

Generating Cooperative System Responses in Information Retrieval Dialogues

Markus Fischer

Information Technology Research
Institute, University of Brighton (ITRI)
markus.fischer@itri.bton.ac.uk

Elisabeth Maier

German Research Center for Artificial
Intelligence GmbH (DFKI)
maier@dfki.uni-sb.de

Adelheit Stein

Integrated Publication and Information
Systems Institute (GMD-IPSI)
stein@darmstadt.gmd.de

Abstract

This paper describes the Corinna system which integrates a theoretical approach to dialogue modeling with text generation techniques to conduct cooperative dialogues in natural language. It is shown how the dialogue model COR can be augmented by adding discourse relations as an additional level of description which is particularly valuable for the generation of dialogue acts.

1 Introduction

Text planning and generation become more and more important as components of intelligent multimodal user interfaces. It has been proposed (see, e.g., Stein and Thiel, 1993) to model interaction with an intelligent information system as a conversation between two participants. The conversation metaphor applies to natural language as well as to graphical and multimodal user interfaces, if an appropriate model is provided. Employing features of human-human communication into interface design (even if only to a small extent), the system can respond to the user in a more natural way. Natural communication does not necessarily imply natural language, but undoubtedly, language plays an important role in conversation. Theories and tools developed in the areas of Natural Language Processing (NLP) and Natural Language Generation (NLG) can help making models for conversations (or dialogues, as we will call it for the rest of the paper) more adequate and richer in their descriptive power.

The speech-act oriented dialogue model COR (*CO*nversational *R*oles) has been developed at GMD-IPSI (Sitter and Stein, 1992). It covers the genre of information seeking human-computer interactions as opposed to human-human conversations and other (less restricted) genres such as spontaneous conversations, narratives, arguments, etc. Due to its strong focus on the pragmatics of a dialogue, leaving semantics aside, COR is a domain independent conversation model. It is general in that it covers the basic illocutionary aspects and role expectations of cooperative information-seeking exchanges.

In the search for ways to enhance the COR model further, we found the Rhetorical Structure Theory (RST) to be a very good candidate. This theory – initially developed for monologues only – has been formalized and integrated into many text generation and text planning systems, one of which has been developed at IPSI (cf. Bateman *et al.*, 1991). Some empirical studies were carried out to find out possible points of integration (Fischer,

1993). The results were very promising and hence used to specify a dialogue manager of a prototypical information system, called Corinna. This dialogue system uses automatically generated natural language as its major interaction modality. The incorporation of RST within the COR dialogue model served as an important parameter within the text generation process.

The remainder of the paper is divided into two main parts: the theoretical part concludes with the description of our own approach of integrating COR and RST (section 2.2), which is then elaborated and exemplified in the second part of the paper (section 3) presenting the basic features of the Corinna system.

2 Theoretical Framework

2.1 Related Work

Research activities in several areas, such as NLG, dialogue modeling, information retrieval and multimedia interfaces played an important role in motivating our work. Two streams of research were particularly interesting in our context: on one hand the incorporation of dialogue models in natural language systems, on the other hand the extension of RST and its application for (not exclusively natural-language-based) dialogue systems. The following gives some examples of work done in these two areas.

As part of the Communal project (cf. Fawcett *et al.*, 1988), which includes generation as well as understanding of natural language, a dialogue model called SFM (Systemic Flowchart Model) was developed. It uses a discrimination network to describe situations and actions that can occur in a dialogue. Due to the fact that many different speech acts (based on Searle, 1969) and speech act sequences were to be considered, the network is quite complex. Attempts were also made to integrate this model with RST (cf. Fawcett and Davies (1992) and section 2.1.2).

A system which is capable of performing dialogues with a user on the basis of speech, was proposed by Smith, Hipp and Biermann (1992). Its domain is the maintenance of electrical appliances, and the emphasis in this approach lies on (nested) communicative goals, and concepts such as intentional, attentional and linguistic structures (Grosz and Sidner, 1986).

Another system for the treatment of spoken dialogues is reported in Bilange (1991). The approach, which has been developed in the framework of the SUNDIAL project,

is based on the assumption that dialogues can best be described by means of a multi-level approach. The author distinguishes four levels: a domain-specific transaction level, an exchange level, an intervention level modeling initiative, reaction and evaluation, and finally a level consisting of dialogue acts. The system has been developed for the domain of flight reservations.

Plan-oriented approaches for dialogue modeling are described in Litman and Allen (1987) and Lambert and Carberry (1992). Both approaches distinguish domain or task, problem-solving and discourse levels. Knowledge from the various levels is employed to solve the task of plan recognition in dialogues. Connections do not only exist between the various levels but also between elements within one level. These links are modeled either as *discourse plans* which follow the course of the interaction (e.g. CONTINUATION, CLARIFICATION and TOPIC-SHIFT, see Litman and Allen (1987)) or as *discourse actions* that link an utterance with the context. A similar distinction of various levels of representation is made in O'Donnell (1990), except that the links between various elements of his discourse or exchange model are not made explicit.

The EES project (Explainable Expert Systems) is the basis for the theoretical and practical work of Moore and Paris (1989), Moore and Swartout (1990), and Carenini and Moore (1993). A central goal in EES was the creation of a flexible explanation module for expert systems communicating in natural language, allowing the user to ask questions about explanations given by the system and generating appropriate system responses. Strategies incorporating parameters such as context and focus were used to disambiguate the user's utterances. To address the intentions behind utterances and communicative goals, concepts of RST (cf. Mann and Thompson, 1987) and Speech Act Theory were exploited. The focus, however, is on flexible explanation dialogues and, unlike COR-RST, not on modeling information-seeking dialogues as complex "negotiations" with flexible ways to withdraw and reject dialogue contributions.

The Intelligent Documentation Advisory System (IDAS), developed by Reiter, Mellish and Levine (1992) represents an attempt to use dynamically generated natural language in the framework of an information retrieval system using hypertext techniques. To obtain information, the user clicks on the object under consideration and then chooses one out of a list of request options displayed by the system. However, there is no dialogue model at all, the system only allows simple query - answer cycles.

The concepts of speech acts in combination with RST are used in several multimedia presentation systems. Among the first systems following this approach were the WIP system developed at DFKI (see André and Rist (1993)) and the system developed by Maybury (1991). However, neither the WIP project nor Maybury's system use high-level dialogue structures. As pointed out in Arens *et al.* (1993) global structures are necessary for establishing overall coherence in the context of multimedia interfaces. COR-RST takes this into account. It focuses on Natural Language Generation, yet allows the extension to multimodal dialogue acts.

