Verbmobil
Translation of Face-To-Face Dialogs

Wolfgang Wahlster
German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
D-6600 Saarbrücken 11, Germany
Phone: +49 681 302 5252 or 2363
Fax: +49 681 302 5341
E-mail: wahlster@dfki.uni-sb.de

Abstract

Verbmobil is a long-term project on the translation of spontaneous language in negotiation dialogs. We describe the goals of the project, the chosen discourse domains and the initial project schedule. We discuss some of the distinguishing features of Verbmobil and introduce the notion of translation on demand and variable depth of processing in speech translation. We describe the role of anytime modules for efficient dialog translation in close to real time.

The long-term vision behind the project Verbmobil is a portable translation device that you can carry to a meeting with Speakers of other foreign languages and it will translate what you say for them.
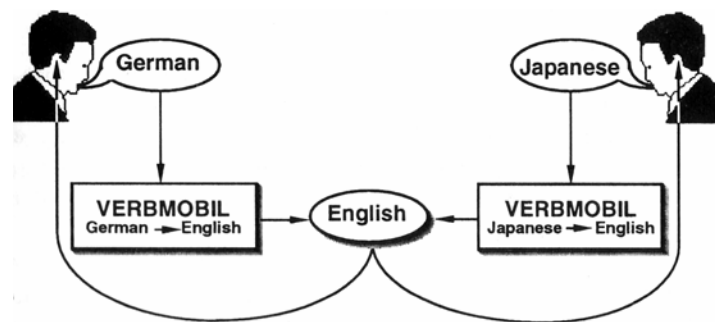

Fig. 1: English as the Common Dialog Language in Verbmobil

This very ambitious scientific goal will be persued in a series of well-defined project phases. The first versions of Verbmobil will provide translation on demand for the two participants who have a passive knowledge of English but of which neither is a fluent speaker. We assume that most of the dialog will be conducted in English as a common dialog language. This is a realistic assumption for most international technical or business discussions. But for uncommon words or phrases, complex constructions and critical segments of the negotiation dialog the participants may want to switch back to their native language. This means that they need translation help and therefore turn to their Verbmobil devices.

In the course of the conversation each dialog partner can activate his version of Verbmobil (eg. German-to-English or Japanese-to-English translation) and signal that he is now speaking in his native language (eg. German or Japanese), and that what he says should be translated into English (see Fig. 1).

This means that there are three input modes for Verbmobil:

1)   Both dialog participants speak English with a German or Japanese accent. In this case, no translation is necessary, but Verbmobil has to follow the conversation and extract context information for subsequent translation tasks. This is an extremely difficult problem, since the input can be ill-formed in many ways, so that various phonetic and grammatical constraints have to be relaxed in order to cope with the foreign accent and unusual constructions. Often Verbmobil will extract only a very shallow discourse using keyword spotting or other partial analysis techniques.

2)   In the course of an utterance, a participant switches from English as the common dialog language back to German or Japanese as his native language. In this case, Verbmobil must generate a translation that fits with the context of the English sentence fragment. For example, if a German participant says "Let's meet again in June ↓außer am Pfingstmontag↑" Verbmobil should produce "except on Whit Monday" to complete the English fragment correctly (the arrows ↓ and ↑ indicate that the speaker has signalled the code switching to Verbmobil).

3)   The participant speaks in his own language and Verbmobil will translate his utterance into English. In this case, Verbmobil must generate an appropriate approximation of the communicative intent of the input, in close to real time. In many situations, Verbmobil will be able to find translations that preserve most but not necessary all of the content of the original, since translation is inescapably a matter of compromise.

Clarification subdialogs play an important role in the conversational setting discussed above, since the dialog partners are no fluent speakers of English and Verbmobil is an imperfect understander and translator. In the Verbmobil project, two types of clarification subdialogs are studied (see Fig. 2):

1) Clarification subdialogs between the participants are conducted in English. There are two variants of this type of subdialog: both dialog partners use English or Verbmobil translates their utterances from their native language into English.

