# Speech Understanding by Alternating Control: a Semantic Network Based Approach*

**B. Seestaedt, F. Kummert, G. Sagerer**

Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik
Postfach 8640, D-4800 Bielefeld 1, Federal Republic of Germany
Tel.: +49–(0)521–106–5329, Fax: +49–(0)521–106–2992

**Abstract**

*This paper regards control facilities of a practical tested speech understanding system based on a generalized semantic network conception. The system can significantly enhance results of signal recognition and therefore produces more likely a correct interpretation. Elementary control processes driven by data or from the knowledge base and their alternation are demonstrated to explain this capacity. The problem independent system kernel can be adapted for a variety of pattern recognition tasks.*

**Keywords:** *Speech Processing, Speech Understanding*

* This research was supported by the German Research Foundation (DFG). Only the authors are responsible for the contents of this publication.

## 1 Introduction

The dicussed speech understanding system works on the basis of an acoustic signal recognition phase. As a result of this phase a number of n best evaluated word hypotheses is preprocessed. In general, the word recognition can not provide ideal hypotheses, which are defined as assignments from signal segments to lexicographic units. The Knowledge restricted in special discours fields (train information in the follow) and generalized linguistic knowledge have to be used in a speech understanding system which starts and controls competition and combination processes between the supposed word hypotheses. A machine has to recognize the speakers intention and to generate an adequate response concentrating on those parts of the spoken sentences relevant for its meaning. Modelling the functional roles of such parts within nodes and a variety of their relations as links a semantic network is usefull to create the computational knowledge base at design time. The knowledge base motivates several control processes for the linguistic analysis. On the other hand, speakers emphasize some words more than others with respect to pragmatical relevance. Therefore these entries fall often under the best words [3]. This motivates a lexically controlled process started from the bottom level of the network. An short analysis exemplifies in section 3 the alternating phases.

## 2 Knowledge Base and Process Model

The **semantic network** conception [4] has the capacity to integrate different knowledge levels in one framework. For imagination the levels can be 3D represented as half-landings spanned by three axies (Figure 1a). Each axis stands for one of the possible link types. The depicted part_relation/specialization plane corresponds to the well known syntagmatic/paradigmatic categorization. In the case of linguistic analysis e.g. a **specialization** points from 'word category' to 'noun' and a **part_relation** from 'nominal phrase' to 'noun'. By a **concretization** the control flow from plane to plane can be directed. At runtime the first accessed plane is the hypothesis level from which can be traced up to the plane of the goal concepts. Choosing a notation of prefixes for concept names by `H_` , `SY_`, `S_`, `P_` and `D_` for the levels of hypotheses, syntax, semantic, pragmatic and dialog, the concretization links are labeled by such prefixes. A goal concept which models the

speakers intention to request a train information can be denoted by P_TRAIN_INFORMATION, against a goal concept P_TRAIN_CONNECTION models the case that the arrival of a train is the departure of an other. The last concept as describing the more general case points to the other by a specialization link. SY_concepts are the syntactical constituents, e.g. SY_NP as noun phrase, the S_concepts connotate functional roles corresponding to the deep cases and the verb frames of Fillmore [1 ]. Verb frames were indicated by VF_ prefixes. The instantiation of an S_VF_concept by an admissible verb hypothesis is a central point of the analysis.

The **process model** is to be defined as a set of basic processes of creation/replacement of nodes/links within a three-folded working area. These processes start in the work memory from copies of the **concepts** in the knowledge area (Figure 1 b). The main process is the **instantiation** or creation of an **instance** I() as a successor of such a copy denoted as **modified concept** M() in the work memory if all part and concrete nodes linked with M() are already instantiated. In Figure 2a the initializing process is depicted to access a word from the hypotheses set which has to instantiate the start node M(SY_NPR). Such a start node is to choose with respect to the application field. Processing in the work memory culminates in climbing the network layers up to the instantiation of a goal node. It can be regarded as a step-by-step **expansion** of the runtime net using a set of 6 rules (defined only in terms of problem independent network categories). The main processes in the search space are creations of nodes of the **search tree** summarizing subsequent expansion series and branching of nodes with respect to designed alternatives of the instantiation process. Finding an optimal search path through the branches is controlled by a judgement vector within an A* strategy [2]. In the priority scale of judgements linguistic consistency has the highest position.

