Gefördert vom

bmb+f

Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Verbmobil

Verbundvorhaben

Insights into the Dialogue Processing of VERBMOBIL

Jan Alexandersson
Norbert Reithinger
Elisabeth Maier

DFKI GmbH

Vm

REPORT 191
März 1997

März 1997

Jan Alexandersson
Norbert Reithinger
Elisabeth Maier

DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany

Tel.: (0681) 302 - 5347/5346
Fax: (0681) 302 - 5341
e-mail: {alexandersson,reithinger,maier}@dfki.un-sb.de

Gehört zum Antragsabschnitt: 10

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01IV101K/1 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei den Autoren.

ISSN 0949-6084

Zusammenfassung

We present the dialogue module of the speech-to-speech translation system VERBMOBIL. We follow the approach that the solution to dialogue processing in a mediating scenario can not depend on a single constrained processing tool, but on a combination of several simple, efficient, and robust components. We show how our solution to dialogue processing works when applied to real data, and give some examples where our module contributes to the correct translation from German to English.

1 Introduction[1]

The implemented research prototype of the speech-to-speech translation system VERBMOBIL (Wahlster, 1993; Bub and Schwinn, 1996) consists of more than 40 modules for both speech and linguistic processing. The central storage for dialogue information within the overall system is the dialogue module that exchanges data with 15 of the other modules.

Basic notions within VERBMOBIL are *turns* and *utterances*. A turn is defined as one contribution of a dialogue participant. Each turn divides into utterances that sometimes resemble clauses as defined in a traditional grammar. However, since we deal exclusively with spoken, unconstrained contributions, utterances are sometimes just pieces of linguistic material.

For the dialogue module, the most important dialogue related information extracted for each utterance is the so called dialogue act (Jekat et al., 1995). Some dialogue acts describe solely the illocutionary force, while other more domain specific ones describe additionally aspects of the propositional content of an utterance.

Prior to the selection of the dialogue acts, we analyzed dialogues from VERBMOBIL's corpus of spoken and transliterated scheduling dialogues. More than 500 of them have been annotated with dialogue related information and serve as the empirical foundation of our work.

Throughout this paper we will refer to the example dialogue partly shown in figure 1. The translations are as the deep processing line of VERBMOBIL provides them. We also annotated the utterances with the dialogue acts as determined by the semantic evaluation module. ``//'' shows where utterance boundaries were determined.

We start with a brief introduction to dialogue processing in the VERBMOBIL setting. Section 3 introduces the basic data structures followed by two sections describing some of the tasks which are carried out within the dialogue module. Before the concluding remarks in section 8, we discuss aspects of robustness and compare our approach to other systems.

[1]This paper is a reprint from the Proceedings of the Fifth Conference on Applied Natural Language Processing, 31 March - 3 April, 1997, Washington, D.C.

2 Introduction to Dialogue Processing in VERBMOBIL

In contrast to many other NL-systems, the VERBMOBIL system is mediating a dialogue between two persons. No restrictions are put on the locutors, except for the limitation to stick to the approx. 2500 words VERBMOBIL recognizes. Therefore, VERBMOBIL and especially its dialogue component has to follow the dialogue in any direction. In addition, the dialogue module is faced with incomplete and incorrect input, and sometimes even gaps.

When designing a component for such a scenario, we have chosen not to use one big constrained processing tool. Instead, we have selected a combination of several simple and efficient approaches, which together form a robust and efficient processing platform.

As an effect of the mediating scenario, our module cannot serve as a "dialogue controller" like in man-machine dialogues. The only exception is when clarification dialogues are necessary between VERBMOBIL and a user.

Due to its role as information server in the overall VERBMOBIL system, we started early in the project to collect requirements from other components in the system. The result can be divided into three subtasks:

- we allow for other components to *store* and *retrieve* context information.
- we draw *inferences* on the basis of our input.
- we *predict* what is going to happen next.

