Ontologies and the Structure of Dialogue

David Milward
Linguamatics Ltd
david.milward@linguamatics.com

Martin Beveridge
Advanced Computation Laboratory
London Research Institute
Cancer Research UK
martin.beveridge@cancer.org.uk

Abstract

In this paper we will consider relatively simple dialogues, but in domains which involve multiple tasks and services, and concepts of different granularity. We re-examine the notion of focus of attention, and show how ontological information combined with information states can shed new light on the distinctions between linguistic and intentional structure. Elements of this work have been implemented in spoken dialogue systems for home information and control, and for a system that advises on appropriate action for doctors examining patients with suspected breast cancer.

1 Introduction

In current commercial dialogue systems, domain knowledge tends to be incorporated into the dialogue scripts, or used within very tight bounds e.g. via specific database queries at particular points in the dialogue. This kind of approach becomes costly as the domain becomes more complex. Some more recent systems provide a cleaner separation between domain knowledge and generic dialogue interaction rules. The most prominent system of this kind is AT&T's HMIHY system (Abella and Gorin, 1999), where a task inheritance hierarchy is kept separate from generic dialogue "motivators" (such as "missing information" or "clarification"). The task hierarchy encodes information that e.g. a "billing method" is either a "collect call", a "calling card" or a "billing number".

The approach we describe here follows the same motivation as the HMIHY system, but emphasises general ontological knowledge, of which a task hierarchy is a part. By 'ontology' here we mean simply a network of concepts and instances which are related to each other by semantic links. Ontologies, in particular those based on description logics, have been argued to be "the solution of first resort for all problems related to terminology and semantics" because they "occupy the sweet-spot between maximal expressive power on the one hand and computer tractability on the other" (Ceusters and Smith, 2003).

The systems we describe are for two specific and rather different domains. The first is home information and control. In this domain, the amount of domain knowledge required is not huge, but it is highly dynamic. It cannot be fixed in advance by the dialogue designer, or even after the installation of the system in the home. For example, the user can register for new services and add or move devices. In the second domain of cancer, the domain knowledge is more fixed, but is also much larger. Encoding this into a dialogue script by hand was therefore not an option, so a pre-existing medical ontology was used instead.

In sections 2 and 3 we show how is-a relations and part-whole relations can be used to handle clarification questions, and to influence the sequencing of a dialogue. We then relate this work to Grosz and Sidner's (1986) work on intentions and the linguistic structure of dialogue, and to the notion of focus of attention, as used for text generation by McCoy and Cheng (1991).

2 Clarification Dialogues

During a dialogue, the user may provide a response to a prompt that does not match any of the expected range of responses (as defined by the application domain) but is a hypernym or hyponym of an expected response (Milward and Beveridge, 2003). In this case the system should be able to discover the relation between the expected terms and the user response in order to resolve the discrepancy. In the case where the user's reply was a hypernym of an expected term, then their reply can be considered underspecified and the system can issue clarification questions in order to obtain a more detailed reply. Consider the following example from the cancer domain:

(1) S: Do you have a family history of chronic disease?
[*system expects: lung cancer, leukemia, sarcoma,…*]
U: Yes, <u>cancer</u>
S: What type of cancer?
U: <u>Lung cancer</u>.

In this example the system expected a specific disease but the user replied with a more generic term. The system therefore formulated a more specific question in order to elicit an answer at the required level of the disease ontology. Hypernyms are similarly treated in a generic fashion in the AT&T system (Gorin et al., 2002) via the use of a disambiguation operation.

In the case where the user's reply was a hyponym of an expected term then their reply can be considered over-specified and the system can (a) find a more general related term which matches the expected responses in order to answer the current question and (b) avoid asking subsequent more specific questions that have already been answered. For example

(2) S: Have you had any chronic diseases?
[*system expects: cancer, hypertension, diabetes, …*]
U: Yes, <u>Leukemia</u>
S: ??What kind of cancer?

In this example the system expected a more generic answer to the question (such as "cancer"), but the user replied with a more specific term. The system, however, can match the answer to the question on the basis that leukemia is a cancer, which is a chronic disease. The system therefore avoids asking what kind of cancer since this has already been answered and therefore would be pragmatically ill-formed (indicated by "??").

In the home information and control domain the system uses an is-a hierarchy for tasks to allow users to skip directly to a task instance or a more specific task class.