## 3 Linguistic Analysis by Alternating Control

Analyzing the hypotheses set of the signal equivalent of the sentence '*Ich will nach Leuven fahren*' ('I want to go to Leuven') the exemplary control processes are represented by Figures 2,3,4. Figure 2a shows one of the possible initializing processes selecting 'proper noun' as starting word category. Clearly, all matched word hypotheses start an own instantiation process summarized in an own search tree node. The instantiation stops at the concept level of a prepositional phrase M(SY_PP) because of the lack of a preposition instance. After this the pragmatic level is approached by expansion. M(P_ARRIVAL),-M(P_DEPARTURE) are created as alternatives followed by branching in the search tree. Figure 2b exemplifies a basic process **hypothesis prediction** activated in this case with respect to pragmatic consistency, e.g. '*in*' and '*nach* instances are created and the search tree has to be branched. This process relaxes the earlier hypotheses set also by predicting positional information, e.g.'*nach Leuven*'. The control alternates between bottom-up and top-down by activating a further instantiation as shown in Figure 2c and stops because of the context dependency of the S_GOAL concept which models a deep case. The process conceptualized for such states is the creation of a **partial** (or intermediate) **instance** Ip(S_VF_ or P_VF_) which activates a hypothesis prediction for the deep case generating verb. Figure 3 summarizes the complex process of **verb frame hypothesis prediction** by the bigger or

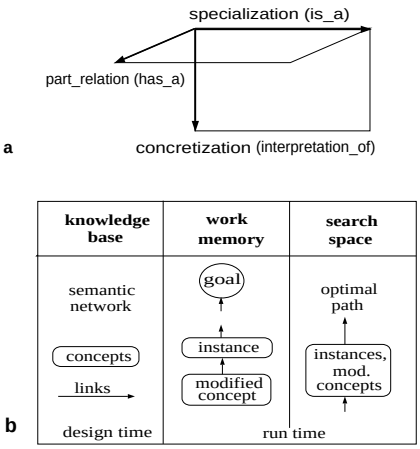


Figure 1 a) axis system spanning the network; b) three-folded work field of analysis

doubled arrows integrating an underlying process of hypothesis prediction for a 'modal verb'. The double arrows indicate the alternating of the top down process from `M(S_VF_FAHREN)` up to the creation of verb instances and the bottom up process creating `Ip(S_VF_FAHREN)`. The thinner arrows indicate an expansion intermediately started from `M(S_VF_FAHREN)` which continues the context dependance to te pragmatic level. The content of a **search tree node** can be characterized as subnet of the work memory after reducing the replaced nodes/links, e.g. `M(),Ip()`. In this sense the Figures 3,4 represent search tree nodes. The last Figure summarizes the following processes: at first a pronoun hypothesis prediction is activated from `Ip(S_VF_FAHREN)` alternating expansion up to the complete instantiation of the syntactical context node. In this state the partial instantiation of the corresponding pragmatical context was activated. The arrows of oppositional direction indicate that control works at different levels. At last `M(P_ARRIVAL)` can be instantiated from where the goal node level is directly approachable. From the two approachable goals `P_TRAIN_INFORMATION`, `P_TRAIN_CONNECTION` the latter, being the more general one, has to be instantiated. Considering the signal segment derived from the included word hypotheses a sufficient covering of the speech signal can be proved. If necessary a further hypotheses prediction is to be started to instantiate the goal concept linked by specialization.

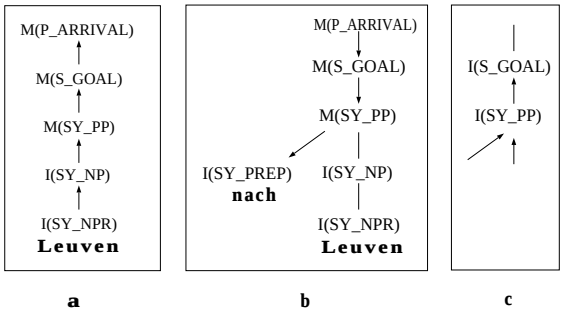


Figure 2 shows the initializing phase in three steps: a) hypothesis expansion up to pragmatic level; b) hypothesis prediction; c) further instantiation