Moreover, within VERBMOBIL there are different processing tracks: parallel to the deep, linguistic based processing, different shallow processing modules also enter information into, and retrieve it from, the dialogue module. The data from these parallel tracks must be consistently stored and made accessible in a uniform manner.

Figure 2 shows a screen dump of the graphical user interface of our component while processing the example dialogue. In the upper left corner we see the structures of the *dialogue sequence memory*, where the middle right row represents turns, and the left and right rows represent utterances as segmented by different analysis components. The upper right part shows the *intentional structure* built by the plan recognizer. Our module contains two instances of a *finite state automaton*. The one in the lower left corner is used for performing clarification dialogues, and the other for visualization purposes (see section 7). The *thematic structure* representing temporal expressions is displayed in the lower right corner.

A01: Tag // Herr Scheytt.
(GREET, INTRODUCE_NAME)
(*Hello, Mr Scheytt*)

B02: Guten Tag // Frau Klein // Wir müssen noch einen Termin ausmachen // für die Mitarbeiterbesprechung.
(GREET, INTRODUCE_NAME, INIT_DATE, SUGGEST_SUPPORT_DATE)
(*Hello, Mrs. Klein, we should arrange an appointment, for the team meeting*)

A03: Ja,// ich würde Ihnen vorschlagen im Januar,// zwischen dem fünfzehnten und neunzehnten.
(UPTAKE, SUGGEST_SUPPORT_DATE, REQUEST_COMMENT_DATE)
(*Well, I would suggest in January, between the fifteenth and the nineteenth*)

B04: Oh // das ist ganz schlecht. // zwischen dem elften und achtzehnten Januar bin ich in Hamburg.
(UPTAKE, REJECT_DATE, SUGGEST_SUPPORT_DATE)
(*Oh, that is really inconvenient, I'm in Hamburg between the eighteenth of January and the eleventh,*)

...

A09: Doch ich habe Zeit von sechsten Februar bis neunten Februar
(SUGGEST_SUPPORT_DATE)
(*I have time afterall from the 6th of February to the 9th of February*)

B10: Sehr gut // das paßt bei mir auch // Dann machen wir's gleich aus // für Donnerstag // den achten // Wie wäre es denn um acht Uhr dreißig //
(FEEDBACK_ACKNOWLEDGEMENT, ACCEPT_DATE, INIT_DATE, SUGGEST_SUPPORT_DATE, SUGGEST_SUPPORT_DATE, SUGGEST_SUPPORT_DATE)
(*Very good, that too suits me, we will arrange for it, for thursday, the eighth, how about half past eighth*)

A11: Am achten // ginge es bei mir leider nur bis zehn Uhr // Bei mir geht es besser nachmittags .
(SUGGEST_SUPPORT_DATE, SUGGEST_SUPPORT_DATE, ACCEPT_DATE)
(*on the eighth, Is it only unfortunately possible for me until 10 o'clock, It suits me better in the afternoon*)

B12: gut // um wieviel Uhr sollen wir uns dann treffen ?
(FEEDBACK_ACKNOWLEDGEMENT, SUGGEST_SUPPORT_DATE)
(*good, when should we meet*)

A13: ich würde ähm vierzehn Uhr vorschlagen // geht es bei Ihnen.
(SUGGEST_SUPPORT_DATE, REQUEST_COMMENT_DATE)
(*I would suggest 2 o'clock, is that possible for you?*)

B14: sehr gut // das paßt bei mir auch // das können wir festhalten
(ACCEPT_DATE,ACCEPT_DATE, ACCEPT_DATE)
(*very good, that suits me too, we can make a note of that*)

...