In the following two sections the two theoretical approaches (COR and RST) that were most influential for our work will be presented. Especially the COR model

will be described in detail, because it is essential for understanding our approach.

2.1.1 The Conversational Roles Model (COR)

In the field of information retrieval (IR) the interactive and communicative aspects of IR have only recently been emphasized (cf. for example, Belkin and Vickery, 1985; Belkin *et al.*, 1993). There exist approaches to distinguish various types of information retrieval strategies and tactics (cf. Bates, 1979), task hierarchies and global phases of the interaction. However, no elaborate interaction models are provided in this field (except simplistic iterative question-answer models). In the area of conversational analysis and discourse theory, on the other hand, we find various discourse and dialogue models which address local dialogue structures (e.g., Fawcett *et al.*, 1988; Grosz and Sidner, 1986, 1990; Reichman, 1985).

To be able to design a flexible dialogue system which can engage in cooperative information-seeking dialogues we use the "Conversational Roles" model (COR) developed by Sitter and Stein (1992). It has been used to design the interface of a multimedia information system, called MERIT (cf. Stein *et al.*, 1992). The COR model was originally influenced by the "Conversation for Action" (CfA) model (Winograd and Flores, 1986) which was applied to design computer-aided human-human interactions.

By adopting basic concepts of speech act theory and existing discourse models, and extending the CfA model for the situation of information-seeking human-computer interactions, the COR model shows the following features:

- it depicts the interaction as a cooperative two-party "negotiation" where commitments (to supply information or meta-information) can be made, retracted or rejected;
- it permits mixed-initiative dialogues and is flexible enough to describe all possible - even extremely complex - interaction patterns (this includes the temporary role changes of information seeker/ information provider, which frequently occur in highly vague task settings such as information-seeking);
- it provides the means for an explicit representation of the dialogue history in an abstract form, i.e., disregarding the interaction mode (graphical, linguistic, mixed).

According to COR the two participants (A and B) have dialogue goals and pursue specific conversational tactics to achieve these goals. The speaker's and addressee's mutual expectations about possible responses and about the subsequent course of the dialogue are essential. The model can be represented as a recursive state-transition network (see figure 1). The network defines the full potential of all possible interactions where successfully completed dialogue contributions/ acts end in specialized *states* (circles). *Transitions* (arcs) represent the various types of dialogue acts: e.g., REQUEST, OFFER, and INFORM can exactly be mapped onto Searle's basic "illocutionary types": directives, commissives, and assertives (cf. Searle, 1979); the other generic acts belong to the same categories, but they are less significant because they are merely responsive, e.g., PROMISE is a commissive act, but it adopts the conditions of action expressed by the preceding REQUEST (cf. Sitter and Stein, 1992). The

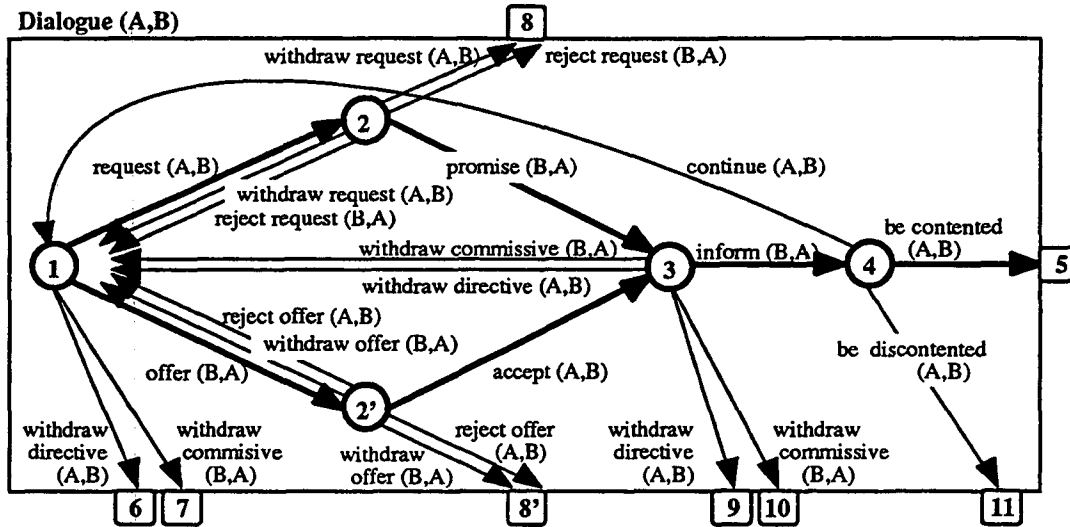


Figure 1: The DIALOGUE net of the COR model

traversal of the graph stops in states which are marked by squares. State $\langle 5 \rangle$, for instance, is reached when the information seeker (A) has expressed contentment with the given information and quits the dialogue. States $\langle 6 \rangle$ to $\langle 11 \rangle$ are also terminal states, but here the information need could not be satisfied. Note that a dialogue which ends in one of these states can be well-formed, cooperative and complete (e.g., B rejects a request of A, because the requested information could not be retrieved).

The bold arcs leading from state $\langle 1 \rangle$ to state $\langle 5 \rangle$ denote two "idealized" courses of the interaction which follow the basic role-expectations or role assignments. The two initiative acts (REQUEST, OFFER) typically establish new conditions of action, whereas the subsequent acts are reactive and do not introduce new conditions (PROMISE, ACCEPT, INFORM, and BE-CONTENTED are all "expected" in that they are positive responses to the preceding acts). Really encountered information-seeking dialogues, however, often do not follow such a simple, linear conversational development. Directive acts (REQUEST, ACCEPT) can be rejected by the addressee, commissions (OFFER, PROMISE) are often withdrawn. Both rejections and withdrawals (called "alternative" acts or responses) can either lead back to state $\langle 1 \rangle$, where the dialogue is entered again (begin of a new dialogue cycle), or to a terminal state (definite REJECT, WITHDRAW).

Another way of departing from the linear course of interaction is even more important. So far we have only considered "atomic" dialogue acts which are not further decomposed. But consider the following situations which often occur in information-seeking interactions: if the meaning of an utterance (atomic act) has not been understood by the addressee, or, if she needs additional information to be able to proceed, an embedded clarification dialogue might be necessary. In order to resolve this problem, the transitions in figure 1 must not be interpreted as atomic acts but as structured dialogue contributions. The extended COR model, therefore, defines basically two types of subnetworks: figure 2 displays the net of an INFORM contribution; figure 3 shows a representative net for all other types of contributions (here: REQUEST

as an example). Thus, recursion is taken into account.