(3) S: Which service do you want?
[*system displays the following options: entertainment, control, settings*]
U: Cinema booking
S: ??What kind of entertainment?

During a dialogue, the user may also provide a response to a prompt that does not match, and is not a hypernym/hyponym of, any expected response, but is associated with one or more of the expected responses by a non-is-a link that imposes an ordering on concepts, e.g. mereological ("part-whole") relations, or topological relations such as "in". In this case the system needs to discover the associative relation between the expected terms and determine whether the supplied term is more or less specific than the expected terms according to the ordering imposed by the relation. Hence, the system can issue clarification questions in order to obtain a more specific reply, in the same way as before. For example:

(4) S: Where does it hurt?
[*system expects: elbow, wrist, shoulder,…*]
U: In my <u>arm</u>.
S: Where in your arm?
U: In my <u>elbow</u>.

In this example the system initially expected a more specific body-part than the user supplied, but recognized that the supplied term "arm" was related to the expected terms in a part-whole hierarchy and was more general then the expected terms

3 Sequential Structure

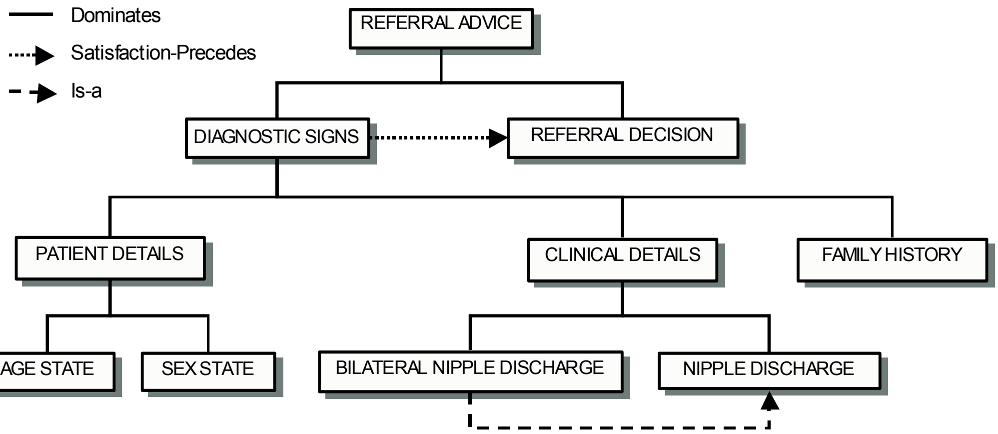

Figure 1. Intentional and informational relations for breast cancer referrals

In system-initiated dialogue, questions are grouped together at design time. Even in systems allowing some mixed initiative, this rarely affects the ordering of subsequent questions. Consider, the following dialogue:

(5) S: What is the patient's sex?
U: Female with some nipple discharge
S: What is the patient's age?
U: Fifty five
S: And is it a bilateral nipple discharge?
U: No

A standard form filling dialogue system might allow more than one answer to be given at once, but the order of the questions does not change. The Philips system (Aust et al., 1995) did deal with the case where the value for the new slot is underspecified, allowing an immediate clarification question, but did not deal with the more general case where two slots are related to each other. We can achieve much more natural dialogues if we cluster questions dynamically according to the user utterance. Consider the following:

(6) S: What is the patient's sex?
U: Female with some nipple discharge
S: And is it a bilateral nipple discharge?
U: No
S: What is the patient's age?
U: Fifty five

This appears to be a much more natural exchange, with the system immediately asking the follow-on question concerning nipple discharge. Questions which elaborate a previous question, either by asking about a particular attribute, or by asking for more specific information (determined by hyponymic, mereological or topological relations) will ideally appear straight afterwards. We achieve this behaviour by specifying what information the task requires, but not providing a strict ordering. The dialogue manager chooses the precise ordering according to the current state of the dialogue and its own domain knowledge.

