A Model for Robust Processing of Spontaneous Speech by Integrating Viable Fragments*

Karsten L. Worm
Universität des Saarlandes
Computerlinguistik
D-66041 Saarbrücken, Germany
worm@coli.uni-sb.de

Abstract

We describe the design and function of a robust processing component which is being developed for the Verbmobil speech translation system. Its task consists of collecting partial analyses of an input utterance produced by three parsers and attempting to combine them into more meaningful, larger units. It is used as a fallback mechanism in cases where no complete analysis spanning the whole input can be achieved, owing to spontaneous speech phenomena or speech recognition errors.

1 Introduction

In this paper we describe the function and design of the *robust semantic processing* component which we are currently developing in the context of the Verbmobil speech translation project. We aim at improving the system's performance in terms of coverage and quality of translations by combining fragmentary analyses when no spanning analysis of the input can be derived because of spontaneous speech phenomena or speech recognition errors.

2 The Verbmobil Context

Verbmobil (Wahlster, 1997) is a large scale research project in the area of spoken language translation. Its goal is to develop a system that translates negotiation dialogues between speakers of German, English and Japanese in face-to-face or video conferencing situations. The integrated system developed during the first project phase (1993-96), the *Research Prototype*, was successfully demonstrated in autumn 1996 (Bub et al., 1997). The final *Verbmobil Prototype* is due in 2000.

Verbmobil employs different approaches to machine translation. A semantic transfer approach (Dorna and Emele, 1996) based on a deep linguistic analysis of the input utterance competes with statistical, example based and dialogue act based translation approaches.

The spoken input is mapped onto a *word hypothesis graph* (WHG) by a speech recognizer. A prosody component divides the input into segments and annotates the WHGs with prosodic features. Within the semantic transfer line of processing, three different parsers (an HPSG-based chart parser, a chunk parser using cascaded finite state automata, and a statistical parser) attempt to analyse the paths through the WHG syntactically and semantically. All three deliver their analyses in the VIT format (see 3). The parsers' work is coordinated by an *integrated processing* component which chooses paths through the WHG to be analysed in parallel by the parsers until an analysis spanning the whole input is found or the system reaches a time limit.

Since in many cases no complete analysis spanning the whole input can be found, the parsers produce partial analyses along the way and send them to the *robust semantic processing* component, which stores and combines them to yield analyses of larger parts of the input. We describe this component in section 5.

The relevant part of the system's architecture is shown in Figure 1.

3 The VIT Format

The VIT (short for *Verbmobil Interface Term*) was designed as a common output format for the two alternative and independently developed syntactic-semantic analysis components of the first project phase (Bos et al., 1998). Their internal semantic formalisms differed, but both had to be attached to a

* The author wishes to thank his colleagues Johan Bos, Aljoscha Burchardt, Björn Gambäck, Walter Kasper, Bernd Kiefer, Uli Krieger, Manfred Pinkal, Tobias Ruland, C. J. Rupp, Jörg Spilker, and Hans Weber for their collaboration. This research was supported by the German Federal Ministry for Education, Science, Research and Technology under grant no. 01 IV 701 R4.

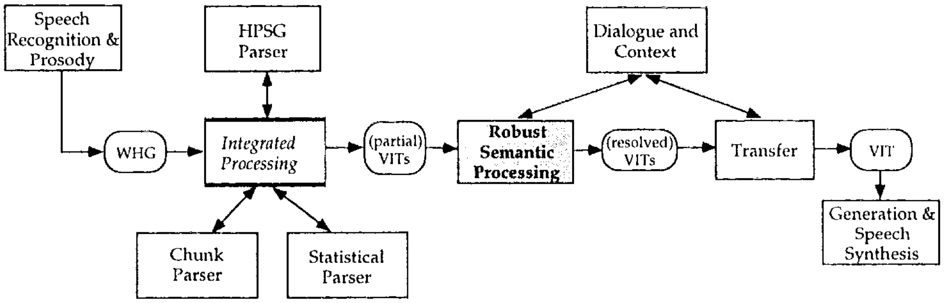

Figure 1: Part of the system architecture.