In the figures 2 and 3 "A: request" and "A: inform" denote atomic acts, whereas a suffix notation indicates structured contributions, for example: ASSERT(A,B) or DIALOGUE(B,A, solicit context information). The traversal of the INFORM net is quite simple: A's INFORM act can be followed by a subdialogue initiated by B (e.g., by a REQUEST such as a clarifying question), or, if B does not need additional information, state $\langle c \rangle$ is reached immediately (jump). Thus, A's ("nuclear") INFORM act might be sufficient, whereas the ("satellitic") subdialogue is optional and depends on B's decision.

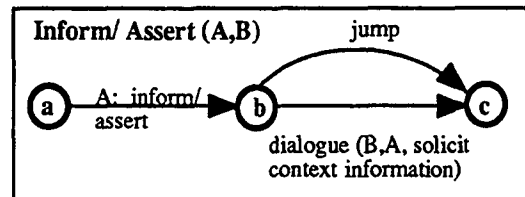


Figure 2: The INFORM net

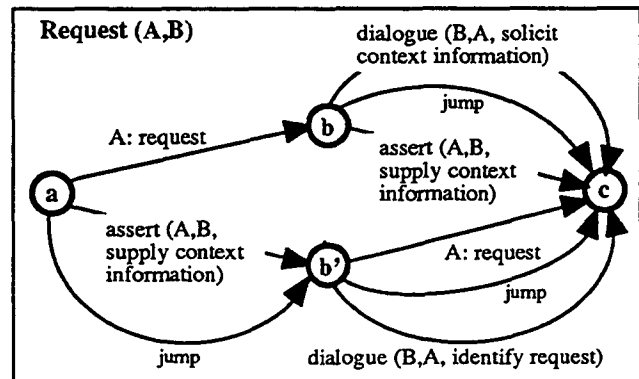


Figure 3: The REQUEST net

The transition net given in figure 3 is more complex but

follows the same principles. The subdialogues are also optional; A has two possibilities to start in state $\langle a \rangle$:

- A may start with a nuclear act (here, a REQUEST for information, such as a query to the database) and has then the opportunity to supply additional context information (e.g., an assertion to explain, specify or illustrate the request). If A does not give this context information voluntarily, B may enter a subdialogue to solicit the required context information. Thus, the ASSERT transition and the DIALOGUE leading from state $\langle b \rangle$ to $\langle c \rangle$ have a similar function, both being optional, i.e., satellitic.
- A may start with an ASSERT (entering the subnet type displayed in figure 2) to give context information concerning her request; she may add the explicit REQUEST immediately afterwards, or may skip the explicit utterance (jump), in case she believes that B is able to infer the intended request from the context. If B is not able to identify the request, B has the option to initiate a clarification DIALOGUE.

The COR model focuses on the illocutionary aspects of the conversation and abstracts away from the specific propositional content of dialogue contributions. However, it has been recognized that COR in its first version was only a partial model which had to be further enhanced by addressing rhetorical and semantic aspects (cf. Maier and Sitter, 1992; Stein and Maier, 1993). This was recently verified by Fischer (1993) who used the COR model to analyze a corpus of real dialogues between humans (information seekers communicating with information brokers to prepare a database search).

2.1.2 Rhetorical Structure Theory (RST)

Among the theories for modeling discourse, the RST – Rhetorical Structure Theory (cf. Mann and Thompson, 1987) is the theory most exploited for natural language processing, particularly for natural language generation. RST is a theory which describes the structure of written monologues. One of the most basic assumptions of RST is that coherence can be modeled by means of named relations which hold between adjacent text units. Such relations can be used to structure texts by iteratively applying relations thereby composing complex text units out of smaller ones.

Another assumption of RST is that, in general, a relation imposes an asymmetric structure on the connected text units. For a given pair of related text units, the so-called *nucleus* corresponds to the unit which contains highly relevant information, while the *satellite* carries less significant information; the satellite can be either substituted or left out without significantly changing the overall meaning of the discourse.

Since RST relations have been specified in a semi-formal way, this theory was a good candidate for a computational specification of coherence and later for an implementation of text planning and text generation systems.

Recent attempts have been made to use this theory also for modeling dialogues, in particular for modeling both the connections within and between various dialogue contributions which is in contrast to approaches which only use RST to model links within a dialogue contribution like, e.g., Moore and Paris (1993). Nearly all approaches (see, for example, Fawcett and Davies, 1992, and

Daradoumis, 1993) are in the area of human-computer interaction, where a generation component is responsible for the automatic production of system utterances. Another approach, which was developed for the domain of information-seeking dialogues, is reported in Maier and Sitter (1992). The authors showed that in such dialogues a specific subset of relations, the so-called *interpersonal relations* (see Maier and Hovy (1991)), are used. This classification of relations is based on three types of meaning as distinguished in Halliday (1985): ideational, interpersonal and textual meaning. Ideational meaning is the representation of experience of the world. Interpersonal meaning refers to what the speaker or writer does in order to address the goals of the recipient. Textual meaning, finally, relates pieces of discourse to the context and indicates how the discourse structure has to be interpreted. Interpersonal relations, therefore, share the behavior that they mainly address features of the discourse participants. Among these relations we find, for instance, JUSTIFICATION, where the satellite provides reasons why the speaker or the listener should carry out actions specified in the nucleus, or EVALUATION, where the satellite presents a subjective account of the information given in the nucleus.

2.2 Integration of COR and RST

2.2.1 Corpus Analysis

To find out how COR and RST can be integrated for modeling information-seeking dialogues, a corpus of dialogue transcripts – obtained from Prof. Saracevic, Rutgers University, New Jersey – was carefully analyzed. The transcribed dialogues were conversations between a person seeking information and an information broker specialized in database search.

The transcripts contained oral communication with frequent syntactical mistakes, incomplete and halfway reformulated sentences. Our approach was not to try to model these attributes of dialogues. Instead, the utterances were adapted to match written, error-free text.

First, a COR analysis was carried out, resulting in a segmentation of the transcribed dialogues into *acts*, *contributions* and whole dialogue *cycles*. Then these dialogue elements were assigned an illocutionary point and a nucleus or satellite status.

