An Extension of Rhetorical Structure Theory for the Treatment of Retrieval Dialogues

Elisabeth Maier
Department KOMET
GMD-IPSI
Dolivostr. 15
D-W-6100 Darmstadt
Germany
maier@ipsi.darmstadt.gmd.de

Stefan Sitter
Department COGITO
GMD-IPSI
Dolivostr. 15
D-W-6100 Darmstadt
Germany
sitter@ipsi.darmstadt.gmd.de

Abstract

A unification of a speech-act oriented model for information-seeking dialogues (COR) with a model to describe the structure of monological text units (RST) is presented. This paper focuses on the necessary extensions of RST in order to be applicable for information-seeking dialogues: New relations are to be defined and basic assumptions of RST have to be relaxed. Our approach is verified by interfacing the dialogue component of an intelligent multimedia retrieval system with a component for natural language generation.

1 Introduction and Problem

Approaches to discourse organization especially in the area of computational linguistics are oriented towards the treatment of either monologues or dialogues. Only recently efforts have been reported to develop models which cover both types of discourse (e.g., Sitter & Stein, 1992; Fawcett & Davies, 1992). Our paper is a contribution to this research topic providing a unified model to describe coherence in dialogues in a computational framework.

Background for our work is the development of an intelligent information-retrieval system MERIT (Stein et al., 1992), which makes use of natural language as one of the modalities for system-user interaction. The system integrates the text generation system developed in the KOMET project (Bateman et al., 1991) with an implementation of a speech-act oriented dialogue model called COR (*co*nversational *r*oles).

In this paper we focus on the *theoretical* part of the system integration and discuss the consequences resulting from the application of the monologically-oriented Rhetorical Structure Theory (RST. see Mann & Thompson, 1987) on dialogues. We demonstrate our concepts using a few excerpts from one example dialogue.

2 State of the Art

Since we want to tie the two research strands together - work on dialogue structure and approaches focusing on the treatment of texts - we discuss trends in both fields and also provide a basis for the description of our approach.

2.1 Models of dialogue structure

Many classical systems which are interfaced with natural language systems, like explanation components, database access modules, tutoring systems, etc. (for reviews, see Perrault & Grosz, 1986 and McCoy et al., 1991), lack a model of the information-seeking dialogue and have the simple underlying conception of a dialogue as iteration of adjacent query-answer pairs. More recent work gives an explicit account of the dialogue – i.e., its thematic structure, its relation to an external task, types of failure, etc. (Reichman, 1985; Grosz & Sidner, 1986; Carberry, 1985). Carberry deals with information-seeking problems in a dialogue between the user as the information-seeker and the system as the information-provider. She presupposes that information-seeking takes place in the context of a defined task, and that the user has a plan for his task which (1) can in principle be formulated and (2) has well-defined gaps and misconceptions. The job of the system is to recognize the plan and to assist in plan execution, i.e. provide necessary missing information, inform about hidden obstacles. In case of misconceptions, it also assesses the relevance of the user's questions and uncovers false presuppositions.

Unfortunately, in many realistic situations it seems too restricting an assumption that the information-seeker should be able to verbalize his plan or even have a plan (McAlpine & Ingwersen, 1989; Belkin & Vickery, 1985). If meaningful structures can be construed in information-seeking dialogues in highly vague task settings, this must be done without strong reference to a domain structure.

Winograd and Flores (Winograd & Flores, 1986) give a different account of dialogue structure. The authors argue on philosophical grounds that only on the level of interactional conventions, which specify how to express and negotiate behavioral expectations and commitments, can interactions be formally described. As an example they use two part-

ners', A and B, negotiation of a task which B has to fulfill (*Conversation for Action*, in the following called CfA). The process of negotiation is represented as the traversal of a state-transition network summarizing all possible chainings of A's and B's dialogue actions like 'Request', 'Promise', 'Reject', 'Withdraw', etc., ignoring the contents of the actions. The authors do not treat the computer as a dialogue partner, but as a medium for the structuring of inter-human interaction.

Moore's system (Moore, 1989) for the generation of explanations in the framework of expert systems is able to deal with follow-up questions of users who do not understand parts or all of the system's explanations. Her model differs from the approach proposed by Winograd and Flores insofar as the system plans explanations depending on the user's questions and on the communicative goals which were responsible for the generation of previous system contributions. The system also makes assumptions about the lack of user knowledge and generates clarifications. Unfortunately, the system is restricted to the treatment of a small set of speech acts and to the modeling of a subset of possible interactions in explanatory dialogues. This model therefore can profit from an integration with a more extensive model of human-machine interaction like, e.g., COR - an enterprise we describe in section 3.

