gefördert vom
Bundesministerium
für Forschung
und Technologie

Verbmobil
Verbundvorhaben

Treatment of Incomplete Dialogues in a Speech-to-Speech Translation System

Norbert Reithinger, Elisabeth Maier, and Jan Alexandersson

DFKI GmbH

Report 79
April 1995

April 1995

Norbert Reithinger, Elisabeth Maier, and Jan Alexandersson

DFKI GmbH
Stuhlsatzenhausweg 3
66123 Saarbrücken

Tel.: (0681) 302 - 5346
e-mail: `reithinger@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.

Treatment of Incomplete Dialogues in a Speech-to-Speech Translation System

Norbert Reithinger, Elisabeth Maier, Jan Alexandersson*

DFKI GmbH, Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
{reithinger,maier,alexandersson}@dfki.uni-sb.de

Abstract

For the speech-to-speech translation system VERBMOBIL the dialogue component has the task of providing contextual information for other VERBMOBIL subcomponents and to follow the flow of the dialogue. Since VERBMOBIL operates on demand, only parts of the dialogue are processed using syntactic and semantic knowledge. Therefore, robustness with respect to incomplete structures is a basic requirement. We show the performance of the statistic prediction module and the planning module under such conditions.

1 Introduction

VERBMOBIL combines the two key technologies speech processing and machine translation. The goal of this project is the development of a prototype for the translation of spoken dialogues between two persons who want to find a date for a business meeting (for more detail on the objectives of VERBMOBIL see [Wahlster, 1993]). A special characteristic of VERBMOBIL is that both participants are assumed to have at least a passive knowledge of English which is used as intermediate language. Translations are produced *on demand* so that only parts of the dialogue are processed. If VERBMOBIL is inactive, shallow processing by a keyword spotter takes place which allows the system to follow the dialogue at least partially.

*This work was funded by the German Federal Ministry for Education, Research, Science and Technology (BMBF) in the framework of the VERBMOBIL project under grant 01IV101K/1. The responsibility for the contents of this study lies with the authors.

This report also appears in the Proceedings of the ESCA Workshop on Spoken Dialogue Systems, 30.5.-2.6. 1995, Hanstholm, Denmark.

Dialogue processing in VERBMOBIL differs from systems like SUNDIAL and EVAR [Andry, 1992, Niedermair, 1992, Mast *et al.*, 1992] in two important points: (1) VERBMOBIL *mediates* the dialogue between two human dialogue participants; the system is not a participant of its own, i.e. it does not control the dialogue as it happens e.g. in the flight scheduling scenario of SUNDIAL; (2) VERBMOBIL processes maximally 50 % of the dialogue contributions in depth, i.e. when the 'owner' of VERBMOBIL speaks German only. The rest of the dialogue can only be followed by a keyword spotter.

2 The Dialogue Component – an Overview

The dialogue component within VERBMOBIL has four major tasks:

- to support the speech recognition and the linguistic analysis when processing the speech signal. Top-down predictions made by the dialogue modul restrict the search space of other VERBMOBIL components [Young *et al.*, 1989, Andry, 1992].
- to provide contextual information for other VERBMOBIL components [Ripplinger and Caroli, 1994, LuperFoy and Rich, 1992].
- to follow the dialogue when VERBMOBIL is off-line by using a keyword spotter. This device scans the input for a small set of predetermined words which are characteristic for certain stages of the dialogue. From the recognized words, the most probable speech act is inferred. Keyword spotting is currently rather unreliable.
- to control clarification dialogues between VERBMOBIL and its users.

The abovementioned requirements cannot be met when only a single method of processing is used. Therefore we chose a hybrid 3-layered approach (see fig. 1) where the layers differ with respect to the type of knowledge they use.

We share with other approaches (e.g. [Niedermair, 1992]) the assumption that the intentions of dialogues can be described by speech acts. From the analysis of the VERBMOBIL corpus, we defined 18 speech acts for the domain of appointment scheduling (for a detailed description see [Maier, 1994]). They are shown as part of the dialogue model in figure 2. The main part at the top describes the expected sequence of speech acts, while the subnetwork at the lower left side describes spech acts for digressions which can occur at every stage in the dialogue.