2) Clarification subdialogs between Verbmobil and one participant are conducted in the native language of the respective dialog partner.
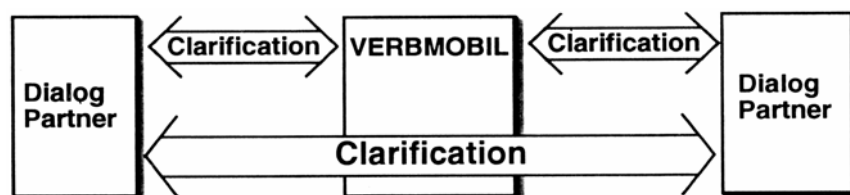

Fig. 2: Two Types of Clarification Subdialogs

The Project Goals

There are four distinguishing features of the Verbmobil approach:

- speaker-adaptive recognition of spontaneous speech
- negotiation dialogs in face-to-face situations
- portable translation device that can be tailored to the individual user and to specified application domains
- three language scenario (English, German, Japanese) with English as a dialog language, ensuring system transparency and user acceptance.

In contrast to previous projects on speech translation (cf. [2], [5]) Verbmobil does not deal with telephone conversations but with face-to-face dialogs in a small meeting room. In face-to-face dialog translation we can exploit the fact that information passes between the participants not only on the linguistic channel but also on various nonverbal and paralinguistic channels. The hearer can merge information from the translation with information from gestural motions of the hands, fingers, head and eyes, eyeblinks, eyebrows movements, change of body posture and orientation. The research program includes some empirical investigations of translation and interpreting as done by humans in similar situations.

Verbmobil does not deal with read speech input, but with incrementally produced spontaneous dialog contributions. Such utterances are rarely well-formed, since speakers make errors and correct them. Verbmobil has to deal with false starts, aborted phrases, speech repairs, hesitations, interjections, self-correction phrases and many other characteristic features of spontaneous speech (see Fig. 3).

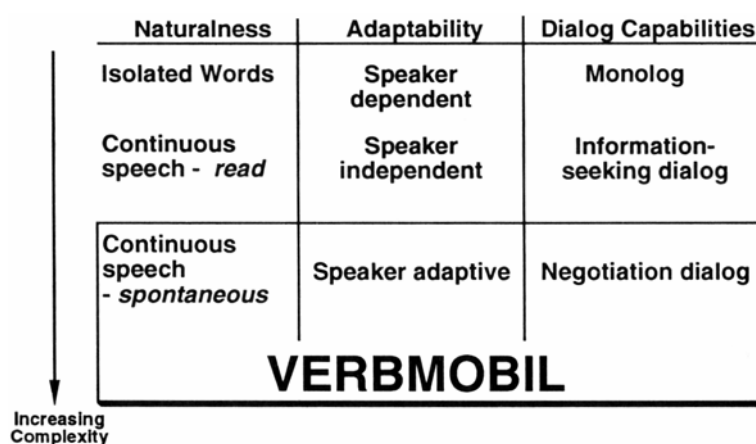| Naturalness	Adaptability	Dialog Capabilities
Isolated Words	Speaker dependent	Monolog
Continuous speech - *read*	Speaker independent	Information-seeking dialog
Continuous speech - *spontaneous*	Speaker adaptive	Negotiation dialog

VERBMOBIL

Increasing Complexity

Fig. 3: Challenges of Language Technology

In the discourse situation studied for the initial demonstrator the dialog partners discuss a possible date for their next meeting using a calendar in front of them. After the development of the initial demonstrator, the domain of discourse will be extended considerably for the first research prototype. Two negotiation tasks will be considered for the research prototype (see Fig. 4).