2.2 Models of monological discourse

Various approaches for the description of discourse structures emphasize the conventionalized order of discourse segments. Among them are approaches like macrostructures (vanDijk & Kintsch, 1983), grammar-like descriptions for specific genres (Rumelhart, 1975) or schemata (McKeown, 1985; Paris, 1987).

While approaches of this family – of which we consider also the dialogue model CfA a member – describe the sequence of elements in an interaction, they do not give an account of *how* they are related. The recipient of the information has to recognize why the information is presented in the given sequence. A model like RST is able to model this feature of texts, i.e. it makes use of constructs – the relations – which model the semantics of the links between text segments.

RST provides means to represent the structure of *monological texts* hierarchically. In RST, Mann and Thompson defined an open set of relations which are used to describe the semantics of the links between units of texts. Such text units are segments of discourse, the minimal length of which is one proposition. Each pair of text units connected by a relation is again considered to be one unit. A basic assumption of RST is that relations impose an asymmetrical structure on two connected texts units: One unit is more important than the other – it cannot be removed without changing the core meaning of the text. This text unit is called the *nucleus* ($\mathcal{N}$). The segment of text which is of less importance is the *satellite* ($\mathcal{S}$).

The set of relations proposed in RST has been augmented recently by a large number of relations taken from other approaches (e.g. Hobbs, 1990; Sanders, Spooren & Noordman, 1990; Iri, McMillan & Merz, 1990). A classification of these relations (Maier & Hovy, 1991) distinguishes three types: *ideational*, *interpersonal* and *textual* relations. Ideational relations capture links which are concerned with experiential knowledge, while textual relations are used to signal text-internal links, i.e. links which refer to segments of text instead of segments of text-external knowledge. Interpersonal relations are of special importance here since they take the specific features of the discourse participants into account. Among the interpersonal relations are such which affect the ability of the reader (e.g. ENABLEMENT), his willingness to do something (e.g. MOTIVATION), his beliefs (e.g. EVIDENCE), etc. A subtaxonomy for interpersonal relations is given in figure 1 (from Maier & Hovy, 1991). We assume that dialogic discourse implies the need for more relations of this type which leads to the extension of this part of the network. In this paper we show that the application of RST on dialogues leads to an extension of the set of interpersonal relations (section 3.2) and the change of at least one basic assumption of RST (section 3.3).

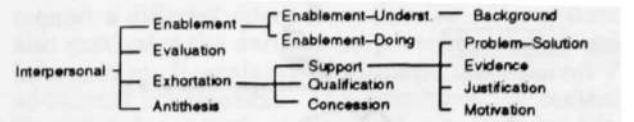

Figure 1: Taxonomy of interpersonal rhetorical relations (taken from Maier & Hovy, 1991).

3 Approach

3.1 The COR Dialogue Model

The COR model describes dialogic information-seeking processes, involving two dialogue partners A and B with the roles of information-seeker and information-provider, respectively (Sitter & Stein, 1992). There are two "ideal" courses of action:

- A formulates his request, B promises to answer and answers the request, A expresses contentment.
- B offers to provide some information (assuming that he has sufficient knowledge about A's information need), A accepts the offer, B provides the information, and A expresses contentment.

These two courses are *expected* in the sense that they match the role expectations which are adequate for the information-seeking situation. In fact, many everyday information-seeking exchanges already follow this expected course, e.g. requesting the time. The state-transition network in figure 2 presents the "expected" courses of action (bold arrows). In more problematic information-seeking situations beyond such short exchanges – e.g., those involving the use of information systems – there are reasons for deviations from the expected course. Information re-