The three subcomponents of the dialogue module are

A Statistic Module The task of the statistic module is the prediction of the following speech act, using knowledge about speech act frequencies in our training corpus. [Reithinger, 1995] shows a prediction accuracy of approximately 80% for the three best predictions. Training and test data

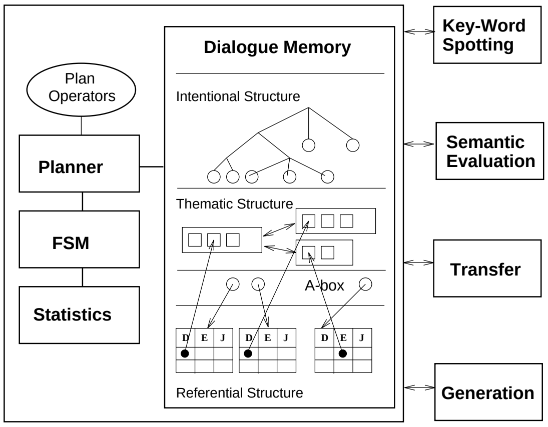

Figure 1: Architecture of the dialogue module

are selected from currently over 200 annotated dialogues of the VERBMOBIL corpus.

A Finite State Machine (FSM) The finite state machine describes the sequence of speech acts that are admissible in a standard appointment scheduling dialogue as described by our dialogue model. The FSM checks the ongoing dialogue whether it is compatible with the model.

A Planner The hierarchical left-to-right planner processes speech acts making extensive use of contextual knowledge. Since processing is sensitive to inconsistencies, robustness and backup-strategies are the most important features of this component.

Both the finite state machine and the planner are used to implement the possible speech act sequences as defined in the dialogue model. While incomplete dialogues are not part of the language defined by the FSM, the planner has specific strategies to cope with such phenomena (see section 5).

During dialogue processing tree-like structures are built which mirror the structure of the dialogue as pursued so far. The dialogue memory consists of three structural layers: (1) an intentional structure representing dialogue phases and speech acts as occurring in the dialogue, (2) a thematic structure representing the dates being negotiated, and (3) a referential structure keeping track of lexical realizations.

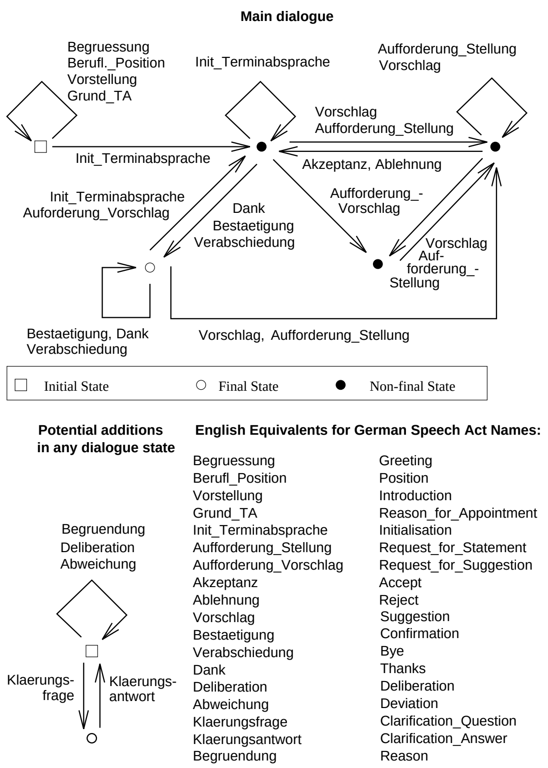

Figure 2: A dialogue model for the description of appointment scheduling dialogues

3 Predicting Incomplete Dialogues

The prediction algorithm is based on the dynamic interpolation formula of [Jellinek, 1990], adapted for speech act processing [Reithinger, 1995]. Using conditional speech act frequencies, we compute the most probable follow up speech acts. These predictions are e.g. used by the analysis components to narrow down the search space.

Since VERBMOBIL currently tracks only 50% of the dialogue, we compared the performance of the prediction algorithm when applied to complete and incomplete dialogues. Figure 3 shows hit rates of the predictions for two experiments. TS1 uses 52 complete dialogues as training material and 81 as test data. TS2 uses only the contributions of one dialogue partner of the same 52 dialogue as training material and the contributions of one speaker of 177 dialogues as test data. As can be seen, prediction accuracy decreases about 5% to 7% for the incomplete dialogues.