single transfer module. The need for a common output format is still present, since there are three alternative syntactic-semantic parsing modules in the new Verbmobil system, all of which again produce output for just one transfer module.

```
(1) vit(vitID(sid(1,a,ge,0,20,1,ge,y,
              semantics),
            [word(montag,13,[l116]),
             word(ist,14,[l117]),
             word(gut,15,[l110])]),
      index(l113,l109,i104),
      [decl(l112,h105),
       gut(l110,i105),
       dofw(l116,i105,mon),
       support(l117,i104,l110),
       indef(l111,i105,l115,h106)],
      [ccom_plug(h105,l114),
       ccom_plug(h106,l109),
       in_g(l112,l113),
       in_g(l117,l109),
       in_g(l116,l115),
       in_g(l111,l114),
       leq(l114,h105),leq(l109,h106),
       leq(l109,h105)],
      [s_sort(i105,time)],
      [],
      [num(i105,sg),pers(i105,3)],
      [ta_mood(i104,ind),
       ta_tense(i104,pres),
       ta_perf(i104,nonperf)],
      []
    )
```

The VIT can be viewed as a theory-independent representation for underspecified semantic representations (Bos et al., 1996). It specifies a set of discourse representation structures, DRSs, (Kamp and Reyle, 1993). If an utterance is structurally ambiguous, it will be represented by one VIT, which specifies the set of DRSs corresponding to the different readings of the utterance.

Formally, a VIT is a nine-place PROLOG term. There are slots for an identifier for the input segment to which the VIT corresponds, a list of the core semantic predicates, a list of scopal constraints, syntactic, prosodic and pragmatic information as well as tense and aspect and sortal information. An example of a VIT for the sentence *Montag ist gut* ('Monday is fine') is given in (1).

4 Approaches to Robustness

There are three stages in processing where a speech understanding system can be made more robust against spontaneous speech phenomena and recognizer errors: *before*, *during*, or *after parsing*. While we do not see them as mutually exclusive, we think that the first two present significant problems.

4.1 Before parsing

Detection of self corrections on transcriptions before parsing has been explored (Bear et al., 1992; Nakatani and Hirschberg, 1993), but it is not clear that it will be feasible on WHGs, since recognition errors interfere and the search space may explode due to the number of paths. Dealing with recognition errors before parsing is impossible due to lack of structural information.

4.2 During parsing

Treating the phenomena mentioned during parsing would mean that the grammar or the parser would have to be made more liberal, i. e. they would have to accept strings which are ungrammatical. This is problematic in the context of WHG parsing, since the parser has to simultaneously perform two tasks: *Searching* for a path to be analysed and *analysing* it as well.

If the analysis procedure is too liberal, it may already accept and analyse an ungrammatical path when a lower ranked path which is grammatical is also present in the WHG. I. e., the search through the WHG would not be restricted enough.

5 Robust Semantic Processing

Our approach addresses the problems mentioned after parsing. In many cases the three parsers will not be able to find a path through the WHG that can be assigned a complete and spanning syntactic-semantic analysis. This is mainly due to two factors:

- spontaneous speech phenomena, and
- speech recognition errors.

However, the parsers will usually be able to deliver a collection of partial analyses — each covering a part of a path through the WHG.

The goal of the *robust semantic processing* component in Verbmobil-2 is to collect these partial analyses and try to put them together on the basis of heuristic rules to produce deep linguistic analyses even if the input is not completely analysable.

We speak of robust semantic processing since we are dealing with VITs which primarily represent semantic content and apply rules which refer to semantic properties and semantic structures.

The task splits into three subtasks:

1. *Storing* the partial analyses for different WHG (sub)paths from different parsers;
2. *Combining* partial analyses to yield bigger structures;
3. *Choosing* a sequence of partial analyses from the set of hypotheses as output.

These subtasks are discussed in the following subsections. Section 5.4 contains examples of the problems mentioned and outlines their treatment in the approach described.

5.1 Storing Partial Analyses

The first task of the robust semantic processing is to manage a possibly large number of partial analyses, each spanning a certain sub-interval of the input utterance.

The basic mode of processing — store competing analyses and combine them to larger analyses, while avoiding unnecessary redundancy — resembles that of a chart parser. Indeed we use a chart-like data structure to store the competing partial analyses delivered by the parsers and new hypotheses obtained by combining existing ones. All the advantages of the chart in chart parsing are preserved: The chart allows the storage of competing hypotheses, even from different sources, without redundancy.

Since the input to the parsers consists of WHGs rather than strings, the analyses entered cannot refer to the string positions they span. Rather they have to refer to a time interval. This means also that the chart cannot be indexed by string positions, but is indexed by the time frames the speech recognizer uses. This makes necessary slight modifications to the chart handling algorithms.