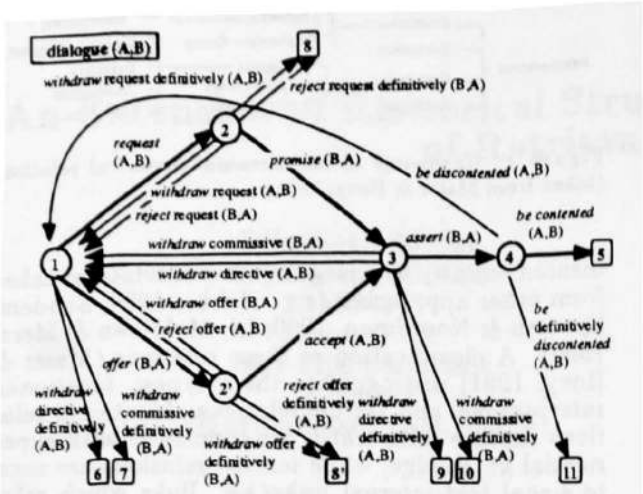

Figure 2: A network representing role expectations in information-seeking dialogues (from Sitter & Stein, 1992). The initial state is $<1>$. The transitions in the network are made by performing the dialogue contributions. The parameters A and B correspond to the speaker and the addressee of the contribution respectively.

quested may not be available; information provided may not satisfy the information-seeker. These examples lead to the introduction of further paths in the network. These additional paths either lead back to the initial state $<1>$ in figure 2, if the dialogue is to be continued, or to final states besides state $<5>$. We call these acts *alternative*. Alternative acts reject or withdraw role expectations prevalent in the dialogue situation. E.g., after A's asking a question, B's promise to answer the question is expected, but also B's rejection or A's withdrawal of the question may happen under certain conditions.

The network so far resembles the state-transition network used by Winograd and Flores as an example of their CfA (Winograd & Flores, 1986). However, we interpret the transitions in figure 2 not as atomic *acts* (a request, an answer, etc.), but as possibly extended dialogue *sections* that are subordinated to illocutionary functions (to arrive at a *mutually understood* request, answer, etc.).[1] We will refer to them as dialogue *contributions*.

A contribution can be subdivided into two components. The first has the purpose to express the illocutionary function of the whole section – i.e., instantiate a role expectation or a response to a role expectation (acceptance or rejection). The second delivers some kind of *contextual information* for it, which may become necessary if the first threatens to fail. *Context* here refers to the presence or absence of the conditions which must be fulfilled to render a dialogue contribution successful.[2] E.g, for a question to succeed, there must be some agreement that its answering helps the information-seeker's goals, that the dialogue partner knows the meaning of the words used, etc. These conditions justify behavioral expectations; severe disagreement on them motivates deviation from the expected course of action.

Example 1 shows a 'request' dialogue contribution, consisting of the request proper and contextual information (modified from figure 3 (U = user, S = system)):

Example 1:
U: Tell me about EC-funded projects with enddate after 1992 dealing with 'illocutionary models'. (Request)
'Illocutionary models' is, very roughly, like 'dialogue modeling'. (Contextual information)

Instead of U's supplying contextual information, S might have initiated an embedded dialogue:

Example 2:
U: Tell me about EC-funded projects with enddate after 1992 dealing with 'illocutionary models'. (Request)
S: What do you mean by 'illocutionary models'? (Request for contextual information)
U: It is, very roughly, like 'dialogue modeling'. (Contextual information)

Both variants – U's voluntarily supplying contextual information to his own dialogue contribution (example 1) and S's dialogically exploring the context of U's contribution (example 2) – have the function of increasing the probability of success of the contribution. To cover this similarity, both variants are considered 'request' contributions.

The part of a dialogue contribution addressing the role expectations, labeled as illocutionary types ('Request', 'Reject Request', etc. in figure 2), is called the *nucleus*. The other part addressing the context is called the *satellite*. These terms have been adopted from the analysis of monological texts by means of RST (section 2.2). Often, as in the examples presented throughout this paper, the nucleus is an atomic act not further decomposed by COR. The satellite is either an assertion or an embedded dialogue, or it may be omitted. A dialogue exploring the context of the contribution is itself an information-seeking dialogue and follows the same rules as the top-level retrieval dialogue (see example 2).

Figure 3 shows a dialogue section, mainly elaborating on an initial user request as in example 2. On the left of figure 3, the decomposition of the dialogue is shown. (For more details, see Sitter & Stein, 1992.)

[1] The network, together with further networks for the contributions (Sitter & Stein, 1992), is interpreted as an ATN (augmented transition network).

[2] These conditions include Austin's (Austin, 1962) and Searle's (Searle, 1969) "felicity conditions", which subsume mainly *conditions of the speaker*. However, like Wunderlich (e.g., Wunderlich, 1976), we go beyond these and also include conditions the *addressee* must fulfill and which can be decided about only in the course of the succeeding interaction. Therefore, Wunderlich introduces the term *successfulness* ("Erfolgreichsein") in addition to felicity.

