gefördert vom
Bundesministerium
für Forschung
und Technologie

Verbmobil
Verbundvorhaben

# A Multi-Dimensional Representation of Context in a Speech Translation System — A Practical Approach

Elisabeth Maier

DFKI GmbH

Vm Report 64
April 1995

April 1995

Elisabeth Maier

DFKI GmbH
Stuhlsatzenhausweg 3
66123 Saarbrücken

Tel.: (0681) 302 - 5347
e-mail: maier@dfki.uni-sb.de

Das diesem Bericht zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministers für Forschung und Technologie unter dem Förderkennzeichen 01 IV 101 K/1 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei den Autoren.

# A Multi-Dimensional Representation of Context in a Speech Translation System — A Practical Approach

Elisabeth Maier*

DFKI GmbH, Stuhlsatzenhausweg 3, D-66123 Saarbrücken
maier@dfki.uni-sb.de
Phone: +49 681 302 5347, Fax: +49 681 302 5341

April 6, 1995

**Abstract**

In this paper we show how the notion of context has been refined in order to fulfill the requirements posed by a natural language processing system. We describe the context model of the speech translation system VERBMOBIL which has been derived from constraints given by various system components. In VERBMOBIL context is stored in the *dialogue memory* which is incrementally constructed by the dialogue processing component. We describe our context model and give sample representations. We also relate our model to state of the art approaches in linguistics.

## 1 Introduction

It is a widely agreed-upon fact that many decisions in natural language (NL) systems are influenced by contextual information. This is of particular im-

*This paper also appeared in the Proceedings of the IJCAI-95 Workshop "Context in Natural Language Processing". The paper has been significantly improved through input by Norbert Reithinger, Jan Alexandersson and Wolfgang Finkler. Thanks go also to Martin Klesen who implemented the dialogue memory.

portance in NL systems which have to take efficiency and robustness of processing into account and which therefore have to come to fast and reliable solutions. In this paper we present the contextual representation of such a system, VERBMOBIL, which has been developed to produce on-demand translations of contributions in spoken negotiation dialogues. Since VERBMOBIL is a strongly application-oriented system, the development of our context model aimed not so much at determining the full potential of features which characterize the context in which the negotiation dialogue takes place; rather, the design of the context model was guided by practical considerations like the following:

- Which decisions in a speech-to-speech translation system depend on contextual information?
- Which decisions can be rendered more efficient by additionally taking contextual information into account?
- Which information has to be available in a context representation in order to enable such decisions?
- How must the context be represented to support the various system components in an adequate way?

In this paper we show how these questions have been answered in the VERBMOBIL framework. We first discuss which requirements and features had to be taken into account when designing our context representation (section 2). Then we outline our notion of context which is derived from these requirements and which served as basis for the development of our context model (section 3). At the same time we contrast our point of view with some relevant theoretical approaches. In section 4 we present our multi-functional 3-layered model of context, giving examples for the contextual representation of fragments taken from our dialogue corpus. After a discussion of the computational adequacy of our model (section 5.1) we compare it with state of the art work in computational linguistics (section 5.2). We conclude and give an outlook on future extensions (section 6).

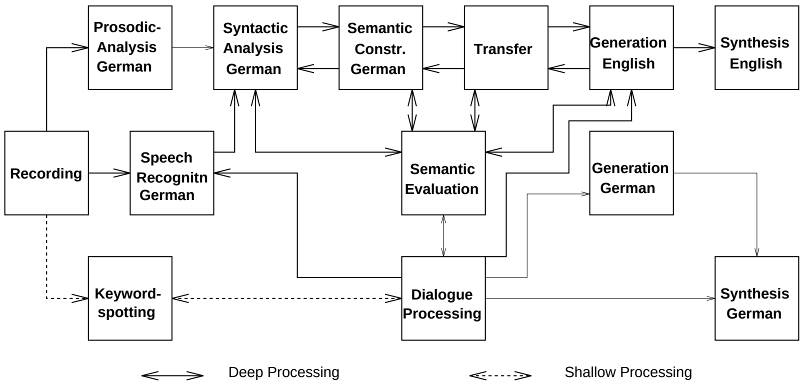


Figure 1: Architecture and control flow of the VERBMOBIL system.