Abbildung 1: An example dialogue

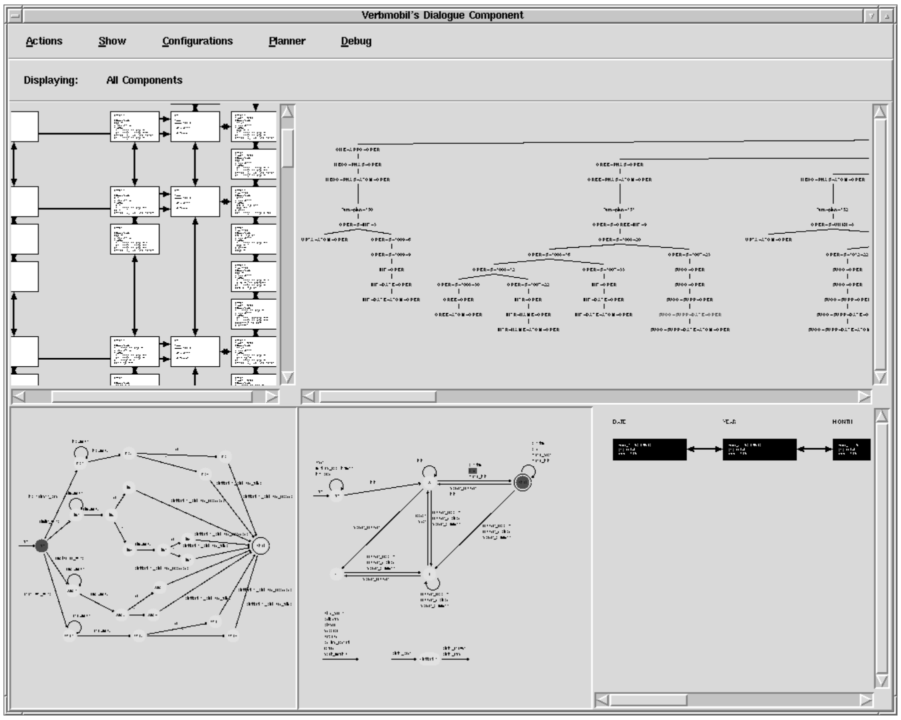

Abbildung 2: Overview of the dialogue module

3 Maintaining Context

As basis for storing context information we developed the *dialogue sequence memory*. It is a generic structure which mirrors the sequential order of turns and utterances. A wide range of operation has been defined on this structure. For each turn, we store e.g. the speaker identification, the language of the contribution, the processing track finally selected for translation, and the number of translated utterances. For the utterances we store e.g. the dialogue act, dialogue phase, and predictions. These data are partly provided by other modules of Verbmobil or computed within the dialogue module itself (see below).

Figure 3 shows the dialogue sequence memory after the processing of turn B02. For the deep analysis side (to the right), the turn is segmented into four utterances: *Guten Tag // Frau Klein // Wir müssen noch einen Termin ausmachen // für die Mitarbeiterbesprechung*, for which the semantic evaluation component has assigned the dialogue acts Greet, Introduce_Name, Init_Date, and Suggest_Support_Date respectively. To the left we see the results of one of the shallow analysis components. It splits up the input into two utterances *Guten Tag Frau Klein // Wir müssen ... die Mitarbeiterbesprechung* and assi-

