MULTILEVEL SEMANTIC ANALYSIS IN AN AUTOMATIC SPEECH UNDERSTANDING AND DIALOG SYSTEM

Ute Ehrlich
Lehrstuhl für Informatik 5 (Mustererkennung)
Universität Erlangen-Nürnberg
Martensstr. 3, 8520 Erlangen, F. R. Germany

ABSTRACT

At our institute a speech understanding and dialog system is developed. As an example we model an information system for timetables and other information about intercity trains.

In understanding spoken utterances, additional problems arise due to pronunciation variabilities and vagueness of the word recognition process. Experiments so far have also shown that the syntactical analysis produces a lot more hypotheses instead of reducing the number of word hypotheses. The reason for that is the possibility of combining nearly every group of word hypotheses which are adjacent with respect to the speech signal to a syntactically correct constituent. Also, the domain independent semantic analysis cannot be used for filtering, because a syntactic sentence hypothesis normally can be interpreted in several different ways, respectively a set of syntactic hypotheses for constituents can be combined to a lot of semantically interpretible sentences. Because of this combinatorial explosion it seems to be reasonable to introduce domain dependent and contextual knowledge as early as possible, also for the semantic analysis. On the other hand it would be more efficient prior to the whole semantic interpretation of each syntactic hypothesis or combination of syntactic hypotheses to find possible candidates with less effort and interpret only the more probable ones.

1. Introduction

In the speech understanding and dialog system EVAR (Niemann et al. 1985) developed at our institute there are four different modules for understanding an utterance of the user (Brietzmann 1984): the syntactic analysis, the task-independent semantic analysis, the domain-dependent pragmatic analysis, and another module for dialog-specific aspects. The semantic module disregards nearly all of the thematic and situational context. Only isolated utterances are analyzed. So the main points of interests are the semantic consistency of words and the underlying relational structure of the sentence. The analysis of the functional relations is based on the valency and case theory (Tesniere 1966, Fillmore 1967). In this theory the head verb of the sentence determines how many noun groups or prepositional groups are needed for building up a syntactically correct and semantically consistent sentence. For these slots in a verb frame further syntactic and semantic restrictions can also be given.

2. Semantic and Pragmatic Consistency

Semantic Consistency

The semantic knowledge of the module consists of lexical meanings of words and selectional restrictions between them. These restrictions are possible for a special word, for example the preposition 'nach' ('to Hamburg') requires a noun with the meaning LOCation. In the case of a frame they are for a whole constituent; for example, the verb 'wohnen' ('to live in Hamburg') needs a prepositional group also with the meaning LOCation.

The selectional restrictions are expressed in the dictionary by the feature SELECTION. The semantic classes (features) are hierarchically organized in a way, so that all subclasses of a class also are accepted as compatible. For example, if a word with the semantic class CONcrete is required, also a word with the class ANImate (a subclass of CONcrete) or with the class HUman (a subclass of ANImate) is accepted.

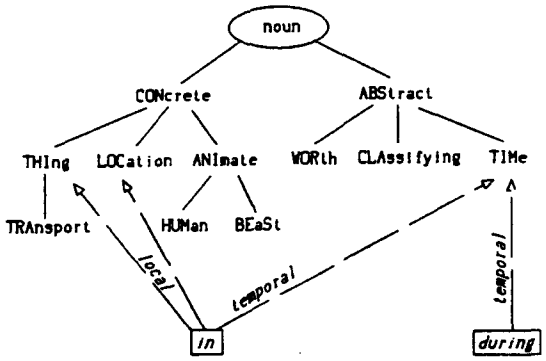


Fig. 1: Semantic classification of nouns (part)

In Fig. 1 a part or our semantic classification system for nouns is shown. For each preposition or adjective there can be determined with which nouns they could be combined. That is done by selecting the semantic class of the head noun of a noun group or prepositional group. For example 'in' in its temporal meaning can be used with nouns as

Fig. 2 shows, how this system could be used to solve ambiguities.