# 2 Contextual Requirements in a Speech Translation System

Before we discuss how various VERBMOBIL components depend on the availability of contextual information we briefly sketch architecture and function of the system.

## 2.1 The VERBMOBIL Architecture

Since VERBMOBIL provides translation on demand only, the system distinguishes two processing modes (see figure 1): (1) *Deep Processing* when one of the dialogue participants requests a translation. In this case the input goes through phases of speech recognition, syntactic and semantic analysis, dialogue processing, transfer, re-generation and synthesis, delivering spoken output in the target language; (2) *Shallow Processing* when both dialogue participants interact in the same language without the necessity for translation. In order to follow the dialogue superficially a key word spotter examines the input for cue words which are characteristic for certain dialogue steps.

In both cases it is the task of the dialogue component to monitor the progress of the dialogue: at any point of interaction the dialogue module must be able to determine in which state the dialogue currently is and how the interaction progresses.

The underlying dialogue model, which describes the expected actions of the participants in an appointment scheduling dialogue, is composed of dialogue acts (see e.g. [Bunt, 1981]). Our model can be compared to numerous state of the art approaches for dialogue systems, as e.g. the dialogue models of the systems EVAR [Mast *et al.*, 1992] and SUNDIAL [Bilange, 1991]. A model based on dialogue acts seems to be an appropriate approach also from the point of view of machine translation and of transfer in particular: While in written discourse sentences can be considered the basic units of transfer, this assumption is not valid for spoken dialogues. In many cases only sentence fragments are uttered, which often are grammatically incomplete or even incorrect. Therefore different descriptive units have to be chosen. In the case of VERBMOBIL these units are dialogue acts.

These acts are composed into larger dialogue segments, i.e. into dialogue strategies and dialogue phases. A dialogue consists of three phases: (1) an *introductory phase*, where the discourse participants greet each other, introduce themselves, provide information about their professional status and introduce the topic of conversation, e.g. the necessity to find a date for a business meeting; (2) a *negotiation phase*, where the discourse participants repeatedly offer possible time frames, make counteroffers, refine the time frames, reject offers and request other suggestions; and (3) a *closing phase*, in which a final agreement is reached; in this phase either the dialogue is terminated or the dialogue partners begin the negotiation of another appointment.

In our dialogue model the three phases are composed of 18 dialogue acts which have been derived from a corpus of transcribed appointment scheduling dialogues (for more details see [Maier, 1994]).

## 2.2 Contextual Information Needed by VERBMOBIL Components

In the current version of the VERBMOBIL prototype the following subcomponents rely on context information:

### Transfer

Both the nature of the underlying dialogue act and the phase to which an utterance belongs contribute to the determination of translational equivalents. Below we give two examples from our corpus of appointment scheduling dialogues (Example 1(a) and 1(b)):

**Example 1(a)** – utterance mps1_1_16:

```
wo sollen wir uns denn treffen {seos} (REQUEST_FOR_SUGGESTION)
<Pause> geht es bei Ihnen {quest} (REQUEST_FOR_STATEMENT)
```

**Example 1(b)** – utterance fsp2_1_02:

```
/h/ (..) die zweite (wie sieht's aus) Oktober-Woche (SUGGESTION)
wie sieht 's bei Ihnen aus {quest} {seos} (REQUEST_FOR_STATEMENT)
```

