

Syntactic Analysis in Speech Understanding

G.Th. Niedermair
ZT ZTI INF 322
SIEMENS AG

1. Abstract

Within the SPICOS project, which is part of the joint research-project 'Speech Understanding Systems', a prototype of a dialogue-system was successfully implemented, which allows question- ing a database in continuous speech. The answers are output in synthesized speech. For this SPICOS prototype we have developed a powerful and yet flexible approach for syntactic and semantic analysis. The analysis has successfully been adapted to different acoustic outputs, which vary with respect to the quality and the quantity of the hypothesized words of the utterance. The stages of the linguistic analysis and problems resulting from the different outputs of the acoustic modules will be pointed out in the paper. The system deals with a vocabulary of about 1000 words of German and covers the type of sentences typical for data-base queries, such as imperatives, wh-questions and sentence questions.

2. Analysis components

The linguistic analysis interfaces two different acoustic search strategies. One is purely bottom up and exhaustive, the other is top down oriented and uses a finite state network of coarse semantic classes in order to filter out illformed chains of words. The first approach comes up with a number of word hypothesis, that is 50 - 100 times the words in the sentence, the second with the most likely chain of words. Both are transformed into a graph, consisting of nodes for the words and edges pointing to its possible predecessors. The syntactic analysis is followed by a transformation into a logic representation based on model theoretic semantics. The expressions are evaluated through the database and a proper answer is generated according to the type of input question. The answer-string is transformed into synthesized speech. (For system architecture see also (Dreckschmidt 87))

2.1. Three-Step Syntax-Analysis

The whole syntactic analysis is based on a modified chart parsing algorithm as described in Winograd (1983). The modifications concern efficiency measures in order to cope with the huge number of chart-entries and a speed up of search within these.

Given the output of the acoustic analysis as a graph of words, the syntactic analysis is carried out in three steps

1) : the analysis of nominal constituents (NP's, PP's, etc.)

To do this the output is discriminated according to syntactic categories

2) : verb phrase analysis with a gapping mechanism

All possible verb-groups, whose parts may be at different places in the sentence, are built up.

3) : sentence analysis on the basis of binary dependency trees

Here we use the full verb information right from the beginning of sentence analysis. This is especially useful in German, where the meaning carrying part of the verb is often moved behind its objects. This enables us to use the full case-frame information of the particular verb when attaching any nominal-groups during the sentence analysis and thereby avoiding quite a number of paths, which only later would die. For a detailed description see (Niedermair 1986)

This type of analysis is only possible and necessary since in this demonstration-model SPICOS I we have no continuous interaction between acoustic and linguistic analysis. Acoustic analysis is finished before linguistic analysis starts. Although we recognized the necessity for a closer interaction between acoustics and linguistics, this was first designed in order to have a common, clear cut interface to both approaches. Above that it gave an insight into what type of restrictions the grammar has to model when it is applied to hypotheses, which are not already filtered by any kind of syntactic, semantic knowledge. When, in the future, there is a closely linked interaction between acoustic recognition and syntactic analysis, the grammar has to contain these restriction as well, because then it not only verifies given chains of words but also serves the task of deducing hypotheses for succeeding word candidates on the basis of a partial structures. Another point, that has proved to be most important is to make use of semantic feature-restriction in parallel to syntactic restrictions.

- The Grammars

The grammars used for all three parts of the syntactic analysis are uniform. They can be viewed as context-free augmented phrase structure grammars.

They contain

- a phrase structure part,
- a condition-part, that carries out features checks on the parts
- a carry-over part, which guides the inheritance of the features
- a semantic description, which states the composition of the natural language expression into a formal logic expression.

The nominal- as well as the sentence- grammar make use of the caseframe-lexicon, which is both for nouns and verbs. The test is supplied with features of the phrase in question. They are compared to the restriction-parts in the case frame of the head that is to be attached to. If successful, it delivers an identification of the realized deep-case. This contributes to build up the predicate-argument structure of a sentence.

2.2. Semantic analysis

Parallel to the syntactic grammar rules there is a set of semantic rules, which, at two levels - called formal (ELF) and referential (ELR) - are used to build first surface oriented, then data-base oriented formal logic expressions, representing the full meaning of the sentence. It implies two important principles: first, it is compositional, meaning that complex expressions are derived from smaller subexpressions; second it works rule-to-rule. This implies that for each syntactic rule there has to be at least one semantic rule saying how the surface expressions, captured by this rule has to be represented logically. This could lead to a multiplication of identical phrase structure rules if according to the semantic content of a particular syntactic category (e.g. nouns) the same rule gets different formal semantic representations. For a detailed description of the formalism see (Bunt 1985).