Taking these results, an RST analysis was performed. The existing segmentation into acts, contributions and dialogue cycles was used to create the text spans that make up the constituents – nuclei and satellites – of the RST analysis. In our analyses no major problems were encountered by applying RST to dialogues, even though RST was developed for monologues only.

2.2.2 Relations in the basic COR model

There is a relatively small number of typical RST relations that connect pairs of dialogue acts of the *basic* COR model which does not incorporate the recursive structure for subdialogues. The most important ones are SOLUTIONHOOD, EVALUATION, EVALUATION* and BACKGROUND. Note that EVALUATION* is a newly defined relation that inherits aspects of EVALUATION; it will be described later in this section.

How these relations apply to illocutionary acts as defined in the COR model is shown in a sample dialogue and its RST analysis (figures 4 and 5). As can be seen there, in the first phase of the dialogue an act of type REQUEST (for information) gets acknowledged positively by a PROMISE (to search for information and to present what has been found). The nucleus of the EVALUATION relation holding between the two acts is the REQUEST.

User	(REQUEST)	How many EC-funded projects dealing with Recombination are there?
System	(PROMISE)	Let's see.
System	(INFORM)	There are 15 projects.
User	(BE-CONTENTED)	Okay.

Figure 4: A sample dialogue

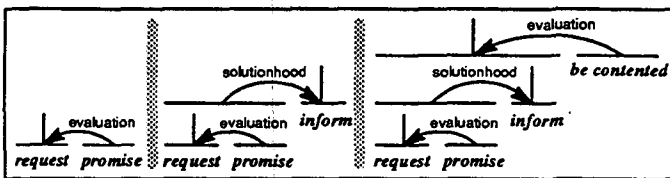


Figure 5: Development of a dialogue as a RST diagram

In the second stage, the requested information is given. REQUEST and PROMISE become satellite of a new relation which has an INFORM act as its nucleus. The suitable relation here is SOLUTIONHOOD, since the INFORM carries the answer to the REQUEST. Finally, the appropriateness of the provided data is confirmed by an act of type BE-CONTENTED. The recursively determined nucleus of the whole dialogue turn is the INFORM act. This is in line with what can be expected for information retrieval dialogues: the presentation of information to be looked for is the most central part of the whole dialogue. Initiative acts (REQUEST, OFFER) are also important, but they merely open the structure span which is completed when the INFORM act is given, or when it is REJECTED or WITHDRAWN.

Several dialogue cycles which appear within one level of dialogue are usually connected by the BACKGROUND relation (not shown here, see Fischer (1993) for examples).

2.2.3 Expectation and Nuclearity

In the COR model, roles – or rather expectations of specific role behavior – are essential. Some acts are expected, while others are not; the latter ones are called *alternative*. The following describes how the expectation of certain dialogue acts influence their status as a constituent within an RST relation, i.e., whether they are considered the nucleus or the satellite of the relation.

On the dialogue level (see figure 1), the acts REQUEST, OFFER, and INFORM are most important. They actually contribute to the progression of the dialogue insofar as they actively model the negotiation of information. Other acts are merely evaluations of these three acts, they can either be positive or negative. The positive ones include ACCEPT, PROMISE, and BE-CONTENTED. The negative ones are WITHDRAW, REJECT, and BE-DISCONTENTED.

The acts ACCEPT, PROMISE, and BE-CONTENTED are all *expected* ones. In our corpus, between any of these acts

and their respective preceding acts the (positive) EVALUATION relation holds. Since the EVALUATION relation defines the constituent that contains the *evaluating* expression as the *satellite* of the relation and the *evaluated* expression as the *nucleus*, this behavior matches features of the COR model: The evaluated dialogue acts are REQUEST, OFFER, or INFORM which make significant propositional contributions to the dialogue. For example, in the REQUEST-PROMISE pair of dialogue acts, REQUEST is the nucleus and PROMISE the satellite – the latter one being the expected positive acknowledgement.

Evaluations that are *not expected* are WITHDRAW, REJECT, and BE-DISCONTENTED. They give the dialogue an alternative turn. This means that the evaluation is of high relevance and overrides the importance of previous acts. In order to model this fact by means of RST, an alternative evaluation relation has to be introduced, which defines the *evaluating* expression to be the *nucleus* of the relation, in contrast to its definition as given above. We called this relation EVALUATION* because it resembles EVALUATION, except that it swaps the roles of the involved constituents.

2.2.4 Context relations for the extended COR model

The *extended* COR model contains complex dialogue contributions with a recursive structure (see figures 2 and 3). These complex acts have two constituents. One of these constituents expresses the illocutionary point of the whole contribution. Therefore, it takes on the role of the nucleus. Examples are: A: INFORM, A: REQUEST, A: OFFER. The other constituent is either an ASSERT contribution (atomic or complex) or a DIALOGUE to negotiate the contextual information (see also section 2.1.1).

The fact that ASSERT and DIALOGUE serve the same purpose in COR – namely to provide context information – also had to be modeled adequately in terms of RST. This was achieved by simply assigning the INFORM act to be the nucleus of the whole DIALOGUE. Both ASSERT and INFORM may contain the same proposition (contextual information). The only difference is the way to get to this state: either A gives the information voluntarily (ASSERT) or B initiates a sub-dialogue to ask for the information (DIALOGUE).

Concerning the question about the types of relations typically holding between the nuclear act – carrying the illocutionary point of the compound contribution (REQUEST, for example) – and the accompanying satellitic (assertive) act, our analyses resulted in finding two distinct types: (1) The first type is for additional information that is needed in order to answer a question or to understand a certain statement. In our genre of dialogues, this information is obtained by applying *Information Retrieval Tactics*. (2) The second type is for supplementary information that explains the underlying reasoning behind an act made by a dialogue participant. We call this information type *meta-information*. The two relation types outlined above are now described in more detail.

Context relation type 1: Information Retrieval Tactics In information-seeking dialogues, it is very uncommon that questions can be answered immediately. In most cases additional information is necessary. One way

to do this is to use *Information Retrieval Tactics*. Their purpose is the procurement of data that is needed to successfully answer a request. In the transcribed dialogues, the tactics were employed by the information broker. An intelligent information retrieval system should also be able to handle at least some of these tactics. Tactics include replacement or addition of search terms, pursuing or neglecting search paths; for a detailed collection see Bates (1979).

The short dialogue shown in figure 6 exemplifies the use of the tactic *SUPERTERM*, which is used to replace a specific term by a more general one. The system cannot give an appropriate answer to the original request for all known EC-funded projects involving "NL Generation" because that search item is not contained in the database. In order to get some relevant entities, it tries to extend the search, by replacing the original search term by a superterm. Therefore, it initiates a subdialogue asking the user to provide another term, and the user chooses "Natural Language".