In both cases the translation of the prepositional phrase (PP) *bei Ihnen* can only be decided when the context is taken into account: while in Example 1(a) the PP has a locational interpretation Example 1(b) refers to the availability of the dialogue partner at the time indicated. Therefore the first occurrence of the PP has to be translated with *at your place*; the second occurrence corresponds to *Would that suit you?*. Whether the preposition has a locational meaning can in our example be decided on grounds of the context. In Example 1(a) the negotiation of a time has been concluded and the dialogue is already in the closing phase, where negotiations usually concern locations. The utterance preceding the ambiguous sentence introduces a location as potential focus when the specification of a place for a meeting is requested. In Example 1(b) the dialogue is still in a phase where a time is being negotiated so that the locational reading of the PP can be excluded. From the discussion of the Examples we can conclude that for the VERBMOBIL system context has to contain at least (a) a representation of the content included in an utterance, (b) a description of the dialogue act standing behind an utterance, (c) a representation of the dialogue phase to which an utterance belongs.

### Generation

After the determination of the translational equivalents the generation component of VERBMOBIL has to produce an utterance which fulfills the same function as the corresponding utterance in the source language. The use of contextual information serves as one means to establish this functional equivalence. The discussion of the following dialogue fragment (see Example 2) clarifies this principle:

**Example 2** – utterance mps1_1_15:

```
schoen dann machen wir 's so {period}
```

Various contextual aspects can contribute to determine the correct target language expression for this utterance: without any information which illocutionary force stands behind this utterance, i.e. whether the sentence is a CONFIRMATION (possibly following an ACCEPTANCE made previously by the other dialogue participant) or a REQUEST_FOR_STATEMENT (possibly following a SUGGESTION) it is difficult to decide between the two possible translations *Okay, then let's do it like that.* or *Should we do it like that?*. Also, a contextual knowledge is necessary to identify the referent for the pragmatic anaphora `so` (engl.: *like that*). Knowledge of this type can be used to prevent the generation of ambiguous expressions where the original expression in the source language is unique[1].

### Key Word Spotting

In order to achieve more reliable results the key word spotter is trained on the recognition of a limited set of key phrases. Since the set of most significant key phrases differs for every dialogue act, various sets are activated depending on the dialogue state which is expected next. These expectations are determined by the dialogue component. In order to do so previous dialogue steps have to be taken into account.
To clarify this point we discuss the processing of the dialogue fragment given in Example 3, where the utterance `DE6` is followed by means of the key word spotter only[2].

**Example 3:**

```
DE4: #oh ja, gut, nach meinem Terminkalender <Pause>, wie waers im
     Oktober?# (SUGGESTION)
VM5: just lookin at my diary, I would suggest October. (SUGGESTION)
DE6: <Pause> I propose from  Tuesday the fifth/- <Pause> no,
     Tuesday the fourth to Saturday the eighth, those five days?
     (SUGGESTION)
```

While processing the dialogue a statistical subcomponent of the dialogue module predicts the most-likely follow-up dialogue acts on the basis of what has been said before. Allowing two predictions for each dialogue step we get the following results for the above dialogue fragment:

[1] It is not a goal of VERBMOBIL to disambiguate expressions in the source language; instead, the preservation of ambiguity across source and target language is aimed at.

[2] `DE` indicates the German speaker, `VM` the translation provided by VERBMOBIL, `EL` the English speaker and `#` the request for a translation.

```
EL3: INITIALISATION Prediction: (SUGGESTION REQUEST_FOR_SUGGESTION)
DE4: SUGGESTION Prediction: (ACCEPT SUGGESTION)
DE6: SUGGESTION Prediction: (ACCEPT SUGGESTION)
```

We find that after processing contribution `DE4` the two dialogue acts ACCEPT and SUGGESTION can be expected. The key word spotter accordingly scans the input for phrases which typically occur in utterances which belong to these two classes. Since *propose* serves as key word for the dialogue act SUGGESTION this act is recognized correctly[3].

In future versions of the VERBMOBIL system also other components, as e.g. analysis and synthesis of prosody, speech recognition and parsing, will exploit contextual information in order to increase robustness and efficiency. These issues are subject to future research.

