-formula; more precisely it is a constraint on the form of the described (base-language) formulae, namely that they contain at this place a subformula, which in the case of (4) must have e and x amongst its variables. It is important to note that *unknown_rel* is *not* a second order variable, and it is *not* something that simply gets replaced by a predicate symbol of the same arity. Rather, *unknown_rel* is a constraint more like the $=_q$-constraints, constraining the form of the described formulae. It is anaphoric, because the subformula that is to be inserted at this point in the described formula is not known by the grammar, but must be provided by the context. Procedurally, the relation can simply be understood as a signal for attention to a resolution mechanism; we will describe this mechanism below in Section 3.

PP-fragments differ only in that a further relation (corresponding to the preposition) is present. We implemented our semantic analysis within the wide-coverage English Resource Grammar (ERG, see `http://lingo.stanford.edu`), in which functional prepositions like that in (5) are represented in the MRS, even though they might eventually translate into the trivial (i.e., always true) predicate $\top$ in the base language.[3]

(5) A: On what time did we agree for the meeting? — B: On 3 pm.

We exploit this feature here, representing B's reply in (5) as shown in (6). Note that the handle of the EP corresponding to the preposition is constrained to be subordinate to that of the *unknown_rel* (i.e. $h_2 \leq h_3$), not just $=_q$. This is because this *unknown_rel* can resolve to a complex formula containing not only the predicate for the phrasal verb that selects for *on*, but possibly also other scope-

[2] The authors do not provide such a semantics for MRSs in (Copestake et al., 1999), but it is relatively straightforward to do so, for example along the lines of (Egg et al., 2001), or, as we use in our system, (Asher and Lascarides, in press). Also note that we do not make any assumptions about the base language and its logic here; the descriptions are compatible with it being static first order predicate logic, or a dynamic logic like DRT (Kamp and Reyle, 1993).

[3] This is independently motivated because it makes the grammar monotonic (i.e., every lexical item introduces an EP, and the LF of a mother-node contains all those of its daughters), a property that is desirable for generation (Shieber, 1986). We will make use of this feature below in section 2.3.

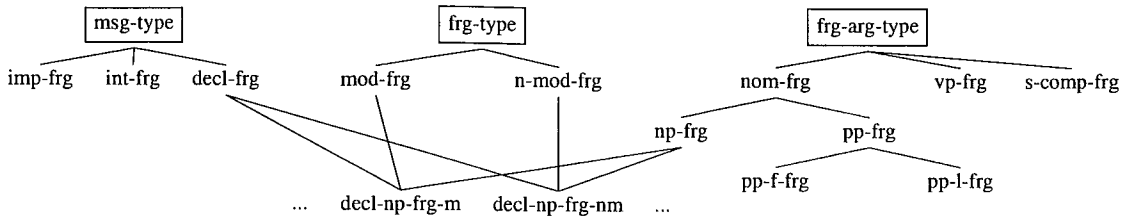


Figure 1: An extract of the construction hierarchy for fragments

bearing elements that aren't quantifiers, eg. in the context of a question like "what time might Peter agree on?".

(6) $\langle h, e, \{h{:}prpstn\_rel(h_1),$
$h_2{:}unknown\_rel(e,x),$
$h_3{:}\_on\_rel\_s(e',x),$
$h_6{:}def\_rel(x,h_8,h_9),$
$h_{10}{:}numbered\_hour\_rel(x,15)\},$
$\{\ h_1 =_q h_2, h_8 =_q h_{10}, h_2 \leq h_3\}\ \rangle$

2.2 A Grammar of Fragments

The grammar rules that produce these MRSs are relatively straightforward. Fragments are treated as phrases,[4] possibly modified by adverbs. As (7) shows, only scopally modifying adverbs are allowed.

(7) A: When shall we meet?
B: Maybe at 3 pm on Sunday. / *Quickly at 3pm on Sunday.

We will allow both sentential modification, where an adverb attaches to an S[*frag*], and 'VP'-modification, where the adverb is selected by the fragment rule (eg. "not 3pm"). Semantically, the difference amounts to whether the whole proposition (which is to be resolved) is modified or only the *unknown_rel*.

In a pseudo phrase-structure notation, the rules are of the form 'S-frag $\rightarrow$ (ADV) XP'. We formalise this in a version of HPSG that allows *constructions* (Sag, 1997), i.e. phrase-types that make a semantic contribution. The fragment-signs are organised along three dimensions, as shown in Figure 1: the message type, i.e. whether they resolve to a proposition, a question or a request; whether they are modified by an adverb or not; and what the type of the argument is (i.e., whether it's for example an NP-fragment or a PP-fragment).

Figure 2 gives an NP-fragment sign in a tree-style notation. It shows how the CONTent of the fragmental sentence (an MRS in AVM notation) is a combination of the contribution of the construction (C-CONT) and the content of the fragment-phrase, with the index of that phrase (5) being the argument to the *unknown_rel* coming from C-CONT.

Note that since the semantic representations do not need any information about the context of the utterance, the grammatical rules do not need to assume any additional machinery that is not independently motivated in the ERG. Further, since these rules are implemented in a wide-coverage grammar they are shown to be compatible with the analyses of a wide variety of linguistic constructions.

2.3 Discourse Update and Resolution of Underspecification

2.3.1 SDRT

As we said earlier, we use SDRT to compute the pragmatically preferred update of the context with new utterances. The co-dependent tasks of computing speech acts and goals and resolving semantic underspecification are tackled as a by-product of computing this update. This update is formulated within a precise nonmonotonic logic, in which one computes the *rhetorical relation* (or equivalently, the speech act type) which connects the new information to some antecedent utterance. This speech act places constraints on content and possibly even on the form of their arguments, and also on the so-called *speech act related goals* or SARGs (these are the goals which are conventionally associated with utterances of various forms; see (Asher and Lascarides, 1998) for details). These constraints serve to resolve semantic underspecification.

The rhetorical relations which are relevant here are:[5] *Q-Elab*(α, β), Question Elaboration, where β is a question where any possible answer to it elaborates a plan for achieving one of the SARGs of α, as in (2); and *IQAP*(α, β), Indirect Question Answer Pair, where α is a question and β conveys information from which the questioner can infer a direct answer to α, as in (1).[6] Note that these speech

[4] This goes back to (Morgan, 1973); explicit rules can be found in (Barton, 1990). We ignore for now more complicated examples like 'A: Does John devour or nibble at his food? — B: Oh, John devours.'

[5] RUDI computes other speech acts as well, such as *plan-correction* and *plan-elaboration*, but currently only for full sentences and not for fragments.

[6] We only look at fragments that are *direct* answers here; this is a subclass of *IQAP* where the direct answer follows trivially from β.

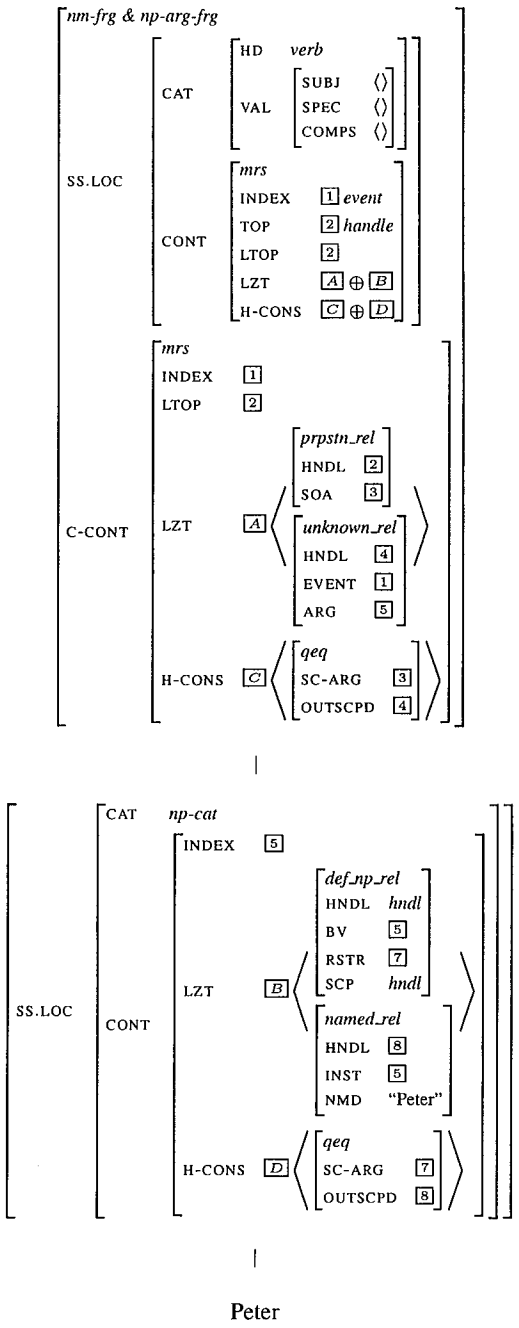


Figure 2: "Peter" as a declarative fragment.

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

These speech acts are computed via default rules; those for *Q-Elab* and *IQAP* are given below. In these rules, $\langle \tau, \alpha, \beta \rangle$ means β is to be attached to α with a rhetorical relation (α and β label bits of content) where α is part of the discourse context τ; α :? means that α is an interrogative, and $A > B$ means *If A then normally B*:[7]

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

Q-Elab stipulates that the default role of a question is to help achieve a SARG of a prior utterance. IQAP stipulates that the default contribution of a response to a question is to supply information from which the questioner can infer an answer. Thus inferences about speech acts, and hence about implicit content and goals, can be triggered (by default) purely on the basis of sentence moods.

One main tenet of SDRT is that it is desirable to separate the logic of information content from that of information *packaging* (the logic in which constructing logical form takes place), because the former will be at least a first-order language, and hence undecidable, whereas constructing logical forms should be a decidable undertaking (Asher and Lascarides, 1998 & in press). MRSs can be seen as part of the latter logic. Since the underspecification arising from fragments can be resolved on this description level, we are only concerned here with MRSs and can neglect the DRSs they are meant to describe in this theory.

2.3.2 Fragmental Questions and Answers

We now address resolving the underspecification indicated by *unknown_rel*. In particular, we argue that there are certain constraints on the *form* of fragments that stand in an *IQAP*-relation to a prior utterance, whereas *Q-Elab*-fragments need satisfy only *semantic* constraints. These different constraints can be motivated with a look at the semantics of the speech acts given above. *Q-Elab*s constrain their arguments on the level of discourse plans, i.e. on a purely semantic level, whereas *IQAP*s connect the utterances in a tighter way, given the semantics of answerhood with its function-application aspect. Note that these constraints are not mutually exclusive, and so an utterance can stand in both of these relations to the context.

Syntactic Parallelism in Fragmental Answers We begin with a look at complement questions like (9) below. Intuitively, one can say that there is a 'hole' in such questions, marked syntactically by the *wh*-phrase and semantically by a variable (be that bound by a λ-operator, as in (Groenendijk and Stokhof, 1984) or by a quantifier, as in the ERG).

[7] (Asher and Lascarides, 1998) show that these rules can be derived from a precise model of rationality and cooperativity.

(9) A: What date did we agree on for the meeting?
B: Not (on) next Monday.

This initially suggests that to resolve the content of the fragment, one could attempt to do syntactic reconstruction, 'plugging' the syntactic structure of the fragment into the (syntactic) 'hole' in the question (cf. (Morgan, 1973)). Unfortunately, as G&S attest, such a strategy fails for some cases; eg. for (9) above: "we agreed on not next Monday for the meeting." is not a well-formed sentence.

On the other hand, G&S also attest that a purely semantic reconstruction, where the semantic representation of the fragment is 'plugged into' the (semantic) 'hole' in the question, is also unsatisfactory. Certain grammatical idiosyncrasies seem to persist beyond sentence boundaries. This can also be shown with (9). The preposition in the question is generally considered to be semantically empty and to only serve a grammatical function. Nevertheless, only this particular semantically empty preposition can occur in a fragmental answer to this question (although it is optional). If all we have is a logical form of the fragment phrase, we cannot express this restriction on the *form* of that phrase.[8] Moreover, we couldn't rule out such functional PPs as answers to NP-questions, since these PPs are not distinguishable semantically from NPs.

Another English example with which this syntactic parallelism can be shown is (10) (from G&S, p.300). Here the fragmental answers must be of the syntactic category required by the verb in the question (VP[*bse*] and VP[*inf*], respectively), even though the semantic objects denoted by these VPs presumably are of the same type.[9]

(10) a. A: What did he make you do?
B: Sing
b. A: What did he force you to do?
B: To sing.

We can now formulate a preliminary version of the constraint on fragmental *IQAP*s:[10]

(11) IQAP-frag: $IQAP(\alpha,\beta) \wedge frag(\beta) \rightarrow syn\text{-}par(\alpha,\beta) \wedge resolve(\alpha,\beta)$

In words, if a fragmental utterance β answers an antecedent question α, then a certain (yet to be specified) syntactic-parallelism has to hold between α and β (note that this means that syntactic information must be accessible to this rule), and the semantic information contained in α must be sufficient to resolve the underspecification in β.

G&S set up the syntactic parallelism as identity of category between *wh*-phrase and fragment-phrase. However, this analysis cannot be straightforwardly extended to questions where the *wh*-phrase is an adjunct rather than a complement, such as "When shall we meet?". The *wh*-phrase of this question is a PP and yet acceptable short answers to it include NPs such as "three pm". Further, it seems questionable to stipulate that the *wh*-phrase in a question like "what time should we meet at?" (which can be answered with a PP-fragment) is syntactically a PP. Thus strict identity of syntactic category of question and answer seems too strong. We will realise the syntactic parallelism with a combined syntactic / semantic constraint on resolved fragments, as described in Section 3. Our analysis of short answers to both complement-questions (e.g., "What time did we agree on?") and adjunct-questions (e.g., "When did we agree to meet?") will be uniform.

One use of Fragmental Questions Fragmental questions can be used to help further the purpose of a dialogue: (12) gives examples of successful and of unsuccessful fragmental *Q-Elab*s.

(12) A: Let's meet next week.
B: (OK.) Monday? / On Monday? / #3 pm? / #Peter? / #Week 3 of term?

To rule out the infelicitous replies in (12) we do not need to look at the *form* of the antecedent; complying with the *semantic* constraints imposed by *Q-Elab* is enough. In this domain, the speech act related goal of an utterance α is to identify a time to meet within some time interval (which we call SARG_α). So if a question β proposes a time t_β that is included in SARG_α (i.e., $temp\_inc(\text{SARG}_\alpha, t_\beta)$ holds), then all answers to β will elaborate a plan to achieve α's goal, as required: some answers restrict the search to t_β; while the others rule out t_β from the search. Hence, the only constraint on *Q-Elab* we need here is the following (see (Schlangen et al., 2001) for further details of the interplay between speech acts, semantics and goals):

(13) Q-Elab:
$Q\text{-}Elab(\alpha,\beta) \rightarrow temp\_inc(\text{SARG}_\alpha, t_\beta)$

All the infelicitous replies in (12) are now ruled out: the first because of uniqueness constraints on antecedents to anaphoric definites like "3pm"; the second because it can't resolve to a question which satisfies the constraint (13). "Week 3 ..." is also

[8] As we said earlier, such semantically empty prepositions are in fact represented in the MRSs produced by the ERG; MRSs however are *descriptions* of logical forms, and keeping these empty prepositions in is a way of retaining syntactic information. We will exploit this in the resolution strategy described below.

[9] For a further discussion of the exact extent of this parallelism see (Schlangen, 2002).

[10] Given the semantics of the relation *resolve* explained below, *IQAP*-fragments are all direct answers.

ruled out as a *Q-Elab*, since the time interval referred to is *identical* to the SARG of the antecedent. It is nevertheless a coherent reply, standing in a relation of *Clarification* to A's utterance, which we do not deal with in this paper. G&S analyse clarification questions, but to distinguish between this speech act and the speech act of *Q-Elab* would require them to complicate their *grammatical* analysis considerably.[11]

3 Implementation

3.1 Overview of the system

Reflecting the modularity of the underlying theory, RUDI divides the update process into several stages. We only give a brief overview of the system here, referring the interested reader to (Schlangen et al., 2001) for details on the algorithm. We have made some changes to the set-up (besides adding resolution of fragments), though, which will be described in some more detail below.