The semantic analysis also comprises the context-free translation of these formulas into data-base dependent expression and their evaluation. The referential expression has the same syntax as the formal expression, where the constants of the formal language are assigned their referential equivalents according to the types specified for the application. The formal representation can be seen as representing the meaning of the sentence independent of the area of application, whereas the referential expressions link this abstract meaning to a particular meaning for the database.

2.3. Answer-Generation

Before the evaluation of the referential expressions is carried out the type of question is taken into account in order to update the user model and to define a proper pattern for the answer. For simple imperatives the retrieved values are inserted into patterns, for wh-questions and y/n-questions the answer-sentences are syntactic transformations of the input sentence into declarative structures with appropriate insertions of the values.

3. Different acoustic inputs

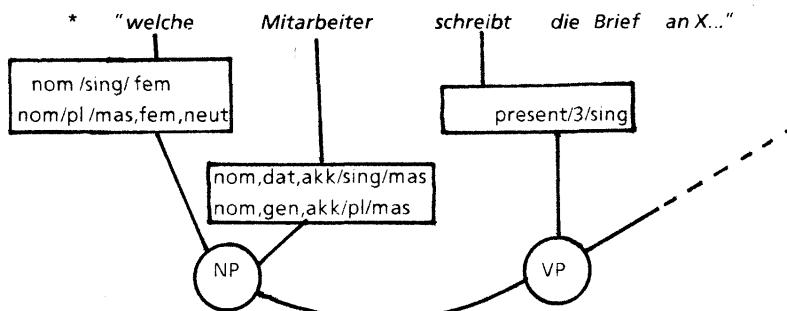
The system had basically to deal with two different kinds of input, according to the different methods in acoustic analysis and search.

3.1. Optimal Sentences

The first approach delivers, by dynamic programming methods, the optimal sentence that the acoustic search could recognize. This is supported by a network roughly modelling the syntactic and semantic restrictions in the domain (see Mergel, Paesseler 1987). Due to the crudeness of the underlying language model the optimal sentence can contain mistakes. The output delivers no alternatives or second best sentences. Therefore in case of such mistakes in recognition the syntactic / semantic analysis has no chance to recover from such mistakes. Mistakes in this approach of course occur mainly because the network serving as a filter for the acoustic recognition is more permissive than the grammar of the linguistic analysis is. This is especially so with non-local dependencies between words or grammatical categories. To model these in a finite state network would blow up such a network to an intractable size for any larger portion of natural language. The grammar can only cope with these mistakes by loosening the restrictions on the feature parts of the phrase structure rules where morpho- syntactic variations are not crucial to the meaning of the sentences.

Nevertheless one has to be careful here too, because it does make a difference of course to the semantic interpretation whether a noun-group is e.g. singular or plural. Take the example:

assume acoustic output:



The question arises whether one should determine the morphology of the whole NP (*welche mitarbeiter*) according to the morphology of the head noun (which here is undecided) or to the possible combinatorial possibilities with ADJ or DETs, which in 1) would be *nom,acc/pl/masc*, which in turn would not match with the morphology of the VP

Again if the VP-morphology should overrule the morphology of the NP, one would have to assume that the DET has the wrong morphology; or does one assume the NP is right and the VP is wrong? In a similar way the same problem comes up with the second NP (*die Brief*). Regarding this as plural because 'die' seems more different from 'der' or 'den' than 'Brief' is from 'Briefe' (which from an acoustic point of view is not clear at all, since 'briefe' is a two-syllable word if not contracted with the initial vowel of the following word) then the possible combinations are turned into a semantic problem because syntactically one could regard the second NP as subject/sing that fits nicely with the VP and the first NP as object NP in accusative case. Only this is semantically not compatible, since "schreiben" needs a subject with semantic feature '+animate'.

To solve that problem properly the syntactic analysis would have to be able to produce all these variations and measure the acoustic distance of each variant against that of the original solution. Apart from the fact that the acoustic significance for one or the other is doubtful especially in the area of word-endings, it is questionable whether such a procedure would be worth the effort. One way out would be to question the user. However, if the user offers an option it is doubtful whether this would enhance the performance.

A more natural but also more costly way would seem to interrupt analysis at the point where an incongruity is detected and offer the user the alternatives. Since this could happen quite frequently and eventually annoy the user, it is advisable to keep this type of user-interaction to a minimum.

In any case all these solutions presuppose a close link between acoustic and linguistic analysis. Since in the demo-system we received the best matching sentence, for the time being only straight-forward solutions to these problems were implemented.

3.2. Bottom-Up Hypotheses

The second approach works purely bottom-up and delivers, based on acoustic phonetic features and explicit segmentation, a lattice of word hypotheses for a given input sentence. Since we use a syntactic grammar, none of the words are allowed to be missing. This pushes the pruning factor up such that 50 to 150 times the number of words of the original sentence are produced.

