

Clarification Dialogues as Measure to Increase Robustness in a Spoken Dialogue System

Elisabeth Maier Norbert Reithinger Jan Alexandersson

DFKI GmbH

Stuhlsatzenhausweg 3

D-66123 Saarbrücken, Germany

{maier,reithinger,alexandersson}@dfki.uni-sb.de

Abstract

A number of methods are implemented in the face-to-face translation system VERBMOBIL to improve its robustness. In this paper, we describe clarification dialogues as one method to deal with incomplete or inconsistent information. Currently, three types of clarification dialogues are realized: subdialogues concerning phonological ambiguities, unknown words and semantic inconsistencies. For each clarification type we discuss the detection of situations and system states which lead to their initialization and explain the information flow during processing.

1 Dialogue Processing in VERBMOBIL

The implemented research prototype of the speech-to-speech translation system VERBMOBIL (Bub and Schwinn, 1996) consists of more than 40 modules for both speech and linguistic processing. In the system, different processing streams are realized: concurrently with a deep linguistic-based analysis, two methods of shallow processing are realized. On the basis of a set of selection heuristics, the best translation is chosen for synthesis in the target language.

The central system repository for discourse information is the dialogue module. Like all subcomponents of the VERBMOBIL system the dialogue module is faced with incomplete and incorrect input, and with missing information. Therefore we have decided to use a combination of several simple and efficient approaches, which together form a robust and efficient processing platform for the implementation of the dialogue module.

1.1 The Tasks of the Dialogue Component

The dialogue component of the VERBMOBIL system fulfills a whole range of tasks:

- it provides contextual information for other VERBMOBIL components. These components are allowed to store (intermediate) processing results in the so-called dialogue memory (Maier, 1996);
- the dialogue memory merges the results of the various parallel processing streams, represents them consistently and makes them accessible in a uniform manner (Alexandersson, Reithinger, and Maier, 1997);
- on the basis of the content of the dialogue memory inferences can be drawn that are used to augment the results processed by other VERBMOBIL components;
- taking the history of previous dialogue states into account, the dialogue component predicts which dialogue state is most likely to occur next (Reithinger et al., 1996).

The dialogue component does not only have to be robust against unexpected, faulty or incomplete input, it also corrects and/or improves the input provided by other VERBMOBIL components. Among the measures to achieve this goal is the possibility to carry out clarification dialogues.

1.2 The Architecture of the Dialogue Component

The dialogue component is realized as a hybrid architecture: it contains statistical and knowledge-based methods. Both parts work with dialogue acts (Bunt, 1981) as basic units of processing. The statistics module is based on data automatically derived from a corpus annotated with dialogue acts. It determines possible follow-up dialogue acts for every utterance. The plan recognizer as knowledge-based module of the dialogue component incorporates a dialogue model, which describes sequences of dialogue acts as occurring in appointment scheduling dialogues (Alexandersson and Reithinger, 1995).

For the representation of contextual information a dialogue memory has been developed which consists of two subcomponents: the Sequence Memory, which mirrors the sequential order in which the utterances and the related dialogue acts occur, and the Thematic Structure, which consists of instances of temporal categories and their status in the dialogue. Both components are closely intertwined so that for every utterance of the dialogue the available information can be easily accessed.

2 Strategies for Robust Dialogue Processing

The dialogue module has to face one major point of insecurity during operation: the user's dialogue behavior cannot be controlled. While the dialogue module incorporates models that represent the expected moves in an appointment scheduling dialogue users frequently deviate from this course. Since no module in VERBMOBIL must ever fail, we apply various recovery methods to achieve a high degree of robustness. In the plan recognizer, for example, robustness is ensured by dividing the construction of the intentional structure into several processing levels. If the construction of parts of the structure fails, recovery strategies are used. An important ingredient of dialogue processing is the possibility of repair: in case the plan construction encounters unexpected input it uses a set of repair operators to recover. If parts of the structure cannot be built, we estimate on the basis of predictions what information the knowledge gap is most likely to contain.

To contribute to the correctness of the overall system we perform different kinds of clarification dialogues with the user. They will be explained in more detail in the remainder of this paper.

In the current implementation of the VERBMOBIL system, two types of clarification dialogues occur:

- **human-human subdialogues** where a dialogue participant elicits unclear or missing information from his or her dialogue partner. Typical cases occur when a dialogue contribution contains ambiguous information as e.g. in the following dialogue fragment:

A: *What about meeting on Friday?*

B: *Which Friday are you talking about?*

A: *Friday February 28.*

This type of clarification dialogue is processed without any active intervention by the dialogue component: the individual utterances are analyzed and translated by the various processing streams while the dialogue component enters the results into the dialogue memory.

- **human-machine subdialogues** where the machine engages in a dialogue with the user to elicit information needed for correct processing.