# 3 Our Notion of Context

As has been shown in the previous section our notion of context is mostly derived from practical requirements in system development. Many theoretical linguistic models of context, instead, try to collect and organize all factors which characterize the framework / the situation in which a discourse takes place. These models are mostly guided by empirical research examining dialogues and texts from a linguistic / psychological / sociological point of view without taking representational and implementational considerations into account. In this section we will discuss how a practical system-oriented approach can be related to such broader theoretical approaches for context.

As examples for rather broad theoretical models of context we take Halliday's specification of the *context of situation* (see [Halliday, 1985]) and the context model developed by Bunt (see [Bunt, 1994]). The former model has been developed within the framework of Systemic Functional Linguistics. It describes all the factors which constrain the choice of linguistic means in the production of a discourse (i.e. a dialogue or a text). Halliday distinguishes three major classes of features: (1) *field*, which includes a specification of the domain around which the discourse evolves, (2) *tenor*, which subsumes various features related to the discourse participants (e.g. beliefs, knowledge and attitudes of the discourse participants and their social relationships), and (3)

[3] For the set of key words which characterize dialogue acts see [Mast, 1995].

*mode*, which describes the role the discourse is playing, the aims which are followed by means of the discourse and the modalities which are used.

While this model is very remote from the intuitive notion of context, Bunt proposes a model which subsumes both *global aspects* which – in analogy to Halliday's model – describe the environment which influences the realization of a discourse and *local aspects* which describe how the discourse develops. While global aspects tend to remain constant throughout the discourse, local aspects change as the discourse evolves. Like Halliday, Bunt distinguishes context along various levels (i.e. *dimensions*): (1) the *linguistic context*, which represents the surrounding linguistic material; (2) the *semantic context*, which circumscribes aspects related to the underlying task; (3) the *physical context*, which subsumes the physical circumstances under which the discourse takes place; (4) the *social context*, which describes the social potential and the competence of the discourse participants in the given situation; and, finally, (5) the *cognitive context*, which subsumes the beliefs, intentions and plans of the discourse participants.

If we try to position the contextual features required in VERBMOBIL using the models discussed above we can draw the following conclusions: (1) for the time being the global aspects of context do not have to be considered since they do not change significantly within the given application; in particular, the VERBMOBIL system is supposed to be trained on its owner, so that both the description of this user and also the specification of the domain remain constant. The VERBMOBIL application can be described as a specific instantiation of a context of situation. (2) the contextual information required by VERBMOBIL subcomponents can be related to Bunt's local, dynamic aspects of context. The correspondencies between our types of contextual information and the various dimensions will be detailed in section 4.

In the following section we will give a description of the contextual representation as developed and implemented within VERBMOBIL. We use a fragment of a typical appointment scheduling dialogue to further clarify our representation and the use of context. The dialogue is fully processed by the current VERBMOBIL prototype.

# 4 The Multi-Dimensional Context Model in VERBMOBIL

In this section we show how the contextual information as required by the various subcomponents of VERBMOBIL are represented. In the design of the dialogue memory which contains the representation of the context we followed the principle of modularity: due to different function and use within the system various types of contextual information have been distinguished and represented separately. In the following we discuss the three types of contextual information, for each type giving an example for the representation of a sample dialogue fragment (see Example 4).

**Example 4:**

```
DE4: #oh ja, gut, nach meinem Terminkalender <Pause>, wie waers im
     Oktober?# (SUGGESTION)
VM5: just lookin at my diary, I would suggest October. (SUGGESTION)
DE6: <Pause> I propose from  Tuesday the fourth to Saturday the
     eighth, those five days? (SUGGESTION)
EL7: oh, that's too bad, I'm not free right then. (REJECT) I could
     fit it into my schedule the week after, from Saturday to
     Thursday, the thirteenth. (SUGGESTION)
```

## 4.1 The Intentional Structure

The intentional structure of the dialogue represents the single intentions which have to be followed to schedule a business meeting. The units of representation are intentions which stand behind the dialogue contributions. They exist along various degrees of abstraction: the top level goal APPOINTMENT_NEGOTIATION can be decomposed into a conventionalized sequence of finer-grained goals, each of which stands for a dialogue phase. Such goals can be decomposed into intentions representing negotiation strategies, like e.g. active or re-active negotiation. The lowest level of abstraction consists of single dialogue acts.