The structure of the system is shown schematically in Figure 3. The input to the system are MRSs coming out of the grammar. As mentioned above, we have modified a wide-coverage HPSG (the ERG) to produce representations for fragments as well. The changes required for our analysis of fragments were moderate, about one hundred lines of code in a grammar comprising several tenthousand lines. This grammar is executed by a parser, the LKB (Copestake, 2002). The initial stage adds to the MRS of the chosen ERG-parse predicates which abstract away from certain semantic details, for example about which actions permit meeting at a certain time and which don't. At the next stage, an utterance in the context is chosen to which the current one can be attached via a rhetorical relation, and this in turn determines which antecedents are available. The preference is to attach to the prior utterance; this default is derivable in the logic of SDRT from assumptions about cooperativity and rationality, but since this derivation is context-independent, the preference is hard-coded here.

The speech act(s) of the current utterance is (are) then inferred non-monotonically from information about the antecedent and the current utterance and axioms like those given above for *IQAP* and *Q-Elab*. Here we deviate from (Schlangen et al., 2001), where we explored to what extent we can make the reasoning about speech-acts monotonic. We decided that for further extensions we need to fully implement the non-monotonic theory specified by SDRT. For this we implemented an automated theorem prover for the fragment of the nonmonotonic logic (Common Sense Entailment) employed in SDRT to compute rhetorical relations. (For details about the theorem prover see (Schlangen and Lascarides, 2002).) The rules (8) for inferring *IQAP* and *Q-Elab* given above can be passed directly to this theorem prover.

The next module, `sa_cnstr`, tests whether the monotonic constraints on the speech acts (for our application, these are given by the rules (11) and (13)) are all satisfied. After satisfying the constraints the SARGs are computed and any remaining underspecification is resolved; in a last step, the context is updated with the resolved representation of the current utterance.

3.2 A Worked Example

We illustrate how fragmental *IQAP*s are resolved in the system with example (1) from the introduction, repeated here as (14). The MRS-representation of the question is shown in (15); that of reply B was already given above in (4).

(14) A: What time on Tuesday is good for you?
B: 3pm. / B′: #At 3pm.

(15)
$$\begin{aligned}\langle h_1, e_2, \{&h_3 : \_which\_rel(x6, h_7, h_5),\\ &h_8 : \_time\_rel(x_6),\\ &h_8 : \_on\_temp\_rel(e_{11}, x_6, x_13),\\ &h_{14} : dofw\_rel(x_{13}, \text{"}TUE\text{"}),\\ &h_{16} : def\_np\_rel(x_{13}, h_{17}, h_{18}),\\ &h_{20} : \_good\_rel(e_2, x_6),\\ &h_{20} : \_for\_rel(e_{22}, e_2, x_{24}),\\ &h_{25} : pron\_rel(x_{24}),\\ &h_{26} : def\_rel(x_{24}, h_{27}, h_{28}),\\ &h_1 : int\_rel(h_{32})\},\\ \{\ &h_7 =_q h_8, h_{17} =_q h_{14},\\ &h_{27} =_q h_{25}, h_{32} =_q h_{20}\}\rangle\end{aligned}$$

In the implementation, the two predicates *syn-par* and *resolve* from constraint (11) are combined into one predicate `resolve` which is computed in three steps: first, the question MRS is transformed into a "λ-abstract", or equivalently, the subformula that will fill the position indicated by the *unknown_rel* is identified, see (16-a); second, this is "applied" to the fragment meaning (i.e., the replacement is made), see (16-b) (the added material is printed indented; h_2 and h'_{20}, e and e'_2, and x and x'_6 are identified); and third, the well-formedness of the result is checked.

This well-formedness constraint is *not* a test for parallelism between wh- and fragment-phrase (see earlier discussion). Rather, we check whether we can "assemble" a well-formed MRS from both utterances: it must be possible to partition the result-

[11] In fact, we do not have a speech act *Clarification* as such in our theory, but rather model the semantic effect of such utterances by a combination of *Elaboration* and *Q-Elab*. The kind of clarifications G&S analyse, namely those where the *content* of the previous utterance is being clarified (rather than the intention behind making it), do exhibit syntactic parallelism like the short-answers discussed above. Eg. "Peter relies on Sandy. — On whom?". Such utterances are *also Q-Elab*s, but the syntactic constraints are on the speech act *Elaboration*.

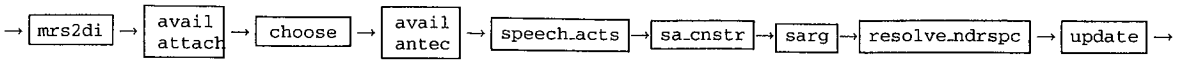


Figure 3: Flowchart of the algorithm

ing MRS so that all partitions are semantic representations for verb-relations, their arguments or possibly their adjuncts. This is the case for our example, as the reader is invited to check. It wouldn't be for answer B' from (14), because here the fragment brings with it a relation corresponding to a preposition that isn't matched in the material coming from the question.

(16) a. $\{\ h_8 : \_on\_temp\_rel(e_{11}, \mathbf{x_6}, x_13),$
$h_{14} : dofw\_rel(x_{13}, "TUE"),$
$h_{16} : def\_np\_rel(x_{13}, h_{17}, h_{18}),$
$h_{20} : \_good\_rel(\mathbf{e_2}, \mathbf{x_6}),$
$h_{20} : \_for\_rel(e_{22}, \mathbf{e_2}, x_{24}),$
$h_{25} : pron\_rel(x_{24}),$
$h_{26} : def\_rel(x_{24}, h_{27}, h_{28})$ $\}$

b. $\{\ h' : prpstn\_rel(h'_1),$
$\mathbf{h'_2} : unknown\_rel(\mathbf{e}, \mathbf{x}),$
$\mathbf{h_{20}} : \_good\_rel(\mathbf{e'_2}, \mathbf{x'_6}),$
$\mathbf{h_{20}} : \_for\_rel(e'_{22}, \mathbf{e'_2}, x'_{24}),$
$h_8 : \_on\_temp\_rel(e'_{11}, \mathbf{x'_6}, x'_{13}),$
$h_{14} : dofw\_rel(x'_{13}, "TUE"),$
$h_{16} : def\_np\_rel(x'_{13}, h_{17}, h_{18}),$
$h_{25} : pron\_rel(x'_{24}),$
$h_{26} : def\_rel(x'_{24}, h'_{27}, h'_{28}),$
$h'_6 : def\_rel(x, h'_8, h'_9),$
$h'_{10} : numbered\_hour\_rel(x, 15)\},$
$\{\ h'_1 =_q h'_2,\ h'_8 =_q h'_{10}\ \}\ \rangle$

This formulation of the well-formedness condition relies on two features of the ERG: first, adjunct questions introduce an underspecified preposition-relation in the MRS, as seen in (17); second, as already mentioned, all lexical items, including semantically empty ones, introduce an EP. This means that in effect we use the syntactic information contained in the MRSs to model the syntactic parallelism G&S noted. This seems to work for English, and makes our implementation relatively simple. We stress here though that this is not a crucial point of the underlying *theory*, which simply demands that some syntactic information is accessible to constraint (11). We discuss other approaches that use such mixed syntactic and semantic representations below in Section 4.

To give another example, "on Monday" is accepted as an answer to "when shall we meet?" (see (17) below) because after applying the question to the fragment, the variable that denotes the Monday will be argument of the *unspec_loc_rel* introduced by "when". The well-formedness constraint involving partitions above forces this underspecified relation to resolve to the EP corresponding to the "on" from the answer.

Our approach predicts that a question like "what did we agree on?", where there is a functional preposition in the question, can be answered both with a PP-fragment, in which case the two preposition-rels are equated, and with just an NP-fragment, which is β-reduced into the argument-position of the preposition. PPs (both with functional and with lexical prepositions) are ruled out as answers to NP-questions because there the well-formedness check fails, since there is no preposition-rel in the question that would match that in the answer.

(17) $\langle h_1, e_2, \{h_1 : int\_rel(h_{20}),$
$h_{15} : meet\_rel(e_2, x_{10}),$
$h_{15} : unspec\_loc\_rel(e_2, x_4),$
$h_5 : which\_rel(x_4, h_8, h_7),$
$h_3 : temp\_rel(x_4),$
$h_{11} : def\_rel(x_{10}, h_{12}, h_{13}),$
$h_9 : pron\_rel(x_{10})\},$
$\{\ h_{20} =_q h_{15}, h_8 =_q h_3, h_{12} =_q h_9\}\ \rangle$

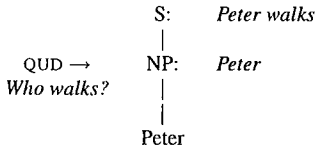
We turn now to the resolution of *Q-Elabs*. As explained above, there is no specific constraint for fragmental realisations of *Q-Elabs*, and so in an example like (18) only the temporal expression is resolved, in the way explained in (Schlangen et al., 2001). B' in the example below is predicted to be incoherent simply because it doesn't provide a temporal expression.

(18) A: Let's meet next week.
B: Tuesday? / B': #Peter?

4 Related Work

As mentioned in the introduction, G&S offer a non-modular approach to the resolution of short-answers (and some other fragmental speech acts). (19) shows a very schematic representation of their approach.

(19)



A grammar rule specific to short-answers directly projects NPs as sentences, with parts of the sentential content coming from a contextual feature QUD (question under discussion). This grammar rule in one go checks the syntactic parallelism and constructs the intended content of the fragment.

We have already given some features that we see as advantages of separating these different tasks

above (see Section 1). From a practical perspective, it seems that a non-modular approach also leads to certain complications. The system described in (Ginzburg and Gregory, 2001), an implementation of G&S, performs unification-operations on (syntactic) representations coming out of the grammar to insert the contextual information *a posteriori*. This means that no standard off-the-shelf parsers can be used in their system.

Another strand of related work concerns the phenomenon of parallelism, which has been studied for a range of both intra- and intersentential constructions; for instance coordination, VP-ellipsis and discourse structure (see references in (Prüst et al., 1994)). (Prüst et al., 1994) deals with the problem of computing which elements of clauses are to be considered parallel on the level of content. The parallelism-constraints in CLLS (Egg et al., 2001) make sure that certain scope-decisions in one representation are tied to that in another representation. Both these notions are complimentary to ours, which concerns certain *syntactic* features of parallel elements. Closest in spirit is (Kehler, 2002), where the question of whether syntactic features play a role is relativised to the speech act.

Interestingly, all theses approaches used mixed syntactic/semantic-representations similar to the one we use in our system. (Prüst et al., 1994) retain information about constituency in their representations (as does (Asher, 1993)), while (Kehler, 2002) keeps all syntactic information to allow reconstruction of syntactic structure. Our proposal lies somewhere in the middle, with some syntactic features above and beyond constituency being required to persist, while others are not being carried over from grammar.

Mixed representations have also been motivated on more practical grounds, for example in (McRoy et al., 1998) and (Milward, 2000), because they support more robust techniques of processing natural language. Their representation formats seem to contain enough information to be useful for our system as well; this is something we might explore in the future.

5 Conclusion and Further Work

We have offered a compositional semantics of fragments and constraints on two speech acts that can be performed with them. We have described how a dialogue system can use these constraints to resolve the intended meaning of fragments. We also have discussed why we think this approach has certain advantages compared to others (eg. G&S). In future work, we will extend the system to cover other speech acts that can be performed with fragments, for example *Elaboration* and *Correction*. We also intend to evaluate the system on a larger scale, by running corpus examples through it.

References

N. Asher and A. Lascarides. 1998. Questions in dialogue. *Linguistics and Philosophy*, 23(2):237–309.

N. Asher and A. Lascarides. in press. *Logics of Conversation*.

N. Asher. 1993. *Reference to Abstract Objects in Discourse*. Kluwer Academic Publisher, Dordrecht.

E. L. Barton. 1990. *Nonsentential Constituents*. John Benjamins, Amsterdam / Philadelphia.

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

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

A. Copestake. 2002. *Implementing Typed Feature Structure Grammars*. CSLI publications.

M. Egg, A. Koller, and J. Niehren. 2001. The constraint language for lambda structures. Journal of Logic, Language, and Information. To appear.

J. Ginzburg and H. Gregory. 2001. SHARDS: Fragment Resolution in Dialogue. In H. Bunt, I. van der Silius, and E. Thijsse, editors, *Proceedings of the 4th International Conference of Computational Semantics*, pages 156–172, Tilburg, The Netherlands.

J. Ginzburg and I. A. Sag. 2001. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications, Stanford.

J. Groenendijk and M. Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam, Amsterdam.

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

A. Kehler. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications.

S. W. McRoy, S. S. Ali, and S. M. Haller. 1998. Mixed depth representations for dialog processing. In *Proceedings of the Cognitive Science Society 1998*.

D. Milward. 2000. Distributing representation for robust interpretation of dialogue utterances. In *Proceedings of the Association for Computational Linguistics 2000*.

J.L. Morgan. 1973. Sentence fragments and the notion 'sentence'. In *Issues in Linguistics*. UIP, Urbana.

H. Prüst, R. Scha, and M. van den Berg. 1994. Discourse grammar and verb phrase anaphora. *Linguistics and Philosophy*, 17:261–327.

U. Reyle. 1993. Dealing with ambiguities by underspecification. *Journal of Semantics*, 10:123–179.

I. A. Sag. 1997. English relative clause constructions. *Journal of Linguistics*, 33(2):431–484.

D. Schlangen and A. Lascarides. 2002. CETP: An automated theorem prover for a fragment of common sense entailment. Informatics Research Report EDI-INF-RR-0119, Edinburgh University.

D. Schlangen, A. Lascarides, and A. Copestake. 2001. Resolving underspecification using discourse information. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Proceedings of BI-DIALOG 2001*, pages 79–93, Bielefeld, June.

D. Schlangen. 2002. A compositional approach to short answers in dialogue. In G. Mann and A. Koller, editors, *Proceedings of the Student Research Workshop at the 40th ACL*, Philadelphia, USA, July.

J. Searle. 1967. *Speech Acts*. CUP.

S. Shieber. 1986. *An Introduction to Unification-based Approaches to Grammar*. CSLI publications.

Context in Abductive Interpretation

Matthew Stone
Computer Science and Cognitive Science
Rutgers University
mdstone@cs.rutgers.edu

Richmond H. Thomason
AI Laboratory
University of Michigan
rich@thomason.org

Abstract

This paper develops a general approach to contextual reasoning in natural language processing. Drawing on the view of natural language interpretation as abduction (Hobbs et al., 1993), we propose that interpretation provides an *explanation* of how an utterance creates a *new discourse context* in which its interpreted content is both *true* and *prominent*. Our framework uses dynamic theories of semantics and pragmatics, formal theories of context, and models of attentional state. We describe and illustrate a Prolog implementation.

1 Problem Statement

Context in discourse and dialogue involves at least two components, *shared information* and *coordinated attention*; each of these components suggests a mechanism through which speakers may frame natural language utterances to fit the context.

- Shared information makes available to the participants a body of facts that characterize the world under discussion (here we mean to be neutral between various characterizations of this availability in terms of common ground, common knowledge, mutual belief, etc.); accordingly, language users may presume that information in an utterance links up with this body of facts, whenever possible (Lascarides et al., 1992; Hobbs et al., 1993).
- Coordinated attention puts certain entities at the forefront of the discussion; accordingly, language users may presume that descriptions in an utterance refer to these entities, whenever possible (Grosz and Sidner, 1986; Grosz et al., 1995; Walker et al., 1997).

How these two potential mechanisms interact is an empirical question—but this question can only be addressed within a framework that describes in precise terms how utterance interpretation interacts with an evolving context of coordinated attention and shared information. Our aim in this paper is to lay out such a framework.

In this brief, initial formulation of the overall theory, we introduce no formal devices that distinguish face-to-face conversation from other genres. Nevertheless, the need for an integrated account of pragmatic context such as ours is clearest in connection with spoken dialogue. Interlocutors in face-to-face dialogue liberally exploit the shared information of their perceptual environment and the coordinated attention of an ongoing real-world collaboration. Empirical studies clearly show speakers drawing on both factors to further efficient communication; for a recent study, see (Brown-Schmidt et al., 2002).

