THE LANGUAGE COMPONENTS IN VERBMOBIL

Hans Ulrich Block

Siemens AG, Corporate Technology Department,
Otto-Hahn-Ring 6, 81730 München, Germany
Hans-Ulrich.Block@mchp.siemens.de

ABSTRACT

This paper gives an overview over the main problems and their solutions in the language components of the Verbmobil speech translation system[12]. Interpretation of spontaneously spoken language has to take into account that syntax and semantics differ from written language, that punctuation is missing, that accent and intonation have effects on the meaning and the translation, that the output of the speech recognizer may be noisy and that speakers produce errors due to distraction. The Verbmobil interpretation and translation components try to attack these problems by means of a grammar for spoken language, heavy use of prosodic information, a syntactic search on word hypothesis graphs and a shallow robust fall back translation device that is used in case the „deep" translation fails.

1. PROBLEMS TO BE SOLVED

Syntactic and semantic processing of spontaneously spoken language is faced with problems that differ dramatically from those posed by the processing of written text. The special problems that arise from spoken input can be grouped into five distinct sets of problems:

1. Spoken language differs from written language both in syntax and semantics [1]. In spoken German you find e.g. constructions like the so called „ellipsis of the Vorfeld" (*Paßt mir nicht so gut* [*That doesn't suit me*]), extraposition of arguments and adjuncts (*Wie sieht es aus am Dienstag? [How about Tuesday?*]) and dislocation of semantic groups (*Ich möchte um 2 Uhr einen Termin machen* [*I'd like to make an appointment at 2 o'clock*.]).

2. There is no punctuation. An utterance like *wie sieht es aus am Dienstag um 17 Uhr geht es nicht* can therefore be translated by either of the following utterances: [*How is it?*] [*On Tuesday at five p.m. it is not possible.*], [*How about Tuesday?*] [*At five p.m. it is not possible.*] or [*How about Tuesday at five p.m.?*] [*Isn't that possible?*].

3. Different sentence stress or intonation may yield a different semantics and a different translation. Whereas the sentence *wir brauchen noch einen TERMIN* should be translated by *we (still) need an appointment*, the same sentence with stress on *noch* (*wir brauchen NOCH einen Termin*) should be translated by *we need another appointment*.

4. The output of the speech recognizer is noisy. Even with good recognizers it appears quite often that the most probable recognition result does not correspond to what the speaker has said, e.g. said *dann bin ich nämlich in Münster*, understood *dann bin ich nehme ich in München.*

5. The speaker's utterances are sometimes errorneous. By „errorneous" we do not mean here cases where a speaker does not obey the rules of a normative grammar. These cases fall under problem 1. What we mean here are errors that arise from distraction of the speaker like false starts, repetitions, stuttering or sentence merging as in *heute geht es bei dir also heute also bei mir geht es heute nicht* [*today it is possible for you so today oh for you so for me it is impossible today*].

Combining a speech recognizer with a commercial translation system makes these problems very apparent. Consider for example the spoken utterance *da geht es bei mir wieder leider nicht dann bin ich nämlich in Münster ich könnte dann wieder ab 28. Mai* taken from the Verbmobil corpus. If we segment this by hand and give each segment to the translation system the output is *Again unfortunately, it doesn't go with me there. I then am namely in Münster. I then could as of 28 May again.* which is not very good English but somehow understandable (problem 1). Without the segmentation the quality of the translation decreases drastically (problem 2): *//geh// there again unfortunately, I am not it with me then namely in Münster I could then again as of 28 May.* Things get even worse if we have the system translate the most probable string produced by the speech recognizer *da geht es*

[1] The work described in this paper was partially funded by the German Federal Ministry for Research and Technology (BMBF) in the framework of the Verbmobil Project under Grant 01IV102AO. The responsibility for the contents of this paper lies with the author.

[2] This paper reports on work done by many people at different sites, namely the DFKI, the IAI and the Universities of Saarbrücken, Stuttgart, Tübingen, and Erlangen-Nürnberg and at Siemens.

bei mir weder leider nicht dann bin ich nehme ich in München ich könnte wenn wieder ab 28. Mai (problem 4): *//geh// with me there //weder// unfortunately, I am not then relieve I I could in Munich if again as of 28 May*. Similar experience can be made with problem 3 and especially problem 5.

2. THE VERBMOBIL SOLUTIONS

The solutions explored in the Verbmobil project mirror to a certain extend the problem groups. We have tried to solve problem 1 by the development of a grammar for spoken German. Problems 2 and 3 are attacked by a substantial integration of prosody recognition and processing. We tried to handle problem 4 by the use of a word hypothesis graph (word lattice) and a linguistic search routine and problem 5 by a „shallow" robust secondary analysis and translation component that combines techniques from speech act detection and information extraction.