For example:

coach

coach.1.1: "railway carriage"
CLASS: TRAnsport, LOCation
coach.1.2: "privat tutor, trainer in athletics"
CLASS: ACting_Person

****

in

in.1.1: "in the evening"
CLASS: DURation
SELECTION: TIMe

in.1.2: "in the room"
CLASS: PLAce
SELECTION: LOCation

Fig. 2: Semantic Interpretation of "in the coach"

Although there are 4 possibilities for combining the words in their different meanings only one possibility (in.1.2 | coach.1.1) is semantic consistent.

At this time no scoring is provided for 'how compatible' a group of words is, only if it is semantically consistent or not.

Pragmatic Consistency

Because of the above mentioned combinatorial explosion it seems to be useful to integrate also at this task-independent stage of the analysis some domain dependent information.

This pragmatic information should be handled with as few effort as possible. On the other side the effect as a filter should also be as good as possible. What is not intended is to introduce here a first structural analysis but to decide whether a group of words pragmatically fit together or not, only dependent on special features of the words itself.

For this reason here it is tried to check the pragmatic consistency of groups of words or constituents and give them a pragmatic priority. This priority is not a measure for correctness of the hypothesis, but determines in which order pragmatically checked hypotheses should be further analyzed. It indicates, whether all words of such a group can be interpreted in the same pragmatic concept, and how much the set of possible pragmatic concepts could be restricted.

In our system the pragmatic (task-specific) knowledge is represented in a semantic network (Brietzmann 1984) as is the knowledge of the semantic module. The network scheme is influenced by the formalism of Stuctured Inheritance Networks (Brachman 1978). In this pragmatic network at the time six types of information inquiries are modelled. Each of these concepts for an information type has as attributes the information that is needed to find an answer for an inquiry of the user. For example, the concept 'timetable information' has an attribute 'From_time' which specifies the range of time during which the departure of the train should be (see Fig. 3). This attribute could linguistically be realized for example with the word 'tomorrow'.

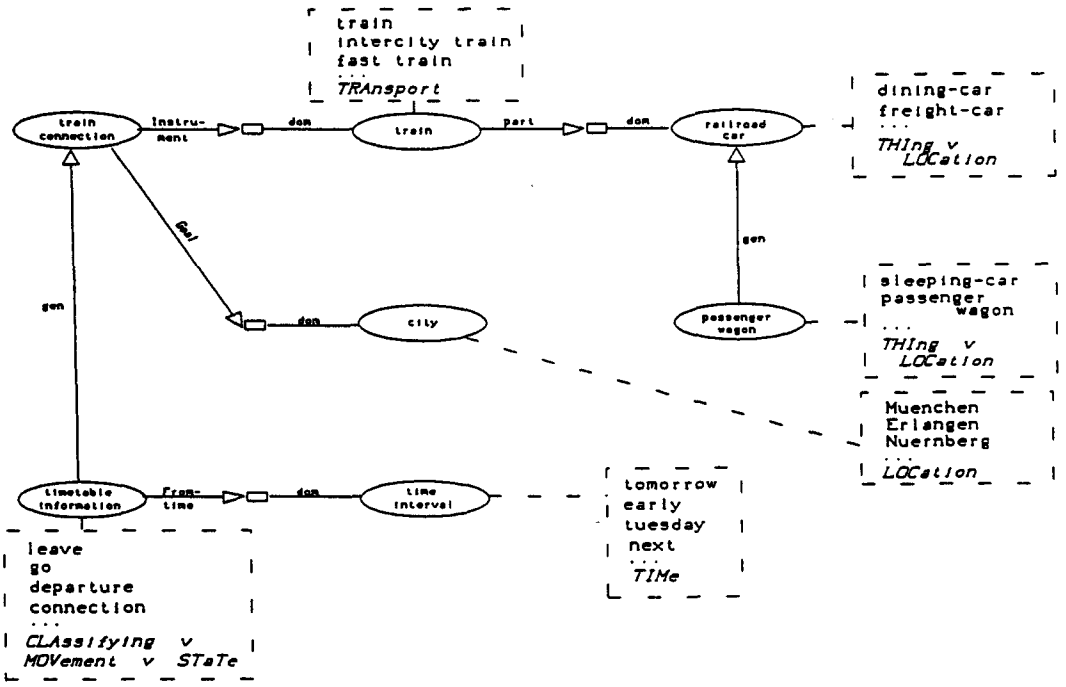


Fig. 3: Pragmatic Network (Part)