1.1 Interpretation as abduction

On the theory of interpretation as abduction articulated in Hobbs et al. (1993), an interpretation is represented as a proof of the logical form (LF) of the utterance from a background knowledge base (KB) and some further new assumptions. For instance, a telegraphic utterance *lube-oil alarm sounded* might be assigned this LF:

(1.1) $\mathit{lube\text{-}oil}(o) \wedge \mathit{alarm}(a) \wedge \mathit{nn}(o,a) \wedge \mathit{sound}'(e,a)$

The interpretation might then be represented as the graph in Figure 1, where arrows indicate inferences from KB facts (about a particular quantity of lube oil o_3 and a particular alarm r_5) to LF conjuncts, and a box indicates a literal that is assumed rather than derived. (Think of the variables in the logical forms in Figure 1 as existentially quantified.) A proof in this framework gives a specific way of linking up the meaning of an utterance with the shared information in the context; in Figure 1, for example, the proof spells out how $\mathit{for}(r_5, o_3)$, which says that the function of alarm r_5 is to monitor oil o_3, justi-

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 169-176.

fies the use of the noun-noun modification structure in the utterance $nn(o, a)$ describing the alarm.

The selection of an interpretation depends on informational preferences, including preferences for merging redundancies, preferences for making certain kinds of assumptions, and preferences for using certain kinds of knowledge. Proofs provide general concrete computational data structures to record such links. They serve as a natural, well structured domain over which to exhibit how these preferences depend on background knowledge. Moreover, computational logic provides increasingly attractive inference algorithms to search for the most preferred interpretation within an interpretation-as-abduction theory (Konrad, 2000; Kohlhase, 2000; Baumgartner and Kühn, 2000; Gardent and Konrad, 2000).

However, Hobbs-style abductive interpretations like the one presented in Figure 1 provide no representation of the changing status of entities in attentional prominence as the context evolves. Take a simple example:

(1.2) She asked her a question.

The LF is as follows.

(1.3) $\mathit{woman}(x) \wedge \mathit{woman}(y) \wedge x \neq y \wedge$
$\mathit{question}(q) \wedge \mathit{ask}'(e', x, y, q)$

The theory offers only two ways that we might single out one woman w_i as the preferred value of x in an abductive proof of (1.3), and so resolve the reference of *she*. We might appeal to knowledge that helps explain why w_i would ask something. Alternatively, following (Lascarides et al., 1992; Hobbs et al., 1993), we might posit further conjuncts to the LF which would require us to infer a rhetorical link between e and the eventualities introduced by prior discourse. But these two strategies are empirically and conceptually problematic: domain information has to be limited and related to the discourse, and LFs need to be specified systematically and compositionally. The obvious rival explanation is simply that w_i is the most salient woman in the context of the utterance, and that *she* refers to the most salient woman (other things being equal). In disallowing this explanation, the interpretation-as-abduction theory cuts off an extremely natural and successful approach to language interpretation.

KB		LF
$\mathit{lube\text{-}oil}(o_3)$	$\rightarrow$	$\mathit{lube\text{-}oil}(o)$
$\mathit{alarm}(r_5)$	$\rightarrow$	$\mathit{alarm}(a)$
$\mathit{for}(r_5, o_3), \mathit{for}(X, Y) \supset \mathit{nn}(Y, X)$	$\rightarrow$	$\mathit{nn}(o, a)$
		$\boxed{\mathit{sound}'(e, a)}$

Figure 1: Interpretation of *lube-oil alarm sounded* in the theory of Hobbs et al. (1993). The proof reflects the substitutions $o = Y = o_3$ (the lube oil) and $a = X = r_5$ (the alarm).

1.2 Context in abductive interpretation

Our proposal is that the interpretation of an utterance is an explanation of how the utterance creates a *new context* in which its content is *both true and prominent*. Thus, we retain an abductive and inferential characterization of interpretation, but we explicitly relativize explanatory goals to particular contexts, and we provide for explicit reasoning about the attentional aspects of those contexts.

Example (1.4) provides a schematic illustration of our approach. We will use the notation $c : g$ to indicate that goal g must be proved in the context associated with term c. To account for the informational and attentional dimensions of the discourse, we use three terms to identify the contexts before and after the utterance: i, a_1 and a_2. Term i describes the informational features of the discourse context as an accumulating body of domain content; the basic LF constrains this context i. Term a_1 identifies the current attentional state of the discourse; for a_1 we prove that y must be *in-focus* (like any pronominal referent), and that x should be particularly central to the discourse (as the referent of a subject pronoun), a property we represent as *in-focus**. (The star records a particular preference for a good explanation of this relationship.) Finally, term a_2 records the new attentional state *after* the utterance is understood; we use the formula $a_1[x < y < q]a_2$ to indicate that a_2 differs from a_1 in that the ranking $x < y < q$ determines the three most salient entities in a_2, with other salient entities from a_1 represented afterwards. Thus an interpretation is an explanation that shows the following:

(1.4) $a_1[x < y < q]a_2 \wedge$
$a_1 : \mathit{in\text{-}focus}^*(x) \wedge a_1 : \mathit{in\text{-}focus}(y) \wedge$
$i : \mathit{woman}(x) \wedge i : \mathit{woman}(y) \wedge x \neq y \wedge$
$i : \mathit{question}(q) \wedge i : \mathit{ask}'(e', x, y, q)$

A proof of (1.4) will spell out how the utterance *she asked her a question* draws on the attentional state a_1 and the shared background i in order to expand the shared background i and create a new attentional state a_2. In the result, i provides the declarative information that she asked her a question, and a_2 places this relationship at the center of attention.

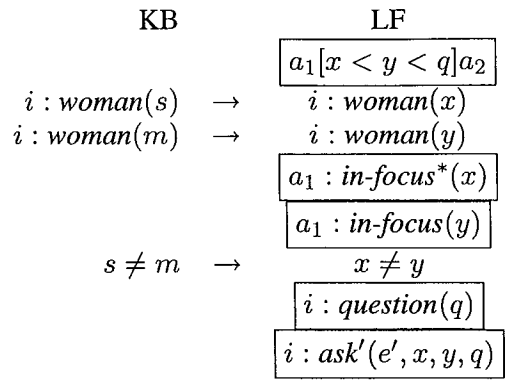


Figure 2: Our proposed interpretation for *she asked her a question.* The proof reflects the substitutions $x = s$ (Susan) and $y = m$ (Mary).

Figure 2 shows an explanation that proves this goal for the case where *she* is Susan (s) and *her* is Mary (m); this would count as an interpretation of the utterance on our proposal. We have in such explanations all of the resources of an interpretation-as-abduction theory to describe the role of knowledge in resolving phenomena of local pragmatics. But we can also model the role of attention in interpretation, in terms of preferences for the knowledge to use and the assumptions to make about a_1. And we have a principled way to describe the dynamics of attention across utterances, using the assumptions we make to characterize a_2. Thus we can represent and investigate hypotheses about the interaction of attentional state and shared information in stitching together interpretations.

1.3 Overview

(Hobbs et al., 1993) makes the case for the generality of the abductive approach to discourse using a large and varied inventory of examples: e.g., metonymy, anaphora and ambiguity resolution, compound noun interpretation, and the recovery of discourse relations. But each of these examples has to be solved separately; the framework of that paper does not work well with dialogues or even extended monologues, since there is no way to deal with changes of priorities. In removing this limitation, the dynamic extension of Hobbs' approach presented here provides a framework with the same versatility, but which is capable of giving an integrated analysis for multiple sentences in succession. Work subsequent to (Hobbs et al., 1993) showed that abduction could also be used as a framework for discourse generation; see (Thomason et al., 1996). Although the ideas are illustrated in this paper only with one fairly simple example, we intend them as a general framework for representing and reasoning about not only the interpretation, but the generation of utterances across an evolving context.

In the rest of this paper, we develop this proposal in detail. We first present a theory of discourse context that covers both semantics and pragmatics by specifying the truth of propositions and attentional preferences for resolving ambiguities. The theory draws on theoretical proposals on the role of context in natural language interpretation, such as Stalnaker (1972) and Kaplan (1978); on accounts of the dynamics of discourse context, such as Lewis (1979), Kamp (1981), and Beaver (1992); and on the ideas that have emerged from developments in formalizing context in AI, such as Buvač (1993) and McCarthy and Buvač (1995). Next, we show how to integrate this model of discourse context into the architecture for weighted abduction developed in Stickel (1991) and used in Hobbs et al. (1993), by relativizing axioms, assumptions and goals to contexts, and by determining the plausibility of an explanation as a function of the pragmatic prominence assigned to possible axioms and assumptions in context. As usual for resolution-based proof search systems, the natural realization of our proof rules gives a simple but effective Prolog implementation of this abductive reasoner. Finally, we illustrate the proposal by describing how this framework describes the resolution of pronouns in a complex discourse in which the resolution depends on linguistic constraints, common-sense reasoning, and the evolving prominence of discourse referents.

2 The Representation of Context

Our representation of context must describe both information and attention.[1] Following Stalnaker (1972), we treat shared information in terms of modal operators with a possible-worlds interpretation. A proposition counts as part of the context when it is true in all possible worlds compatible with the discourse participants' mutual knowledge. To characterize the influence of attention, we will follow Kaplan (1978) in creating an abstract disambiguation space; you can think of an *index*, or point in this space, as a simultaneous choice of appropriate values resolving all the ways in which an expression of the language can be ambiguous. If free

[1]The following paragraphs are an informal statement of ideas presented elsewhere. See Thomason (1999) for formal details.

variables are the only source of ambiguity, then an index is simply an assignment of values to variables. If the language contains indexical expressions, like 'I' and 'now', an index will include an evaluation of these expressions. If the language has ambiguous expressions, the index will determine a disambiguation of these expressions. (The values that an index assigns to parameters are intensions; they are functions from possible worlds to various domains.)

In the AI literature on context, a context does two things: it accesses certain axioms or knowledge sources, and it disambiguates ambiguous expressions. If we treat a context c as a pair $\langle [c], i_c \rangle$ consisting of a modal operator $[c]$ and an index i_c, it is natural to revise the usual notation for contexts as follows: $ist(\phi, [c], i_c)$. In this formula, ϕ denotes a function from indices to sets of possible worlds,[2] $[c]$ is interpreted using a relation R_c over possible worlds, and i_c denotes an index. A world w belongs to the the set of worlds assigned to (2) in a model iff for all w' such that wR_cw', w' belongs to the set of worlds assigned to ϕ on disambiguation i_c.

This representation of context is appropriate for tasks in which information that is distributed across a number of knowledge sources must be integrated for reasoning purposes. The information integrator is equipped with its own context, and also knows how other contexts encode information. It receives messages that are labeled according to their sources, and needs to translate them in order to bring them into a single context for subsequent inference.

In natural language interpretation, the disambiguation problem is quite different; there are competing disambiguations of an utterance, the context is unknown, and the problem is to find a plausible disambiguation. In other words, a large part of natural language interpretation is *to infer the intended context.* Naturally, this reasoning uses information concerning what has been said previously and the mutually perceived environment (including features of the utterance like intonation), and naturally this information is often called the "context" of the interpretation. This terminological clash has created much confusion in attempts to use theories of context in connection with natural language. To avoid this confusion, we will use 'S-context' ('Semantic Context') when we have in mind a pair $\langle [c], i_c \rangle$, and will use 'D-context' ('Discourse Context') for the informational situation of the natural language interpreter.

[2] In Kaplan's terms, ϕ denotes a *character.*

A large part of this situation is a changing set of preferences that are used to establish anaphoric references, to resolve ambiguities, and to fill in implicit meanings. We propose, then, to identify a D-context with a triple $\langle [c], I_c, \prec_c \rangle$, where I_c is a set of indices and $\prec_c$ is a partial order over I_c.

The formalization of a linguistic example for abductive interpretation assigns costs to proofs which are associated with various interpretations of an utterance. This can be represented by a D-context that ranks interpretations according to their abductive costs. The abductive search for a least-cost proof is then equivalent to a search for an interpretation that is maximally preferred in the D-context.

Each utterance is not only evaluated with respect to a D-context; it has the potential to change that context. In what follows we will assume that an utterance can change a D-context in two ways: by adding new information to the background for the next utterance and by altering the preferences that apply to disambiguate it.

3 Abduction for D-Context

Our computational realization of this formalism for reasoning in D-context consists of an algorithm for abductive inference in modal logic and a representation for linking abductive modal proofs with preferences for particular interpretations.

3.1 Abductive modal inference

We couch the declarative part of our contextual reasoning in a modal logic; modal operators of the form $[c]$ are associated with contexts; we also have a distinguished operator $\Box$ which we use to specify axioms that are true in all contexts. We assume atomic formulas as in Prolog, using A to schematize them; we also assume an always true atom $\top$. If κ is a sequence of context operators of the form $[c_0] \ldots [c_n]$ (possibly empty), we use the notation $\kappa(\phi)$ to name the formula $[c_0] \ldots [c_n]\phi$; we use $\kappa \circ \kappa'$ to denote the concatenation of κ and κ'.

We use a fragment of modal logic that extends Prolog in a simple way. Goals G are modalized atomic formulas; clauses P are modalized Prolog clauses whose antecedent and conclusion formulas may themselves be modalized:

$$G ::= \kappa(A) \mid \top$$
$$P ::= \kappa'(\kappa_1(A_1) \ldots \kappa_m(A_m) \rightarrow \kappa''(H)) \mid$$
$$\Box\kappa'(\kappa_1(A_1) \ldots \kappa_m(A_m) \rightarrow \kappa''(H))$$

The interpretation of this language is the usual Kripke semantics, with each $[c]$ treated as a K

modality and $\Box$ treated as an S4 modality with an accessibility relation containing that of each $[c]$; see e.g., Fitting and Mendelsohn (1998).

In this fragment, we can streamline the prefixed tableau inference method of Fitting and Mendelsohn (1998) with simplified representations of possible worlds suitable for reasoning with definite clauses. Each *query* can be simplified to the form $\kappa : A$, where κ is a modality and A is atomic formula. In resolution, we must match the head H of our clause with the query atom A, but we must also match the context κ of the query with the context $\kappa' \circ \kappa''$ or $\Box \circ \kappa' \circ \kappa$ of our clause ($\Box$ goes proxy for any modality κ^*). Baldoni et al. (1998) show that this strategy suffices for sound and complete inference in this modal fragment.

To refine this procedure into an abductive inference algorithm, we introduce *marked* queries, which consist of a query together with a designation *resolved*, *assumed* or *unsolved*; now, as Stickel does, we can describe a partially-constructed abductive proof by a list of each marked query derived and its current proof status. In this list, a *resolved* or *assumed* query all of whose predecessors are also *resolved* or *assumed* has in fact been proved (abductively) from the common premises. This justifies a new factoring rule to derive a subsequent query by reusing this proof.

The resulting algorithm constructs abductive explanations by applying any of four rules nondeterministically to transform one list of marked queries into another until all the formulas in the list are either resolved or assumed. These rules can be implemented immediately as Prolog clauses, by using Prolog variables and unification to represent object-level variables in formulas and by using Prolog search to manage nondeterminism in the application of rules. (In our implementation, we have gone beyond this only in testing for conflicts, in assigning costs to proofs, and in using bounds on cost to impose an iterative-deepening search regime.)

Suppose the current state is represented as a list $L = Q_1, \ldots, Q_n$ with $Q_i = \kappa : A$ is the first (leftmost) query marked *unsolved*. We can transform the state as follows:

- *Truth.* If A is $\top$, derive a new state exactly like L except that the label of Q_i is *resolved* rather than *unsolved*.
- *Assumption.* Derive a new state exactly like L except that the label of Q_i is *assumed* rather than *unsolved*.
- *Resolution.* Select a clause R of the form P or $\Box P$, where P is
$$\kappa'(\kappa_1(A_1) \ldots \kappa_m(A_m) \rightarrow \kappa''(H))$$
with its variables renamed, if necessary, so that it has no variables in common with L. Suppose H and A are unifiable with most general unifier σ, $\kappa = \kappa^* \circ \kappa' \circ \kappa''$, and unless R is $\Box P$ then κ^* is empty. Then derive the new state:
$$Q_1\sigma, \ldots, Q_{i-1}\sigma,$$
$$\kappa^* \circ \kappa' \circ \kappa_1 : A_1\sigma[\textit{unsolved}], \ldots,$$
$$\kappa^* \circ \kappa' \circ \kappa_m : A_m\sigma[\textit{unsolved}],$$
$$\kappa : A\sigma[\textit{resolved}], \ldots, Q_n\sigma$$
- *Factoring.* Suppose some element Q_s describing a query $\kappa : H$ precedes Q_i in L, and H and A are unifiable with most general unifier σ. Then derive a new state suppressing the duplicate proof of Q_i:
$$Q_1\sigma, \ldots, Q_{i-1}\sigma, Q_{i+1}\sigma, \ldots, Q_n\sigma$$

An abductive proof for a query Q is a sequence of lists whose initial element is $Q[\textit{unsolved}]$, in which each later list is obtained from its predecessor by the application of one of these transformations, and which terminates in a list none of whose elements are marked *unsolved*. An abductive proof determines an *answer* to the query, a pair $\langle Q_0, \Delta \rangle$ consisting of the found instantiation Q_0 of Q (determined from the final element of the terminal list) together with the postulated assumptions Δ (a multiset consisting of a formula $\kappa(A)$ for each element of the form $\kappa : A[\textit{assumed}]$ in the terminal list).