User	(REQUEST)	Show me all EC-funded projects dealing with NL Generation.
		<i>Subdialogue to solicit contextual information, to</i>
System	(REQUEST)	Can you tell me a more general term than NL Generation?
		<i>Subdialogue to solicit contextual information</i>
User	(REQUEST)	Why?
System	(INFORM)	I ask you because NL Generation is unknown.
User	(BE-CONTENTED)	I see.
User	(INFORM)	How about Natural Language.
System	(BE-CONTENTED)	OK.
System	(INFORM)	Here is what I found. <i>(System presents table of results.)</i>
System	(OFFER)	Would you like to see projects of a particular company?
User	(ACCEPT)	Yes, ...

Figure 6: A dialogue example

Information retrieval tactics can be integrated into RST without significant problems. They are subsumed by existing RST relations, such as *GENERAL-SPECIFIC* and *ABSTRACT-INSTANCE*. We have observed that most common tactics used in Information Retrieval are in fact forms of (the very broad) *ELABORATION* relation. A more exhaustive analysis regarding the mapping of tactics to RST relations is required for a complete picture.

Context relation type 2: Meta-Information

Apart from the contextual information concerned with the contents of the database, there is a second type of context which deals with underlying reasoning behind utterances. Making these explicit means to make this underlying reasoning transparent to the dialogue partner. An information retrieval system can enhance confidence of the information seeker by giving this meta-information. There are several RST relations available for connecting assertive acts with meta-information to the nuclear act. Examples are *CAUSE*, *PURPOSE* and *INSTRUMENT*.

The RST analysis of the sample dialogue given above is shown in figure 7. It shows examples for tactics (*SUPERTERM*) and meta-information (*CAUSE*). Note that the context relations are connecting dialogue acts (*REQUESTS* in this example) with complete (sub-) dialogues,

to be precise: the respective *INFORM* act within these sub-dialogues. Indeed, this consequently means that at some stage of the dialogue, no real relation holds between two adjacent dialogue acts, e.g., two consecutive *REQUESTS*. For the text generation processes, however, this does not involve a drawback. The *COR* model itself transposes the RST structure given for the whole sub-dialogues to single dialogue acts.

3 The Corinna System

In order to create an information system based on the integrated model of *COR* and *RST*, several aspects had to be considered which go beyond the previously outlined theoretical framework.

The models *COR* and *RST* are of descriptive nature; they are not sufficient to actually *construct* a dialogue between user and system. To do this, a dialogue manager also needs further semantic information which constitutes the propositional content of the dialogue act to be generated. This semantic knowledge depends on and is provided by the genre (information seeking), the application area (information retrieval) and the actual domain.

As domain we chose the content of the *CORDIS* databases, which represent information about EC-funded research projects, participating organizations and strategic funding programs. The databases are offered online by *ECHO*, a database host of the Commission of the European Community. They have been transformed to and integrated into a relational database system which is publicly available. For the implementation of *Corinna* a way of incorporating this domain knowledge into the dialogue management and the generation of the dialogue acts had to be found.

Natural language was decided to be the main mode of interaction in the *Corinna* system, because it achieves a high level of expressive power. The user is presented a menu of possible dialogue acts (according to the current dialogue state) among which she may choose the most appropriate one. The *PENMAN* generation system was used to produce both the system's and the user's dialogue contributions. An additional modality of deictic gestures was introduced to allow convenient interaction with the system when the type of user contribution cannot be generated, for example when the basic search term and method is selected from a visual presentation.

3.1 Discourse and Knowledge Structures

3.1.1 Concepts and relations

COR focuses on the dialogue pragmatics and omits semantic aspects in order to be domain-independent. As described above, *RST* can be used to introduce an additional level of coherence description between dialogue contributions. Additionally, some of the *RST* relations can be specialized in such a way that they represent relations between *concepts* that constitute the propositional content of dialogue acts.

The propositional level was modeled and implemented in a completely object-oriented way. Concepts of all levels of knowledge (dependent as well as independent from the domain) were classified as classes, instances and attributes. The object-oriented approach makes use of the

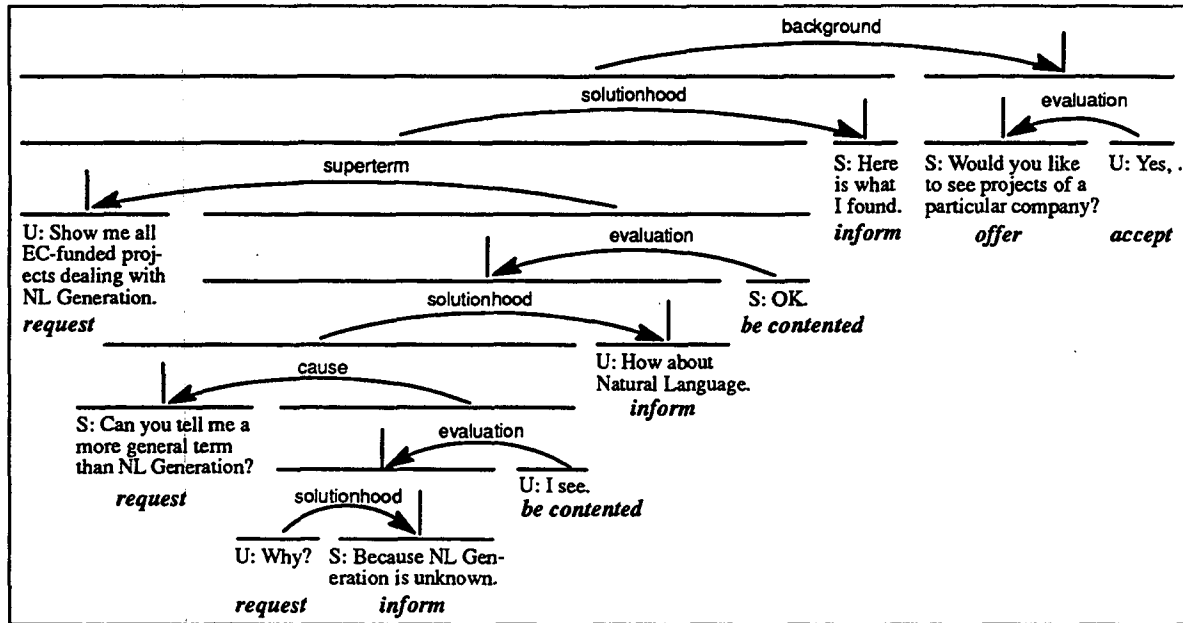


Figure 7: RST analysis of the sample dialogue

notion of "relations". In the Corinna system we focused on the representation of relations on the meta-level of the object-oriented paradigm. Examples include *SUBCLASS*, *CLASS-INSTANCE*, and *INSTANCE-ATTRIBUTE*.