Pred.	*TS1*	*TS2*
1	44.24%	39.99 %
2	66.47 %	61.96 %
3	81.46 %	74.44 %

Figure 3: Prediction hit rates for complete and incomplete dialogues

4 Plan-based Treatment Dialogue Processing

To incorporate constraints and to allow decisions to trigger follow-up actions a plan-based approach has been chosen as one layer of dialogue processing. Planning proceeds in a top-down fashion, i.e. each plan operator is characterized by a goal which can be decomposed into subgoals. These subgoals have to be achieved individually and in a prespecified sequence in order for the whole plan to be fulfilled. Our top-level dialogue goal, for example, is decomposed into three subgoals each of which is responsible for the treatment of one dialogue segment: the introductory phase, the negotiation phase and the closing phase. Iterated application of plan operators can also be specified. In our hierarchy of plan operators the leaves, i.e. the most specific operators, correspond to the individual speech acts. The application of plan operators is mainly controlled by pragmatic and contextual constraints. Among these constraints are, for example, features related to the discourse participants (acquaintance, level of expertise etc.) and features related to the dialogue history (e.g. the occurrence of a certain speech act in the preceding context). Also, our plan operators contain an actions slot, where operations which are triggered after a successful fulfillment of the subgoals are specified, as e.g. the construction of a dialogue memory.

5 Treatment of Unexpected Input

One of the main tasks of the planner is to cope with incomplete knowledge or with speech act information that does not coincide with a given dialogue state. In order to recover from such a state repair techniques have been incorporated into the dialogue planner. We currently use both statistical and knowledge-based techniques exploiting the statistical model and the dialogue memory.

The dialogue model specified in the networks models all speech act sequences that can be usually expected in an appointment scheduling dialogue. In case unexpected input occurs repair techniques have to be provided to recover from such a state and to continue processing the dialogue in the best possible way. The treatment of these cases is the task of the dialogue planner.

Two different repair techniques have been developed to cope with cases where unexpected input occurs, i.e. when the incoming speech acts are not compatible with the dialogue model. The first method exploits knowledge available from the statistical component; the second method relies on a subset of plan operators, the so-called *repair-operators*.

The first method for error recovery is based on the hypothesis that the attribution of only one speech act to a given utterance is insufficient and that an utterance has more than one speech act reading. If an incompatible speech act is encountered the statistical component is looked up in order to find out whether a statistically relevant speech act exists which is able to bridge the previous and the current (incompatible) speech act. If such a speech act can be found and if the insertion of this speech act renders the dialogue compatible a multiple reading is proposed and the speech act is inserted.

```
Planner: -- Processing AUFFORDERUNG_VORSCHLAG
Planner: -- Processing VORSCHLAG
Planner: -- Processing BESTAETIGUNG

Trying to find a speech act to bridge
VORSCHLAG and BESTAETIGUNG ...

Possible insertions and their scores:
((AKZEPTANZ 62419))

Testing AKZEPTANZ for compatibility with surrounding
speech acts ...

The current speech act BESTAETIGUNG has an additional
reading of AKZEPTANZ:
BESTAETIGUNG -> AKZEPTANZ BESTAETIGUNG !

        Warning -- Repairing...
Planner: -- Processing VERABSCHIEDUNG
```

Figure 4: Statistical Repair in a Sample Dialogue.

To clarify this method we give an example for repair as occurring in a dialogue of our corpus (see figure 4). The trace shows how the incompatible speech acts VORSCHLAG (proposal) and BESTAETIGUNG (agreement) can be connected inserting a speech act of type AKZEPTANZ (acceptance). Since in our corpus frequent cooccurrences of the speech acts AKZEPTANZ and BESTAETIGUNG in one utterance can be found while there are no examples for a simultaneous occurrence of VORSCHLAG and AKZEPTANZ we give BESTAETIGUNG a multiple reading.