In the following we focus on this latter type of clarification dialogues. In our current system we only implemented clarification dialogues where the potential user of VERBMOBIL is likely to have sufficient expertise to provide the information necessary for clarification; where the problems presented to the user require too much linguistic expertise we consider different recovery strategies (e.g. the use of defaults). The following types of clarification dialogues are incorporated in our system¹:

1. dialogues about phonological similarities (`similar_words`) which cope with possible confusions of phonetically similar words like *Juni* vs. *Juli* (engl.: *June* vs. *July*);
2. dialogues about words unknown to the system, in particular unknown to the speech recognizers (`unknown_words`);
3. dialogues about inconsistent or inexistent dates (`inconsistent_date`), e.g. *um 16 Uhr am Vormittag* (engl.: *at 16 hours in the morning*) or *am 30. Februar* (engl.: *on February 30*).

If all of the above types of clarification dialogues are enabled all the time they tend to occur too often. Empirical studies have shown that interruptions of a dialogue – as is the case in clarifications – put additional stress on the users and have a negative influence on performance and acceptance (Krause, 1997). Therefore, we implemented the possibility to selectively enable and disable the various types of clarification dialogues.

In the following chapter we explain how the various types of clarification dialogues are processed.

3 Processing Clarification Dialogues

3.1 Processing Flow

In the deep processing mode spoken input is sent through components for speech recognition, syntactic and semantic treatment, transfer, tactical generation and speech synthesis. The processing results of the morphological, syntactic and semantic components are continuously monitored by the dialogue component. For every utterance `utt_id` and for each type of clarification dialogue the dialogue component sends a message to the central

¹In the remainder of this paper words printed in teletype font indicate full or partial system messages.

control component of the VERBMOBIL system indicating whether a clarification dialogue has to be executed or not (<x utt_id> or <no_x utt_id>, where x is either `similar_words`, `unknown_words`, or `inconsistent_date`).

If a subdialogue has to be carried out, the clarification mode is switched on (`clarification_dialogue on`) and the processing flow of the system is changed. Depending on the clarification type x, a synthesized message is sent to the user, informing him/her of the necessity and reason for a clarification dialogue. A list of options for recovery is presented. In order to minimize processing errors the options the user can choose from are formulated as yes/no questions; a yes-/no recognizer with a recognition rate of approx. 100 % developed specifically for this purpose processes the user's response. If the user chooses an option that allows a continuation of the dialogue it is used to modify the system's intermediate results; the utterance `utt_id` and the updated message are sent to the control module (`clarification_dialogue_succeeded utt_id <modified-message>`), the system switches back into the normal processing mode (`clarification_dialogue off`), and computation is resumed using the modified data. If the user finds none of the presented options appropriate, the user is requested to reformulate the original utterance, the control component is informed of a failure of the subdialogue (`clarification_dialogue_failed utt_id`) and the clarification dialogue is switched off (`clarification_dialogue off`).

To ensure robustness for clarification dialogues we have added a counter to measure the time elapsed since a system request (e.g. the presentation of options to choose from). If the user does not respond within a given time frame, the system assumes a negative answer, which leads to a failure of the subdialogue and the request for a reformulation of the initial utterance. All clarification types mentioned in this paper are fully implemented. All three subdialogue types follow this uniform processing scheme.

3.2 Phonological Similarities

The dialogue system has access to a list of words that are often confused on the basis of a high degree of phonological similarity. Not all of the word pairs included in this list are intuitive candidates for an average VERBMOBIL user. Examples are e.g. the German word pairs *Halle - fahren* or *Modus - Morgen*. We compiled a subset of this list that contains only word pairs that are plausible for a user who has no phonological expertise. This list includes word pairs like e.g. *Sonntag - sonntags* (engl.: *Sunday -*

sundays) or *fünfzehn - fünfzig* (engl.: *fifteen - fifty*).

If the word string processed by the syntactic/semantic components contains a member of this word list the dialogue initializes the generation of a system message that points out the potential confusion to the user. If for example the original input sentence is *Wie wär's Sonntag?* (engl.: *How about Sunday?*) the system triggers the message *VERBMOBIL hat eine mögliche Verwechslung erkannt. Meinen Sie die Angabe 'Sonntag'?* (engl.: *VERBMOBIL encountered a possible ambiguity. Do you mean the word 'Sunday'?*). Depending on the answer of the user either the proposed word is accepted or the remaining other candidate is proposed. The chosen word is then inserted into the intermediate processing result, so that the translation later contains the word chosen by the user.

3.3 Unknown Words

The speech recognizers of the VERBMOBIL system are able to recognize input as unknown to the system; if such a fragment is encountered the symbol UNK_ followed by the SAMPA-transcription (SAM, 1992) of the fragment (e.g. <UNK_maI62> for the unknown spoken input *Maier*) is inserted into the output of the recognizers. In our domain, unknown words often refer to names, e.g. of locations or persons. The user is asked to confirm this assumption. A message including a synthesized version of the word's SAMPA transcription is presented to the user, e.g. *Handelt es sich bei maI6 um einen Namen?* (engl.: *Is maI6 a name?*). If this assumption is confirmed, syntactic processing is continued treating the fragment as a name. The SAMPA transcription is later included in the output of the English generator and synthesized accordingly. Further syntactic and semantic information is not elicited since such knowledge is irrelevant for a satisfactory treatment of names.