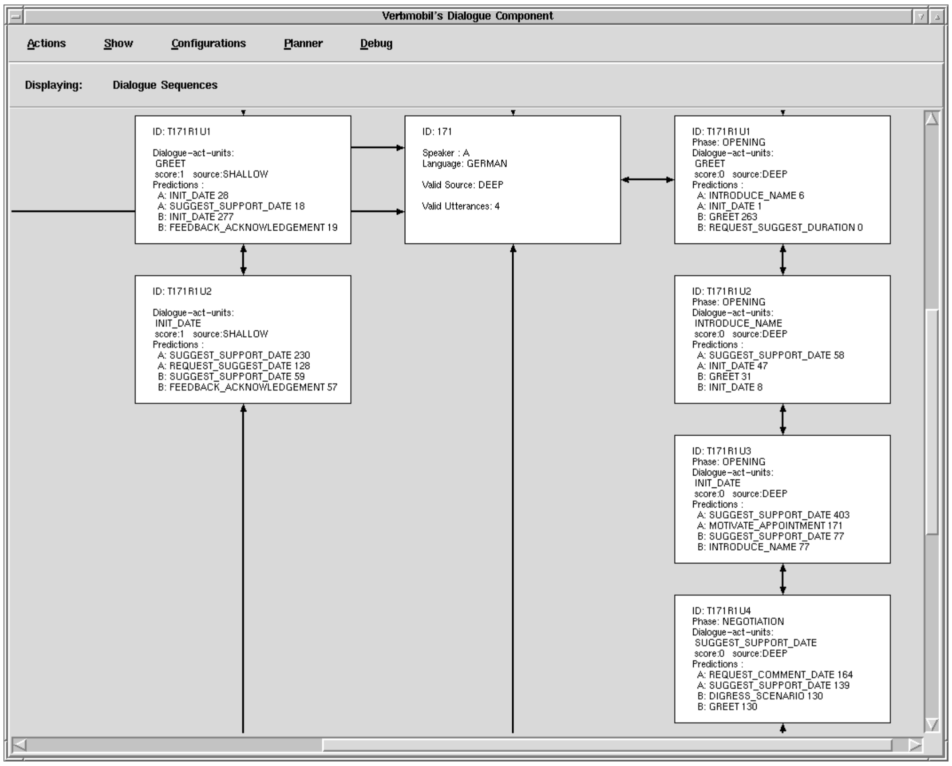

Abbildung 3: A part of the sequence memory

gns the dialogue acts GREET and INIT_DATE.

The need for and use of this structure is highlighted by the following example. In the domain of appointment scheduling the German phrase *Geht es bei Ihnen?* is ambiguous: *bei Ihnen* can either refer to a location, in which case the translation is *Would it be okay at your place?* or, to a certain time. In the latter case the correct translation is *Is that possible for you?*. A simple way of disambiguating this is to look at the preceding dialogue act(s). In our example dialogue, turn `A13`, the utterance *ich würde ähm vierzehn Uhr vorschlagen* (*I would hmm fourteen o'clock suggest*) contains the proposal of a time, which is characterized by the dialogue act SUGGEST_SUPPORT_DATE. With this dialogue act in the immediately preceding context the ambiguity is resolved as referring to a time and the correct translation is determined.

In our domain, in addition to the dialogue act the most important propositional information are the dates as proposed, rejected, and finally accepted by the users of VERBMOBIL. While it is the task of the semantic evaluation module to extract time information from the actual utterances, the dialogue module integrates those information in its *thematic memory*. This includes resolving relative time expressions, e.g. *two weeks ago*, into precise time descriptions, like "23rd week of 1996". The information about the dates is split in a specialization hierarchy. Each date to be negotiated serves as a root, while the nodes represent the information about years, months, weeks, days, days of week, period of day and finally time. Each node contains also information about the attitude of the dialogue participants concerning this certain item: proposed, rejected, or accepted by one of the participants.