Space limitations prevent us from describing the workings of the algorithm and correctness results in further detail here.

3.2 Preferences in D-contexts

Any nontrivial example of discourse provides alternative interpretations; we therefore need preferences in order to choose among alternative explanations. Stickel treats these overall preferences as a function of the assumptions and axioms used in a proof. We generalize this idea to provide for the dependence of these preferences on a dynamically evolving discourse context.

Assumption costs are specified indirectly by propagating values through the proof. The initial query is annotated directly with an assumption cost, while costs for further queries are determined by annotating rules so that each subgoal G_j is associated with an *assumability function* f_j:

$$(\Box)\kappa'(G_1 \,/ f_1 \wedge \ldots \wedge G_m \,/ f_m \rightarrow \kappa''(Q))$$

When such a rule is instantiated to values x and resolved against a query with cost c in a context $\kappa^* \circ \kappa' \circ \kappa''$, each new subgoal G_j determines a query with an assumption cost calculated as

$$c \times f_j(\kappa^* \circ \kappa', x)$$

Each axiom is also annotated with a function f_r, where $f_r(\kappa^* \circ \kappa', x)$ is a cost for the use of the axiom in the explanation. To get the overall cost for an answer $\langle Q, \Delta \rangle$, we sum the assumption costs associated with the elements in Δ, together with axiom costs for each resolution step in the proof.

To illustrate the context-sensitivity of assumption costs, take the familiar (nonlinguistic) case of the WET-LAWN, which can be explained either by RAIN or by SPRINKLER. The relevant rules are these.

(3.1) $\Box$(SPRINKLER $/ f_1 \rightarrow$ WET-LAWN)
(3.2) $\Box$(RAIN $/ f_2 \rightarrow$ WET-LAWN)

Here, we represent the costs associated with using rules (3.1) and (3.2) as functions $f_1(\kappa)$ and $f_2(\kappa)$ of the reasoning context. Thus, even though we allow these axioms to apply in all contexts, in a context κ where $f_1(\kappa)$ is relatively high and $f_2(\kappa)$ relatively low, (3.2) will lead to RAIN as the preferred explanation of WET-LAWN. When $f_1(\kappa)$ is relatively low and $f_2(\kappa)$ is relatively high, SPRINKLER will be the preferred explanation.

In discourse interpretation, it is vital not only that assumability functions take different values across contexts, but also that those values can be established dynamically, as each sentence draws on the pragmatic state set up by the interpretation of its predecessors. To enable this, we provide the abductive reasoner with a high-level specification of the behavior of assumability functions in new contexts—an abstract characterization of the discourse preference dynamics.

4 An Example

We will use anaphora resolution to illustrate the approach to context in natural language interpretation articulated in the previous sections. Consider the following three-sentence text.

(4.1) Susan met Mary.
She asked her a question.
She answered no.

On the natural interpretation, it is Susan who asked Mary a question, and Mary who answered no.[3]

In our model, this interpretation is an explanation of how discourse (4.1) updates an evolving context. As with (1.4), we describe this in terms of a term i for the information in the discourse and terms a_0, a_1 and a_2 for the changing attentional state of the discourse. From the first sentence, we learn that Susan met Mary, and update attention so that Susan and Mary are salient, by proving

(4.2) $a_0[u < v]a_1 \wedge i : \mathit{susan}(u) \wedge i : \mathit{mary}(v) \wedge i : \mathit{meet}'(e, u, v)$

The second sentence is analyzed as (1.4); the third is analyzed along similar lines:

(4.3) $a_2[z]a_3 \wedge a_2 : \mathit{in\text{-}focus}^*(z) \wedge i : \mathit{woman}(z) \wedge i : \mathit{answer}'(e'', z, q', \mathit{no})$

Figure 3 diagrams a proof corresponding to the intended interpretation of this discourse. Note how the commonsense inference (4.4) helps to explain the answer.

(4.4) If an event E in which X asks Y something has just occurred, an event E' in which Y gives X the answer *no* is a strong possibility; this overrides salience considerations based on grammatical relations only.

By using this in the proof of Figure 3, we need only assume a Hobbs-style default $\mathit{etc}_{q-a}(e', e'')$.

With such representations, it is straightforward to describe general criteria which favor the interpretation of Figure 3 over all alternatives, and to realize those criteria as preferences so that our abductive theorem prover automatically identifies the intended interpretation for (4.1). Our criteria (4.5) and (4.6) correspond to the forward-looking center ranking of Centering Theory and its preference for continue transitions in discourse (Walker et al., 1997).

(4.5) A sentence with a feminine subject and direct object induces a context in which the subject is most 'she'-salient (i.e., resolutions of 'she' to the subject take a lower cost in the induced context) than the direct object, and the direct object is more 'she'-salient than anything else.

(4.6) An interpretation in which the subject refers to a more salient entity is to be preferred to one where it refers to a less salient one (in the absence of more compelling considerations).

[3] Although (4.1) involves only a single speaker, the same tradeoffs in interpretation appear in dialogue, as this example illustrates:

A: Did Susan meet Mary? B: Yes. A: Did she invite her? B: Yes. And she accepted.

With rules to deal with the context-changing effects of questions, our techniques could formalize this reasoning, too.

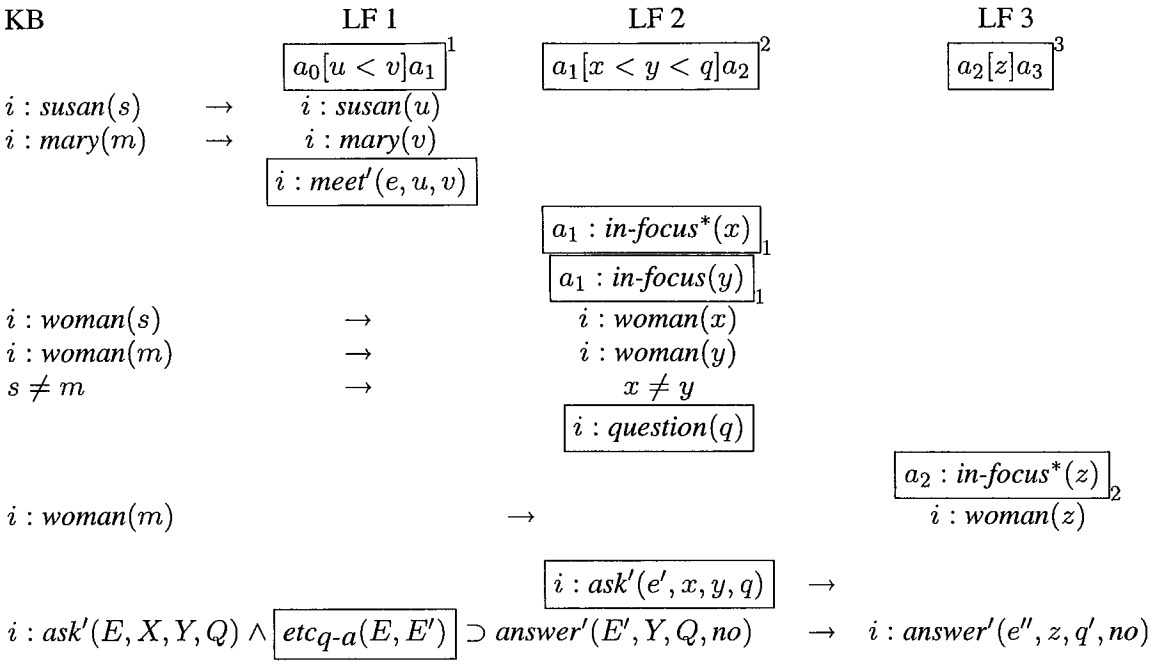


Figure 3: Figure 3: Intended abductive interpretation for discourse (4.1). Subscripts indicate assumptions whose costs depend on the correspondingly superscripted context description. The proof reflects the substitutions $u = x = X = s$ (Susan), $v = y = Y = z = m$ (Mary), $Q = q = q'$ (the question), $E = e'$ (the asking) and $E' = e''$ (the answering).

They yield to the commonsense criterion of (4.4).

Abductive proof search explores a space of interpretations in which the variables x, y and z are unified either to s or to m. The resolution of x and y in the second sentence is independent of that of z in the third; so let's begin with the second. Inconsistency eliminates the possibilities where $x = y$. Now because by (4.5) s is more salient than m in a_1, we get different cost-factors $w_{x=s} = w_{y=s} < w_{x=m} = w_{y=m}$ for different *in-focus* assumptions. Meanwhile, the goals created according to (4.4) assign a cost u_x to assuming $in\text{-}focus^*(x)$ for the subject. This exceeds the cost u_y for assuming $in\text{-}focus(y)$ for the object: $u_y < u_x$. The overall cost is therefore minimized when we resolve x to Susan and y to Mary (at $w_{x=s} \times u_x + w_{y=m} \times u_y$). Turning now to the resolution of z, the context contains an event of Susan asking Mary a question. The key alternative here is whether we take z to be Susan, assuming the answering outright with some cost u_c, or we take z to be Mary, using the asking event to derive Mary's answer with reduced cost u_m. According to (4.4) and (4.6), the decrease in cost from u_c to u_m outweighs the additional cost of chosing Mary, the less salient 'she', for z.

5 Conclusions and Future Work

The changing interpretation of pronouns against an evolving semantic and pragmatic context in discourse has been studied from many linguistic and computational perspectives. The relation of the more computationally oriented work, such as Centering Theory (Walker et al., 1997), to general theories of context is not entirely clear; and the literature provides no general mechanism for integrating reasoning based on linguistic structure and pragmatic context with other information to capture coherence or commonsense inferences. On the other hand, treatments of anaphora resolution couched in terms of semantic approaches such as the Structured Discourse Representation Theory used in Lascarides et al. (1992), account only for the connection between semantics, world knowledge and discourse coherence. In place of general concepts of salience, they offer a taxonomy of specific rhetorical and informational connections that justify apparently salient interpretations—which can seem quite cumbersome. Recent extensions of the interpretation-as-abduction theory (Konrad, 2000; Kohlhase, 2000; Baumgartner and Kühn, 2000; Gardent and Konrad, 2000) seem to suffer from the same drawbacks. The problem is compounded in Hobbs et al. (1993),

where context change is not properly represented at all. Our treatment of pronoun resolution in (4.1) allows different kinds of information to be specified simply and separately, but to be combined straightforwardly in reasoning about a changing context.

We believe that the general approach we have articulated and formalized in this paper should enable a theoretically informed investigation of tradeoffs between information and attention in discourse. Promising cases for formalization include the information-dependent models of attention of Kehler (2001), where inferred rhetorical connections change but do not eliminate attentional preferences; and context-dependent representations of apparently inconsistent utterances, like Strawson's (1952) *He didn't jump*, in which a speaker's appeal to coordinated attention requires us to understand an utterance as rejecting claims we would otherwise have taken as shared.

6 Acknowledgments

This material is based on work supported by the National Science Foundation under Grants No. IRI–9314961, "Integrated techniques for natural language generation and interpretation" and 9818322, "A laboratory for interactive applications of computational vision and language".

References

Matteo Baldoni, Laura Giordano, and Alberto Martelli. 1998. A modal extension of logic programming: Modularity, beliefs and hypothetical reasoning. *Journal of Logic and Computation*, 8(5):597–635.

Peter Baumgartner and Michael Kühn. 2000. Abducing coreference by model construction. *Journal of Language and Computation*, 2(1):175–190.

David Beaver. 1992. The kinematics of presupposition. Technical Report LP–92–05, ILLC, University of Amsterdam.

Sarah Brown-Schmidt, Ellen Campana, and Michael K. Tanenhaus. 2002. Reference resolution in the wild: On-line circumscription of referential domains in a natural interactive problem-solving task. In *Proceedings of the Cogntive Science Society.*

Saša Buvač. 1993. Propositional logic of context. In *Proceedings of AAAI*, pages 412–419.

Melvin Fitting and Richard L. Mendelsohn. 1998. *First-Order Modal Logic.* Kluwer.

Claire Gardent and Karsten Konrad. 2000. Interpreting definites using model generation. *Journal of Language and Computation*, 1(2):175–190.

Barbara J. Grosz and Candice L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204.

Barbara J. Grosz, Arivind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):227–253.

Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142.

Hans Kamp. 1981. A theory of truth and semantic representation. In J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof, editors, *Formal Methods in the Study of Language.* Mathematisch Centrum, Amsterdam.

David Kaplan. 1978. On the logic of demonstratives. *Journal of Philosophical Logic*, 8:81–98.

Andrew Kehler. 2001. *Coherence, Reference and the Theory of Grammar.* CSLI.

Michael Kohlhase. 2000. Model generation for discourse representation theory. In *14th European Conference on Artificial Intelligence.*

Karsten Konrad. 2000. *Model Generation for Natural Language Interpretation and Analysis.* Ph.D. thesis, Universität des Saarlandes. Available at http://www.ags.uni-sb.de/~konrad.

Alex Lascarides, Nicholas Asher, and Jon Oberlander. 1992. Inferring discourse relations in context. In *Proceedings of ACL*, pages 1–8.

David K. Lewis. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(3):339–359.

John McCarthy and Saša Buvač. 1995. Formalizing context (expanded notes). Available from http://www-formal.stanford.edu/buvac.

Robert C. Stalnaker. 1972. Pragmatics. In Gilbert Harman and Donald Davidson, editors, *Semantics of Natural Language*, pages 380–397. D. Reidel.

Mark Stickel. 1991. A prolog technology theorem prover: a new exposition and implementation in prolog. Technical Report 464, SRI International.

P. F. Strawson. 1952. *Introduction to Logical Theory.* Methuen.

Richmond Thomason, Jerry Hobbs, and Johanna Moore. 1996. Communicative goals. In Kristiina Jokinen, Mark Maybury, and Ingrid Zukerman, editors, *Proceedings of the ECAI 96 Workshop on Gaps and Bridges: New Directions in Planning and Natural Language Generation.* Springer.

Richmond H. Thomason. 1999. Type theoretic foundations for context, part 1: Contexts as complex type-theoretic objects. In Paolo Bouquet et al., editors, *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*, pages 352–374. Springer.

Marilyn A. Walker, Arivind K. Joshi, and Ellen Prince, editors. 1997. *Centering Theory in Discourse.* Oxford University Press.

Centering and Pronominal Reference: In Dialogue, In Spanish

Maite TABOADA
Department of Linguistics
Simon Fraser University
Burnaby, B.C. V5A 1S6
Canada
mtaboada@sfu.ca

Abstract

This paper describes an application of Centering Theory (Grosz et al. 1995) to dialogue in Spanish. Centering is a theory of local focus and local discourse coherence. It also relates focus in discourse to choice of referring expression. In this paper, I discuss the analysis of nine task-oriented conversations. The main aim of the paper is to establish links between transition type and choice of pronoun, studying correlations between those two, and examining what other factors might be at play when a general tendency (e.g., of CONTINUE transitions to result in pro-drop) is not followed. I explore whether some of those factors could be related to characteristics of dialogue, such as signaling turn-taking.