	train connection	timetable information	train	railroad car	passenger wagon	city	time interval	pP(w)
when (TIMe)	0	1	0	0	0	0	1	2
does	1	1	1	1	1	1	1	7
the	1	1	1	1	1	1	1	7
next	0	1	0	0	0	0	1	2
train	1	1	1	0	0	0	0	3
leave	1	1	0	0	0	0	0	2
for	1	1	1	1	1	1	1	7
Hamburg	1	1	0	0	0	1	0	3
SENTENCE	0	1	0	0	0	0	0	1

Fig. 4: "When does the next train leave for Hamburg?"

For many words in the dictionary a possible set of pragmatic concepts can be determined. With this property of words for each word a *pragmatic bitvector pbv(w)* is defined. Each bit of such a bitvector represents a concept of the pragmatic network. It therefore has as its length the number of all concepts (at the time 193). In this bitvector a word w has "1" for the following concepts:

1. For concepts that could be realized by the word and all generalizations of that concept.
2. For all concepts and their specializations for which the concepts of 1. can be the domain of an attribute.
3. If the word belongs to the basic lexicon, i.e. the part of the dictionary that is needed for nearly every domain (for example pronouns or determiners), it gets the "1" with respect to their semantic class. For this there exists a mapping function to pragmatic concepts. For example, all such words which belong to the semantic class TIMe (as 2. to the concept 'time interval' which could be realized by these words.
   In many cases (for example determiners) all bits are set to "1".

The pragmatic bitvector of a group of words w1 ... wn is then:

pbv(w1 ... wn) := pbv(w1) AND pbv(w2) ... AND pbv(wn)

The *pragmatic priority pP(w1 ... wn)* is defined as the number of "1" in pbv(w1 ... wn) and has the following properties:

* If the pragmatic priority of a group of words = 0, then the group is pragmatically inconsistent.
* The smaller the priority the better the hypothesis with these words.
* The bits of the pragmatic bitvector determine which pragmatic concept and especially which information type was realized. To make use of contextually determined expectations about the following user utterance the pragmatic interpretation of groups of words can be restricted with:

  pbv(w1 ... wn) AND pbv('timetable information')
  has to be > 0

where pbv('timetable information') is the bitvector for the pragmatic concept 'timetable information' and has the "1" only for the concept itself.
An example for pragmatic bitvectors and priorities pP(w) is given in Fig. 4.

3. Scoring

A main problem in reducing the amount of hypotheses for further analysis is to find appropriate scores, so that only the hypotheses that are 'better' than a special given limit have to be regarded further. In the semantic module different types of scores are used:

* Reliability scores from the other modules.
* A score indicating how much of the speech signal is covered by the hypothesis.
* The pragmatic priority.
* A score indicating how many slots of a case frame are filled. For determining this score a function is used that takes into account that a hypothesis does not become always more probable the more parts of a sentence are realized. Also hypotheses built of only short consitutents (i.e. mostly pronouns or adverbs) are less probable.

4. Stages of Semantic Analysis

At the present time the semantic analysis has three stages.

To demonstrate the analysis here an English example is chosen. It is an invented one for we only analyse German spoken speech. In Fig. 5 the result of the syntactic analysis is shown: all constituents that are one upon another are competing with regard to the speech signal. To find sentences covering at least most of the range of the speech signal there can be only combined groups of constituents together that are not competing to each other.