For example, Figure 1 shows (a fragment of) the information available to the dialogue manager for the breast cancer referrals application (Beveridge and Milward, 2003). Boxes indicate tasks (implemented using the PRO*forma* process specification language (Fox et al., 2003)) with their associated topics. These are related by intentional dominance and satisfaction-precedence relations (Grosz and Sidner, 1986) as well as by ontological relations such as subsumption (Is-a). This is similar to recent approaches to discourse analysis. For example, Moser and Moore stress the need for representations of both intentional and informational relations between discourse segments, where the "informational structure [is] imposed by domain relations among the objects, states and events being discussed" (Moser and Moore, 1993, p. 416). The informational structure would therefore typically include "causal relations of various sorts, set relations, … the relation of identity between domain objects" and so on, creating "a complex network of domain relations that is defined independently of the intentional structure" (Moser and Moore, 1996, p. 417). In our case the informational structure is provided by a domain ontology.

In determining the sequential structure of a dialogue, satisfaction-precedence relations obviously provide the strongest constraint. For example, in Figure 1, the decision task should not be considered until all the enquiries have been completed. In the absence of satisfaction-precedence constraints, the default behaviour is to group siblings together following the dominance relations. Overlaid on this, however, is the effect of ontological relations between topics. For example, the presence of an ontological subsumption relation between bilateral discharge and discharge in Figure 1 causes the dialogue manager to infer a rhetorical *elaboration* relation (Mann and Thompson, 1988) between the related tasks and so overrides the default ordering in two ways. First, the question concerning the satellite task of the elaboration relation (e.g. bilateral discharge) will not be asked until the nucleus task (e.g. discharge) has been completed. This means that topics will "by preference be 'fitted' to prior ones – topics therefore often being withheld until such a 'natural' location for their mention turns up" (Levinson, 1983, p. 313), in this case when the topic of discharge has been introduced, either by the system or by the user. Secondly, once the nucleus task has been completed, the satellite task will have high priority for being considered next. This means that related topics are pursued as soon as possible, whilst they are still relevant, in order to avoid "unlinked topic 'jumps'" (Levinson, 1983, p. 313) later on. Indeed, "the relative frequency of marked topic shifts of this sort is a measure of a 'lousy' conversation" (Levinson, 1983, p. 313) in human-human dialogue.

4 Focus of Attention

In a very simple menu based dialogue system, the dialogue context can be simply the current node in the ontology. However, even if we just allow skipping (as in the case of hyponyms above) it becomes convenient to store the path to the node, not just the node itself, since we may not want to allow the user to back up to nodes that were skipped before, or will expect a differently marked expression e.g. "now go to" vs. "go back to" if they do. If the dialogue does not just consist of traversing a menu tree, but each node may be itself a partial description of a scenario, the intermediate nodes will have their own structure. This gets us to focus stacks (in Grosz and Sidner's terms) i.e. not just a path of atomic nodes, but a path through a series of focus spaces (each of which may contain some objects, properties or relations) which comprises the attentional state, or "focus of attention".

Focus of attention is often seen as closely related to part-whole relationships. For example, the focus changes when the participants are discussing a particular component of an object, or a particular step in a plan. However, authors such as McCoy and Cheng (1991), working in text generation, have also discussed focus of attention moving from a kind of action to a specialisation of the action. This corresponds to moving down an is-a hierarchy, similar to the task or disease clarification examples discussed above. McCoy and Cheng also emphasised that it is not just the objects that must be taken into account, but also the perspective taken on the objects. In the following description, the entity in focus is always the balloon, but the switching of perspective between colour and size makes the text infelicitous:

(7) ??The balloon was red and white striped. Because this balloon was designed to carry men, it had to be large. It had a silver circle at the top to reflect heat. In fact, it was larger than any balloon John had ever seen.

McCoy and Cheng build a focus tree rather than a focus stack. A focus stack could be derived by taking the right frontier of the tree, but keeping the full tree structure is more general, allowing for operations which access nodes that are not on the right frontier. The tree has a fine-grained structure, for example, when describing the person "John" they get the following:

(8) [John [physical [[brown hair] [blue eyes]]]
[interests[[plays football][collects stamps]]]]

In our approach, a dialogue is coherent if each dialogue participant acts independently to preserve coherence to the best of their ability. For example, if they want to ask a clarificatory or elaborative question (according to the participant's ontology) they do so immediately. Coherence is determined by considering the information states of the participants, rather than from the perspective of an independent observer. Thus the decision of the system to ask about bilateral discharge has similar motivation to the close attachment of elaboration utterances in well-structured texts (Mann and Thompson, 1988), but that does not mean that coherent multi-party dialogues will necessarily have elaborations right next to their heads. For example, we predict the following dialogue to be coherent