Introduction

Different theories attempt to account for how anaphoric terms can be linked to their referents, often with the goal of providing anaphora resolution in computational systems. Centering Theory (Grosz et al. 1995), a theory of focus of attention in discourse, has also been applied to the problem of anaphora resolution (Brennan et al. 1987). Usually the perspective is that of a language understanding system, which needs to keep track of referents and pronouns in discourse. In this paper, I proceed from a different point of view, trying to establish not what antecedent a pronoun has, but instead what type of anaphoric term is used in a particular context. Given the transitions proposed by Centering, I examine the anaphoric term (pronoun, clitic, stressed pronoun, noun phrase, etc.) chosen for the backward-looking center of an utterance.

To examine these issues, I studied a corpus of dialogues in Spanish. The paper first presents a brief introduction of Centering, and then the application of the theory to a relatively new area, dialogue, and to a new language, Spanish (Sections 2 and 3). In Section 4, I present the results of the corpus study.

1 Centering

Centering (Grosz et al. 1995) is a theory of focus in discourse, and its relation to choice of referring expressions. It provides rules and constraints to explain how entities become focused as the discourse proceeds, and how transitions from one focus to the next make the discourse coherent. For each utterance in the discourse, Centering establishes a ranked list of entities mentioned or evoked[1], the *forward-looking list* (Cf). The first member in the Cf list is the *preferred center* (Cp). Additionally, one of the members of the Cf list is a *backward-looking center* (Cb), the highest-ranked entity from the previous utterance that is realized in the current utterance.

In Example (1), part of a conversation, the speaker proceeds from having *he* as a focus of attention in (1a), to *we all* in (1b), and back again to *he* in (1c). The changes are indicated by using explicit subjects in both (1b) and (1c), *todos* and *él*, respectively. The centers for each utterance in (1) are provided in (2)[2].

[1] I discuss what "evoked" means in Section 2.

[2] Abbreviations used in the examples: 1/2/3 - first/second/third person; CL - clitic; DAT - dative; NOM - nominative; SG - singular; PL - plural; IMP - imperative; INF - infinitive; FUT - future. Only those examples that illustrate a grammatical point are rendered in a word-by-word gloss. The rest are simply

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 177-184.

(1) CREA, uam480.per001
a. Entonces si le surge un problema,
then if he-CL-DAT arise-3SG a problem
'So if he has any problem come up,'

b. pues todos le vamos a echar una mano,
then all he-CL-DAT go-1PL to give
a hand
'we'll all give him a hand,'

c. pero él es la persona responsable.
but he-NOM is the person responsible
'but he's the one in charge.'

(2) a. Cf: HE, PROBLEM; Cb: HE; Cp: HE
b. Cf: WE ALL, HE; Cb: HE; Cp: WE ALL
c. Cf: HE; Cb: HE; Cp: HE

In addition to the three types of centers, Centering proposes different types of transitions, based on the relationship between the backward-looking centers of any given pair of utterances, and the relationship of the Cb and Cp of each utterance in the pair. Transitions, shown in Table 1, capture the introduction and continuation of new topics. $Cb(U_i)$ and $Cp(U_i)$ refer to the centers in the current utterance. $Cb(U_{i-1})$ refers to the backward-looking center of the previous utterance. Thus, a CONTINUE occurs when the Cb and Cp of the current utterance are the same and, in addition, the Cb of the current utterance is the same as the Cb of the previous utterance. Transitions are ranked according to the demands they pose on the reader. The ranking is: CONTINUE > RETAIN > SMOOTH SHIFT > ROUGH SHIFT. In (1) above, the transition from (1a) to (1b) is a SMOOTH SHIFT, whereas the transition from (1b) to (1c) is a CONTINUE.

	$Cb(U_i)=Cb(U_{i-1})$ or $Cb(U_{i-1})=0$	$Cb(U_i)\neq Cb(U_{i-1})$
$Cb(U_i)=Cp(U_i)$	CONTINUE	SMOOTH SHIFT
$Cb(U_i)\neq Cp(U_i)$	RETAIN	ROUGH SHIFT

Table 1. Transition types.

translated.

With these constructs as a starting point, the main goal of the paper is to show how the choice of anaphoric terms in spoken Spanish can be explained according to the transition type between one utterance and the next. For this purpose, I have carried out a corpus analysis of task-oriented dialogues in Spanish. The speakers communicate in order to complete a task, in this case, finding a date when they can both meet. Before analyzing spoken Spanish, I will describe the application of Centering Theory to dialogue, and to Spanish.

2 Centering in Dialogue

Centering has been applied mostly to monologic discourse. A few studies show how it can be extended to explain the coherence of dialogue. Brennan (1998) studied a corpus of spontaneous dialogues, and observed problems in segmentation and speaker change. Byron and Stent (1998) suggest the following as the main issues regarding the adaptation of Centering to two-party dialogue (multi-party dialogue is yet to be addressed):

1. Utterance segmentation. Utterance boundaries determine which entities are available for the following utterance's Cf list. A related issue is how to segment complex sentences.
2. Speech phenomena. We also need to consider the utterance status of segments before and after pauses and hesitations, and how to account for side sequences or self-repairs (as studied in Conversation Analysis).
3. First and second person pronouns. In dialogue, participants often refer to themselves and to each other through a pronoun in subject position. Since in the ranking of the Cf list, subjects take precedence (at least in English), these subjects will be ranked higher in the Cf list. However, it is not clear whether they are the centers of the utterance.
4. Linearity. Centering depends on the previous utterance to determine the entities and the focus of the current utterance. Since current and previous might have been uttered by different speakers, we should consider whether to use "previous by same speaker" rather than simply "previous".

The model for dialogue adopted here is Byron and Stent's (1998) *Model 1*, that is, a model where both first and second person pronouns are included in the Cf list. In addition, utterances are

consecutive: in the search for Cb_n, only Cf_{n-1} is searched, whether it was produced by the same speaker or not. Byron and Stent (1998) found that this model performed better than models that discarded first and second person pronouns, and models that considered previous or current speaker's previous utterance[3].

The decision to use Model 1 settles issues (3) and (4) above, and seems justified given the data. I found that reference to first and second person was important in the conversations, and often 'I' is mentioned in contrast to 'you' (singular) and to 'both of us', because the speakers discuss what dates one of them is available, versus what dates the other speaker has free, and which dates suit both of them as a unit.

Another general issue in Centering is what to include in the Cf list, and how to interpret the need that entities in that list be realized in the current utterance. The definition of 'realize' depends, according to Walker et al. (1998: 4), on the semantic theory one chooses. This is of particular importance in dialogue because it relies more than monologue on the context outside the text proper. Particularly difficult were decisions having to do with dates and times, and how those are related to each other. In general, I considered only "include" relations (Hurewitz 1998), such that a date was deemed to be part of the previous utterance's focus if it was part of a date range mentioned there. However, when the date was not within the time frame established, it is plausible to think that the hearer had to construct a new model for it. In Example (3), speaker famm[4] proposes the week of the fourth, after having discussed the previous week. However, speaker mjnm returns to the previous week, and mentions Friday, October 1st, i.e. a date not in the week of the fourth. I believe this is a new entity, and cannot be related to the immediately preceding utterance. As it happens, this results in a zero Cb.

(3) famm_06_12: ... quieres tratar la semana de cuatro?
'... do you want to try for the week of the 4th?'
mjnm_06_13: qué te parece el viernes primero de octubre, luego de las once de la mañana?
'what do you think of Friday October 1st, after 11am?'

There are still two dialogue-related issues I have not discussed, namely (1) and (2) above. As for utterance segmentation, I separated compound sentences in their component coordinate clauses, following Kameyama (1998). For subordinates, I separated finite subordinate clauses that have a clear final intonation (4), but not non-finite clauses (5) or those where there is continuing intonation.

(4) a. realmente tengo una reunión <de> desde las diez hasta las doce.
'I actually have a meeting <from> from 10 till 12.'

b. por tanto no creo que sea muy conveniente ese día.
'So I don't think that day is very convenient.'

(5) por favor no te olvides de traer todos los legajos. <para poder este> para tener toda la información a mano.
'please don't forget to bring all the papers. <to be able to uh> to have all the information handy.'

The issue of speech disfluencies is left open at this point. For the present analysis, I have discarded backchannels, repairs, and utterances with no entities in them. On the other hand, I did include side sequences and clarifications. These often result in zero Cbs, at either the beginning or the end boundary of the clarification.

---

[3] Their performance measures were based on (i) number of zero Cbs, (ii) whether the Cb that Centering found corresponded with a loose notion of sentence topic, and (iii) number of cheap vs. expensive transitions. Cheap and expensive refer to inference load on the hearer (Strube and Hahn 1999), based on whether Cp_{n-1}, the expected Cb_n, is actually realized as Cb_n.

[4] Speakers are referred to using their initials, plus an 'f' or 'm' in front, to identify their gender. These conventions are also used to name dialogues (e.g., in Table 2). 'famm_06_12' refers to speaker famm's turn 12 in conversation number 6. In the transcripts, slashes indicate backchannels (/uh huh/), and angle brackets false starts (<de> de). Intonation is represented by orthographic signs (. , ?).

3 Centering in Spanish

Centering is proposed as a cross-linguistic universal that relates participants' focus of attention and choice of referring expression. The transitions, constraints and rules are believed to be common to all languages, because they reflect human processing constraints[5]. The only adjustment necessary from language to language is the ordering of the Cf list, what Cote (1998) calls the *Cf template* for a language.

In English, the Cf template is considered to correspond closely to grammatical function and to linear order. Thus, subjects are ranked higher than objects, and these higher than adverbials. Walker, Iida and Cote (1994) proposed a different ranking for Japanese, which includes topic markers and markers of empathy in verbs, ranked higher than subjects or objects. Di Eugenio (1998) also ranks empathy the highest in her template for Italian, following Turan's (1995) for Turkish.

Spanish is a pro-drop language; subjects do not need to be realized as pronouns if they are known in context. Additionally, it has direct and indirect object clitics (unstressed pronouns). Corresponding stressed object pronouns are possible for animate entities only.

Animacy is a relevant feature in Spanish, in my view. Clitics and reflexive pronouns that refer to participants in the discourse have two characteristics that would make them candidates for a higher ranking: (i) they convey empathy and (ii) they are often placed before the verb, linearly before non-animate direct objects. In (2b), the most salient entities are 'you' and 'me' (captured in the clitic *me*); 'the time' should rank lower in the list.

(6) a./mm/ de todas formas <el> el martes estaré listo.
/mm/ of all ways <the> the Tuesday be-FUT-1SG ready
'Anyway I'll be ready on Tuesday.'

b. fíjame tú la hora
set-IMP-2SG-me-CL you the time
'You set the time (for me)'

[5] Walker (1998) discusses Centering as a model of human memory.

I also ranked experiencers in psychological and perception verbs higher in the scale, following previous research (Turan 1995; Brennan 1998). These are the experiencers in verbs such as 'interest', 'seem' and 'feel'. The experiencer in such verbs is often expressed through a pre-verbal clitic pronoun in Spanish (*me parece*, 'it seems to me'). Ranking animacy higher unifies empathy in psychological and perception verbs, and it also captures the fact that pronouns that refer to participants are placed first in the discourse, thus resulting in linear ordering of the Cf list.

A proposal for a Cf template for Spanish is presented in (7). This is a working version, and it still needs to address other issues: subject-verb inversion (in presentational and other constructions), possessives[6], etc.

(7) Cf template for Spanish
empathy/animacy > subject > animate indirect object > direct object > other

One case that I found especially difficult was the frequent occurrence of *para mí* ('for me') as a benefactive. The speakers often say 'X date is not good for me' or 'For me that's not good'. Following the template above, we could include it as an animate indirect object. However, the emphasis seemed to be different according to whether *para mí* appeared at the beginning or the end of the clause, possibly making this a case for considering word order as well as grammatical function.

4 Centering and Reference in Spoken Spanish

This section presents the results of the corpus analysis, an analysis of nine two-party conversations from the Interactive Systems Lab scheduling corpus[7]. These are conversations between two speakers, with the goal of finding a suitable time to meet. They are grouped by gender of participants, three conversations being between

[6] See Di Eugenio (1998) for a discussion of possessives.

[7] The conversations were recorded by ISL at Carnegie Mellon University. Thanks to ISL and Alex Waibel, its director, for permission to use the data.

	Utterances	Cb=0	Continue	Retain	Smooth shift	Rough shift
FMGL_FMCS_01	52	15	26	2	7	1
FNBA_FCBA_04	33	10	16	3	2	1
FVGC_FSNM_09	18	8	6	1	1	1
MARC_MPHB_02	27	9	9	5	2	1
MJBP_MMBU_04	29	4	13	3	5	3
MRBZ_MCRA_03	34	9	19	2	1	2
FCBA_MEBA_08	24	5	12	4	2	0
FMEM_MEOC_02	24	9	6	6	0	2
MJNM_FAMM_06	30	12	14	1	2	0
Total	271	81	121	27	22	11
% of total transitions (excluding 0, n=181)			66.85%	14.92%	12.15%	6.08%

Table 2. Transition types per conversation.

	Continue		Retain		Smooth shift		Rough shift	
Pro, participant	90	74.38	9	33.33	16	72.73	2	18.18
Stressed pronoun, participant	7	5.79	0	-	2	9.09	2	18.18
Demonstrative pronoun	2	1.65	1	3.70	1	4.55	3	27.27
Stressed IO (*para mí*)	5	4.13	1	3.70	0	-	0	-
Unstressed IO (clitic, *me*)	8	6.61	5	18.52	1	4.55	1	9.09
Pro, non-participant	5	4.13	1	3.70	1	4.55	2	18.18
NP, non-participant	3	2.48	3	11.11	0	-	1	9.09
Adverbial	1	0.83	7	25.93	1	4.55	0	-
Total	121		27		22		11	

Table 3. Choice of referring expression according to transition.

two females, three between two males, and three mixed. Conversations were divided into utterances, and centers coded for each. The conversations had in total 2,858 words and 271 utterances. They were divided in utterances, according to the methodology explained in Section 2, and centers were coded for each utterance. Overall numbers of utterances and transitions are presented in Table 2.

In general for the entire corpus, and in particular for each conversation, CONTINUE is the preferred transition by far (66.85% of all non-zero transitions in the corpus). RETAIN and SMOOTH SHIFT are similar in overall percentage (14.92% and 12.15%, respectively). Finally, ROUGH SHIFT is the least frequent of all transitions, occurring 6.08% of the time.

A number of backward-looking centers were empty, which resulted in 81 zero transitions. A backward-looking center is empty when none of the entities in the previous utterance, U_{n-1}, is repeated in the current utterance, U_n. In (8) the entities of (8a) are not repeated in (8b), because the speaker proceeds from the mention of 'I' and 'you' to 'us'. This was considered to be a new entity in the discourse. Given that Cb for (8b) is empty, the transition is zero. Instances of empty Cb were very common towards the end of the conversations, which consist mainly of good-byes and repetitions of the dates agreed upon. From a structural point of view, a series of empty Cbs

could then indicate that the conversation is nearing its end.

(8) a. fmgl_01_10: sí. okay. /eh/ te llamo por teléfono antes cuando yo salga de mi oficina. /mm/?
'yes. okay. /uh/ I'll call you on the phone before, when I leave my office /mm/?'
Cf: FMGL [*pro*], FMCS [te], PHONE [por teléfono], FMGL [yo], OFFICE [mi oficina]

b. así <[n]> combinamos bien
'That way <[n]> we can coordinate'
Cf: US [*pro*]; Cb: 0; Transition: ZERO

The focus of this paper is the relation of transition type to type of pronoun chosen for the backward-looking center. For each transition pair, I looked at the linguistic realization in the backward-looking center in the second utterance of the pair. That realization was then characterized according to the categories presented in Table 3. The first column for each transition presents raw frequencies, and the second column the percentage of that transition type that was realized in each of the categories. (Although the focus is on choice of pronoun, all the realizations of the Cb were coded.)

The first two categories include reference to participants as subjects, divided in stressed pronoun or *pro*. In (9), the speaker refers to herself with a stressed pronoun in the first sentence of the turn, but continues the reference with a zero pronoun (9b).

(9) a. a ver *yo* estoy de viaje del treinta y uno hasta el miércoles junio dos, el dos de junio,
'Let's see *I* am away from the 31st until Wednesday the 2nd, June 2nd.'

b. o sea que [*pro*] no voy a poder
so [*pro*] no go-1SG to be-able-to-INF
'So (I) won't be able to make it.'