4.1 Local Interpretation of Constituents

A constituent (hypothesized by the syntax module) is checked to see whether the selectional restrictions between all of its words are observed. Only if this is true (i.e. the constituent is semantically consistent), and the constituent is also pragmatically consistent, is it regarded for further semantic analysis.

Selectional restrictions are defined in the lexicon by the attribute SELECTION. For the local interpretation all selectional restrictions that are given by some words in a constituent to some others in the same constituent have to be proved. There are especially restrictions given by words of special word classes which can be combined with nouns and can restrict the whole set of nouns to a smaller set by semantic means, i.e. the prepositions (see the example of Fig. 2), the adjectives or even the numbers. In the above example all constituents with a '*' are rejected.

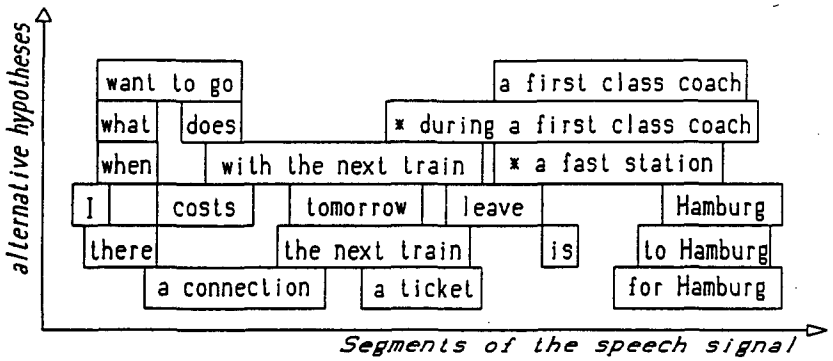


Fig. 5: Constituent hypotheses generated by the syntax module

To give a view about how many syntactic constituents semantically are not correct see Fig. 6. The experiments here shown base on real word hypotheses, but for the syntactic analysis only the best word hypotheses are used (between 35 and 132 for a sentence out of more than 2000). All hypotheses about the really spoken words are added.

number of experiment	syntactic constituents	semantic consistent constituents	rejected constituents
limit	21	18	14 %
0250	192	112	41 %
246a	88	65	26 %
246b	205	104	49 %
5518	280	155	44 %
5520	247	150	39 %
total	1033	604	41 %

Fig 6: Results of the local interpretation

4.2 Pre-Selection of Groups of Hypotheses

The next step is to build up sentences out of the semantic consistent constituents. This is not done by the syntax module because there exist too many possibilities to combine the syntactic constituents to syntactically correct sentences (there exist nearly no restrictions that are independent of semantic features). On the other hand there is always the difficulty with gaps in the speech signal (i.e. not or only with low priority with regard to other hypotheses

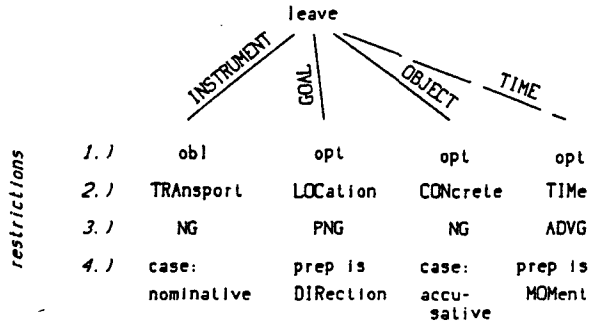


Fig. 7: The case frame or "to leave"

found but really spoken words). For this reason this analysis is done by the semantic module with additional syntactic knowledge.

The analysis is based on the valency and case theory. All verbs, but also some nouns and adjectives are associated with case frames which describe the dependencies between the word itself (i.e. the nucleus of the frame) and the constituents with which it could be combined. Such a case frame describes also the underlying relational structure. The frames are represented in a semantic net (see Brietzmann 1984).