The ideational branch of Maier and Hovy's (1991) meta-functionally motivated taxonomy of RST relations made it particularly easy to augment the hierarchy, achieving a powerful model that uses relations as a unified way to structure knowledge.

3.1.2 Developing the dialogue

Within the process of information seeking, the situations of the search - specifically failure or partial success - direct the dialogue in certain ways. Three main decision criteria have been incorporated in the Corinna system: 1) the requested data must be available in the database, 2) the data must be presentable in such a way that it is of use for the information seeker, and 3) the data has to be relevant with respect to the expressed information need. Depending on these factors, a follow-up dialogue act can be selected. For example, an *OFFER* is only possible when availability and presentability of the data can be granted. If the data is relevant, a follow-up *ACCEPT* is possible. If, on the other hand, the information need is not matched by the offered information, the *OFFER* will have to be answered (i.e., evaluated) by a *REJECT-OFFER*.

Currently, the system is mainly able to judge availability and presentability of data. It is able to reason about the user's information need only in a very limited way. However, this does not lead to a significant drawback, since the user is always given the choice between several alternatives, as for example between an *ACCEPT* and a *REJECT-OFFER*.

Taking into account what has been said above about the integration of COR and RST, the process for the generation of new dialogue contributions also includes the

treatment of RST relations. Given the propositional content of a dialogue act, the determination of a follow-up act is influenced by the RST relation which holds between the two acts. For example, if a relation of the type *meta-information* (see section 2.2.4) has been chosen where contextual information has to be supplied, an appropriate proposition expressing the state or direction of the database search has to be composed. By this way sequences of coherent propositions for dialogue acts are produced.

3.1.3 Access to the database

Typically, an information retrieval system has its main challenge in the treatment of very large amounts of data. The CORDIS databases contain information about some 14,000 projects and 10,000 organizations, persons and other entities.

The object-oriented structure of Corinna enabled us to implement retrieval methods for several types of data storage - internally stored discourse knowledge and database contents represented externally - still having the benefit of a uniform programming interface. Thus, the dialogue manager is transparently accessing two distinct types of data, simplifying the design and implementation of the system significantly.

3.2 Realizing Dialogue Acts

So far the focus was on the underlying models of a dialogue. This is motivated by the fact that the described mechanisms for performing a dialogue based on COR and RST apply to any modality, not necessarily natural language. For the Corinna prototype we chose the linguistic form as interaction mode because it incorporates a high level of expressive power needed for Information Retrieval tasks. Also, it has significant advantages over other interaction modes, as far as the clarification of misunder-

standing and dialogue failure are concerned. Information types required in such dialogue states cannot be presented graphically – language is the preferable interaction mode (cf. also Dilley *et al.*, 1992).

The realization of dialogue contributions is divided into two major steps. First, the internal representation of a dialogue act is transferred into a high-level logical form that is capable of dealing with language-oriented attributes of an utterance, out of which English text is then automatically generated in the second step. For this task we used the PENMAN generator (PENMAN (1989)) with the Sentence Plan Language SPL (Kasper and Whitney (1989)) as intermediate representation.

3.2.1 Sources of Data for Realizing Acts

We want to focus here on three main parameters influencing the process of realizing a dialogue act: *illocution*, *proposition* and *RST relation*.

It has been pointed out earlier that the propositional content and illocutionary point of dialogue acts are managed independently. During the design of Corinna, it became evident that also the realization process can take place quite independently, as far as the creation of the SPL plan is concerned. The flexibility needed for natural language utterances is achieved by combining various intermediate logical representations of realizations for propositions and illocutions. Several typical linguistic attributes for the illocutionary acts in the COR model can be assumed in order to allow management in the context of a limited computer program. For example, many forms of REQUEST incorporate the use of imperative forms like “tell me”, “show me”, etc.

In cases where contextual information is to be given for an assertive dialogue act the RST relation can make significant contributions towards a coherent realization of the two utterances involved. For example, relations like PURPOSE or INSTRUMENT can be signaled by markers such as “in order to”, “for”, etc.

Taking together 1) the fragmentary realization plan of the proposition on a purely assertive level, 2) the generated fragments for an illocutionary act of COR, and 3) the RST relation that connects this act to the previous act, a suitable SPL statement for PENMAN can be composed. The final result is an English sentence, which is passed to the user interface.

3.2.2 An example

The following simplified example illustrates the various steps in the planning of a contribution. We assume the scenario as given in figure 6 where the system has just asked for a SUPERTERM of “NL Generation”. In this example the user wants to know the reason of the request and the system gives the corresponding answer. The SPL statement for this utterance is given in figure 8. Three distinct types of information were used to compute this sentence plan: (1) The RST relation CAUSE was selected, since the user asked the question “Why?”. (2) The domain of this relation is the verbal action that was performed by the system previously in the dialogue: The REQUEST for the more general term. (3) The motivation for making the utterance is the situation in the database – the search term is unknown. This fact is represented

by a propositional item describing the state of the search. Note that the system’s response is a dialogue act of type INFORM. As it has been stated earlier, the acts of type INFORM become the nucleus of the whole dialogue, which means that the (sub)dialogue takes on the role of an assertion after its completion. In case of a subdialogue, this assertion then carries the accompanying contextual information for another act one dialogue level up - in our case this is the REQUEST of the system, asking for a replacement term, and the contextual information is the CAUSE of that question. The RST relation plays then an important role in creating the coherence between the original *question* of the system and the contextual information, which is the *reason*, why it asked this question.

```
(connecting-relation / rst-volitional-cause
:domain ;; The system has asked for a
        ;; superterm of "NL Generation"
(act / addressee-oriented-verbal-process
:lex ask
:sayer speaker
:addressee hearer)
:range ;; The system informs the user
        ;; about the reason why it asked.
(proposition / property-ascription
:tense present
:domain (concept / object
         :name NL-Generation)
         :range (q / quality
                :lex unknown)))
```

Figure 8: Sentence plan for “I ask you, because NL Generation is unknown”

3.3 Interaction with the system

It has been pointed out previously that we aimed at having a mainly language-driven interface. Additionally to the generation of the system’s utterances we decided to also generate the dialogue contributions of the user. That is, there is no Natural Language Understanding module in Corinna; instead, in each dialogue situation where it is the turn of the user, Corinna generates several suggested utterances, from which the user can then choose the one that suits her best in the context of the current dialogue situation. We believe that this does not pose a severe limitation, the dialogue model based on COR-RST, as well as the semantic knowledge about the search in the database allow predictions about how the dialogue might be continued.