(8) S: What are the patient details?
U: Female, severe nipple discharge, family history of cancer
S: And, is it a bilateral nipple discharge?
U: Yes
S: Can you give details of the family history?
U: Her mother was diagnosed with ovarian cancer at age forty-five

For the respondent, the grouping of 'female', 'severe nipple discharge' and 'family history' is acceptable since these are all pieces of information relevant to the referral decision and the respondent does not know at this stage that the system requires further information about the discharge. This suggests that the linguistic structure of dialogue arises from the interaction between the intentional/informational structures of the dialogue participants, but need not correspond to either. In contrast, in monologue, since there is a single intentional/informational structure, we would expect the speaker to elaborate immediately at the appropriate points. We would therefore predict the following to be infelicitous:

(9a) ??I have a female patient. She has nipple discharge. She has a family history of cancer. She suffers from bilateral discharge. Her mother was diagnosed with ovarian cancer at age forty-five.

In both monologue and dialogue, the intentional/informational structure can be violated by using marked constructions. For example, in the following monologue (Florian Wolf p.c.) the phrase "as far as family history is concerned" marks the fact that we are revisited an earlier topic.

(9b) I have a female patient with nipple discharge and a family history of cancer. The nipple discharge is bilateral. As far as family history is concerned, her mother was diagnosed with ovarian cancer at age fourty-five.

The discussion so far suggests that for dialogue systems we need a context which is at least as detailed as the focus trees of McCoy and Cheng (1991). For example, the dialogue context corresponding to the dialogue in example (6) might be as follows:

(10) [Diagnostic Signs [Patient Details [Sex=Female][Age=55]] [Clinical Details [Nipple Discharge=Yes] [Bilateral Nipple Discharge=No]]]

Attachment without any marked syntax or cue words can be performed not only at the right frontier, but at any nodes newly introduced by the other dialogue participant, as in example (8) above, where both "family history of cancer" and "severe nipple discharge" require further clarification.

Traversal down is-a hierarchies is represented similarly. For example, the structure corresponding to the user choosing "cinema booking" as an entertainment option is as follows:

(11) [Service [Entertainment [Cinema_Booking [Film=?][Time=?][No.People=?]]]

The context provides an appropriate abstraction of the dialogue history and current user/system goals. This, together with known ontological relationships, allows the system to decide on the next move. It should not be confused with a more detailed dialogue history which would be necessary for e.g. pronominal anaphora. As we have discussed, this might be differently structured. For example, a single value filled in for the "time" slot in the context above may have been established after several discontinuous turns, first specifying e.g. "morning" then "10 o'clock".

5 Implementation

Most of the components in the theoretical approach outlined above have been implemented, but currently not all within a single system. The dialogues 1 - 3 and 6 are all real dialogues with the systems.

Focus of attention is used for generating prompts, for interpreting utterances in context, and for restricting the possible hypotheses of the speech recogniser. For example, in the home information demonstrator, the initial focus of attention at the topmost "service" node results in the question: "which service do you require?". At this point the speech recogniser grammar contains the expected service options, and possible hypernyms or hyponyms. The assumption is that users will say something that will be coherent with respect to the current state of the dialogue, or one of a limited number of marked utterances (e.g. "back to the top"). In order to generate the grammars we use the system's ontology. Strictly according to the approach above, it should be according to the user's ontology not the system. However, in both of the domains considered we can provide the system with a rich ontology for which it is reasonable to assume the user's ontology will be a subset.

In the cancer demonstrator, grammars are generated for the currently open nodes as well as for the right frontier. Furthermore, grammars are generated dynamically at the start of each dialogue segment in order to ensure the language model is consistent with the current high-level context. In the home domain, a narrowing of the focus of attention is achieved by going down not just the part-whole and is-a hierarchies, but also through the 'in' hierarchy (e.g. from kitchen to cooker). This position in the hierarchy provides a situation in which we can evaluate definite references (c.f. Milward 1995). For example, if the current position in the hierarchy is a particular room in the house, "the light" in "turn on the light" will be taken to be the light in this room.

6 Relation to Other Work

As described earlier our approach here has similar motivations to those behind the design of AT&T's HMIHY system (Abella and Gorin, 1999), except that HMIHY only makes use of a task hierarchy whereas we extend this to include domain ontological knowledge also.