Note that the appointment scheduling task is a subtask of both scenarios considered for the research prototype. The domains chosen deal with linguistically ordinary language, so that the linguistic knowledge sources can simply be extended when the domain is scaled up. In all conversational settings studied in the Verbmobil project the subject matter is limited and the aims of the dialog partners are known in advance. We take it that both dialog partners come to a meeting in a spirit of cooperation and that they are highly motivated to reach a successful conclusion.
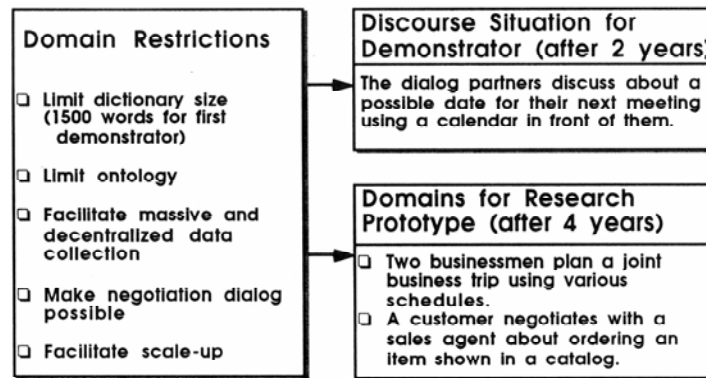
Fig. 4: Discourse Domains for Verbmobil.

Verbmobil channels energy into key areas of language technology and integrates major subfields of advanced information technology like
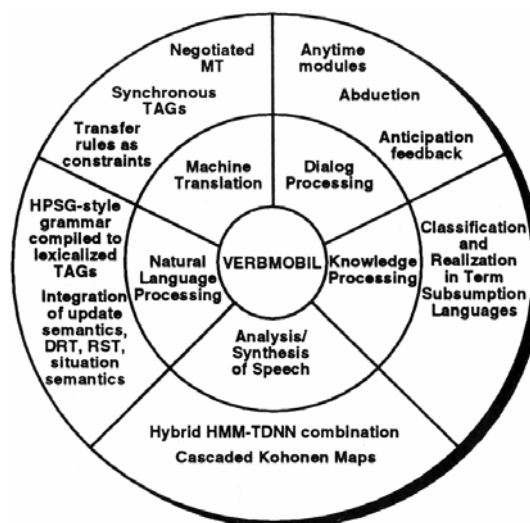

Fig. 5: Integrating Major Subfields of Language Technology

natural language processing, speech recognition and synthesis, machine translation, dialog and knowledge processing (cf. Fig. 5).

Since there in no doubt, that the fact that language is always situated is very important for translation and that a proper translation almost always depends on context, Verbmobil must integrate research on translation with work on dialog processing as well as knowledge representation and reasoning.

Verbmobil is an interdisciplinary attempt to build a face-to-face translation system on the basis of current theories that leading researchers in artificial intelligence, computational linguistics, speech processing, neuro-computing and translation science

would subscribe to. The Verbmobil consortium believes that the scientific foundation of dialog translation technology should never be compromised in the interests of achieving some functionality or speed-ups in the short run by ad hoc techniques, that cannot be generalized and scaled-up.

Anytime Modules for Face-to-Face Dialog Translation

Obviously, there is a tradeoff between run-time and quality of results in systems for face-to-face dialog translation. Verbmobil's analysis should not be deeper than necessary, its translation should be as shallow as possible, and its generation process should start as soon as possible.
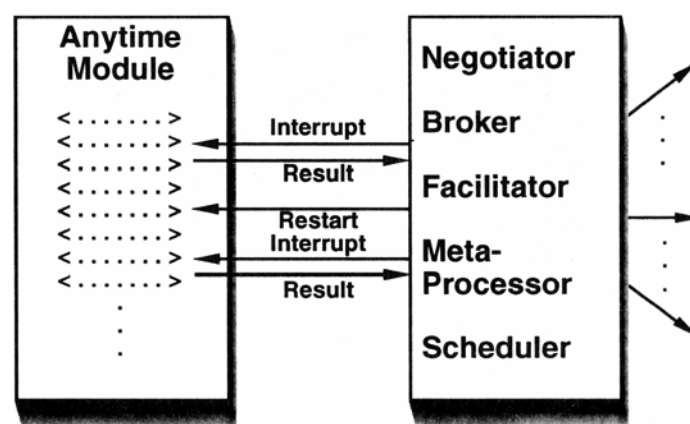

