

speaker independent kernel; d. a basis grammar of German for spontaneous speech, based on linguistic knowledge, with deep and shallow semantic analysis; e. spoken dialogs for clarification between the user and the VM-system in situations with recognition or translation problems; f. semantic transfer for German to English and Japanese to English; g. speech generation for English, and for German paraphrases; h. over 70% approximately correct translations in the end-to-end evaluation in the domain of appointment scheduling; i. pure processing time $< 6 \cdot$ real time (length of the speech signal).

1.3 The second phase

The second phase concentrates on robust and direct translations of spontaneous speech for the language pairs German and English, and German and Japanese. In the second phase, VM will be implemented on a central speech server that can be used over ISDN-telephone lines, ATM-based telecooperation services, or GSM mobile telephones. The server will be able to identify the language used at each input channel, and to recognize, translate and generate the relevant output.

Compared with VM-I, VM-II will have the following additional specifications:

- Multi lingual: It is supposed to translate spontaneous speech in different languages.
- Multi functional: It should be quickly adaptable to different dialog domains.
- Multi medial: It should offer translation aids for international multimedia applications.
- Mobile: With the use of language servers it should also be usable with mobile telephones.
- Multi-party functionality: It should be usable not only in dialog situations, but also in telecooperation systems with many dialog partners.

1.3.1 Financing the project

The first phase of the project (1993–1996) was financed with 64,9 million German marks (DM) by the German Federal Ministry of Education, Science, Research and Technology (BMBF). In addition, 31 million DM were contributed by partners from industry. The second phase of the project (1997–2000) is financed with 50,2 million DM by the BMBF, and with 20,4 million DM by the industrial partners [2].

2 THE ARCHITECTURE OF VERBMOBIL

VM contains basically the following components. Due to time limits, a more detailed description of the dif-

ferent components implemented in the VM system will be given in the final version of this paper.

2.1 Speech recognition

The speech recognizers used in VM are statistical Acoustic input is

VM-I included for the first time a real time speaker independent recognizer for spontaneous speech with a high recognition rate.

VMI was able to process extremely long input sequences (sentences) that are used in negotiation dialogs, as opposed to command input in speech controlled systems.

2.2 Prosodic segmentation

VM is the only system that uses prosodic information for translation at different levels of processing, e.g. syntactic, semantic, shallow or stochastic processing. Furthermore, prosodic segmentation introduces information about sentence boundaries that facilitates the syntactic processing for about 92%, and the disambiguation process for about 96%.[4]

2.3 Shallow analysis and translation

The shallow translation splits up a word sequence into communication units. The meaning of the communication units is offered in the translation, independent of the exact linear sequence these units in the original language.

2.4 Stochastic translation

Stochastic translation is based on a bilingual corpus and a statistical model that is trained on this corpus. The model predicts for a word sequence of the input language a word sequence of the target language.

2.5 Deep analysis and translation

In the component of deep translation, the input word sequence is analyzed syntactically and semantically. The linguistic structures are then transferred to the target language.

2.6 Example-based translation

The component for example based translation stores all sentences or parts of sentences in a database. Input sentences are

2.7 Generation

VM includes a very efficient generator that is based on a reversible HPSG-grammar for English. The mean generation time of this generator for one sentence is

0,7 seconds.

3 EVALUATION

The VM system is evaluated every six months. The evaluation of VM-II concentrates on the user perspective. "Real users" are involved in examining the systems properties and usability for arranging appointments or journeys. Available measures from such experiments is the dialog success rate, and the time consumption.[6]

Furthermore, VM-II uses a hierarchical evaluation method, that also involves evaluation of single modules, as well as a turn-end-to-end evaluation. The quality of the input and the output is evaluated on a linguistic basis, analyzing not only the information transmission rate in the transfer process, but also the syntactic and acoustic quality of the input and the output.[1]

Due to time limits, a more detailed description of the evaluation methods can only be presented in the final version of this paper.

4 REFERENCES

- [1] Čavar, D. 1998. *End-to-End Evaluation des Verbomobil Herbstdemonstrators 1998*. Mscr. University of Hamburg.
- [2] Karger, R. & W. Wahlster. 1997. Verbomobil: Multilinguale Verarbeitung von Spontansprache. *Künstliche Intelligenz* 4.97, 41–45.
- [3] Menzel, W. & J. J. Quanz. 1997. Linguistische Verarbeitung im Maschinellen Dolmetschen: Syntax, Semantik, Transfer. *Künstliche Intelligenz* 4.97, 19–25.
- [4] Nöth, E. et al. 1997. Spracherkennung und Prosodie. *Künstliche Intelligenz* 4.97, 14–19.
- [5] Wahlster, W. 1993. Verbomobil: Translation of face-to-face dialogs. In *MT Summit IV*, Kobe, Japan.
- [6] Jost, U. & W. Menzel. 1998. *Evaluation in Verbomobil*. Mscr. University of Hamburg.