Lascarides and Asher (1999) similarly make use of both intentional and discourse relations in order to interpret (as opposed to generate) dialogue. They employ a Question Elaboration relation *Q-Elab*(α, β) which "holds if β is a question whose answers all specify part of a plan to bring about an SARG [Speech Act Related Goal] of α". This is demonstrated in (12) below (Schlangen and Lascarides, 2002) in which *Q-Elab*(U_1, U_2) because all possible answers to U_2 specify part of a plan to bring about the SARG of U_1 (to arrange to meet next week).

(12) [U_1] A: Let's meet next week
[U_2] B: (OK.) Thursday at three pm?

This relation is therefore intentionally-based in that its definition refers to partial satisfaction of goals (rather like the dominance relation of Grosz and Sidner (1986)). However, this relation doesn't seem to be applicable to the examples we have described so far, e.g. (13) below.

(13) [U_1] S: What is the patient's sex?
[U_2]U: Female and she has some nipple discharge
[U_3] S: And is it a bilateral nipple discharge?
[U_4] U: no

Here, we don't seem to be able to claim that U_3 is a coherent continuation of U_2 because *Q-Elab*(U_2, U_3). The SARG of U_2 is presumably that S believe that the patient is female and has some nipple discharge, and the answers to U_3 don't appear to specify part of a plan to achieve that goal. In fact U_3 implicitly indicates that the goal of U_2 has already been achieved.

The elaboration relation that we have used is instead an informational relation, similar to the subject-matter elaboration relation of RST (Mann and Thompson, 1988). This means that we consider that U_2 elaborates U_1 if U_2 presents additional detail about the situation or some element of subject matter (e.g. a particular entity) introduced in U_1. Hence U_3 ELABORATE U_2 in (13) above by virtue of that fact that they refer to the common entity 'nipple discharge'.

Ginzburg (in press) also uses discourse relations such as elaboration to order the contents of QUD (Questions Under Discussion (Ginzburg, 1995)). This is used to account for the order in which questions are typically answered. However, only successive queries within a single turn are considered, such as (14) below (adapted from Ginzburg (in press)).

(14) [Q_1] A: Who have you invited?
[Q_2] Have you invited Jill?
B: yes
A: Aha
B: I'm also inviting…

Here the elaboration relation between Q_1 and Q_2 leads to the expectation that Q_2 will be answered before Q_1 and hence Q_2 should be maximal in QUD after A's initial utterance. Such an approach does not really apply to the examples we have discussed here, however, as we are not dealing with successive queries in a single turn.

In (13) above, for example, the system's QUD would be updated with a question Q_1 regarding the patient's sex at U_1, then when the answer A_1 is received in U_2 a question $?A_1$ ("whether A_1") would be added to QUD but not pursued (since A_1 is accepted) and so both it and Q_1 would be downdated from QUD. Similarly, at U_3 QUD will be updated with a question Q_2 regarding whether or not the nipple discharge is bilateral, and after the answer A_2 in U_4 the question $?A_2$ will added, but both $?A_2$ and Q_2 will then be downdated from QUD since A_2 is accepted. Hence, there is never more than one question (plus a ?A question regarding the answer, which is never pursued) in QUD at any given time, and so the ordering of QUD does not seem to account for the coherence of U_3 as a continuation of U_2.

In fact, for both of the systems discussed here (in which answers are assumed to always be accepted) QUD would need to contain at most a main question, and a single other help-type question should the user ask for help (in which case the help question would be maximal in QUD until the system provided an answer and it was downdated).

In order to account for the coherence of (13) we need instead to impose an ordering over pending questions, but these do not form part of the Dialogue Game Board (DGB) in Ginzburg (1995; in press). Instead, they are presumably part of the dialogue participant's (DP's) unpublicised mental situation UNPUB-MS(DP).

The GoDiS system (Larsson et al., 2001), a dialogue system based on Ginzburg's QUD, does represent pending questions via a PLAN field in their information state which specifies "a list of dialogue actions that the agent wishes to carry out" (Larsson et al., 2001, p.1). However, our dialogue context corresponding to the PLAN field is more structured than this: it is a tree (rather than a list) and it is structured, not only according to a task hierarchy, but also following ontological relations. In GoDiS, for example, the only alteration of the sequence of actions in PLAN is to accommodate the user answering more than one question at a time. In our approach, however, we re-order the pending actions in the dialogue context according to ontological relations between them and the user's last utterance. This ensures that the system's next planned action is maximally relevant to the ongoing dialogue topic.