Participants can also refer to themselves through indirect objects realized as clitics, as in (10) and (11), or with stressed pronouns preceded by a preposition: *para mí* ('for me'), *a mí* ('to me'), *a ti* ('to you'), as in (12). Pronouns can also be in first person singular (13).

(10) *me* parece lo mejor dejarlo para la otra semana,
me seem-3SG the best leave-INF-it-CL for the other week
'I think it's better to leave it till the next week,'

(11) /eh/ a qué hora *te* viene mejor?
/uh/ at what time *you* come-3SG better?
'What time is it better for you?'

(12) *para mí* esto está ideal Cati.
for me that be-3SG ideal Cati.
'For me that's ideal, Cati.'

(13) <pero> pero no *nos* alcanza el tiempo porque /eh/ tenemos hasta las cinco de la tarde.
<but> but not *us* reach-3SG the time because /uh/ have-1PL until the five of the afternoon.
'But that's not enough time, because we (only) have until 5pm.'

The next few categories include entities other than the participants, circumstances such as dates and places. These can also be realized as zero pronouns, as in (14), where *el jueves* (Thursday), the subject of *parece*, is implicit.

(14) a. ...me viene mejor el jueves, <pero> por ejemplo empezar a las dos de la tarde.
me come-3SG better the Thursday <but> for instance start-INF at the two of the afternoon.
'Thursday is better for me, <but> for instance to start at 2pm.'

b. qué te parece [*pro*=el jueves]?
what you seem-3SG
'What do you think (of Thursday)?'

In Spanish, it is not common to refer to a non-animate entity with a stressed personal pronoun (*ello*, 'it'). In the corpus, the pronoun of choice for these entities is a demonstrative (15). Finally, reference can also be made through NPs (16), or PPs (17).

(15) y así podemos hacer *eso*, y ya.
and so can-1PL do-INF *that*, and already.
'And then we can do *that*, and that's it.'

(16) bueno. *el dieciséis* está bien.
okay. *the sixteen* is good.
'Okay. *The 16th* is fine.'

(17) qué tal *pause* está tu horario *en esta siguiente semana, del ocho al doce.*
how *pause* is-3SG your schedule *in this following week, from the eight to the twelve.*
'How's... your schedule *in this coming week, from the 8th to the 12th.*'

As for the relation of transition and anaphoric term, and as was expected, CONTINUE transitions are frequently realized through pro-drop. I divided the use of pro-drop according to whether the dropped pronoun referred to a participant in the conversation or to some other entity (dates and places). In 74.38% of all CONTINUE, the Cb referred to a participant, and did so without an explicit pronoun. A few Cbs, 4.13% of CONTINUE, referred to another entity in the discourse. The rest of the categories are small in number, perhaps with the exception of a stressed pronoun to refer to a participant (5.79%). The use of a stressed pronoun in some of these cases might have a role in turn-taking (see below).

In the RETAIN transitions, again the most frequent realization was a *pro*, but two other categories are interesting. First of all, dative clitics (*me, te*, 'me', 'you') appear often. These appear sometimes in the first utterance of a new turn. In (18b) below, speaker fmcs addresses speaker fmgl for the second time. In (18c), the other speaker takes the turn, addressing fmcs, but making reference to herself ('you didn't tell *me* what Monday...'). Since the reference to fmgl is the only connection between the two turns, that is the Cb for (18c).

(18)a. fmcs_01_05: ... qué te parece?
'...what do you think?'

b. fíjate tu horario, a ver qué tal te viene.
'Check your schedule, and see whether that's good for you.'

c. fmgl_01_06: bueno. <primer> en primer lugar, no me dijiste qué lunes o martes o miércoles.
'Okay. First of all, you didn't tell me what Monday, or Tuesday, or Wednesday.'

SMOOTH SHIFT transitions very frequently result in a *pro* for the Cb of the utterance. Because of the situated nature of the conversations, it is not necessary to use a stressed pronoun to clarify referents, even when a shift in focus of attention takes place.

It is interesting to note that some of the stressed pronouns that referred to participants occurred right after a change of turn. There were in total four of these (one in CONTINUE, one in SMOOTH SHIFT and two in ROUGH SHIFT). For instance, in (19), a CONTINUE, speaker fcba signals turn-yielding by asking a question of the other speaker. Speaker fnba then starts a turn by making reference to herself with a stressed subject pronoun, *yo*.

(19) fcba_04_01: ... puedes reunirte conmigo en mayo?
'... can you meet with me in May?'

fnba_04_02: a ver yo estoy de viaje del treinta y uno hasta el miércoles junio dos, el dos de junio,...
'Let's see, I'm away from the 31st to Wednesday June 2, June 2nd,...'

This phenomenon opens up the possibility of establishing points of contact between Centering and Conversation Analysis (e.g., Sacks et al. 1974). There might also be a connection between transition type and turn-taking. Example (20) shows the last two utterances of a long turn in which the speaker presents his available dates, all resulting in CONTINUE transitions. The very last utterance in the turn introduces the other speaker with an imperative[8], but the speaker retains a reference to himself in the clitic *me*. This is the first RETAIN in the sequence, preparing for the turn change.

[8] Subjects of imperatives (i.e., the addressee) are considered entities in the discourse.

(20) viernes puedo todo el día.
'Friday I can all day.'

entonces mientras para no hacértelo más difícil, dime si puedes uno de estos días.
'So in the meantime so that this doesn't get more difficult, let me know if you can one of these days.'

Other phenomena that cannot be discussed here include null objects and their relationship to evoked entities, and the status of clausal subjects. I would also like to explore the relationship of stressed indirect objects (*para mí*) to their unstressed counterparts (clitics). The stressed versions are usually grammatically optional, and they can appear towards the beginning or the end of the clause.

Conclusion

This paper has presented an application of Centering Theory to dialogue in Spanish. A corpus analysis of nine task-oriented conversations shows certain correlations between transition type and choice of referring expression. I have also discussed some of the difficulties of applying Centering to dialogue, and how to establish the Cf template for Spanish. Future directions of this work include analysis of other corpora, especially written data, in order to provide a general characterization of certain phenomena in Spanish such as zero pronoun, clitic doubling, and choice of definite/indefinite article. Another extension will compare these results to both spoken and written English.

Further research is also needed to establish the adequacy of the proposed Cf template for Spanish. A measure would be its success in resolving anaphora.

The final aim of the study is to formalize the relationship between anaphoric terms and transitions, so that the formalization can be used in a computational application.

References

Brennan, S. (1998) Centering as a Psychological Resource for Achieving Joint Reference in Spontaneous Discourse. In Walker et al. (eds.): 227-249.

Brennan, S., M. W. Friedman and C. Pollard (1987) A Centering Approach to Pronouns. *Proceedings ACL 87*: 155-162.

Byron, D. and A. Stent (1998) A Preliminary Approach to Centering in Dialog. *Proceedings ACL 98*: 1475-1477.

Cote, S. (1998) Ranking Forward-Looking Centers. In Walker et al. (eds.): 55-69.

CREA (2002) *Corpus de Referencia del Español Actual.* Real Academia Española (www.rae.es).

Di Eugenio, B. (1998) Centering in Italian. In Walker et al. (eds.): 115-137.

Grosz, B., A. Joshi and S. Weinstein (1995) Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 2 (21): 203-226.

Hurewitz, F. (1998) A Quantitative Look at Discourse Coherence. In Walker et al. (eds.): 273-291.

Kameyama, M. (1998) Intrasentential Centering : A Case Study. In Walker et al. (eds.) : 89-112.

Sacks, H., E. Schegloff and G. Jefferson (1974) A Simplest Systematics for the Organization of Turn-Taking in Cconversation. *Language* 50: 696-735.

Strube, M. and U. Hahn (1999) Functional Centering-Grounding Referential Coherence in Information Structure. *Computational Linguistics* 25 (3): 309-344.

Turan, Ü. (1995) *Subject and Object in Turkish Discourse: A Centering Analysis.* Ph.D. thesis, U. of Pennsylvania.

Walker, M. (1998) Centering, Anaphora Resolution and Discourse Structure. In Walker et al. (eds.): 401-435.

Walker, M., M. Iida and S. Cote (1994) Japanese Discourse and the Process of Centering. *Computational Linguistics* 20 (2): 193-232.

Walker, M., A. Joshi and E. Prince, eds. (1998) *Centering Theory in Discourse*. Oxford: Clarendon.

Walker, M., A. Joshi and E. Prince (1998) Centering in Naturally Occurring Discourse: An Overview. In Walker et al. (eds.): 1-28.

Formalising Hinting in Tutorial Dialogues

Dimitra Tsovaltzi
Computational Linguistics
University of Saarland
Germany

Colin Matheson
Language Technology Group
University of Edinburgh
Scotland

Abstract

The formalisation of the HINTING PROCESS in Tutorial Dialogues is undertaken in order to simulate the Socratic teaching method. An adaptation of the BE&E annotation scheme for Dialogue Moves, based on the theory of SOCIAL OBLIGATIONS, is sketched, and a taxonomy of hints and a selection algorithm is suggested based on data from the BE&E corpus. Both the algorithm and the tutor's reasoning are formalised in the context of the INFORMATION STATE theory of dialogue management developed on the TRINDI project. The algorithm is characterised using UPDATE RULES which take into account the student model, and the tutor's reasoning process is described in terms of CONTEXT ACCOMMODATION.

1 Theoretical Background

1.1 Information State Update

The formalisation of the hinting process uses the Information State (IS) approach, in which dialogue is seen in terms of the information that each participant has at every point in the interaction. The formal account thus accords with the proposals put forward on the TRINDI project[1] (Bohlin et al. (1999), Larsson et al. (1999)).

The IS representation consists of fields containing different kinds of information about the dialogue.[2] This information is updated after each new utterance using update rules of various types. The latter include details of how each field, or information attribute, in the IS should be updated, if at all.[3] The background theories that inform the IS update formalisation assumed here are the obligation-driven approaches discussed in Matheson et al. (2000) and Traum and Allen (1994), and Context Accommodation as described in Cooper et al. (2000), Kreutel and Matheson (2000), and Larsson et al. (2000).

1.2 Discourse Obligations

The theory of obligations introduces the notion of discourse and social obligation as a way of analysing some of the social aspects of interactions and provides an explanation for behaviour that other theories do not predict. It is an augmentation to the representation of the INTENTIONS of dialogue participants that attempts to capture the natural flow of conversation. There are different genres of obligation-based research, as evidenced by the above references.

Treating tutorial dialogues in terms of obligations is an intuitive way of analysing and predicting some specific kinds of dialogue behaviour that do not seem to follow the rules of everyday discourse. For example, applying the Shared Plans theory (see Rich and Sidner (1998)) to tutorial dialogues would soon prove problematic since the tutor does not follow the principle of co-operativity which is central to the theory. A prerequisite for achieving goals, according to Shared Plans, is that the dialogue participants should be as clear as they can about their beliefs and plans, when the one thing that the tutor avoids doing in the hinting process is exactly that. For example, SharedPlans do not explain why the tutor in utterance T[2] in example 1 below does not just tell the student what a sine-wave is, which she could easily do.

[1] Telematics Applications Programme, Language Engineering Project LE4-8314.

[2] For examples of some relevant fields see (2) below.

[3] A commented example of an update rule can be found in (4 below.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Pages 185-192.

According to the obligation-driven framework, on the other hand, intentions are necessary but not the only driving force behind an utterance. For example, only if one considers it the student's obligation to address the tutor's questions and follow her directives can one interpret the total lack of overt signals from the student that he intends to cooperate, such signals being normally central to collaboration. In the context of the overall obligations the tutor knows what the student's intentions are, and will be able to interpret his actions correctly because the tutorial dialogue genre will not permit any other behaviour, since it is the student's obligation to follow the tutor's directives.

2 Obligations and Dialogue Moves

2.1 Dialogue Moves in the BE&E Corpus

An informal analysis of the dialogue moves that are observed in the BE&E corpus has been sketched taking into account considerations based on obligation theory as outlined above. Moves that do not contain any particular interest for the representation of obligations have generally been adopted from the BE&E annotation scheme. The BE&E (Basic Electricity and Electronics) corpus consists of recorded human-to-human dialogues that were carried out via a computer keyboard. The student performs actions in a lab, represented by a GUI, towards a specific target such as measuring current. The tutor observes the actions and intervenes when necessary, or when the student asks her to.[4] The move HINT is given here as an example for obvious reasons.

2.2 The Hint Move

It is clear from the corpus that the student is obliged to address the tutor's utterances, whereas the tutor hardly ever answers questions directly, contrary to the norm outside the genre. Instead, she gives HINTS which withhold the answer. A hint is a tutorial strategy that aims to encourage active learning. It can take the form of eliciting information that the student is unable to access without the aid of prompts, or information which he can access but whose relevance he is unaware of with respect to the problem at hand. Alternatively, a hint can point to an inference that the student is expected to make based on knowledge available to him, which helps the general reasoning needed to deal with a problem (Hume et al., 1996).

Partial answers from the tutor discharge the obligation to address the student's questions. There are examples in the BE&E corpus where the tutor explicitly states her method or the student shows he is aware of it, such as "Very good. You answered your own question" or "I'll give you another hint.", from the tutor and " I need another hint", from the student. The initiation of hints can be due to various reasons:

(i) the tutor observes that the student is not making any progress in the task, or that he is taking steps in the wrong direction

(ii) the student asks a question and the tutor does not want to answer it directly

(iii) the student gives the wrong answer or asks the wrong question in response to a tutor's question

Figure 1 illustrates some of the points briefly discussed above. The tutor gives a few hints to try to make the student follow her reasoning (in T[2], T[4], T[6], and T[7]). She realises, however, that the student does not remember the lesson very well; he is so bad at interpreting her hints that she is forced to give explanations about basic concepts (T[9]). She therefore asks him to read the lesson again (T[11]), not wanting to just give the answers away.

3 The Hinting Strategies

Hint is one of the moves that appear in tutorial dialogues. Although the surface structure of hints is heterogeneous, there appears to be an underlying structure common to different categories of hints, and we undertook the formalisation based on these perceived regularities. The taxonomy of strategies formalising the hinting process in our model entails the following hints: Pragmatic Hint, Relevant Concept, Following Step, Speak to Answer, Encourage, Rephrase Question, Logical Relation, Move Backwards, More General Concept, Spell Out the Task, Explain Fundamental Principles of Domain, Narrow Down Choices, Explain, and Point to Lesson. The names are intended to

[4] Also see http://www.hcrc.ed.ac.uk/~jmoore/tutoring/BEE_corpus.html

S[1]: I have no idea what a sinewave is. Was this covered in the tutorial?
T[2]: Yes, remember the wave that represented alternating current in the lesson?
S[3]: I think i remember it being represented as a ~ on the ammeter control panel
T[4]: OK, that's true about the multimeter's function dial. But do you remember
S[5]: Nope
T[6]: a graph of a wave in the lesson that represented alternating current? (20 sec later)
T[7]: Do you remember reading about frequency and amplitude and all that?
S[8]: I'm not sure. Is this a trick question to see if you can get me to invent a memory?
T[9]: No, this is not a psychology experiment. :) I'm just trying to see how much you remember. A sinewave starts out at 0 and increases to the maximum amplitude then decreases past 0 in the negative direction and then returns to 0 again. Does any of this ring a bell?
S[10]: It really doesn't. But I think I'm following your explanation.
T[11]: Go ahead and reread the lesson.

Figure 1: Example tutorial dialogue

be as descriptive of the content as possible, and should in some cases be self-explanatory. Some of the strategies are not real hints (for example Point to the Lesson), but they have been included in the taxonomy because they are part of the general hinting process. Although the strategies were derived by looking at data from a specific domain, namely Basic Electricity and Electronics (BE&E), the aim was to abstract from the particular characteristics of that domain and produce a domain-independent taxonomy, to the extent that this is possible.

The hinting strategy can be realised either by giving a small clue which the tutor deems necessary to initiate the desired reasoning process, or by eliciting the clue itself, that is by prompting the student to produce it. These bottom-level moves are called `informs` and `t-elicits` respectively.