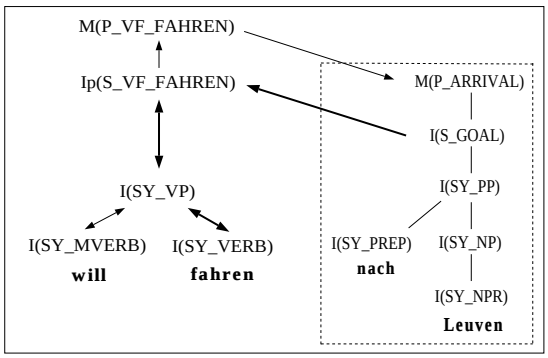


Figure 3 shows resolution of context dependencies on semantic and pragmatic level

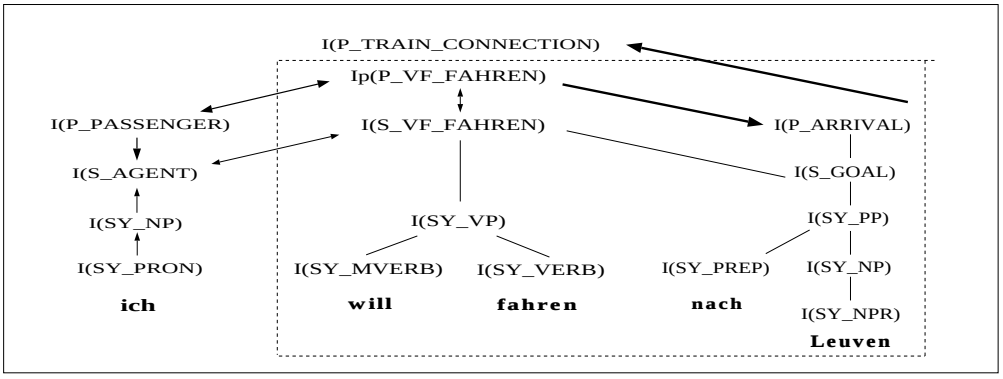


Figure 4 shows expansion up to a goal instance after resolution of the corresponding semantical and pragmatical contexts

| test type | mean cpu time in sec. | correct interpretation | incomplete interpretation | interpretation error | stopped analysis |
|---|---|---|---|---|---|
| T0 | 37 | 60 | 21 | - | - |
| T50 | 68 | 56 | 15 | 5 | 5 |
| T100 | 133 | 43 | 14 | 15 | 9 |
| T200 | 205 | 15 | 10 | 29 | 27 |

Table 1

## 4 Test results

The system was tested using a set of 81 spoken sentences on the basis of 1071 lexical entries and using a DEC RISC station 5000 of 24 MByte work memory. A test T0 was restricted to the word hypotheses equivalent to the words that were really spoken. Table 1 summarizes the results of T0 and test variants T50,T100,T200 using the n best hypotheses (number n indicated after 'T'). The evaluation of the results has to take into account a detection rate of spoken words smaller than 50 % in the optimal word chain. Cases of incomplete interpretation were caused by the run time criteria that stops processing if 80 % of the signal are covered. But the speakers intention was reconstructed such that a correct machinal response could be given. Therefore, a significant enhancement of the acoustic recognition results can be attested resulting in more correct interpretations.

## References


[1] Ch. Fillmore. A case for case. In E. Bach and R. T. Harms, editors, *Universals in Linguistic Theory*, pages 1–88. Holt, Rinehart and Winston, New York, 1968.

[2] F. Kummert. *Flexible Steuerung eines sprachverstehenden Systems mit homogener Wissensbasis*. PhD thesis, Technische Fakultät der Universität Erlangen-Nürnberg, 1991.

[3] E. Nöth. *Prosodische Information in der automatischen Spracherkennung — Berechnung und Anwendung*. PhD thesis, Technische Fakultät der Universität Erlangen-Nürnberg, 1989.

[4] G. Sagerer and F. Kummert. Knowledge based systems for speech understanding. In H. Niemann, M. Lang, and G. Sagerer, editors, *Recent Advances in Speech Understanding and Dialog Systems*, pages 421–458. Springer Verlag, Berlin, Heidelberg, New York, 1988.