2.1 German Syntax

The German grammar of the Verbmobil project is defined in the unification based formalism *Trace & Unification Grammar* (TUG) [2]. The basis of TUG is a context free grammar augmented with feature equations and special rule types for so called „movement rules" that give TUG a descriptive power higher than context free grammars. Prior to parsing or generation, the grammar is compiled into a format that is more suitable for efficient processing. In order to find a solution to problem 1, grammar development was guided by the insight that, though many constructions in spoken german do not obey the rules of standard German, they are nevertheless regular and can be described by a formal grammar [1]. So, for example, the so called „Vorfeldellipse", a construction in which any pronominal argument or adjunct in the Vorfeld of a sentence can be dropped in certain intersentential contexts in spoken German (*ist etwas schlecht bei mir* instead of *das ist etwas schlecht bei mir*) cannot only be easily integrated into a linguistically well motivated grammar of German but also gives further evidence for the correctness of the core description of German syntax. Other constructions like extraposition of arguments and adjuncts follow a similar kind of description that is compatible with general linguistic assumptions about german syntax.

Vorfeld	Verb	Mittelfeld
bei mir	*ist*	*das etwas schlecht* [bei mir]
etwas schlecht	*ist*	*das* [etwas schlecht] *bei mir*
das	*ist*	[das] *etwas schlecht bei mir*

Movement into Vorfeld

0 Vorfeld	Verb	Mittelfeld
[das]	*ist*	[das] *etwas schlecht bei mir*

Vorfeldellipse Movement into Vorfeld

Ellipsis of the Vorfeld: To form a German main clause any sentence element can be "moved" to the Vorfeld position. Any pronominal form in the Vorfeld can be deleted.

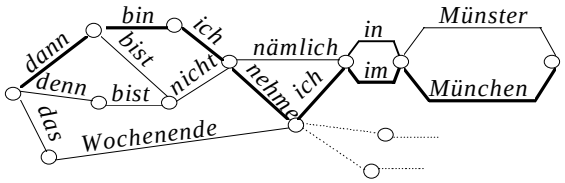
Figure 1

2.2 Linguistic search

In order to handle minor word recognition errors and solving problem 4, Verbmobil uses word hypothesis graphs (WHG) as the general interface between speech recognition and linguistic analysis. These WHGs are processed by a syntactic A*-search algorithm that finds the most probable path through the WHG according to acoustic and trigram or bigram language model scores that forms a grammatical sentence according to the grammar [3]. To reduce complexity the rest costs for each node of the WHG are computed prior to the A*-search and equal prefixes are cached. Ungrammatical prefixes lead to an early exclusion of subgraphs. It has been shown [3] that this acoustic/linguistic interface can increase the sentence recognition rate. Suppose that for the utterance *dann bin ich nämlich in Münster* [*I am in Münster then*] the most probable acoustic path would be the ungrammatical utterance *dann bin ich nehme ich im München* [*then am i take i in the München*]. The path actually found by the linguistic search is *dann bin ich nämlich in München* [*I will actually be in Munich then*] as the prefix *dann bin ich nehme* is not a grammatically valid prefix in German. Accordingly, the prefixes expanded in the WHG in figure 2 are

dann +
dann bin +
dann bin ich +
dann bin ich nehme -
dann bin ich nämlich +
dann bin ich nämlich in +
dann bin ich nämlich in München +

where „+" means „valid prefix" and „-" means „invalid prefix".



A simplified word lattice: The acoustically most probable path is in bold face. The use of the grammar during search forces the search to abandon this path on the word *nehme* thus cutting the subgraph ... out of the search.

Figure 2

Note that the recognized utterance still contains a wrong interpretation of the city name (*München* instead of *Münster*) which of course cannot be solved by the syntactic processing as both words fall into the same syntactic and semantic class. It must be noted that the success of this kind of acoustic/linguistics interaction depends heavily on the fact that the spoken utterance forms indeed a path in the WHG and that the rank of this path is not too high. Otherwise there is a high probability that the linguistic search will find a different

grammatically valid path or will give up due to time constraints.

2.3 Integration of prosodic information

Perhaps the most distinguishing feature of the Verbmobil system is its heavy use of prosodic information [4]. Prosodic information is used to solve problems 2 and 3. As the detection and basic linguistic processing of prosodic syntactic clause boundaries (PSCB) is described in another paper in this conference [5], I will only focus on the effect that in the integrated approach chosen not only the prosodic information helps the parser to choose the right segmentation, but that also the grammar helps to undo PCSBs in cases where they a ungrammatical. For example, if in an utterance like *am PSCB Montag kann ich nicht* [*on Monday, it is impossible for me*] there is a strong probability for a PSCB perhaps because the speaker has made a prosodically unclear pause after *am* the grammar would force the search to select the path without the PSCB because it does not allow clause boundaries between preposition and noun.