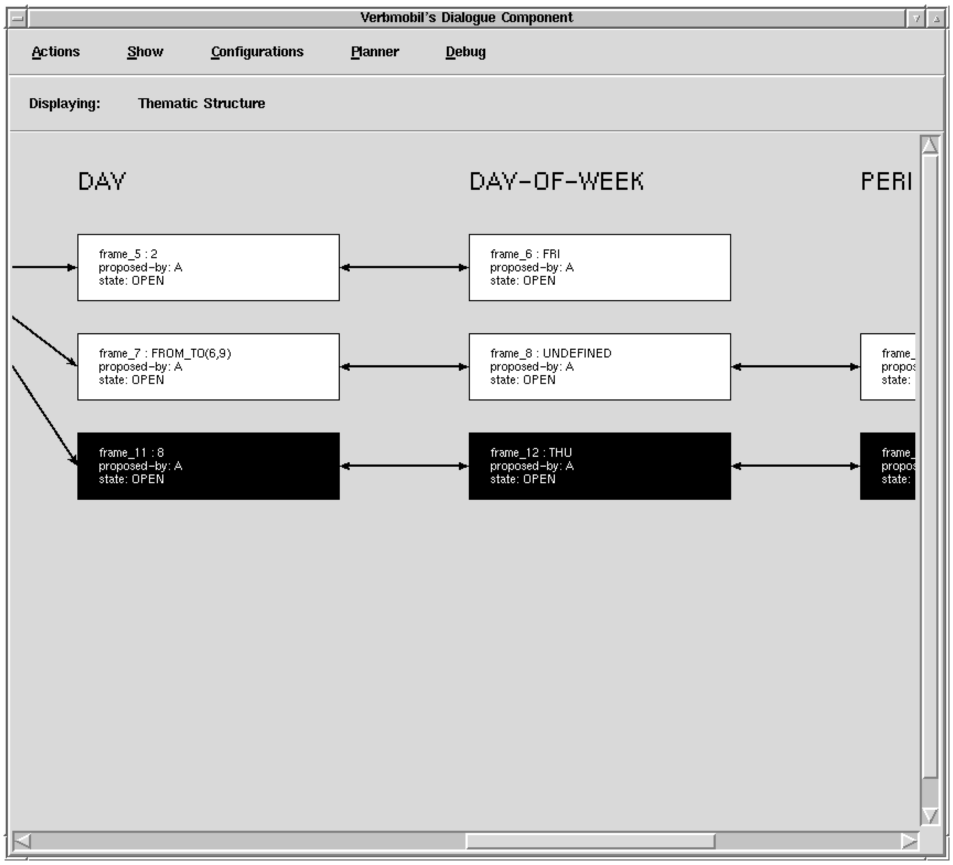

Abbildung 4: Day/Day-of-Week detail of the thematic structure

Figure 4 shows parts of the thematic structure after the processing of turn `B10`. The black boxes stand for the date currently under consideration. Thursday, 8., is the current date agreed upon. We also see the previously proposed interval from 6.-9. of the same month in the box above (`FROM_TO(6,9)`).