In cases where no statistical solution is possible plan-based repair is used. When an unexpected speech act occurs a plan operator is activated which distinguishes various types of repair. Depending on the type of the incoming speech act specialized repair operators are used. The simplest case covers speech acts which can appear at any point of the dialogue, as e.g. DELIBERATION and clarification dialogues (KLAERUNGSFRAGE and KLAERUNGSANTWORT). We handle these speech acts by means of repair in order to make the planning process more

```
Planner: -- Processing VORSCHLAG
Planner: -- Processing KLAERUNGSFRAGE
          Warning -- Repairing...

Planner: -- Processing KLAERUNGSANTWORT
Planner: -- Processing DELIBERATION
Planner: -- Processing VORSCHLAG
```

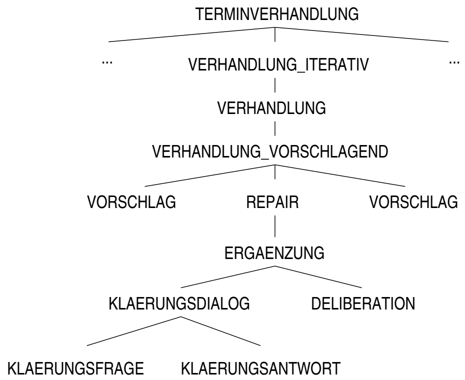

Figure 5: Plan-based Repair in a Sample Dialogue.

efficient: Since these speech acts can occur at any point in the dialogue the planner in the worst case has to test for every new utterance whether it is one of the speech acts which indicate a digression. To prevent this the occurrence of one of these speech acts is treated as an unforeseen event which triggers the REPAIR operator.

As consequence of fulfilling one of the repair subgoals REPAIR activates the plan operator ERGAENZUNG. This operator finally is responsible for the treatment of all types of deviations which may also occur iteratively. This plan operator also suspends the planning process to insert speech acts of that type into the intentional structure and then resumes the dialogue. In figure 5 we give a partial trace and the corresponding intentional structure from another sample dialogue where such a repair case occurs.

In all other cases, for example if a speech act of the introduction phase occurs during the closing phase, it is assumed that the current phase (i.e. the closing phase) has been terminated irregularly and that the dialogue proceeds from the current speech act.

6 Conclusion

In this paper we described a three-layered approach for processing dialogues in a speech-to-speech translation system. In contrast to conventional dialogue modules our system has to be able to cope with inconsistent and incomplete input. To this end specific repair techniques have been developed which allow the system to follow the dialogue even when unexpected input occurs. Statistical and knowledge-based methods are combined in order to guarantee both efficient and robust dialogue processing. Future developments concern the treatment of multiple speech acts within one utterance, the addition of statistical weights to the plan operators in order to make planning more efficient, and the extension of the scenario to the domain of travel planning.

References

References

[Andry, 1992] Francois Andry. Static and dynamic predictions: a method to improve speech understanding in cooperative dialogues. In *Proceedings of the International Conference on Spoken Language Processing*, pages 639–642, Banff, October 1992.

[Jellinek, 1990] Fred Jellinek. Self-Organized Language Modeling for Speech Recognition. In A. Waibel and K.-F. Lee, editors, *Readings in Speech Recognition*, pages 450–506. Morgan Kaufmann, 1990.

[LuperFoy and Rich, 1992] Susan LuperFoy and Elaine A. Rich. A three tiered discourse representation framework for computational discourse processing. Technical report, MITRE Corporation and MCC, 1992.

[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.

[Niedermair, 1992] Gerhard Th. Niedermair. Linguistic Modelling in the Context of Oral Dialogue. In *Proceedings of International Conference on Spoken Language Processing (ICSLP'92)*, volume 1, pages 635–638, Banff, Canada, 1992.

[Reithinger, 1995] Norbert Reithinger. Some Experiments in Speech Act Prediction. In *AAAI 95 Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, 1995.

[Ripplinger and Caroli, 1994] Bärbel Ripplinger and Folker Caroli. Konzept-basierte Übersetzung in Verbmobil. Technical report, IAI Saarbrücken, May 1994.

[Wahlster, 1993] Wolfgang Wahlster. Verbmobil – Translation of face-to-face dialogs. In *Proceedings of the Fourth Machine Translation Summit*, Kobe, Japan, July 1993.

[Young *et al.*, 1989] Sheryl R. Young, Wayne H. Ward, and Alexander G. Hauptmann. Layering predictions: flexible use of dialog expectation in speech recognition. In *Proceedings of IJCAI '89, Detroit*, 1989.