The processing of accent and intonation is similar to that of PSCBs. Just like PSCBs, the probability for sentence accent and intonation is encoded in each edge of the WHG. For accent two value slots (has accent/has no accent) and for intonation three value slots (rise,fall,mid) are provided by the prosody recognition module. To transport this information into the syntactic and semantic processing modules, during the linguistic search this information is mapped into a so called prosodic word form (PROSWOF). There are 10∗10∗3 (=300) different PROSWOFS corresponding to 10 different values for accent and boundary (computed from their probabilites) and the values *rise*, *fall*, or *prog* (mid) for intonation. Thus a PROSWOF has the form a[0-9]_g[0-9]_i{rise,fall,prog}. In the lexical entries for these PROSWOFS the values are copied to grammatical features, e. g.

lexicon(a3_g4_irise,proswof:0) |
0:accent = 3,
0:boundary = 4,
0:intonation = rise.

In the grammar, for each terminal category a rule is introduced that expands the syntactic category to the lexical category and the PROSWOF, e. g.

adv_pr:0 → adv:1, proswof:2 |
0:accent = 2:accent,
0:boundary = 2:boundary,
0:intonation = 2:intonation.

By means of these rules the prosodic information becomes accessible to the further semantic processing (for semantic processing in Verbmobil, see [6]). The semantic component decides for example that an ambiguous sentence like *kommen sie in mein Büro* will be interpreted as imperative if the value of the intonation feature is *fall* and as a yes/no question if its value is *rise*. Accordingly, the translation will be *Come to my office* or *Do you come to my office?*

The accent feature is mapped in the semantic interpretation to an accent predication over the label of the semantic predicate that corresponds to the stressed word, as depicted in figure 3.

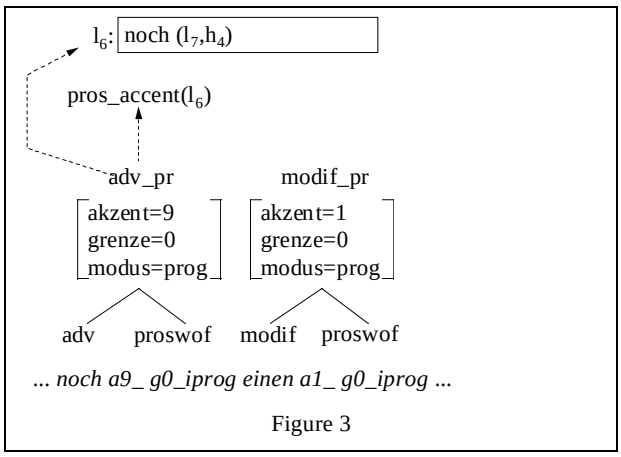


Figure 3

By means of semantic interpretation rules this accent predication is mapped to scope information. So, in the sentence *ich will NOCH einen Termin ausmachen*, *noch* gets scope over *einen* which then in turn is mapped to *another* by the transfer rules.

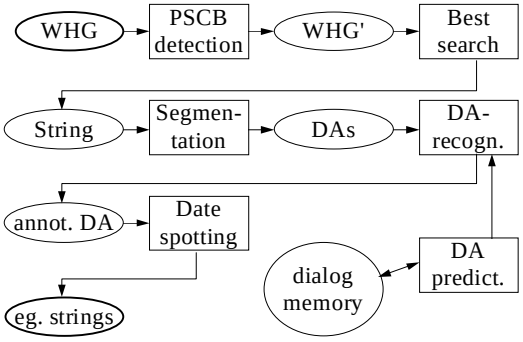
2.4 Shallow translation

During the project it turned out that the approach taken for parsing was not sufficiently robust to handle illformed or badly recognized utterances. Especially, corrupted input and WHGs that do not contain a grammatical path could not be translated correctly. For these reasons, Verbmobil additionally provides a new robust translation mechanism that provides translations that are less detailed but often sufficient in the dialog context. This approach is called „dialog act based reductionist translation“. It is based on the observations that (1) the scheduling domain is very restricted and can (2) be reduced to a small set of so called dialog acts (DA) like *greeting*, *proposal*, *rejection* etc. that convey the main pragmatical meaning of an utterance and that (3) human interpreters also very often reduce the translation to the information bits that are most important in a given dialog context. The hypothesis for the reductionist translation approach then is that a contextually adequate translation can be achieved by dialog act recognition and partial parsing of main information bits. The dialog act (speech act) of an utterance can be detected by means of simple models. As the application domain of Verbmobil is restricted to the scheduling task these dialog acts can be given a narrow interpretation s. t. *proposal* means „propose a date“, *rejection* means „reject a proposed date“ etc. and the information bits are restricted to date expressions such as *on monday at six o'clock*. Once a dialog act is detected we can search the input for appropriate date expressions with very simple grammars. The translation can then be provided by predefined patterns and the translations of the date expressions.