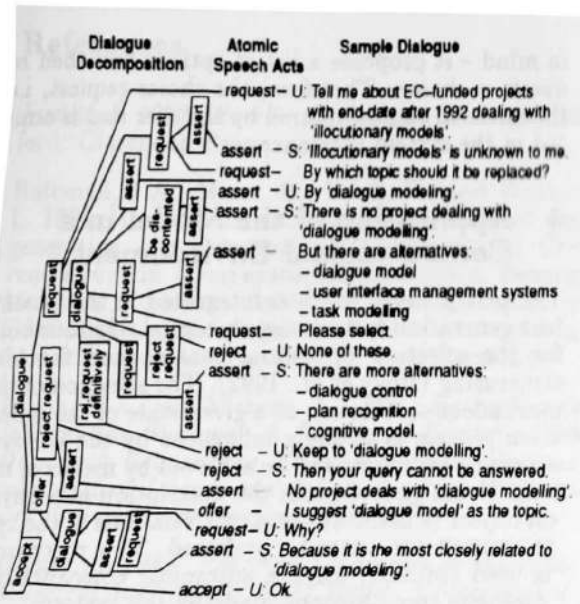

Figure 3: An example MERIT *dialogue section negotiating a user request with its decomposition according to* COR. MERIT *provides information about* EC-funded projects.

3.2 Extension of the Set of Interpersonal Relations

Based on Winograd and Flores' approach (Winograd & Flores, 1986), COR is a model of the level of role expectations in an information-seeking dialogue. We abstract from the content of the dialogue (the thematic level) and leave a description of this to a thematic model, upon which we pose only very low demands (e.g., it need not be a plan-based description of the domain). Therefore, we explicitly permit failures and problematic situations in the dialogue; the dialogue model can organize the means to deal with them interactively.

Still underdeveloped is an account of the various types of failures and problematic situations. Possible reasons for deviations should be classified and adequate means for their expression sought.

As mentioned above we assumed that the application of RST for the description of dialogues may lead to the extension of the set of necessary relations, especially of *interpersonal* relations which typically address discourse participants. The analysis of our sample dialogue brought about such an extension. In the following we give two examples from figure 3 for such "new" relations.

Example 3:
S: 'illocutionary models' is unknown to me.
(Contextual information)
By which topic should it be replaced? (Request)

The first proposition gives an explanation for the – otherwise not understandable – reaction of the system. I.e. the system starts a subdialogue in order to repair a potential failure of the whole dialogue. The information requested by the user ('illocutionary models') is not available, which is the reason why the system comes up with the alternative strategy to request a different query from the user. The system also motivates the decision by providing contextual information: It explains why it asks for different / additional information. The nucleus – the central illocution for the whole fragment – is contained in the second proposition, i.e. it is exactly this system request. The semantics of the relation resembles the RST definition for

BACKGROUND:

constraints on $\mathcal{N}$:
$\mathcal{R}$ (the reader) won't comprehend $\mathcal{N}$ sufficiently before reading text of $\mathcal{S}$
constraints on $\mathcal{S}$: none
constraints on the $\mathcal{N}$ + $\mathcal{S}$ combination:
$\mathcal{S}$ increases the ability of $\mathcal{R}$ to comprehend an element in $\mathcal{N}$
the effect:
$\mathcal{R}$'s ability to comprehend $\mathcal{N}$ increases

The new relation has a more restrictive and dialogue-oriented meaning which can be defined as follows:

BACKGROUND-FOR-REQUEST:

constraints on $\mathcal{N}$:
$\mathcal{R}$ won't comprehend the relevance of the 'Request' specified in $\mathcal{N}$ without being given a reason for it
constraints on $\mathcal{S}$:
$\mathcal{S}$ is an atomic 'Assert' speech act
constraints on the $\mathcal{N}$ + $\mathcal{S}$ combination:
$\mathcal{S}$ increases the ability of $\mathcal{R}$ to comprehend why $\mathcal{N}$, i.e. the request, was uttered by specifying the reason for it
the effect:
$\mathcal{R}$'s ability to comprehend the relevance of $\mathcal{N}$ and his willingness to follow the request in $\mathcal{N}$ increase