Figure 9 shows a sample interaction. The screen is divided into two halves. In the upper one, the instantiated dialogue history is displayed and updated after each contribution. The lower half of the interface is reserved for presentation of the database queries and their results. In the dialogue state as represented by the snapshot, the system has just provided the table of results. The user then has the choice to select one out of three possible acts: BE-CONTENTED (“OK”), CONTINUE (“This is not sufficient”) and BE-DISCONTENTED (“Let’s stop this dialogue”).

Note that, depending on the situation of the dialogue, a participant – system or user – may utter two acts in a row, as opposed to simple system–user alternations. For example, after the user is contented with the answer of the system about why it asked for a term replacement,

CORINNA - Cooperative Retrieval Interface for Natural Language Acts

```

User      Show me all projects with a title like NL GENERATION.
System    Can you tell me a term that is more general than NL GENERATION?
User      Why?
System    I am asking you, because NL GENERATION is unknown.
User      Okay
System    NATURAL LANGUAGE
User      NATURAL LANGUAGE is okay.
System    The table contains all relevant entities.
System

User's choice
OK
This is not sufficient.
I do not like that. Let's stop this dialogue.
User

```

ACRONYM	PROGRAM	ACRONYM	REFERENCE	TITLE	START-DATE	END-DATE
ACQUILEX	ESPRIT 2	3030		ACQUISITION OF LEXICAL KNOWLEDGE FOR NATURAL LANGUAGE PROCESSING SYSTEMS	Jun 1 1989 12:00:00:000AM	Mar 31 1992 12:00:00:000AM
ACORD	ESPRIT 1	393		CONSTRUCTION AND INTERROGATION OF KNOWLEDGE BASES USING NATURAL LANGUAGE TEXT AND GRAPHICS	Jan 1 1985 12:00:00:000AM	Dec 31 1989 12:00:00:000AM
DANDELION	ESPRIT 3	DANDELION		DISCOURSE FUNCTIONS AND DISCOURSE REPRESENTATION: AN EMPIRICALLY	Sep 1 1992 12:00:00:000AM	Aug 31 1995 12:00:00:000AM

Dialogue Driver

Wed 9 Feb 5:14:43 Fischer CL PENMAN: User Input

Figure 9: A sample interaction with Corinna

("Okay"), she then makes a contribution to the dialogue one level up – levels are represented by indentation of contributions – which was a term the system had asked for ("Natural Language"). A similar phenomenon can be observed in the last two utterances of the system. To decide whether the user is allowed to utter two acts in a row, the system applies a set of (relatively simple) heuristics; they are giving the user preference in guiding the dialogue.

4 Summary

The Corinna system is the result of an integration of two models: The COR model and RST. Compared to the single theories the integrated model has a number of significant advantages: (1) The scope of RST could be extended from monologues to (information seeking) dialogues – without changing major assumptions of the original theory; (2) RST adds an additional level of coherence to the existing COR model. While COR focuses on illocutions, dialogue acts and their sequential order, RST provides means to describe the semantics of the links between the single acts; (3) The description of dialogues in terms of COR-RST gives rise to a systematic approach to dynamically generate dialogue contributions for either of the participants. Corinna, as a prototypical implementa-

tion of the combined theories, uses PENMAN to generate natural language utterances. It accesses a database typical for the information retrieval task (CORDIS).

Of course there are many possibilities for improvement. First, the flexibility and power of the generated utterances can be further increased. The plans for the dialogue contributions could be further parameterized, also the knowledge base covering the current situation of the search in the database needs additional concepts. On the side of Information Retrieval, more tactics could be implemented making the search more efficient.

We think that the outlined theoretical framework is powerful enough to go beyond the state of the system as it is implemented so far, and we will continue this research especially in the context of concrete application areas.

Acknowledgements

We would like to thank Stefan Sitter for support and inspiring discussions. We also gratefully acknowledge the comments of three anonymous reviewers who made many helpful suggestions. Part of the work of the second author was funded by the German Ministry for Research and Technology (BMFT) under contract 01 IV 101 k/1 (VERBMOBIL).