Fig. 7 shows an example. The word "to leave" has one obligatory actant with the functional role INSTRUMENT and two optional actants (GOAL and OBJECT). Beside the actants there exist the adjuncts which could be combined with nearly every verb. In the example there is shown only TIME for that is very important for our application, the information about intercity trains. There are different types of restrictions:

1. the information if the actant is obligatory or optional
2. the semantic restriction for the nucleus of the constituent
3. the (syntactic) type of the constituent
4. these are features that exist especially in German: the case of a noun group, for prepositional groups a set of prepositions that belong to a certain semantic class or a special preposition.

If only 1.) and 2.) is used, at least the in Fig. 8 shown sentences could be hypothesized for the example.

First experiments have shown that it is nearly impossible to use only the network formalism for finding sentences because of the combinatorial explosion. On the other hand the process of instantiation does not cope with the posibility that also the nucleus of a case frame will not be found always. Therefore the pre-selection is added to handle these problems.

The idea is to seek first for groups of constituents which could establish a sentence. What should be avoided is that the same group of hypotheses is analyzed in several different contexts and that too many combinations have to be checked. So the dictionary is organized in a way that all actants of all frames with the same semantic restriction and the same type of constituent are represented as one class. These classes are than grouped together to combinations which can appear together in at least one case frame. Each combination has in addition the information in which case frame it can appear.

want to go :

(AGENT | ... | TIME | GOAL)

1) I | want to go | tomorrow | to Hamburg.
2) I | want to go | tomorrow | for Hamburg.
3) I | want to go | tomorrow | Hamburg.

a ticket :

(... | EXPLICATION)

4) a ticket | to Hamburg
...

the next train :

(... | GOAL)

7) the next train | to Hamburg
...

costs :

(MEASURE | ... | OBJECT)

10) what | costs | a ticket to Hamburg
11) what | costs | the next train to Hamburg
12) what | costs | Hamburg

a connection :

(... | GOAL)

13) a connection | to Hamburg
...

there ... is :

(... | OBJECT)

15) there | a connection | is | to Hamburg
...

does ... leave :

(TIME | ... | INSTRUMENT | ... | GOAL | OBJECT)

17) when | does | the next train | leave | to Hamburg
18) when | does | with the next train | leave | to Hamburg
19) when | does | the next train | leave | | Hamburg
20) when | does | the next train | leave | for Hamburg
...

Fig. 8: Sentence hypotheses

With this last information a found group of words could also be accepted if the nucleus is not found. It is even possible to predict a set of nuclei. These could be used as top-down hypotheses for the syntax module or the word recognition module.

For example for "to leave":

INSTRUMENT --> NG-Tra

GOAL --> PNG-Loc

OBJECT --> NG-Con

The combinations are then:

(NG-Tra)
(NG-Tra PNG-Loc)
(NG-Tra NG-Con)
(NG-Tra PNG-Loc NG-Con)
(PNG-Loc NG-Con)

These combinations do not say anything about sequential order, for, in German, word-order is relatively free. The last possibility is regarded although such a sentence would be grammatically incomplete (the INSTRUMENT slot is obligatory) to cope with the fact that not all uttered words are recognized by the word recognition module. To reduce the number of combinations the second combination will be eliminated because the class TRAnsport is a specialization of CONcrete (see Fig. 1) and the combination is then also represented by the last possibility. So there arise ambiguities that have to be solved in the last step of the analysis, the instantiation of frames.

If this method is applied to a dictionary that contains all of the words used in the above example the result is the following list of combinations (instead of 14 possibilities, if nothing is drawn together):

(NG-Con) --> go, cost, leave
(NG-Abs) --> cost, there_is
(PNG-Loc) --> ticket, train, go
(PNG-Loc NG-Con) --> go, leave
(PNG-Loc NG-Tra NG-Loc) --> leave
(NG-Wor Ng-Thi) --> cost

During the first stage of the analysis the semantic consistent constituents are sorted to the above used classes (see Fig. 9) so that a constituent is attached to all classes with which it is semantically compatible and agrees with respect to the constituent type.