The definition of BACKGROUND-FOR-REQUEST is more restrictive than BACKGROUND insofar as the satellite explains why a *request* is made in the nucleus. Compared to the definition of BACKGROUND, constraints on the satellite are available refining its semantics. BACKGROUND-FOR-REQUEST therefore can be considered a subtype of BACKGROUND, which is an *interpersonal* relation. By hypothesizing such a new relation we already extend the set of relations concerned with features of the discourse participants. — Another new relation can be found in the following segment of our sample dialogue:

Example 4:
S: But there are alternatives:
- dialogue model
  user interface management systems
- task modeling.
(Background to request for choice)
Please select. (Request for choice)

The satellite provides a list of options the user has to choose from. The nucleus contains the request proper to select an option. The user is not able to carry out the action specified by the request if the satellite is not given. The new definition can be specified as:

BACKGROUND-FOR-CHOICE-REQUEST:

constraints on $\mathcal{N}$:
$\mathcal{N}$ is a choice-request. $\mathcal{R}$ won't comprehend the 'Request' without being given a parameter list which determines which actions are possible as follow-up reactions for the request
constraints on $\mathcal{S}$:
$\mathcal{S}$ is an 'Assert' speech act which consists of a list of action parameters
constraints on the $\mathcal{N}$ + $\mathcal{S}$ combination:
$\mathcal{S}$ increases the ability of $\mathcal{R}$ to comprehend $\mathcal{N}$ and to carry out the requested action
the effect:
$\mathcal{R}$'s ability to follow the choice-request in $\mathcal{N}$ increases

We consider this relation another subtype of BACKGROUND refining the definition given above.

The two new relations are interpersonal according to the taxonomy given in figure 1. They are subtypes of BACKGROUND and are therefore directly subordinated. — The set of relations necessary for the description of information-seeking dialogues is being determined by the examination of a large corpus of dialogues.

3.3 Re-examination of the Basic Assumptions of RST

As pointed out in section 2, the application of RST on describing discourse structures in dialogues is likely to lead to extensions or changes in general assumptions of the theory. In this section we give an example for the relaxation of restrictions made by RST. To show this we make use of the contrast between example 4 above and example 5 taken from our sample dialogue:

Example 5:
S: There are more alternatives:
- dialogue control
- plan recognition
- cognitive model.

(Background to request for choice)

Both examples have in common that they serve one request speech act made by the system, demanding input from the user. Example 4 suggests various alternatives from which the user is supposed to choose one. The system then explicitly utters a demand to select from the set of options. In contrast to this the explicit demand is left out in example 5. This means that the nucleus containing the atomic request is missing. This is contradictory to the assumption made by RST that the nuclei of complex textual structures must not be omitted without losing the meaning of the whole textual entity. The possibility to omit nuclei in spans of discourse has not been found in the monological texts examined in work on RST. The motivation for such a phenomenon, therefore, has to be found in the nature of dialogues: In our sample dialogue session example 5 occurs shortly after example 4 so that the user still has the pattern of interaction with the system in mind – it proposes a list of options and then requests a choice. Therefore, the choice-request, i.e. the nucleus, can be inferred by the user and is omitted in the system utterance.

4 Application of the Model in a Computational Environment

The text planner which is integrated in the KOMET text generation system uses rhetorical relations both for the selection of textual content and for text structuring (Hovy et al., 1992). The selection of the most adequate relation at a given state of the generation process is strongly influenced by the *communicative goal* which is to be achieved by means of the text. If the goal requires the description of a physical object (DESCRIBE-OBJECT), relations of the type ELABORATION are to be preferred. This mechanism is used similarly for the automatic construction of dialogue contributions made by the system.

In contrast to the KOMET text generation system where goals are mostly triggered by text-type specific features, the communicative goals for the generation of dialogue contributions are posted by the interaction manager handling the COR model and by user reactions. The illocutions available at each point of the dialogue have the same functions as goals and therefore influence the rhetorical relations employed. To further constrain which relation is going to be used the text planner has to check whether the knowledge to express the relation is available in the pool of knowledge supplied by the retrieval component; e.g., to express a WHOLE-PART relation the candidate concept to which new information has to be related must be a decomposable object with at least one part specified. The choice of relations is additionally influenced by the context (the dialogue history) in order to prevent the presentation of redundant information. After a relation has been determined the content is selected and the dialogue history is incremented by the newly planned discourse segment.

The modules required for the generation of system contributions in the given framework therefore are: (1) a model for interaction (COR); (2) a representation of communicative goals; (3) a representation of rhetorical relations; (4) an incrementally growing dialogue history; (5) knowledge bases and a knowledge pool capturing the output of the retrieval component. (1), (4) and (5) are specific for the production of dialogues and this is where adaptions of the original text planner have been made.