Frame oriented and semantic grammars are often regarded to be of advantage in speech recognition because they seem to be able to handle the problem of missing words, not detected or pruned away during acoustic analysis. This is deceptive. In one case these grammars simply do not model these words and their syntactic appearance and forget about them, as it is usual in semantic grammars. They assume that these words do not contribute to the meaning of a sentence. This may be true for a particular application, but not necessarily for the others.

In the other case analysis must make do with what it gets, as is usual in case frame grammars, and has to make guesses about the missing bits. Since normally there is more than just one partially filled case-frame, there can also more bits be missing.

Guessing at what is missing gets as arbitrary as allowing for any constituent to be missing in an ordinary phrase structure grammar. (For a frame oriented approach see (Hayes 1986)).

By allowing missing words, one would gain the advantage of cutting down the pruning threshold, but loose again by introducing combinatorial possibilities through gaps at arbitrary places of arbitrary length.

Often caseframes are used for semantic plausibility checks preceded by pure syntactic analysis as in the EVAR-system (see Brietzmann, Ehrlich 1986) and thus suffer from similar problems as described in the SPICOS I system.

The first step in our syntactic analysis combines the words to noun- prepositional - and verb-groups, in order to rule out illformed word-chains. The number of these is between 100 to 800 correctly built constituents per sentence, out of which in a further step possible sentences are built and ordered according to their weight.

In contrast to the former approach the difficulty here is not how to survive with a minimum of restrictions, but just to the contrary. The grammar has to use every possible restriction in order to rule out not permissible combinations. This includes syntactic restrictions as well as semantic plausibility checks, that are carried out in parallel with each rule application. Where grammars, dealing with written input, survive if their rule and feature apparatus suffices to resolve ambiguity and assign the right structure to the sentence, which a priori is assumed to be correct, here the grammar has to explicitly exclude incorrect phrases and block their combination.

To illustrate this assume a simple example :

2)

NP	<--	NOM	NOM	<--	noun
NP	<--	det + NOM	NOM	<--	propername

would suffice for assigning :	NP	<--	(die) Briefe
	NP	<--	John

but would also let: * NP <-- * *der John*

pass, which is not desired. The rules have to prevent these phenomena on a syntactic and semantic level. Additionally restrictions between constituents should be checked at the most early point in time. to avoid the wrong paths This however may have consequences on the formulation of the rules and the resulting syntactic structures.

5.Conclusion

This approach as well as the network approach has shown that linguistic restrictions at recognition time are essential to guide acoustic analysis. Out of the already built partial structures linguistic analysis has to make the proper predictions of possible and maybe even most plausible successors. A grammar, that is of generative power in a syntactic and semantic sense will still be needed, not only for verification but also for prediction. Linguistic analysis can hardly make up for "mistakes" that acoustic analysis produces and which seem 'minor' to humans. On the other hand acoustic analysis is not in a position to detect all words and endings correctly without syntactic and semantic restrictions. Networks seem a comfortable means to model all these restrictions, yet they have the unpleasant tendency to explode in size when they really have to.

The alternative of filtering acoustic data dynamically by syntactic and semantic grammars and at the same time supplying linguistic analysis with reliable data needs a very close interaction between the two processes (see Niedermair 1987). How this interaction can be realized will be one of the major topics for SPICOS II.

References:

- Bunt H. Mass Nouns and Model theoretic semantics. 1985
- Brietzmann, A., Ehrlich, U.,: The Role of semantic Processing in an Automatic Speech Understanding System. Proceed. of the COLING conf. 1986
- Dreckschmidt, G.: The Linguistic Component in the Speech Understanding System SPICOS. in: Tillmann, H., Willée, G., ed.: Analyse und Synthese gesprochener Sprache, Hildesheim 1987
- Hayes, Ph., Hauptmann, A., Carbonell, J., Tomita, M.: Parsing Spoken Language: A Semantic Caseframe Approach..Proceed. of the COLING Conf. 1986
- Mergel, D., Paessler, A.: Construction of Language Models for Spoken Database Queries. ICASP 1987
- G.T.Niedermair, Divided and Valency Oriented Parsing in Speech Understanding, Proceedings of the Coling 1986
- G.T.Niedermair, Merging Acoustics and Linguistics in Speech Understanding, to appear in: Proceedings of the Conference of the NATO Advanced Study Institute, 1987
- Thurmair, G.: Der Einsatz semantischer Verfahren in sprachverstehenden Systemen. in: Tillmann, H., Willée, G., ed.: Analyse und Synthese gesprochener Sprache, Hildesheim 1987
- Winograd T.: Language as a cognitive Process, Vol.I: Syntax. 1983