Fig. 6: Anytime Modules as Coroutines

This means that the major components of the system must work in an incremental mode allowing the immediate processing of parts of a stepwise provided input. These modules will be realized as anytime modules for the sake of resource-bounded processing of discourse.

Anytime modules are modules whose quality of results improve gradually as computation time increases. They yield imperfect but not useless results if interrupted before completion. If an anytime module is restarted, it can improve what it has generated so far.

For Verbmobil anytime modules are needed on various levels of granularity, e.g.

- speech analysis, parsing, transfer, generation
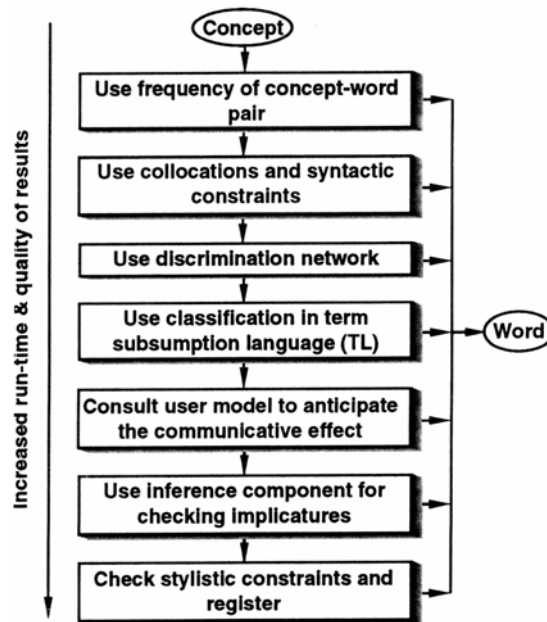- pronoun resolution, focus detection, lexical choice

Fig. 7: Lexical Choice as an Anytime Module

All Verbmobil modules integrate a wide spectrum of layered methods: from simple and low cost to complex and expensive techniques. This can be illustrated by the problem of lexical choice. If lexical choice is implemented as an anytime module, the quality of the results can be measured in terms of the precision of communicating the intended concept in a given situation. The concept-to-word mapping can be achieved by a wide spectrum of techniques from very fast methods using the frequency of concept-word pairs to very elaborate methods like checking possible communicative effects and implicatures.

The concept of anytime modules is tightly connected to the idea of variable depth of processing in a speech translation system. Verbmobil will use a multi-layered semantic representation language, that allows for all kinds of underspecification in the surface-oriented layers. In many cases, ambiguous quantifier scope or PP attachment in the source language need not be resolved before being translated, since a corresponding ambiguity can be captured in the target language. This leads to the new problem of language generation from disjunctive semantic structures.

It is important that each layer of the semantic representation language comes with a specialized inference component, so that even on the level of surface-oriented representations simple inferences can be drawn. While these inferences may be based on primitive rewriting techniques, the inference engine on the more elaborate levels of meaning representation may be a full theorem prover.
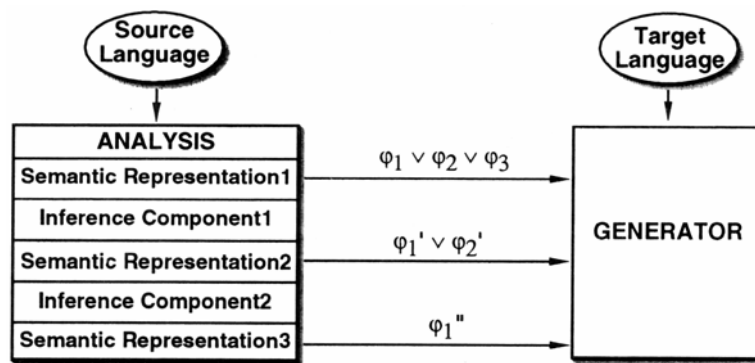
Fig. 8: The Notion of Variable Depth of Processing

The Project Structure

The Verbmobil project is funded by the German Ministry for Research and Technology (BMFT) and an industrial consortium. For the first four years of the project the BMFT funding amounts to 60 Million Deutschmarks.