5.2 Combining Partial Analyses

We use a set of heuristic rules to describe the conditions under which two or more partial analyses should be combined, an analysis should be left out or modified. Each rule specifies the conditions under which it should be applied, the operations to be performed, and what the result of the rule application is. Rules have the following format (in PROLOG notation):

```
[Cond1,...,CondN] --->
        [Op1,...,OpN] & Result.
```

The left hand side consists of a list of conditions on partial analyses, `Cond1` being a condition (or a list of conditions) on the first partial analysis (VIT), etc., where the order of conditions parallels the expected temporal order of the analyses. When these conditions are met, the rule fires and the operations `Op1` etc. are performed on the input VITs. One VIT, `Result`, is designated as the result of the rule. After applying the rule, an edge annotated with this VIT is entered into the chart, spanning the minimum time frame that includes the spans of all the analyses on the left hand side. Examples for rules are given in 5.4.

5.3 Choosing a Result

When no more analyses are produced by the parsers and all applicable rules have been applied, the last step is to choose a 'best' sequence of analyses from the chart which covers the whole input and deliver it to the transfer module. In the ideal case, there will be an analysis spanning the whole input.

Currently, we employ a simple search which takes into account the acoustic scores of the WHG paths the analyses are based on, together with the length and coverage of the individual analyses.

The length is defined as the length of the temporal interval an analysis spans; an analysis with a greater length is preferred. The coverage of an analysis is the sum of the lengths of the component analyses it consists of. Note that the coverage of an analysis will be less than its length iff some material inside the interval the analysis spans has been left out in the analysis; hence length and coverage are equal for the analyses produced by the parsers.[1] Analyses with greater coverage are preferred.

5.4 Examples

The examples in this section are taken from the Verbmobil corpus of appointment scheduling dialogues. The problems we address here appeared in WHGs produced by a speech recognizer on the original audio data.

5.4.1 Missing preposition

Since function words like prepositions are usually short, speech recognizers often have trouble recognizing them. Consider an example where the speaker uttered *Mir wäre es am liebsten in den nächsten zwei Wochen* ('During the next two weeks would be most convenient for me'). However, the WHG contains no path which includes the preposition *in* in an appropriate position. Consequently, the parsers delivered analyses for the segments *Mir wäre es am liebsten* and *den nächsten zwei Wochen.*

These fragments are handled by two rules. The first turns a temporal NP like the second fragment into a temporal modifier, expressing that something is standing in an underspecified temporal relation to the temporal entity the NP denotes:

```
[temporal_np(V1)] --->
   [typeraise_to_mod(V1,V2)] & V2.
```

Then a very general rule can apply that modifier to the proposition expressed by the first fragment:

```
[type(V1,prop),type(V2,mod)] --->
   [apply(V2,V1,V3)] & V3.
```

5.4.2 Self-Correction of a Modifier

Here the speaker uttered *Wir treffen uns am Montag, nein, am Dienstag* ('We will meet on Monday, no, on Tuesday'). The parsers deliver three fragments, the first being a proposition containing a modifier, the second an interjection marking a correction, and the third a modifier of the same type as the one in the proposition. Under these conditions, we replace the modifier inside the proposition with the one uttered after the correction marker:

```
[[type(V1,prop),
  has_mod(V1,M1,ModType)],
 correction_marker(_),
 [type(V2,mod),
  has_mod(V2,M2,ModType)]]
---> [replace_mod(V1,M1,M2,V3)] & V3.
```

5.4.3 Self-Correction of a Verb

In this case, the speaker uttered *Am Montag treffe ... habe ich einen Termin.*, i. e. decided to continue the utterance in a different way than originally intended. The parsers deliver fragments for, among others, the substrings *am Montag*, *treffe*, *habe*, *ich*, and *einen Termin* (all the partial analyses received from the parsers and built up by robust semantic processing are shown in the chart[2] in Figure 2).

Robust semantic processing then builds analyses by applying modifiers to verbal predicates (e. g., analyses 71, 108) and verbal functors to possible arguments (e. g., 20, 106, 47). The latter is done by the following two rules:

```
[type(V1,Type),unbound_arg(V2,Type)]
      ---> [apply(V2,V1,V3)] & V3.
[unbound_arg(V1,Type),type(V2,Type)]
      ---> [apply(V1,V2,V3)] & V3.
```

Note that *einen Termin* is not considered to be a possible argument of the verb *treffe* since that would violate the verb's sortal selection restrictions.

After all partial analyses produced by the parsers have been entered into the chart and all applicable rules have been applied, there is still no spanning analysis (all analyses in Figure 2 are there, except the spanning one numbered 105). In such a case, the robust semantic processing component proceeds by extending active edges over passive edges which end in a chart node in which only one passive edge ends, or all passive edges ending there correspond to partial analyses still missing arguments.