So the problem of finding instances for the above combinations reduces to combining each element of the set of hypotheses attached to one class to each element of the set of hypotheses attached to the second class of the combination, and so on. If one combination comprises another, for example (PNG-Loc) and (PNG-Loc NG-Con), the earlier result is used (the seek is organized as a tree).

Restrictions for combining are given by the fact that two hypotheses cannot be competing with regard to the speech signal and by the fact that the found group of words has to be pragmatically consistent.

To complete these groups there is also tried to find temporal adjuncts to each of them (out of the original group and the so found new groups only the best will be furthermore treated as hypotheses). As temporal adjuncts there will be used all constituents which are compatibal with the semantic class TIMe and chains of such constituents with length of not more than 3 (for example "tomorrow | morning", "tomorrow | morning | at 9 o' clock"). Up to now no more information is used but in the future there will be a module that chooses only in the dialog context interpretable chains of temporal adjuncts.

With this second step of semantic analysis in Fig. 8 all sentences but 3, 11 and 18 are hypothesized. 3 and 17 are rejected because the constituent type is not correct, 11 is not pragmatically compatibal. All sententces in Fig. 8 satisfy the semantic restrictions.

There have been made also experiments that consider in addition simple rules of word order. They cannot be very specific because in German nearly each word order is allowed, especially in spoken

NG-Abs	NG-Con	NG-Loc	NG-Thi	NG-Tra	NG-Wor	PNG-Loc
	a first class coach		a first class coach	a first class coach		
what	what	what	what	what	what	
	the next train		the next train	the next train		
	I					
	Hamburg	Hamburg				
						to Hamburg
						for Hamburg
a connection						
	a ticket		a ticket			

Fig. 9: Constituents sorted to actant-classes

speech. But nethertheless the experiments so far indicate that about a third of all groups are rejected with this criterion (for example the sentence 15 in Fig. 8).

All found groups of hypotheses get the above mentioned scores and are ordered with regard to it.

Results

The results here presented are based on the following utterances (for the conditions of the experiments see also section 4.1):

246a Welche Verbindung kann ich nehmen? (Which connection should I choose?)

246b Hat dieser Zug auch einen Speisewagen? (Has this train also a dining-car?)

0250 Ich moechte am Freitag moeglichst frueh in Bonn sein. (I want to be at Bonn on Friday as early as possible.)

5518 Er kostet zehn Mark. (It costs ten marks).

5520 Wir moechten am Wochenende nach Mainz fahren. (We want to go to Mainz at the weekend.)

Fig. 10 shows how many groups of hypotheses were found dependent on the number of word hypotheses per segment in the speech signal (each segment represents one phon). The experiments here have been made by using as restrictions for the combinations

1. the semantic classes and the type of the constituents (without pbv)
2. the semantic classes, the type of the constituents and pragmatic attributes using the pragmatic bitvectors (with pbv)
3. the same conditions as in 2., but in addition some word order restrictions are checked (word order).

The really spoken utterances are always found but in some cases with a very bad score with respect to competing hypotheses. The main reasons for this result and the often high number of hypotheses are:

* The analysis of the time adjuncts is too less restrictive. Therefore in the future there will be only used constituents or chains of constituents that can really be interpreted in the dialog context as a time intervall or a special moment. So hypotheses as 'yesterday | then | tommorow' or 'at nine o' clock | next year' no longer are accepted. The referred time should also lie in the near future (because of our application).
* Anaphora could fill (nearly) each slot in each frame (similar as the constituent 'what' in Fig. 9). On the other hand they are often very short. So they appear in many combinations with other constituents. For an anaphoric constituent must have a referent which it represents (for example the constituent 'it' in 5518 could possibly refer to 'ticket'), such constituents should