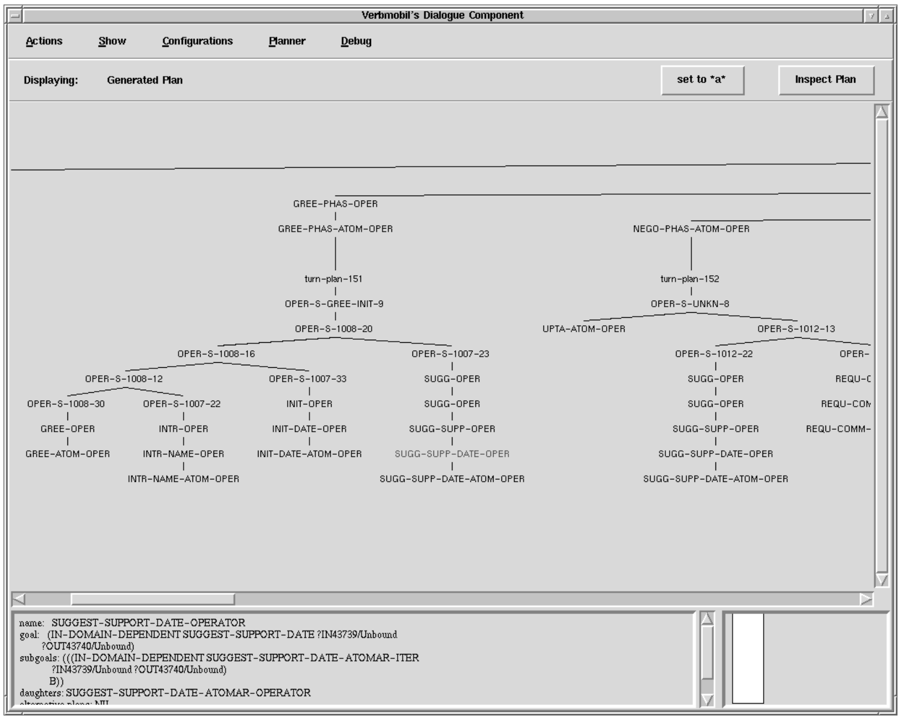

Abbildung 5: Intentional structure for two turns

4 Inferences

Besides the mere storage of dialogue related data, there are also inference mechanisms integrating the data in representations of different aspects of the dialogue. These data are again stored in the context memories shown above and are accessed by the other VERBMOBIL modules.

Plan Based Inferences

Inspecting our corpus, we can distinguish three phases in most of the dialogues. In the first, *the opening phase*, the locutors greet each other and the topic of the dialogue is introduced. The dialogue then proceeds into *the negotiation phase*, where the actual negotiation takes place. It concludes in *the closing phase* where the negotiated topic is confirmed and the locutors say goodbye. This phase information contributes to the correct transfer of an utterance. For example, the German utterance *Guten Tag* is translated to "Hello" in the greeting phase, and to "Good day" in the closing phase.

The task of determining the phase of the dialogue has been given to the plan recognizer (Alexandersson, 1995). It builds a tree like structure which we call the intentional structure. The current version makes use of plan operators both hand coded and automatically derived from the VERBMOBIL corpus. The method used is transferred from the field of grammar extraction (Stolcke, 1994). To contribute to the robustness of the system, the processing of the recognizer is divided into several processing levels like the "turn level" and the "domain dependent level". The concepts of turn levels and the automatic acquisition of operators are described in (Alexandersson, 1996).

In figure 5 we see the structure after processing turns `B02` and `A03`. The leaves of the tree are the dialogue acts. The root node of the left subtree for `B02` is a `GREE(T)-INIT-...` operator which belongs to the greeting phase, while the partly visible one to the right belongs to the negotiation phase.

In the example used in this paper we are processing a "well formed" dialogue, so the turn structure can be linked into a structure spanning over the whole dialogue. We also see in figure 3 how the phase information has been written into the boxes representing the utterances of turn `B02` as segmented by the deep analysis.

Thematic Inferences

In scheduling dialogues, referring expressions like the German word *nächste* occur frequently. Depending on the thematic structure it can be translated as *next* if the date referred to is immediately after the speaking time, or *following* in the other cases. The thematic structure is mainly used to resolve this type of anaphoric expressions if requested by the semantic evaluation or the transfer module. The information about the relation between the date under consideration and the speaking time can be immediately computed from the thematic structure.

The thematic structure is also used to check whether the time expressions are correctly recognized. If some implausible dates are recognized, e.g. April, 31., a clarification can be invoked. The system proposes the speaker a more plausible date, and waits for an acceptance or rejection of the proposal. In the first case, the correct date will be translated, in the latter, the user is asked to repeat the whole turn.

Using the current state of the thematic structure and the dialogue act in combination with the time information of an utterance, multiple readings can be inferred (Maier, 1996). For example, if both locutors propose different dates, an implicit rejection of the former date can be assumed.

5 Predictions

A different type of inference is used to generate predictions about what comes next. While the plan-based component uses declarative knowledge, albeit acquired automatically, dialogue act predictions are based solely on the annotated VERBMOBIL corpus. The computation uses the conditional frequencies of dialogue act sequences to compute probabilities of the most likely follow-up dialogue acts (Reithinger et al., 1996), a method adapted from language modeling (Jelinek, 1990). As described above, the dialogue sequence memory serves as the central repository for this information.