For the reduced translation we currently use the following dialog acts and translation patterns.

greeting	*hi!*
introduce	*nice to meet you*
initialisation	*i would like to make a date*
motivate	*it's because of a date*
suggest	*how about meeting <DATE> ?*
accept	*<DATE> is fine with me.*
reject	*<DATE> doesn't suit me.*
give_reason	*<DATE> I'm busy.*
request_suggest_date	*When would it fit you?*
request_suggest_location	*Where shall we meet?*
feedback_acknowledgement	*okay,<DATE>*
feedback_reservation	*oh, no(, not <DATE>*)
clarify_question	*you mean <DATE>?*
clarify	*well, <DATE>*
deliberate	*let's see,<DATE>*
garbage	
thank	*thanks a lot!*
bye	*bye, see you<DATE>*

Figure 4 gives an overview of the architecture of the reductionist translation approach. From a WHG first the acoustically most probable path is detected. The resulting text string is eventually segmented into different instancies of dialog acts. Each segment is annotated by the dialog recognition with the most probable dialog act according to the dialog recognition model and the dialog prediction. From each segment the date spotter detects date expressions and inserts their translations into the translation patterns that correspond to the detected dialog act and produces the english string that is sent to the synthesizer.



Architecture of the reductionist translation
Figure 4

The following table shows the different steps in a example:

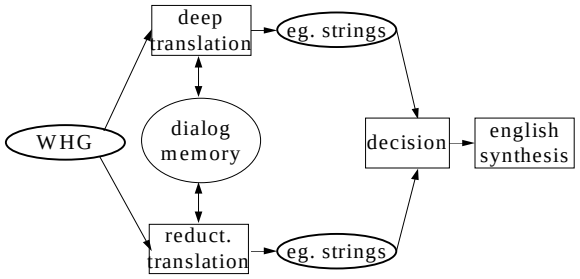
Utterance
ich wollte mit ihnen einen Termin ausmachen wie wär's am fünfzehnten

Speech recognition
gut geben Termin ausmachen BND P erst am fünfzehnten

Dialog act recognition
INIT: gut geben Termin ausmachen
SUGGEST: P erst am fünfzehnten

Date spotting and translation
I want to make a date. How about meeting, say the fifteenth?

The reductionist translation module is integrated in the Verbmobil system as a secondary analysis device. Both „deep" and reductionist translation are processed in parallel. After completion the system decides for one or the other translation [7].



Embedding of the reductionist translation in the Verbmobil system

Figure 5

REFERENCES

[1] Hans Ulrich Block and Stefanie Schachtl, 1995, What a Grammar of Spoken Dialogues has to deal with. in: Hayer, G. and H. Haugeneder, *Language Engineering.* Wiesbaden (Vieweg).

[2] Hans Ulrich Block and Stefanie Schachtl, 1992, Trace and Unification Grammar, *Proc. COLING 92*, pp. 87-93.

[3] Ludwig A. Schmid, 1994, Parsing Word Graphs Using a Linguistic Grammar and a Statistical Language Model, *Proc. ICASSP 1994*, Vol. 2, pp. 41-44.

[4] H. Niemann, E. Nöth, A. Kießling, R. Kompe, A. Batliner, 1997, Prosodic Processing and its use in Verbmobil. *Proc. ICASSP-97* (to appear).

[5] R. Kompe, A. Kießling, H. Niemann, E.Nöth, A. Batliner, S. Schachtl, T. Ruland, H. U. Block, 1997, Improving Parsing of Spontaneous Speech with the Help of Prosodic Boundaries, *Proc. ICASSP-97* (to appear).

[6] J. Bos, B. Gambäck, Ch. Lieske, Y. Mori, M. Pinkal, K. Worm, 1996, Compositional Semantics in Verbmobil. *Proc. Int. Conf. on Computational Linguistics*, Vol. 1, pp. 131-136.

[7] W. Wahlstser, T. Bub, A. Waibel, 1997, Verbmobil: The Combination of Deep and Shallow Processing for Spontaneous Speech Translation. *Proc. ICASSP-97* (to appear).