7 Conclusions

In this paper we have shown how ontological information can be used for clarification dialogues and to order questions to maximise coherence. This has extended the use of rhetorical relations from their traditional role in text analysis and generation to multi-party spoken dialogues, and has started to explore the distinctions between monologue and dialogue. We have implemented parts of this work in spoken dialogue systems for home information and control, and a system that advises doctors on whether to refer patients with suspected breast cancer to a specialist

Acknowledgements

The work on medical systems described in this paper has been funded by the European Union under the 5th Framework Project Homey (http://turing.eng.it/pls/homey/). The work on home control has had support from the EU 6th Framework Project, TALK (http://www.talk-project.org) and the UK Department of Trade and Industry Next Wave Programme. We would like to thank our partners in Homey and TALK for many useful discussions.

References

Alicia Abella and Allen Gorin. 1999. Construct Algebra: Analytical Dialog Management. *Proc. of ACL*, June, Washington D.C., USA.

Harald Aust, Martin Oerder, Frank Seide, and Volker Steinbiss. 1995. The Philips Automatic Train Timetable Information System. *Speech Communication*, 17:249-262.

Martin Beveridge and David Milward. 2003. Combining Task Descriptions and Ontological Knowledge for Adaptive Dialogue. In V. Matoušek and P. Mautner (Eds.) *Proc. of Text, Speech and Dialogue (TSD'03)*, September, České Budějovice, Czech Republic, pp. 341 – 348.

Werner Ceusters and Barry Smith. 2003. *A Realist Ontology for Natural Language Understanding*. Abstract, Colloquium on Knowledge Management, Erasmus University, Nijmegen, The Netherlands (http://www.landcglobal.com)

John Fox, Martin Beveridge, and David Glasspool. 2003. Understanding Intelligent Agents: Analysis and Synthesis, *AI Communications*, 16:139–152.

Jonathon Ginzburg. 1995. Interrogatives: Questions, Facts and Dialogue. In S. Lappin (Ed.) *Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.

Jonathon Ginzburg. In press. Querying and Assertion in Dialogue. *Questions and the Semantics of Dialogue*, Chapter 4. Forthcoming from CSLI Publications and University of Chicago Press.

Kathleen McCoy and Jeanette Cheng. 1991. Focus of attention: Constraining what can be said next. In C.L. Paris, W.R. Swartout, and W.C. Mann (Eds.) *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer Academic Publishers, pp.103-124.

Barabara Grosz and Candace Sidner. 1986. Attention, Intention and the Structure of Discourse. *Computational Linguistics*, 12(3):175-204.

Staffan Larsson, Robin Cooper, and Stina Ericsson. 2001. menu2dialog, In *Proc. of the IJCAI'01 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, August, Seattle, USA.

Alex Lascarides and Nicholas Asher. 1999. Cognitive States, Discourse Structure and the Content of Dialogue. *Proc. of the Amstelogue Workshop on the Semantics and Pragmatics of Dialogue*, May, Amsterdam, The Netherlands.

Stephen Levinson. 1983. *Pragmatics*. Cambridge University Press, Cambridge, UK.

William Mann and S. Thompson. 1988. Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. *Text*, 8(3):243-281.

David Milward. 1995. Integrating Situations into a Theory of Discourse Anaphora. In *Proc. of the 10th Amsterdam Colloquium* pp. 519-538.

David Milward and Martin Beveridge. 2003. Ontology-Based Dialogue Systems. In *Proc. of the IJCAI'03 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, August, Acapulco, Mexico.

Megan Moser and Johanna Moore. 1993. Investigating Discourse Relations. *Proc. of the ACL Workshop on Intentionality and Structure in Discourse Relations*, Columbus, USA.

Megan Moser and Johanna Moore. 1996. Toward a Synthesis of Two Accounts of Discourse Structure. *Computational Linguistics*, 22(3):409-419.

David Schlangen and Alex Lascarides. 2002. Resolving Fragments Using Discourse Information. *Proc. of the Edilog Workshop on the Semantics and Pragmatics of Dialogue*, Sept., Edinburgh, UK.