In this example, this applies to the node in which edges 1 and 71 end, which both are missing the two arguments of the transitive verb *treffe*. Application of the proposition modification rule mentioned in Section 5.4.1 to the modifyer PP *am Montag* has led to an active edge still looking for a proposition. This is now being extended to end at the same node as the two passive edges missing arguments.

[1] The chunk parser may be an exception here since it sometimes leaves out words it cannot integrate into an analysis.

[2] The analyses in the chart are numbered; the numbers in square brackets indicate the immediate constituents an analysis has been built from by robust semantic processing. I. e., analyses with an empty list of immediate constituents have been produced by a parser.

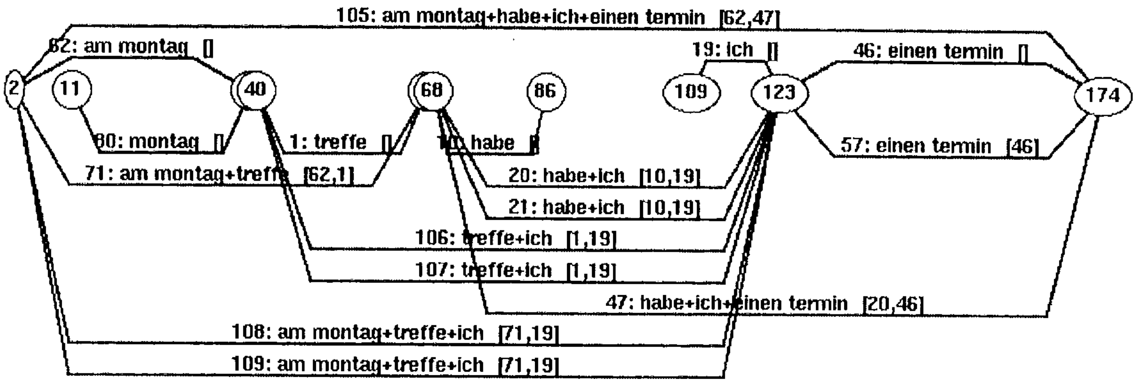

Figure 2: The chart for *Am Montag treffe . . . habe ich einen Termin.*

There, it finds an edge corresponding to a proposition, namely edge 47, which had been built up earlier. The result is passive edge 105 spanning the whole input and expressing the right interpretation.

6 Related Work

An approach similar to the one described here was developed by Rosé (Rosé, 1997). However, that approach works on interlingual representations of utterance meanings, which implies the loss of all linguistic constraints on the combinatorics of partial analyses. Apart from that, only the output of one parser is considered.

7 Conclusion and Outlook

We have described a model for the combination of partial parsing results and how it can be applied in order to improve the robustness of a speech processing system. A prototype version was integrated into the Verbmobil system in autumn 1997 and is currently being extended.

We are working on improving the selection of results by using a stochastic model of VIT sequence probabilities, on the extension of the rule set to cover more spontaneous speech phenomena of German, English and Japanese, and on refining the mechanism for extending active edges to arrive at a spanning analyses.

References

John Bear, John Dowding, and Elizabeth Shriberg. 1992. Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proc. of the 30 th ACL*, pages 56–63, Newark, DE.

Johan Bos, Björn Gambäck, Christian Lieske, Yoshiki Mori, Manfred Pinkal, and Karsten Worm. 1996. Compositional semantics in Verbmobil. In *Proc. of the 16th COLING*, pages 131–136, Copenhagen, Denmark.

Johan Bos, Bianka Buschbeck-Wolf, Michael Dorna, and C. J. Rupp. 1998. Managing information at linguistic interfaces. In *Proc. of the 17th COLING/36th ACL*, Montréal, Canada.

Thomas Bub, Wolfgang Wahlster, and Alex Waibel. 1997. Verbmobil: The combination of deep and shallow processing for spontaneous speech translation. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 71–74, München, Germany. IEEE Signal Processing Society.

Michael Dorna and Martin C. Emele. 1996. Semantic-based transfer. In *Proc. of the 16th COLING*, pages 316–321, Copenhagen, Denmark.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic.* Kluwer, Dordrecht.

Christine Nakatani and Julia Hirschberg. 1993. A speech-first model for repair detection and correction. In *Proc. of the 31 th ACL*, pages 46–53, Columbus, OH.

Carolyn Penstein Rosé. 1997. *Robust Interactive Dialogue Interpretation.* Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA. Language Technologies Institute.

Wolfgang Wahlster. 1997. Verbmobil: Erkennung, Analyse, Transfer, Generierung und Synthese von Spontansprache. Verbmobil-Report 198, DFKI GmbH, Saarbrücken, June.