3.4 Semantic Inconsistencies

If a user tries to propose nonexistent or inconsistent dates, this is signaled to the dialogue component by the semantic module. If possible, this module also proposes alternative dates. The message `clarify_date([dom:31,moy:apr],[dom:30,moy:apr])` for instance, which is sent from the semantic evaluation component to the dialogue module, indicates both that April 31 is an inconsistent date and that the user might have meant April 30. The message is coded in terms of a time description language developed within VERBMOBIL. It allows to specify temporal information using temporal categories (e.g. DAY-OF-MONTH (DOM) or MONTH-OF-YEAR (MOY)) and instances of these categories (e.g. APRIL

(APR)). Upon receipt this information it is transformed into natural language and presented to the user: *Die Angabe 31. April existiert nicht. Meinen Sie die Angabe 30. April?* (engl: *The date 'April 31' does not exist. Do you mean April 30?*) If the user chooses the alternative date, it is passed on to the relevant components and the resulting translation includes the correct date.

4 Related Work

Various approaches have been proposed to cope with problems of unexpected, wrong or missing input: (Allen et al., 1996) decided to choose the most specific possible option when the system is confronted with ambiguities. To handle this problem the TRAINS system tries to recognize and exploit corrections included in follow-up dialogue actions. (Qu et al., 1997) describe a method to minimize cumulative error in the ENTHUSIAST system. To this end dialogue context, statistical information, and grammar information are taken into account to process and predict dialogue states, where non-contextual information is preferred over contextual information when processing conflicts occur.

While clarification dialogues are common in human-machine dialogues (see e.g. (Eckert and McGlashan, 1993)), they are a rather recent development in systems that support computer-mediated interactions. To our knowledge the VERBMOBIL prototype is the first system that uses repair methods, defaults and clarification dialogues to recover from problematic system states.

5 Conclusion and Future Work

In this paper we presented a first approach to achieve robust processing in VERBMOBIL using clarification dialogues. We presented three problems that can be resolved using clarification: phonological ambiguities, unknown words and semantic inconsistencies. In the next prototype of the VERBMOBIL system we will additionally incorporate methods to resolve lexical and referential ambiguities. Also, we will tailor the interaction to different user classes.

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 the VERBMOBIL software integration team – in particular Thomas Bub, Andreas

Klüter, Stefan Mertens, and Johannes Schwinn – for their valuable help.

References

- Alexandersson, J. and N. Reithinger. 1995. Designing the Dialogue Component in a Speech Translation System – a Corpus Based Approach. In *Proc. of TWLT-9*, Enschede, Netherlands.
- Alexandersson, J., N. Reithinger, and E. Maier. 1997. Insights into the Dialogue Processing of VERBMOBIL. In *Proc. of ANLP-97*, pages 33–40, Washington, DC.
- Allen, J.F., B.W. Miller, E.K. Ringger, and T. Sikorski. 1996. A Robust System for Natural Spoken Dialogue. In *Proc. of ACL-96*, Santa Cruz, CA.
- Bub, Thomas and Johannes Schwinn. 1996. Verbomobil: The evolution of a complex large speech-to-speech translation system. In *Proc. of ICSLP-96*, pages 2371–2374, Philadelphia, PA.
- Bunt, H. C. 1981. Rules for the Interpretation, Evaluation and Generation of Dialogue Acts. In *IPO Annual Progress Report 16*, pages 99–107, Tilburg University.
- Eckert, W. and S. McGlashan. 1993. Managing spoken dialogues for information services. In *Proc. of EUROSPEECH-93*, Madrid, Spain.
- Krause, Detlev. 1997. Using an Interpretation System - Some Observations in Hidden Operator Simulations of VERBMOBIL. In E. Maier, M. Mast, and S. LuperFoy, eds., *Dialogue Processing in Spoken Language Systems*, Springer, Heidelberg.
- Maier, E. 1996. Context Construction as Subtask of Dialogue Processing - the VERBMOBIL Case. In A. Nijholt, H. Bunt, S. LuperFoy, G. Veldhuijzen van Zanten, and J. Schaake, eds., *Proc. of TWLT-11*, pages 113–122, Enschede, Netherlands.
- Qu, Y., B. Di Eugenio, A. Lavie, and C. P. Rosé. 1997. Minimizing Cumulative Error in Discourse Context. In E. Maier, M. Mast, and S. LuperFoy, eds., *Dialogue Processing in Spoken Language Systems*, Springer, Heidelberg.
- Reithinger, N., R. Engel, M. Kipp, and M. Klesen. 1996. Predicting Dialogue Acts for a Speech-To-Speech Translation System. In *Proc. of ICSLP-96*, pages 654–657, Philadelphia, PA.
- SAM Final Report. ESPRIT Project 2589 (Multilingual Speech Input/Output Assessment). See <http://www.phon.ucl.ac.uk/home/sampa/home>.