The sequence memory in figure 3 shows in addition to the actual recognized dialogue act also the predictions for the following utterance. In (Reithinger et al., 1996) it is demonstrated that exploiting the speaker direction significantly enhances the prediction reliability. Therefore, predictions are computed for both speakers. The numbers after the predicted dialogue acts show the prediction probabilities times 1000.

As can be seen in the figure, the actually recognized dialogue acts are, for this turn, among the two most probable predicted acts. Overall, approx. 74% of all recognized dialogue acts are within the first three predicted ones.

Major consumers of the predictions are the semantic evaluation module, and the shallow translation module. The former module that uses mainly knowledge based methods to determine the dialogue act of an utterance exploits the predictions to narrow down the number of possible acts to consider. The shallow translation module integrates the predictions within a Bayesian classifier to compute dialogue acts directly from the word string.

6 Robustness

For the dialogue module there are two major points of insecurity during operation. On the one hand, the user's dialogue behaviour cannot be controlled. On the other hand, the segmentation as computed by the syntactic-semantic construction module, and the dialogue acts as computed by the semantic evaluation module, are very often not the ones a linguistic analysis on the paper will produce. Our example dialogue is a very good example for the latter problem.

Since no module in VERBMOBIL must ever crash, we had to apply various methods to get a high degree of robustness. The most knowledge intensive module is the plan recognizer. The robustness of this subcomponent is ensured by dividing the construction of the intentional structure into several processing levels. Additionally, at the turn level the operators are learned from the annotated corpus. If the construction of parts of the structure fails, some functionality has been developed to recover. An important ingredience of the processing is the notion of *repair* – if the plan construction is faced with something unexpected, it uses a set of specialized repair operators to recover. If parts of the structure could not be built, we can estimate on the basis of predictions what the gap consisted of.

The statistical knowledge base for the prediction algorithm is trained on the Verbmobil corpus that in its major parts contains well-behaved dialogues. Although prediction quality gets worse if a sequence of dialogue acts has never been seen, the interpolation approach to compute the predictions still delivers useful data.

As mentioned above, to contribute to the correctness of the overall system we perform different kinds of clarification dialogues with the user. In addition to the inconsistent dates, we also e.g. recognize similar words in the input that will be most likely exchanged by the speech recognizer. Examples are the German words for *thirteenth* (*dreizehnter*) and *thirtieth* (*dreißigster*). Within a uniform computer–human interaction, we resolve these problems.

7 Related Work

In the speech-to-speech translation system Janus (Lavie et al., 1996), two different approaches, a plan based and an automaton based, to model dialogues have been implemented. Currently, only one is used at a time. For Verbmobil, (Alexandersson and Reithinger, 1995) showed that the descriptive power of the plan recognizer and the predictive power of the statistical component makes the automaton obsolete.

The automatic acquisition of a dialogue model from a corpus is reported in (Kita et al., 1996). They extract a probabilistic automaton using an annotated corpus of up to 60 dialogues. The transitions correspond to dialogue acts. This method captures only local discourse structures, whereas the plan based approach of Verbmobil also allows for the description of global structures. Comparable structures are also defined in the dialogue processing of Trains (Traum and Allen, 1992). However, they are defined manually and have not been tested on larger data sets.

8 Conclusion and Future Work

Dialogue processing in a speech-to-speech translation system like Verbmobil requires innovative and robust methods. In this paper we presented different aspects of the dialogue module while processing one example dialog. The combination of knowledge based and statistical methods resulted in a reliable system. Using the Verbmobil corpus as empirical basis for training and test purposes significantly improved the functionality and robustness of our module, and allowed for focusing our efforts on real problems. The system is fully integrated in the Verbmobil system and has been tested on several thousands of utterances.

Nevertheless, processing in the real system creates still new challenges. One problem that has to be tackled in the future is the segmentation of turns into utterances. Currently, turns are very often split up into too many and too small utterances. In the future, we will have to focus on the problem of “glueing” fragments together. When given back to the transfer and generation modules, this will enhance translation quality.