The elements of these various levels of abstraction construct a tree-like representation of the intentions included in the dialogue. While the root of such a hierarchical structure represents the top-level goal of the dialogue the leaves correspond to single dialogue acts. Figure 2 shows the representation

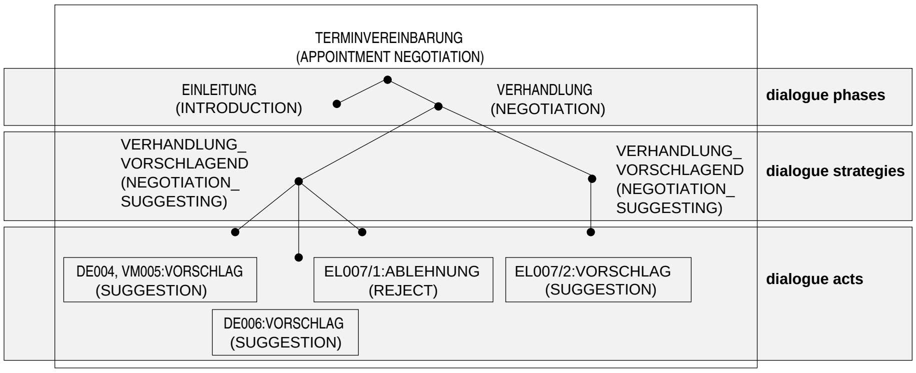


Figure 2: The intentional structure of the sample dialogue.

of the intentional structure of the dialogue given above. It represents two phases of the negotiation dialogue, the INTRODUCTION and the beginning of the NEGOTIATION. The dialogue partner DE follows an active negotiation strategy (NEGOTIATION_SUGGESTING) making suggestions (DE4 and DE6) thereby triggering a reaction by the other user (EL7/1). The active role then is taken over by the English speaking user who makes a counterproposal (EL7/2).

In terms of Bunt's model our intentional structure can best be compared to the local aspect of the social and the cognitive dimension: the intentional structure describes for any dialogue stage which of the possible communicative acts has been chosen; it also monitors the plans and intentions of the users.

In VERBMOBIL the intentional structure is exploited to find the functionally correct translational equivalents. This structure is accessed both by the transfer and the generation component (see section 2.2). Also, dialogue acts are predicted on the basis of the dialogue acts used previously[4].

[4]Dialogue act predictions are made on the basis of a statistical model which has been trained on the dialogue act sequences as occurring in our corpus. The statistical model is updated as the discourse proceeds. For more details see [Reithinger and Maier, 1995].

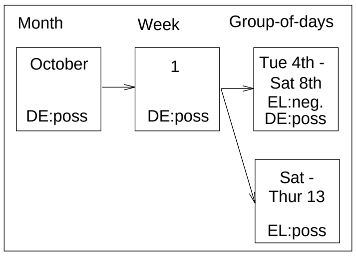


Figure 3: The thematic structure of the sample dialogue.

## 4.2 The Thematic Structure

The thematic structure of the dialogue memory monitors the progress of the dialogue with respect to its propositional content. In particular, it represents the temporal objects being mentioned together with their evaluation by the speakers (time frame possible – POSS; time frame impossible – NEG). Temporal objects are instantiations of concepts like months, weeks, days, and refinements thereof. The temporal objects are related by means of links which indicate that the temporal object at the links' target is a refinement of the object represented as it source. That way a hierarchical structure of time frames is constructed. In terms of Bunt's model this structure can best be described as the local aspect of the semantic dimension of context, which contains task-specific discourse material. In figure 3 we show the thematic structure for Example 4.