Acknowledgements

We would like to thank Adelheit Stein for inspiring discussions and for valuable comments on a previous draft of this paper.

References

Austin, J. 1962. *How to do things with words.* Oxford: Clarendon Press.

Bateman, J.A.; Maier, E.; Teich, E.; and Wanner, L. 1991. Towards an architecture for situated text generation. In *International Conference on Current Issues in Computational Linguistics*, Penang, Malaysia. Also available as technical report of GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, West Germany.

Belkin, N.J., and Vickery, A. 1985. *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-Based Systems*, volume 35 of *Library and Information Research Report.* University Press, Cambridge/GB.

Carberry, S. 1985. *Pragmatic modeling in information system interfaces.* PhD thesis, University of Delaware.

Fawcett, R.P., and Davies, B. 1992. Monologue as turn in interactive discourse: Towards an integration of exchange structure and rhetorical structure theory. In *Proceedings of the 6th International Workshop on Natural Language Generation.* Springer.

Grosz, B.J., and Sidner, C.L. 1986. Attention, intentions and the structure of discourse. *Journal of Computational Linguistics*, 12(3).

Hobbs, J.R. 1990. *Literature and Cognition.* CSLI Lecture Notes.

Hovy, E.; Lavid, J.; Maier, E.; Mittal, V.; and Paris, C. 1992. Employing knowledge resources in a new text planner architecture. In *Proceedings of the 6th International Workshop on Natural Language Generation*, April 1992.

Iri, V.; McMillan, D.; and Merz, T. 1990 S-relators. Technical report, University of Zagreb.

Maier, E., and Hovy, E. 1991. A metafunctionally motivated taxonomy for discourse structure relations. In *Proc. of the 3rd European Workshop on Natural Language Generation*, Judenstein, Austria, March 1991.

Mann, W.C, and Thompson, S.A. 1987. Rhetorical structure theory: A theory of text organization. In Livia Polanyi, ed., *The Structure of Discourse.* Ablex Publishing Corporation.

McAlpine, G., and Ingwersen, P. 1989. Integrated information retrieval in a knowledge worker support system. In *Proceedings of SIGIR*, pages 48 – 57, Cambridge/MA, June 1989.

McCoy, K.F.; Moore, J.D.; Suthers, D.; and Swartout, B. eds. 1991. *AAAI-91 Workshop on Comparative Analysis of Explanation Planning Architectures*, Anaheim, July 1991.

McKeown, K.R. 1985. *Text generation - Using discourse strategies and focus constraints to generate natural language text.* Cambridge University Press.

Moore, J.D. 1989. *A Reactive Approach to Explanation in Expert and Advice-Giving Systems.* PhD thesis, University of California, Los Angeles.

Paris, C.L. 1987. Combining discourse strategies to generate descriptions to users along a naive/expert spectrum. In *Proceedings of IJCAI*, pages 626–632.

Perrault, C.R., and Grosz, B.J. 1986. Natural-language interfaces. *Annual Review of Computer Science*, 1.

Reichman, R. 1985 *Getting computers to talk like you and me. Discourse context, focus, and semantics (an ATN model).* Cambridge, MA: Bradford.

Rumelhart, D.E. 1975. Notes on a schema for stories. In Bobrow and Collins, editors, *Language, Thought and Cognition.* Academic Press, Orlando.

Sanders, T.J.M.; Spooren, W.P.M.S.; and Noordman, L.G.M. 1990. Towards a taxonomy of coherence relations. *Discourse Processes.*

Searle, J.R. 1969. *Speech Acts.* Cambridge/GB: University Press.

Sitter, S., and Stein, A. 1992. Modelling the illocutionary aspects of information-seeking dialogues. *Information Processing and Management*, 28(2):165 – 180.

Stein, A.; Thiel, U.; and Tissen, A. 1992. Knowledge based control of visual dialogues. In *Proceedings of AVI '92 (Advanced Visual Interfaces)*, Rome, Italy, May 1992. Singapore: World Scientific Press.

van Dijk, T.A., and Kintsch, W. 1983 *Strategies of Discourse Comprehension.* Academic Press, Inc.

Winograd, T., and Flores, F. 1986. *Understanding computers and cognition.* Ablex Publishing Company.

Wunderlich, D. 1976. *Studien zur Sprechakttheorie.* Suhrkamp.