Future work includes also more training and the ability to handle sparse data. Although we use one of the largest annotated corpora available, for purposes like training we still need more data.

Acknowledgements

This work was funded by the German Federal Ministry of Education, Science, Research 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. We thank our students Ralf Engel, Michael Kipp, Martin Klesen, and Paula Sevastre for their valuable contributions. Special thanks to Reinhard for Karger's Machine.

References

Alexandersson, Jan. 1995. Plan recognition in VERBMOBIL. In Mathias Bauer, Sandra Carberry, and Diane Litman, editors, *Proceedings of the IJCAI-95 Workshop The Next Generation of Plan Recognition Systems: Challenges for and Insight from Related Areas of AI*, pages 2–7, Montreal, August.

Alexandersson, Jan. 1996. Some Ideas for the Automatic Acquisition of Dialouge Structure. In Anton Nijholt, Harry Bunt, Susann LuperFoy, Gert Veldhuijzen van Zanten, and Jan Schaake, editors, *Proceedings of the Eleventh Twente Workshop on Language Technology, TWLT, Dialogue Management in Natural Language Systems*, pages 149–158, Enschede, Netherlands, June 19-21.

Alexandersson, Jan and Norbert Reithinger. 1995. Designing the Dialogue Component in a Speech Translation System – a Corpus Based Approach. In *Proceedings of the 9th Twente Workshop on Language Technology (Corpus Based Approaches to Dialogue Modelling)*, Twente, Holland.

Bub, Thomas and Johannes Schwinn. 1996. Verbmobil: The evolution of a complex large speech-to-speech translation system. In *Proceedings of ICSLP-96*, pages 2371–2374, Philadelphia, PA.

Jekat, Susanne, Alexandra Klein, Elisabeth Maier, Ilona Maleck, Marion Mast, and J. Joachim Quantz. 1995. Dialogue Acts in VERBMOBIL. Verbmobil Report 65, Universität Hamburg, DFKI Saarbrücken, Universität Erlangen, TU Berlin.

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

Kita, Kenji, Yoshikazu Fukui, Masaki Nagata, and Tsuyoshi Morimoto. 1996. Automatic acquisition of probabilistic dialogue models. In *Proceedings of ISSD-96*, pages 109–112, Philadelphia, PA.

Lavie, Alon, Lori Levin, Yan Qu, Alex Waibel, Donna Gates, Marsal Gavalda, Laura Mayfield, and Maite Taboada. 1996. Dialogue processing in a conversational speech translation system. In *Proceedings of ICSLP-96*, pages 554–557, Philadelphia, PA.

Maier, Elisabeth. 1996. Context Construction as Subtask of Dialogue Processing - the VERBMOBIL Case. In Anton Nijholt, Harry Bunt, Susann LuperFoy, Gert Veldhuijzen van Zanten, and Jan Schaake, editors, *Proceedings of the Eleventh Twente Workshop on Language Technology, TWLT, Dialogue Management in Natural Language Systems*, pages 113–122, Enschede, Netherlands, June 19-21.

Reithinger, Norbert, Ralf Engel, Michael Kipp, and Martin Klesen. 1996. Predicting Dialogue Acts for a Speech-To-Speech Translation System. In *Proceedings of International Conference on Spoken Language Processing (ICSLP-96)*, pages 654–657, Philadelphia, PA, October.

Stolcke, Andreas. 1994. *Bayesian Learning of Probabilistic Language Models.* Ph.D. thesis, University of California at Berkeley.

Traum, David R. and James F. Allen. 1992. A "Speech Acts" Approach to Grounding in Conversation. In *Proceedings of International Conference on Spoken Language Processing (ICSLP'92)*, volume 1, pages 137–140.

Wahlster, Wolfgang. 1993. Verbmobil–Translation of Face-to-Face Dialogs. Technical report, German Research Centre for Artificial Intelligence (DFKI). In Proceedings of MT Summit IV, Kobe, Japan, July 1993.