First, speaker DE introduces October as a possible month for an appointment. In utterance DE6 this time frame is specified further by giving a time span consisting of a sequence of days. From this specification the proposed week can be inferred. While the date has been proposed by speaker DE, it has been rejected by EL who then makes a counterproposal, which leads to the creation of a new temporal object, which is also an instance of GROUP-OF-DAYS.

The hierarchical representation of temporal objects has a number of significant advantages: it can be exploited to determine which objects are referentially accessible, which objects can still be moved into focus and are therefore still open for negotiation. Also, information about the possibility to meet within a certain time frame can be percolated within the thematic structure. If a day is marked as possible for one dialogue participant, this

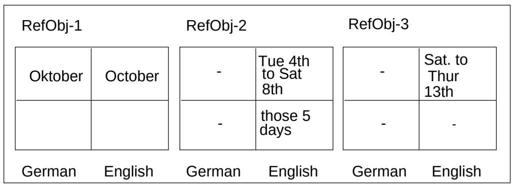


Figure 4: The referential structure of the sample dialogue.

feature also holds for all superordinate time frames, i.e. for the week, the month, etc. of which the day is part (*upwards inheritance*). An example is shown in figure 3, where the proposal for a group of days, `Saturday, until Thursday the 13th`, is inherited upwards so that the possibility information is inserted in both the superordinated instances. Similarly, the information that a certain month is not a viable option can be handed downwards to all subordinated time frames mentioned before (*downwards inheritance*). Information of type *poss* initiates upward inheritance while *neg* calls for downward inheritance.

Within VERBMOBIL the thematic structure is used to support inferences, which depend on the availability of propositional information in relation with the dialogue act it serves. This information is required both by the transfer and the generation component in order to constrain their decisions.

### 4.3 The Referential Structure

The referential structure represents the various realizations which have been used to linguistically express the same object. The realizations of one conceptual entity are stored within a so-called *Referential Object*. Within these objects the linguistic realizations are distinguished with respect to the language in which they have been expressed (i.e. German and English). In particular, referential objects include first mentionings of objects together with their lexical variations or referential expressions used in later stages of the dialogue. In Bunt's model such information types are part of the local aspect of the linguistic dimension.

Figure 4, which shows the referential structure of the sample dialogue, contains three referential objects, one representing realizations of the month proposed and the two others covering group-of-days being negotiated.

The availability of the referential structure is crucial for the generation component; by being able to identify which linguistic expressions refer to the same conceptual entity the generation component has a basis for decisions at the level of microplanning, as e.g. for the generation of anaphoric expressions or for lexical variation.

### 4.4 Combining the Structures into a Multi-Layered Model

In order to gain a context model which fulfills all the requirements posed by the various subcomponents of VERBMOBIL we combined the three types of structure into a multi-functional, multi-dimensional representation.

By combining the three types of structures – two tree-like structure types and a list of referential objects – we achieve a complex net-like structure. The complete contextual representation of our sample dialogue is given in figure 5.

The combination of the three dimensions allows to infer at any point (1) which time frames have been negotiated; (2) which linguistic means have been used for expressing these objects; (3) which of the dialogue participants has been mentioning the time frame under consideration (4) which attitude this participant had with respect to that time frame; (5) which utterances the time frame has been used in; and (6) which dialogue stage an utterance belongs to.