References

- [André and Rist, 1993] Elisabeth André and Thomas Rist. The design of illustrated documentation as a planning task. In Mark T. Maybury (ed.), *Intelligent Multimedia Interfaces*. pages 94-116, AAAI Press/MIT Press, Menlo Park, California, 1993
- [Arens et al., 1993] Yigal Arens, Eduard Hovy and Susanne van Mulken. *A tree traversing prototype that allocates presentation media*. ISI, University of Southern California, 1993
- [Bateman et al., 1991] John Bateman, Elisabeth Maier, Elke Teich and Leo Wanner. Towards an architecture for situated text generation. In *Proc. of the International Conference on Current Issues in Computational Linguistics*, Penang, Malaysia, 1991
- [Bates, 1979] Marcia Bates. Information search tactics. *Journal of the American Society of Information Science*, 30:205-214, 1979
- [Belkin et al., 1993] Nicholas Belkin, Colleen Cool, Adelheit Stein, and Ulrich Thiel. Scripts for information-seeking strategies. In *Case-Based Reasoning and Information Retrieval. Papers from the AAAI Spring Symposium Series*. AAAI Technical Report SS-93-07, pages 8-17, 1993
- [Belkin and Vickery, 1985] Nicholas Belkin, and A. Vickery. *Interaction in information systems*. The British Library, London, 1985
- [Bilange, 1991] Eric Bilange. A task independent oral dialogue model. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics (EACL-91)*. pages 83-88, Berlin, Germany, April 1991
- [Carenini and Moore, 1993] Giuseppe Carenini and Johanna Moore. Generating explanations in context. In *Proc. of the 1993 International Workshop on Intelligent User Interfaces*. pages 175-192, ACM Press, New York, 1993
- [Daradoumis, 1993] Thanasis Daradoumis. Building a dynamic RST-based grammar of coherent interactive dialogues. In Michael Zock, Giovanni Adorni, and Giacomo Ferrari, editors, *Proceedings of the Fourth European Workshop on Natural Language Generation*. pages 91-101, Pisa, Italy, April 1993
- [Dilley et al., 1992] Stephan Dilley, John Bateman, Ulrich Thiel, Anne Tißen. Integrating natural language components into graphical discourse. In *Proc. of the 3rd Conference on Applied Natural Language Processing*. pages 72-79, ACL, Trento, Italy, 1992
- [Fawcett et al., 1988] Robin P. Fawcett, Anita van der Mije, and Carla van Wissen. Towards a systemic flowchart model for discourse structure. In Robin P. Fawcett and David J. Young, editors, *New Developments in Systemic Linguistics; Volume 2: Theory and Application*. pages 116-143. Pinter Publishers, London, 1988
- [Fawcett and Davies, 1992] Robin P. Fawcett and Bethan Davies. Monologue as a 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*. Berlin, Springer, 1992
- [Fischer, 1993] Markus A. Fischer. *Weiterentwicklung und Implementierung eines Dialogsystems für kooperative Informationssysteme*. GMD-Studien, No 228, Sankt Augustin: GMD, 1994 (in German)
- [Grosz and Sidner, 1986] Barbara J. Grosz and Candace L. Sidner. Attention, intentions and the structure of discourse. *Journal of Computational Linguistics*, 12(3):175-204, 1986
- [Grosz and Sidner, 1990] Barbara J. Grosz and Candace L. Sidner. Plans for discourse. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*. pages 417-444. MIT Press, Cambridge, MA, 1990
- [Halliday, 1985] M.A.K. Halliday. *An introduction to functional grammar*. Edward Arnold, 1985
- [Kasper and Whitney, 1989] Robert Kasper and Richard Whitney. *SPL: A sentence plan language for text generation*. Technical report, USC/ISI, 1989
- [Lambert and Carberry, 1992] Lynn Lambert and Sandra Carberry. Using linguistic, world, and contextual knowledge in a plan recognition model of dialogue. In *Proceedings of COLING-92, Nantes, France*. pages 310-316, 1992
- [Litman and Allen, 1987] Diane J. Litman and James F. Allen. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11:163-200, 1987
- [Maier and Hovy, 1991] Elisabeth Maier and Eduard Hovy. A metafunctionally motivated taxonomy for discourse structure relations. In *Proc. of the 3rd European Workshop on Natural Language Generation*. Judenstein, Austria, March 1991
- [Maier and Sitter, 1992] Elisabeth A. Maier and Stefan Sitter. An extension of rhetorical structure theory for the treatment of retrieval dialogues. In *Proc. of the 14th Annual Conference of The Cognitive Science Society*. pages 968-973, Lawrence Erlbaum Associates, Hillsdale, NJ, 1992
- [Mann and Thompson, 1987] William C. Mann and Sandra A. Thompson. Rhetorical Structure Theory: A theory of text organization. In Livia Polanyi, editor, *The Structure of Discourse*. pages 85-96. Ablex Publishing Company, Norwood, NJ, 1987
- [Maybury, 1991] Mark T. Maybury. Planning Multimedia Explanations Using Communicative Acts. In *Proc. of the 9th National Conference on Artificial Intelligence (AAAI'93)*. pages 61-66, AAAI Press/MIT Press, Anaheim, California, 1991
- [Moore and Paris, 1989] Johanna Moore and Cécile Paris. Planning text for advisory dialogues. In *Proc. of the 27th Annual Meeting of the Association of Computational Linguistics*. pages 26-29, Vancouver, B.C., Canada, June 26-29, 1989
- [Moore and Paris, 1993] Johanna Moore and Cécile Paris. Planning text for advisory dialogues. Capturing intentional and rhetorical information. In *Computational Linguistics*, 19(4):651-694, 1993
- [Moore and Swartout, 1990] Johanna Moore and William Swartout. Pointing: A way towards explanation dialogue. In *Proc. of the 8th International Joint Conference on Artificial Intelligence (AAAI'90)*, Vol. 1. pages 457-464, AAAI Press/MIT Press, Menlo Park, California, 1990
- [O'Donnell, 1990] Michael O'Donnell. A dynamic model of exchange. *Word*, 41(3), December 1990
- [PENMAN, 1989] *The PENMAN User Guide*. Information Sciences Institute, Marina del Rey, California, 1989
- [Reichman, 1985] Rachel Reichman. *Getting computers to talk like you and me*. MIT Press, Cambridge, MA, 1985
- [Reiter, Mellish and Levine, 1992] Ehud Reiter, Chris Mellish and J. Levine. Automatic generation of on-line documentation in the IDAS project. In *Proc. of the 3rd Conference on Applied Natural Language Processing*. Trento, Italy, 1992
- [Searle, 1969] John R. Searle. *Speech acts*. Cambridge University Press, UK, 1969
- [Searle, 1979] John R. Searle. A taxonomy of illocutionary acts. In *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge University Press, Cambridge, GB, 1979
- [Sitter and Stein, 1992] Stefan Sitter and Adelheit Stein. Modeling the illocutionary aspects of information-seeking dialogues. *Information Processing and Management*, 28(2):165-180, 1992
- [Smith, Hipp and Biermann, 1992] R.W. Smith, D.R. Hipp, and A.W. Biermann. A dialogue control algorithm and its performance. In *Proc. of the 3rd Conference on Applied Natural Language Processing*. Trento, Italy, 1992
- [Stein and Maier, 1993] Adelheit Stein, and Elisabeth Maier. Modeling and guiding cooperative multimodal dialogues. In *Human-Computer Collaboration; Reconciling Theory, Synthesizing Practice. Papers from the AAAI Fall Symposium Series*. AAAI Technical Report FS-93-05, pages 107-112, 1993
- [Stein and Thiel, 1993] Adelheit Stein and Ulrich Thiel. A conversational model of multimodal interaction in information systems. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI'93)*. pages 283-288. AAAI Press/MIT Press, Menlo Park, CA, 1993
- [Stein et al., 1992] Adelheit Stein, Ulrich Thiel, and Anne Tißen. Knowledge based control of visual dialogues. In *Proceedings of the 1st International Workshop on Advanced Visual Interfaces (AVI-92)*. pages 138-155. World Scientific Press, Singapore, 1992
- [Winograd and Flores, 1986] Terry Winograd and Fernando Flores. *Understanding computers and cognition*. Ablex Publishing Company, Norwood, NJ, 1986