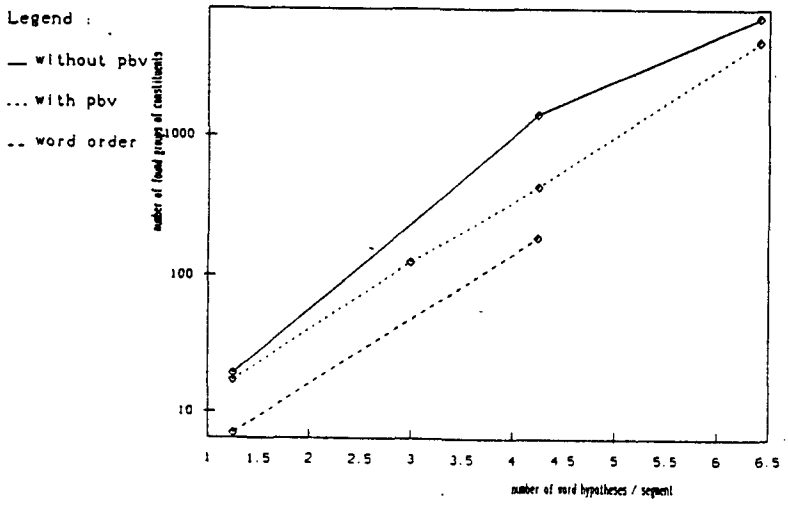


Fig. 10: Results of the pre-selection

obtain the semantic and pragmatic attributes of the possible referents - or, if there are none, should not be regarded for future analysis.

This method will first reduce the number of hypotheses and second will improve the score of a sentence with anaphoric constituents if it was really spoken (or also if it is well interpretable).

4.3 Structural Interpretation

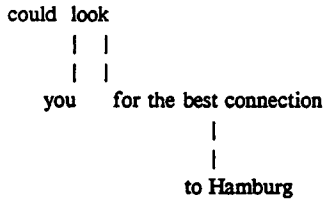
The last step consists in trying to instantiate the found candidates in the semantic network of the module (Brietzmann 1984 and 1986). Here all other selectional restrictions (i.e. especially the syntactic ones) are checked and thus the amount of hypotheses can be reduced a little bit more. Also the ambiguities have to be solved (see above). As a result there are gained instances of frame concepts which are the input for further domain dependent analysis by the pragmatic module.

This step (the instantiation) now is in work. All others are runnable.

5. Conclusion

In this paper a semantic analysis for *spoken* speech is presented. The most important additional problem which arises in comparison to a written input is the combinatorial explosion due to the many word hypotheses produced by the word recognition module. Because of this problem one has to cope with many word ambiguities. For solving these problems we need scores.

Problems arise with time adjuncts and anaphora. Also hierarchically structured sentences cannot be analyzed with the method of pre-selection of groups, for example:
"Could you please look for the best connection to Hamburg?"



Until now two combinations are found but they have bad scores because they cover too less of the speech signal. They cannot be combined together:

Could | you | look | for the best connection
and
for the best connection | to Hamburg

It is planned to expand the pre-selection in a way that also this problem could be solved.

The semantic analysis is implemented in LISP at a VAX 11/730.

REFERENCES

R.J. Brachman: A Structural Paradigm for Representing Knowledge. BBN Rep. No 3605. Revised version of Ph.D. Thesis, Harvard University, 1977.

A. Brietzmann: Semantische und pragmatische Analyse im Erlanger Spracherkennungsprojekt. Dissertation. Arbeitsberichte des Instituts für Mathematische Maschinen und Datenverarbeitung (IMMD), Band 17(5). Erlangen.

A. Brietzmann, U. Ehrlich: The Role of Semantic Processing in an Automatic Speech Understanding System. In: 11th International Conference on Computational Linguistics, Bonn, p.596-598.

H. Niemann, A. Brietzmann, R. Mühlfeld, P. Regel, E.G. Schukat: The Speech Understanding and Dialog System EVAR. In: New Systems and Architectures for Automatic Speech Recognition and Synthesis, R.de Mori & C.Y. Suen (eds). NATO ASI Series F16, Berlin, p. 271-302.

This work was carried out in cooperation with
Siemens AG, München