## 5 Discussion

### 5.1 The Computational Approach

The dialogue component, which is responsible for the incremental construction of the dialogue memory, consists of three submodules:

*a statistics module:*
this subcomponent, which has been trained on a corpus of appointment scheduling dialogues annotated with dialogue acts, computes predictions for follow-up dialogue acts; these predictions are used e.g. to constrain the set of key words to be expected;

*a finite state machine (FSM):*
this subcomponent provides an efficient implementation of our dialogue model

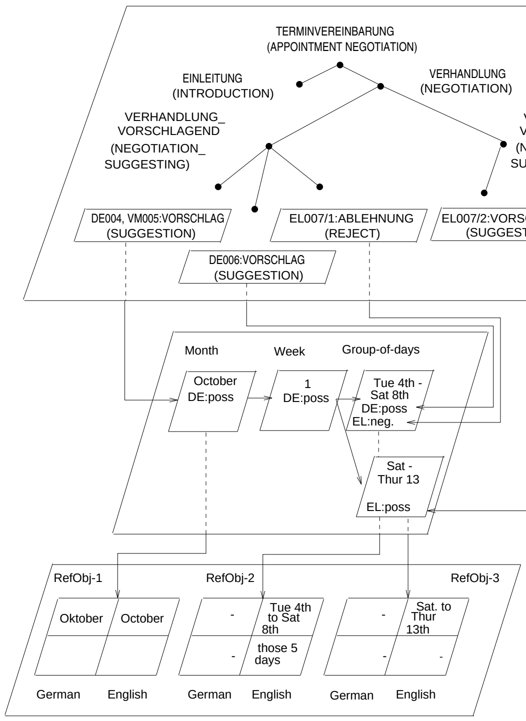


Figure 5: Multi-dimensional contextual representation of the sample dialogue.

which describes the potential sequences of dialogue acts that can be expected in appointment scheduling dialogues. The FSM checks an incoming dialogue act for compatibility with this dialogue model.

*a plan recognizer:*
this module incorporates a knowledge-intensive implementation of the dialogue model. For processing a dialogue, *plan operators* have been defined which are specialized to treat specific events. Plan operator application depends on contextual and pragmatic conditions. Once a plan operator is applied it can initiate follow-up actions, as e.g. an update of parts of the dialogue memory. The plan operators cover various levels of abstraction – at the lowest level plan operators are specialized for the treatment of one single dialogue act.

It has to be noted here that the various types of dialogue acts can be distinguished with respect to their effect on the dialogue memory: some dialogue acts, as e.g. GREETING, INTRODUCTION, THANKS only have a social function and do not contribute to the negotiation of an appointment as such. Therefore only the intentional structure has to be updated with their occurrence. Other dialogue acts typically contain information which concerns dates being negotiated: among these dialogue acts are SUGGESTION, ACCEPT and REJECT. These dialogue acts therefore have to be incorporated both into the intentional and the thematic structure. An update of the referential structure can be induced by all dialogue acts: each utterance potentially contains linguistic material which serves as referent in subsequent steps of the dialogue.

The decision to represent the three types of contextual information separately and to connect them by means of well-defined links has various comoutational advantages: the single representational levels which consist either of tree-like structures or sets of objects can be treated using efficient, standard computational techniques. Nevertheless it is easy to access information of other dimensions by following the connecting inter-dimensional links. Also, the separation of information into various levels allows for highlighting specific structure types if a component does not need the full complexity of our representation.

## 5.2 Comparable Approaches

The identification of the various types of knowledge included in discourse and the adequate representation of this information is of increasing importance in computational linguistics. Among one of the first approaches to examine this question is the work reported in [Grosz and Sidner, 1986]. The authors propose a three-partite separation of discourse knowledge into intentional, attentional and linguistic structure, which can be roughly compared to the intentional, thematic and referential levels of our context representation. In the approach developed by Grosz and Sidner the linguistic structure is concerned with the segmentation of the discourse into distinct discourse segments, the intentional structure refers to purpose and goal which stand behind such a discourse segment, and the attentional structure, finally, addresses the objects the reader has in focus when processing a discourse segment. Each discourse can be described along those three dimensions.