The BMFT commisioned two feasibility studies on the goals of Verbmobil: one from a consortium of German industrial and academic research groups (see [3]) and another from the Center for the Study of

Language and Information (CSLI) in the US (see [1]). Based on the positive recommendations of the two independent studies a detailed project plan and schedule was prepared (see [4]), that formed the basis of a call for proposals in July 1992. An international advisory and review board was appointed by the BMFT consisting of 10 well-known experts in speech, language and translation technology. The scientific review of all submitted proposals was finished at the end of January 1993. The main phase of the project is starting in May 1993.

The project is planned for 8 to 10 years and the first phase of 4 years is structured by 2 major milestones: a demonstrator after 2 years and a research prototype after 4 years (see Fig. 9). The central project coordination task and the implementation of the demonstrator and research prototype will be carried by the German Research Center for AI (DFKI).
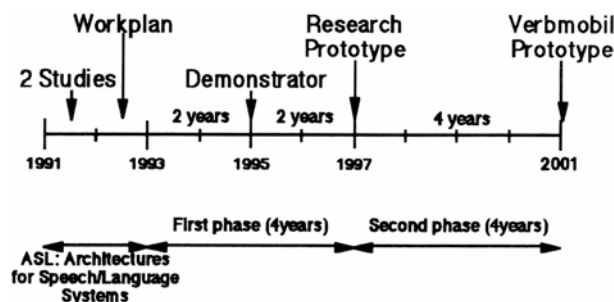

Fig. 9: The Project Schedule for Verbmobil

The success of such an ambitious translation project obviously depends on international cooperation. It is planned to have an intensive collaboration with the ATR Interpreting Telecommunications Research Laboratories in Kyoto. In March 1993, this well-known Japanese center for speech translation research started a new project that will end in March 2000. Like Verbmobil this project deals with the translation of spontaneous dialog language. The funding amounts to 16 billion yen. Data collection, speech modules and linguistic knowledge sources for the Japanese language are the major areas of the planned collaboration. For work packages concerning the English language, cooperations have been prepared with three US research groups: Carnegie Mellon University, CSLI at Stanford University and the International Computer Science Institute (ICSI) at Berkeley.

References:

[1] Kay, M., Gawron, J.M., Norvig, P.: Verbmobil: A Translation System for Face-to-Face Dialog. BMFT Study, CSLI, Stanford Univ., August 1991.

[2] Morimoto, T., Shikano, K., Iida, H., Kurematsu, A.: Integration of Speech Recognition and Language Processing in the Spoken Language Translation System SL-TRANS. In: Proc. of the Intern. Conference on Speech and Language Processing, 1990, p. 921 - 928.

[3] Verbmobil-Consortium: A Portable Translation Device. BMFT Study, Siemens, Munich, August 1991 (in German).

[4] Wahlster, W., Engelkamp, J. (eds.): Scientific Goals and Networks of Work Packages of the Verbmobil Project. BMFT Study, DFKI, Saarbrucken, April 1992 (in German).

[5] Waibel, A., Jain, A.N., McNair, A..E., Saito, H., Hauptmann, A., Tebelskis, J. (1991): JANUS - A Speech-to-Speech Translation System Using Connectionist and Symbolic Processing Strategies. In: Proc. of the 1991 Intern. Conf. on Acoustics, Speech, and Signal Processing, 1991, p.793 -796.