3.1 An Example of a Hinting Strategy

Example (1) below contains an instance of the hint `relevant concept`. In employing this strategy the tutor points to concepts that are relevant to the current problem in order to trigger the information in memory which is required, or in order to correct the student's erroneous reasoning. This can be done by asking a question the answer to which is the relevant concept, or the answer to which is only part of the concept, by asking *about* the relevant concept or by simply mentioning it. The relation of the concept pointed to and the one at hand, whether they are opposite, similar, related to the problem in a similar way, and so on, is made explicit:

(1) **T[1]:** ...What are the instructions asking you to do?
S[2]: Make a circuit between the source(battery) and the resistance (rheostat) and then attach the miliampmeter to measure the resistance in ohms.
T[3]: OK, you're close. But keep in mind that a miliampmeter is a special case of an ammeter. Do you remember what an ammeter measures?
S[4]: Amps?
T[5]: Right, very good...

In example (1) the tutor is trying to make the student see that the meter does not measure Ohms but Amperes. In utterance T[3] she brings up the notion of ammeter, which also measures Amperes. This strategy is analysed here as an example of a `relevant concept` style of hint (formalised in (2) below). She hopes that the student will remember this and will infer that the meter measures Amperes and, thus, current. The ultimate aim of course is that the student will realise that he must use the ohmmeter in order to measure resistance in Ohms, which is the problem at hand. The student follows the hint, as indicated by S[4].

4 Modelling the Hinting Process

In order to model the hinting process an algorithm was derived based on the examples from BE&E corpus. It takes into account the current student answer and the number of wrong answers encountered so far in the dialogue as a means of deciding on the student model and

on which hint to generate. This gives emphasis to the local student model which has been found to be more important than the global one (Freedman et al., 1998). Six categories of student answers according to their performance were judged necessary based on the data: correct, near miss, partially correct (only part of the correct answer was given), grain of truth (there is an indication of some understanding of the problem but the answer is wrong), wrong, misconception (typically confused concepts and suchlike).

The initiation of the hinting process, for example when the student performs a wrong action in the lab, and the conclusion, perhaps when he does not need more help, were also modelled by the algorithm. The algorithm itself was modelled using update rules in the TRINDI format. Section 4.1 includes an example of an update rule whose preconditions and effects were derived based on the algorithm.

The algorithm is a comprehensive guide and we do not claim that it accurately predicts all the dialogues in the corpus. This inaccuracy is largely due to the inconsistency observed in the human tutor behaviour, which is not pedagogically justifiable. We have tried to overcome these inconsistencies by normalising the behaviour based on the theory behind the Socratic method, and hence sacrificing flexibility. Nevertheless, we have no evidence as to whether consistency of an effective method or flexibility is better.

The algorithm takes into account that fact that every time there is a new answer from the student the local student model, the performance in one hinting session, changes. The tutor's hint reveals as much information as needed based on the understanding demonstrated by the current answer. So, when the hinting session has just started it is easier to terminate it, if the student seems to follow. However the hinting strategies used are not very revealing. The tutor will start the hinting session when the student makes a mistake. If she is not certain of the reason behind the mistake, she will do `check for the origin of mistake`. At this level, if the answer is correct, the hinting process will end. If not, the tutor will choose the appropriate hint according to the type of answer the student gives. All hints at this second level are more informative, since the student has already been given a less informative hint and couldn't follow it. After the second hint, the hinting process will continue even if the student gives a correct answer as the tutor is reluctant to let the student carry on by himself. Performance so far has not been good, so she wants to guide the student and make sure he understands the whole task. The hints are yet more helpful.

After three hints in a row, the tutor will require only correct answers in order to continue hinting. In any other case the student model is too bad, since by now the hints are quite revealing and are still not being followed. In order both to avoid frustration on the student part, and to preserve the effectiveness of the Socratic method, the student is given a brief explanation. After the explanation, the tutor will check for understanding again. If a correct answer is still not forthcoming, the student will be referred to the study material, which he obviously has not read properly. The aim is not to teach the material, but to help the student to assimilate it, once read, and to be able to apply it.

4.1 The IS Update Rules: The Basics of the Formalisation

The formalisation of the hinting process follows the approach outlined in Matheson et al. (2000) in using detailed representations of information states and in handling the participants' deliberations using update rules some of which are specific to the particular genre of dialogue. Thus some of the rules contain conditions and effects specific to the hinting process in tutorial dialogues. The conditions must be satisfied in order for the IS to be updated, and the kind of update that will take place is defined by the effects. Example (2) contains an update rule which models the circumstances in which the hinting strategy `relevant concept` should be generated (the notation used in the rule is described in sections 5 and 6 below). This strategy is exemplified in utterance T[3] in (1) above. The student is responding to a `check for the origin of the mistake` in utterance T[1] in (1), but the answer given in S[2] is wrong.[5] Therefore, the tutor points to a

[5]Three kinds of answers are classified as "wrong": genuinely wrong answers, wrong steps in the task and

relevant concept.

(2)

name:	RELEVANT CONCEPT
cond:	in(IS.DH,or_mistake(T))
	in(IS.LM,wrong_answer(S,DA))
effect:	push(IS.INT,ack(T))
	push(IS.INT,rel_concept(T)

The conditions for the hinting formalisation presented here are the ones that must hold according to the algorithm. Briefly, the conditions state that the dialogue history contains a check for the origin of the mistake (or_mistake) and that the latest move is a wrong_answer. The effects are that the tutor has intentions to acknowledge the student's utterance and to perform a relative concept hint. The rule thus implements both the student model, determining if a hint should be generated, and also determines how informative the hint should be, via the effects. The type of hint to be generated depends on the type of answer elicited, and this is also represented in the effects, as shown.

5 The Tutor Model

The notion of CONTEXT ACCOMMODATION has been used in the past to deal with issues such as over-answering (see Cooper et al. (1999)), Question Under Discussion (QUD) ((Ginzburg, 1996)), and for incorporating different plans (Cooper et al. (2000), Kreutel and Matheson (2000), Larsson et al. (2000)). Here we suggest the application of context accommodation to the hinting process. In tutorial dialogues the intelligent system can use context accommodation to decide whether the student is on the right track, or if the hinting process needs to start, by accommodating steps in a predefined order from a given plan.

Three levels of planning are suggested here. At the top level DOMAIN holds a set of plans for all the lab tasks, each consisting of a set of subplans, called preconditions, that the total plan is broken down to.[6] These have to be realised for the goal of the total plan to be fulfilled. Every precondition in turn is broken down into decompositions which represent steps in the possible reasoning process towards achieving the precondition. Therefore, if there is more than one way of reasoning, all the options should be included in the database. The same goes for the preconditions and the plans. The AGENDA holds the subplan at hand, with its preconditions and decompositions, as well as realisations (fixed utterances) for the particular content of every decomposition. In a system that provides a dynamic language generation module, the relevant hint, the decomposition, and all discourse planning information can be passed on, allowing the module to generate an appropriate phrase. A plan of the kind described for the BE&E corpus can be seen in 3. The decompositions (dec) refer only to the last precondition (prec) here.

(3)

header:	Voltage Lab: Measure Voltage(VDC)
prec:	Polarity observed
	Meter on while making measurement
	Meter set to VDC
	Both leads attached
	Meter attached
	Circuit not powered down
	Difference in charge between the two ends of the meter when attached to the circuit
dec1:	vol.difpot.meas
dec2:	vol.meas.so
dec4:	vol.meas.lo
dec5:	vol.difpot.so
dec6:	vol.difpot.lo
dec7:	sou.dif
dec8:	lo.dif
dec9:	sou.ex
dec10:	lo.ex
sub-effects:	Difference in charge between the two ends of the meter when attached to the circuit
total-effect:	Voltage Lab: Measure Voltage(VDC)

Finally, the Dialogue History (DH), which holds all the dialogue acts performed in one hinting session, and the system variable Latest Move (LM) communicate with the AGENDA via the update rules.[7] Based on the latter, either the agenda will be modified, in which case the generation module will be activated, or not, in which case the agenda is only used for passively

no answer at all.

[6] Some of these plans already exist for the BE&E corpus using the same notation.

[7] Update rules are domain-independent but genre-specific.

keeping track of the student's step towards the realisation of the goal of the lab task at hand by popping preconditions off AGENDA and popping plans off DOMAIN.

Update rules are used to model everything described here as well as the accommodation of any preconditions performed and any decompositions which occur in a random order, whenever this is allowed by the task. This is the case only when certain steps, which constitute preconditions for the one currently being performed, have already been performed themselves. For instance, resetting the meter is a precondition of moving any wires when making a measurement. Context accommodation accommodates steps in the task, or the relevant reasoning, which are encountered before they are expected. It captures the fact that the student is following specific points in the tutor's plan and there is no reason to go through the steps that cover them again explicitly.

Example 4 shows the update rule for accommodating random steps, that is, correct steps (or preconditions) that are taken by the student in an order which differs from the order in the tutor's plan:

(4)

name:	**ACCOMMODATION OF RANDOM STEP**
cond:	in(IS.LM,cor_answer(S, Substep)) match_agenda(Substep, Agenda, Domain) not(last_member(AGENDA, cor_answer(S, Substep)))
effect:	push(IS.AGENDA,cor_answer(Substep))

The conditions specified are:

(i) that there is a correct answer that was just performed by the student (that would be the step to be accommodated)

(ii) that the correct answer matches a subplan in AGENDA which is currently the one accommodated by DOMAIN

(iii) that the substep at hand is not the final one in AGENDA.[8]

[8]If it were, another update would fire because it would mean that the reasoning, or task, has been correctly completed.

The effect will be that the substep will be pushed on top of the agenda, and the task can continue from there.

6 An AVM representation

An attribute-value matrix (AVM) representation of the dialogue in example (1) is shown in Figure 2. This represents the course of actions that would produce the last dialogue act by the tutor according to the update rules that formalise the algorithm. For current purposes the fields that are being used are AGENDA, generating the steps to be followed; OBL, holding any actions that have been classified as obligations; INT, which holds the actions to be realised in one turn here, and LM and DH, which stand for Latest Move and Dialogue History respectively, and which as mentioned above are used for capturing some aspects of the student model. For this reason DH includes all the previous dialogue acts, until the system comes across an update rule that specifically tells it to empty the DH.

The Intelligent Tutor, and with it the hinting session, is activated by an information request from the student (S[1]). This is acknowledged explicitly and the hint `relevant concept` points the student in the correct direction (T[2]). With this prompt the student remembers something remotely relevant and gives a grain of truth answer (S[3]). That makes the tutor generate an acknowledgement again and a `speak to answer` hint as a follow-up (T[4]). This action is interrupted (S[5]) and continued again in T[6]. (There is no formal account of interruptions here). What should be a set amount of time goes by and the student does not respond at all, so the tutor's interpretation is that the student does not follow, which for the algorithm is classified as a wrong answer. Hence, a `logical relation` is produced to give more profound guidance (T[7]). The student gives another wrong answer, (S[8]), and that is enough for the tutor to start explaining the current substep (T[9]). After this she checks for the origin of the problem (T[9]) in order to assess the student anew and perhaps generate an appropriate hint. In this case the student gives another wrong answer (S[10]).

All these moves are held in DH, as shown, and as mentioned above LM is the latest move, here the student's wrong answer. Together with DH

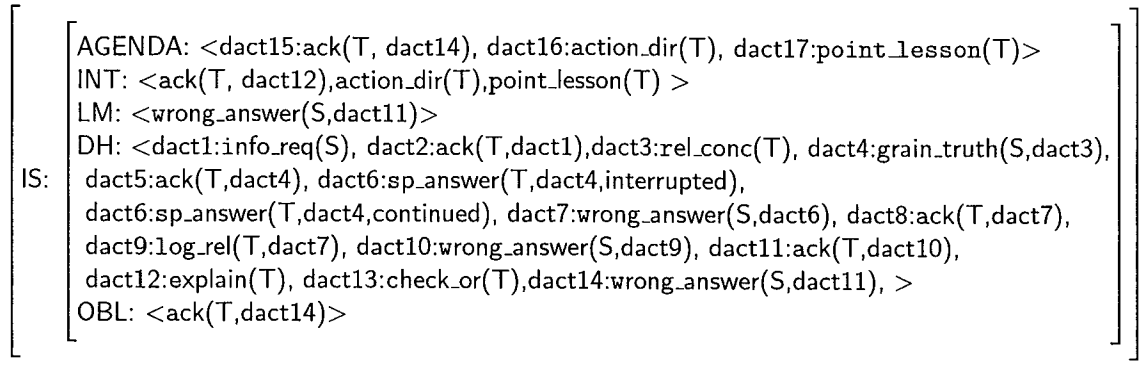


Figure 2: AVM representation of an Information State

they model the student performance, and based on the number of wrong answers the tutor will now direct the student to read the lesson again (T[11]). The student is obliged to do this without negotiation, and the session will stop in the state following the one represented by this particular AVM (after the `point_lesson` intention has been performed). OBL holds the obligation for the tutor to acknowledge the student's answer; here this is done implicitly, but this is enough due to the special obligations. AGENDA holds the acts that must be produced next, based on the update rules, which become the tutor's intentions as represented in INT. These are the moves to be realised in the current turn.

7 Related Work

The BE&E project (Core et al., 2000) also assumes the TRINDI framework, and a plan based approach, but does not allow for partial order in the student steps, modelled here by Context Accommodation. The project employs multiturn tutorial strategies, some of which are motivated by similar theoretical interests to the ones presented here. However, the number of strategies is small and no emphasis is given to the way information is made salient, which is the aim of our taxonomy. The criteria for using one strategy over another are also not clear. Note that Core et al. (2000) contains descriptions of other Intelligent Tutoring Systems that cannot be included here due to lack of space.

Miss Lindquist (Hefferman and Koedinger, 2000) also has some domain specific types of questions that resemble the BE&E strategies in form. Although there is mention of hints, and the notion of gradually revealing information by rephrasing the question is prominent, there is no taxonomy of hints or any suggestions for dynamically producing them.

A detailed analysis of hints can also be found in the CIRCSIM project, and in particular in the work of Hume et al. (1996). This paper has largely been inspired by the CIRCSIM work both for the general planning and for the taxonomy of hints, although the strategies recognised in it are domain specific.

8 Conclusion

An analysis of moves based on obligations and a taxonomy of hints is proposed. An algorithm formalising the hinting process based on the obligations and the taxonomy, and update rules which model the algorithm in accordance with the TRINDI project proposals, have been presented briefly. A suggestion has been put forward for applying context accommodation to hinting and thus modelling some aspects of the Intelligent Tutor. Future considerations include the integration of the move analysis into the hinting process, a reasoner for evaluating and categorising the student answers, a database as described above and, of course, the full implementation of the system.

References

Peter Bohlin, Robin Cooper, Elisabeth Engdahl, and Larsson Staffan. 1999. Information states and dialogue move engines. In Jan Alexanderson, editor, *IJCAI-99 WOrkshop on Knowledge and Reasoning in Practical Dialogue systems.*

Robin Cooper, Staffan Larsson, Colin Matheson, Massimo Poesio, and David Traum.

1999. Coding instructional dialogue for information states. Technical report, University of Gothenburg.

Robin Cooper, Staffan Larsson, Elisabeth Engdahl, and Stina Ericsson. 2000. Accommodating questions and the nature of qud. In *Proceedings G=F6talog2000.*

Mark G. Core, Johanna Moore, and Claus Zinn. 2000. Supporting constructive learning with a feedback planner. Technical report, Human Communication Research Center, University of Edinburgh, 445 Burgess Drive, Menlo Park CA 94025.

Reva Freedman, Zhou Yujian, Michael Glass, Jung Hee Kim, and Martha W. Evens. 1998. Using rule induction to assist in rule construction for a natural-language based intelligent tutoring system. In *Proceedings Twentieth Annual Conference of the Cognitive Science Society*, pages 362–367, Madison.

Jonathan Ginzburg. 1996. Dynamics and the semantics of dialogue. *Language and Computation*, 1.

Neil T. Hefferman and Kenneth R. Koedinger. 2000. Building a 3rd generation its for symbolization: Adding a tutorial model with multiple tutorial strategies. In *Proceedings of the ITS 2000 Workshop on Algebra Learning*, Montreal, Canada.

Gregory D. Hume, Michael A. Joel, Rovick A. Allen, and Martha W. Evens. 1996. Journal of the learning sciences. *Hinting as a Tactic in One-On-One Tutoring*, 5(1):23–47.