Especially in NL generation approaches like the one by Grosz and Sidner have received considerable attention. They have been applied both for the representation of knowledge necessary for planning and realization of texts and also for the representation of the material included (see e.g. [Moore and Pollack, 1993, Hovy, 1993, Maier, 1993]). We exploit a similar approach for the construction of the dialogue memory which contains three separate types; we also provide links to model interdependencies between the various layers.

# 6 Conclusion

In this paper we presented a multi-dimensional model of context which is part of an integrated speech translation system. We showed how our model has been derived from the requirements defined by the various system components. Adopting a computational, application-oriented point of view we tried to provide answers for the following questions:

- *What is context?*
  Within the VERBMOBIL application we define context as a refinement of general, multi-dimensional context models. We identified three structure types each representing one dimension: intentional, thematic and referential structures (see section 4).

- *In which way does context affect interpretation of natural language utterances?*

In section 2 we showed various decisions within transfer, generation and key word spotting which depend on the availability of the above-mentioned types of contextual information.

- *Should the final interpretation of natural language be decontextualized when stored in a knowledge base?*
  As has also been shown in section 2, it is a crucial requirement that the context representation includes both a notion of *progress* to monitor how far the dialogue is advanced and a representation of *referential chains* to allow for the expression of object identity. These requirements can only be fulfilled by means of a *contextualized representation*.

Future versions of our context model will be extended to handle broader domains, as e.g. travel planning. We expect that an extension of our application leads to changes in the design of the thematic structure and in the mechanisms for its construction.

# References

[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.

[Bunt, 1981] Harry C. Bunt. Rules for the Interpretation, Evaluation and Generation of Dialogue Acts. In *IPO Annual Progress Report 16*, pages 99–107, Tilburg University, 1981.

[Bunt, 1994] Harry C. Bunt. Context and Dialogue Control. *Think*, 3:19–31, May 1994.

[Grosz and Sidner, 1986] Barbara J. Grosz and Candace L. Sidner. Attention, intentions and the structure of discourse. *Journal of Computational Linguistics*, 12(3), 1986.

[Halliday, 1985] M.A.K. Halliday. *An Introduction to Functional Grammar.* Edward Arnold, 1985.

[Hovy, 1993] Eduard H. Hovy. In defense of syntax: Informational, intentional, and rhetorical structures in discourse. In Owen Rambow, editor, *Proceedings of the ACL-93 workshop on Intentionality and Structure in Discourse Relations*, Columbus, Ohio, June 1993.

[Maier, 1993] Elisabeth A. Maier. The representation of interdependencies between communicative goals and rhetorical relations in the framework of multimedia document generation. In Owen Rambow, editor, *Proceedings of the ACL-93 workshop on Intentionality and Structure in Discourse Relations*, Columbus, Ohio, June 1993.

[Maier, 1994] Elisabeth Maier. Dialogmodellierung in VERBMOBIL – Festlegung der Sprechhandlungen für den Demonstrator. Technical Report Verbmobil–Memo 31, DFKI Saarbrücken, Juli 1994.

[Mast *et al.*, 1992] M. Mast, R. Kompe, F. Kummert, H. Niemann, and E. Nöth. The dialog module of the speech recognition and dialog system EVAR. In *Proceedings of the International Conference on Spoken Language Processing, Banff, Canada*, pages 1573–1576, 1992.

[Mast, 1995] Marion Mast. Schlüsselwörter zur Detektion von Diskontinuitäten und Sprechhandlungen. Technical Report Verbmobil Memo Nr. 57, Friedrich-Alexander-Universität, Erlangen-Nürnberg, January 1995.

[Moore and Pollack, 1993] Johanna D. Moore and Martha E. Pollack. A problem for RST: the need for multi-level discourse analysis. *Computational Linguistics*, pages 537–544, 1993.

[Reithinger and Maier, 1995] Norbert Reithinger and Elisabeth Maier. Utilizing Statistical Speech Act Processing in VERBMOBIL. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL-95)*, Cambridge, MA, June 1995. forthcoming.