Joern Kreutel and Colin Matheson. 2000. Incremental information state updates in an obligation-driven dialogue model. *Language and Computation*, 0(0):1–32. to appear.

Staffan Larsson, Peter Bohlin, Johan Bos, and David Traum, 1999. *Trindikit 1.0 manual.*

Staffan Larsson, Robin Cooper, and Elisabeth Engdahl. 2000. Question accommodation and information states in dialogue. In *Third Workshop in Human-Computer Conversation*, Bellagio.

Colin Matheson, Massimo Poesio, and David Traum. 2000. Modelling grounding and discourse obligations using update rules. In *Proceedings NAACL 2000*, Seatle.

Charles Rich and Candace L. Sidner. 1998. Collagen: A collaboration manager for software interface agents. *User Modeling and User-Adapted Interaction*, 8(3/4):315–350.

D. R. Traum and J. F. Allen. 1994. Discourse obligations in dialogue processing. In *Proceedings 32nd Annual meeting of the Association for Computational Linguistics (ICSLP-92)*, pages 1–8.

Applications of Lifelong DRT

Gábor Alberti*, Judit Kleiber*, Helga Latyák* and **Helga M. Szabo****
*Linguistics Dept, University of Pecs, Hungary
**National Association of the Deaf, Hungary

After the "international" demonstration of an extension of (van Eijck and Kamp's 1997) Discourse Representation Theory [0], called Lifelong DRT [1], and a ("totally lexicalist" unificational) generative grammar intended to serve as an ideal compositional counterpart of (L)DRT, called Generalized Argument Structure Grammar (GASG) [2], papers on applications to special texts [3–4] and texts in special languages such as Hungarian Sign Language [6] and (Hungarian) child language [5] were/are to be read at national or regional conferences. The planned poster presentation is devoted to the "international" introduction of this project.

The crucial innovation of Lifelong DRT, what provides an effective means for formal analysis of special texts, lies in regarding the interpreter's information state containing the mutual background knowledge shared by "speaker" and interpreter as a gigantic "lifelong" DRS where all pieces of the interpreter's lexical, cultural/encyclopedic and interpersonal knowledge are accessible whose mobilization the exhaustive interpretation of the given text requires.

References

0. van Eijck, J. - Kamp, H. (1997) Representing discourse in context. In: van Benthem, J. & ter Meulen, A. (eds.) Handbook of Logic and Language. Elsevier, Amsterdam, & MIT Press, Cambridge, Mass.
1. Alberti, G. (2000) Lifelong Discourse Representation Structures. Gothenburg Papers in Computational Linguistics 00-5, Sweden, 13–20.
2. Alberti, G. - K. Balogh - J. Kleiber (2002) GeLexi Project: Prolog Implementation of a Totally Lexicalist Grammar. In de Jongh - Zeevat - Nilsenova (eds.) Proceedings of the Third and Fourth Tbilisi Symposium on Language, Logic and Computation, ILLC, Amsterdam, and Univ. of Tbilisi.
3. Alberti, G. - J. Kleiber (2001a) Világok között az Életfogytiglani DRS-ben [Among Worlds in Lifelong DRS]. In A. Kabán (ed.) Funkcionális mondatperspektíva és szövegszerkesztési stratégia [Studies in Text Theory], University Press, Miskolc, 35–46.
4. Alberti, G. - J. Kleiber (2001b) Ellentét, hasonlóság vagy következmény? (Elemzések az Életfogytiglani Diskurzusreprezentációs Elméletben) [Contradiction, Similarity, or Implication? Analyses in Lifelong DRT]. Talk at a Budapest conference, Ellentét és ekvivalencia a jelentésben [Contradiction or Equivalence in Meaning], Oct. 2002.
5. Alberti, G. - H. Latyk (2002) Child Language Analyses in the Framework of Lifelong DRT. Accepted talk at a Budapest conference, Linguistic Socialization, Language Acquisition and Language Disorders, October 7–9.
6. Alberti, G. - H. M. Szab (2002) Discourse-Semantic Analysis of Hungarian Sign Language. Accepted talk at TSD2002, Brno, September 9–12. To appear in its Proceedings, Springer-Verlag.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 193.

Sound and Function Regularities in Nonlexical Interjections in English

Gina Joue and Nikolinka Collier

Recent research presents strong arguments in support of using discourse particles (or discourse markers) in modeling utterances. The use of discourse particles in defining the boundaries of an utterance has been shown in, for example, Heeman (1997). In this presentation, we build on Heeman's claim that discourse particles, including nonlexical interjections, are best processed as utterances rather than merely on the lexical level. We adopt the view of discourse particles as discourse managers that constrain spoken interaction. The research we present focuses on a particular group of discourse particles, nonlexical interjections.

Work on nonlexical interjections have been primarily confined to their role in repetitions, false starts and filled pauses. In most of these models, there is no clear account of how to collapse the sound variants (e.g. ummmm or uhhhhm) of nonlexical interjections to their relevant baseforms or dictionary entry (e.g. um). We believe that the reason for this is the lack of work done on the interdependencies of the sound structure and the functional realisation of nonlexical interjections, and this work is an attempt to contribute to this area of research.

In our analyses, we developed DiSPEL, a discourse particle taxonomy, which defines the function of particles based on the denotation of the particle, the position of the particle within the utterance, and previous discourse moves. With DiSPEL, we annotated the nonlexical interjections in TRAINS 91 (American English) and the HCRC Map Task (Glaswegian English) corpora. The sound and function interdependencies we find are explained with principles drawn from Phonology as Human Behavior (Diver, 1979; Tobin, 1997), a cognitive phonological theory.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 194.

GoDiS—Issue-based dialogue management in a multi-domain, multi-language dialogue system

Staffan Larsson and **Stina Ericsson**
Göteborg University
Box 200, 40530 Göteborg, Sweden
{sl,stinae}@ling.gu.se

GoDiS (Gothenburg Dialogue System) is an experimental dialogue system utilizing dialogue management based on Ginzburg's concept of Questions Under Discussion (QUD). GoDiS is implemented using the TrindiKit, a toolkit for implementing dialogue move engines and dialogue systems based on the Information State approach. While originally built for fairly simple information exchange dialogue, it is being extended to handle action-oriented and negotiative dialogue.

One of the goals of the information state approach is to encourage modularity, reusability and plug-and-play; to demonstrate this, GoDiS has been adapted to several different dialogue types, domains, and languages, including information exchange dialogue in travel agency and autoroute domains, and action-oriented dialogue when acting as an interface to a mobile phone or VCR. GoDiS has also been enhanced with speech input and output, and is able to switch between Swedish and English in an ongoing dialogue.

The GoDiS architecture is an instantiation of the general TrindiKit architecture. The components are the following: the Information State (IS), domain-independent modules, the Dialogue Move Engine (DME), a controller, and three domain-dependent resources: Database, Lexicon, and Domain Knowledge. The TIS is accessed by modules through conditions and operations. The types of the various components of the TIS determine which conditions and operations are available.

The information state used here is basically a modified version of the dialogue game board proposed by Ginzburg. The main division in the information state is between information which is PRIVATE to the agent and that which is SHARED between the dialogue participants. The PLAN field contains a dialogue plan, i.e. is a list of dialogue actions that the agent wishes to carry out. QUD is a stack of questions under discussion. These are questions that have been raised and are currently under discussion in the dialogue.

Adaption to new domains and languages is simplified by the clear separation of domain-dependent and domain-independent components. Given a DME which can handle the dialogue type required by the domain, e.g. information exchange, adaption to a new domain amounts to building resources for that domain. Adaption to a new language only requires adaption of the lexicon resource. Also, resources can be switched on-line, in principle allowing the system to change domain and language during dialogue.

Question and task accommodation allows GoDiS to adapt to the user's utterances, e.g. allowing the user to answer questions before they have been asked, and providing information before any task has been established. Given a simple schema specifying the reimplementation of menu-based user interfaces in the form of dialogue plans, these mechanisms allow a more natural and flexible interface than a simple menu.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 195.

A multi-threaded dialogue system for task planning and execution

Oliver Lemon, Alexander Gruenstein, Laura Hiatt, Randolph Gullett, Elizabeth Bratt, and **Stanley Peters**
CSLI, Stanford University, Stanford, CA 94305

We will demonstrate practical dialogue management techniques for conversations involving planning and execution of concurrent tasks or activities. Conversational context is computed using a "Dialogue Move Tree" and an "Activity Tree" which represent interleaved threads of dialogue about different collaborative activities and their execution status. Dialogue threading also allows the dynamic use of context-sensitive speech-recognition language models—resulting in faster more robust recognition. We shall also demonstrate incremental message selection, generation, and aggregation methods used in the dialogue manager.

The domain of this demonstration is interaction with a robot helicopter, or UAV ('Unmanned Aerial Vehicle'). The same dialogue manager is also being used for intelligent tutoring applications (the MURI Project). The system components are OAA, Gemini, Nuance 8, Festival, and the CSLI dialogue manager. We will demonstrate realtime spoken dialogues, with open-mic speaker-independent recognition, such as the following:

U: Fly to the tower [start thread one]

R: Ok. Now taking off and flying there

U: Our job is to look for a red car. [start thread 2]

R: Ok. I'm looking for it.

R: I have found a red car. Have I found the right car?

. [user does not answer]

R: Have I found the right car? [re-raise question]

U: Forget the car. [cancel pending question]

R: Okay. I have cancelled looking for a red car.

R: I have reached the tower. What shall I do next? [system prompt user]

U: Land at the warehouse. [start thread 3]

R: Okay. Now flying there.

Interesting aspects of the system are its use of multi-threaded dialogue context, domain-independent dialogue moves, support for interleaved planning and execution dialogues, support of context-sensitive speech recognition, and coverage of a range of dialogue phenomena (e.g. commands, questions, repairs, revisions, clarification subdialogues, turn taking, ellipsis, message selection and aggregation).

New in this system (compared to the ACL & SIGdial 2002 demos) are: turn-taking, constraint management dialogues, "undo" moves, and new activities of the device.

Details of the system architecture can be found in proc. SIGdial 2002 and at: `http://www-csli.stanford.edu/semlab/witas/`

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 196.

DMT Analysis of a Difficult Dialogue

William C. Mann
SIL International
bill_mann@sil.org

Dialogue Macrogame Theory is the working framework of Mann's Edilog presentation entitled "Dialogue Analysis for Diverse Situations." It presents an analysis of a simple, wholly cooperative dialogue.

The poster presents a different sort of dialogue, one which includes interaction between near-adversaries.

In 1993 there was a confrontation in Waco, Texas, USA between the FBI (with other government organizations) and a religious commune of a sect known as the "Branch Dravidians." The commune leader, David Koresh, seeks to extend argumentation from a prior conversation. The government representative, Sheriff Harwell, seeks to convey a fresh offer. They "talk past each other" at times, ignore each other's speech at times and yet finally come to an agreement.

This dialogue has been called "hostage negotiation," but that label is quite approximate, since the "hostages" are children of the adults of the commune.

The poster presents the analysis, identifies the Dialogue Macrogames involved, identifies the basic acts that are used. These include both attempts at collaboration and "Unilaterals" which are various kinds of single party acts.

This analysis and others, along with information about the framework, are available on http://www-rcf.usc.edu/~billmann/dialogue.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 197.

Examples of DMT Analysis

William C. Mann
SIL International
bill_mann@sil.org

Dialogue Macrogame Theory is the working framework of Mann's Edilog presentation entitled "Dialogue Analysis for Diverse Situations."

This poster presents analyses of dialogues from other situations. The situations include the Apollo 13 spacecraft emergency, tutoring and travel agent interactions. Others may be included. The framework has also been applied to medical interviews, online manual computer help and to dialogues elicited in various laboratory situations.

All of the analyses, along with information about the framework, are available on http://www-rcf.usc.edu/~billmann/dialogue.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 198.

The SPAAC system for multidimensional annotation of dialogue

Martin Weisser and **Geoffrey Leech**
Lancaster University
Department of Linguistics and Modern English Language
Lancaster University, LA1 4YT, U.K.
m.weisser@lancaster.ac.uk g.leech@lancaster.ac.uk

The SPAAC dialogue annotation system has been developed (under EPSRC grant GR/R371542) primarily for XML speech-act annotation of service dialogues, but in the course of working on it we have broadened the annotation to include six dimensions:

(a) segmentation of dialogue turns into utterances, C-units and discourse markers

(b) tagging of syntactic form (e.g. declarative, interrogative, imperative, fragment)

(c) tagging of topic or subject matter (e.g. in train booking dialogues, topics such as the following are recurrent: address, arrival, cancel, creditcard, date, departure)

(d) tagging of mode (e.g. alternative, condition, probability, expletive)

(e) tagging of polarity positive vs. negative

(f) tagging of speech acts (or dialogue acts), using a taxonomy of 40 tags (accept, ackn, answ, answ-elab, etc).

Dimensions (c) and (d) can be seen as respectively concerned with domain-restricted semantics and domain-free semantico-pragmatics. Dimension (f) provides the key level of annotation, for which other dimensions contribute diagnostic information.

The tagging is semi-automatic: after an initial parsing-assisted structural segmentation of the unpunctuated turns, the annotation tool SPAACy (developed by Weisser) then automatically adds XML mark up to the dialogue and undertakes an analysis of all five dimensions above. The output of SPAACy is then manually post-edited.

The tool has so far been thoroughly tested on a corpus of two kinds of data: (1) train booking dialogues, and (2) telephone operator dialogues. A third kind of dialogue (non-service data from the British National Corpus) will shortly be tested. The system is designed to have generic capabilities to handle a wide range of dialogue types with little or no adaptation, and the plan is to make the resulting annotated corpus available for further dialogue research.

The poster will go into more detail on the rationale behind the SPAAC system, comparing it with other dialogue annotation schemes, and showing samples of annotated data.

The demonstration will show the stage-by-stage operation of the SPAACy software.

Bos, Foster & Matheson (eds): Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), 4-6 September 2002, Edinburgh, UK, Page 199.

A 3-tier Planning Architecture for Managing Tutorial Dialogue*

Claus Zinn, Johanna D. Moore, and Mark G. Core
Division of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK
[zinn|jmoore|markc]@cogsci.ed.ac.uk

There is mounting evidence from cognitive science and intelligent tutoring system 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 unexpected responses, interruptions, and failure of tactics. To provide these capabilities, we have developed a dialogue management framework, inspired by the three-level architectures used in robotics. The integration of a deliberative planner with a dialogue move engine allows the interleaving of planning, execution, and monitoring.

From the software engineering perspective, it is critical that the 3-tier planning architecture and the rest of the dialogue system be as modular as possible to minimize redundancy, and facilitate reusability, maintainability, and extendibility.

Our system, BEETLE (Basic Electricity and Electronics Tutorial Learning Environment) is composed of the following modules, which interact using the Open Agent Architecture (OAA).

- a deliberative planning and execution monitoring module implemented using the Open Planning Architecture (O-Plan) [CT91].
- an NLU module which translates the user's typed English input to a logical form. This is built using the CARMEL NLU toolkit [Ros00], which includes a robust parser and a wide coverage grammar.
- a generation module which synthesises English text from logical forms, implemented in Exemplars [WC98] and the Apache Xalan XSLT stylesheet processors.
- a dialogue manager developed using the TRINDIKIT dialogue system shell [LT00]. 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.
- the Basic Electricity and Electronics Reasoner (BEER), which encodes all of the system's knowledge about the subject matter to be tutored. BEER is written in LOOM, a description-logic-based knowledge representation and reasoning engine.
- a Tcl/Tk-based GUI displaying hypertext lessons, multiple choice questions, and labs.

References

[CT91] K. Currie and A. Tate. O-Plan: the open planning architecture. *Artificial Intelligence*, 52:49–86, 1991.

[LT00] Staffan Larsson and David Traum. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6(3–4):323–340, 2000.

[Ros00] Carolyn P. Rosé. A framework for robust semantic interpretation. In *Proc. of the 1st Annual Meeting of the North American Chapter of the ACL (NAACL-00), Seatle*, 2000.

[WC98] Michael White and T. Caldwell. EXEMPLARS: A practical, extensible framework for dynamic text generation. In *Proc. of the 9th International Workshop on Natural Language Generation*, 1998.

* This research is supported by Grant #N00014-91-J-1694 from the Office of Naval Research, Cognitive and Neural Sciences Division.

Author Index