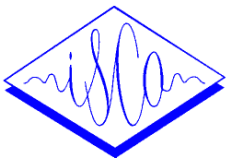

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

SYNTACTIC ANAPHORA RESOLUTION IN A SPEECH UNDERSTANDING SYSTEM

Ute Ziegenhain

Siemens AG ZFE ST SN 74
Otto-Hahn-Ring 6 8000 München 83 Germany*

ABSTRACT

In this paper a way to treat anaphoric processes in a natural language dialogue system is proposed. It makes extensive use of syntactic and semantic information to find the right antecedents for anaphoric expressions. Using the proposed method it can be shown that the right candidate for an anaphoric expression can be found in 89% out of 300 dialogues.

OVERVIEW

The task of the syntactic-semantic anaphor resolution module is to find out possible antecedents of referring expressions and to pass this information to other system modules, e.g.the dialogue module and the module for formal semantic representation. The disambiguation of the referring expressions is achieved as follows. The collecting of anaphoric expressions and possible antecedents takes place in the syntactic APSG grammar of the system. In a special anaphor part, which belongs to each syntactic rule of the grammar, the features needed by the anaphor resolution module are collected and percolated up the parse tree. The sentence node contains thus a list of all antecedent candidates and anaphoric expressions, which is passed on to the resolution module.
A special reference grammar is then used to connect anaphora and their antecedents. The result of the disambiguation process is handed over to the dialogue handler for confirmation.
In case of unsolved ambiguities the user is asked for disambiguation by the dialogue handler, and the result is handed back to the resolution module for updating the antecedent and anaphor lists. The so established link between anaphor and antecedent is passed on to the formal semantic representation module which then replaces the links by the appropriate semantic formulae.
It can be shown that out of a set of about 300 dialogues consisting of about 5-7 turns each 89% of the referential expressions are solved correctly without using any additional information about discourse structure. Moreover the resolution module can easily be adopted to other applications.

THE RESOLUTION MODULE

The anaphor resolution module which is presented here is similar to the one described in [5] insofar as it makes extensive use of syntactic and semantic properties of referring expressions. The resolution module is implemented in the SPICOS II natural language system which enables an on-going dialogue with a data-bank concerning an office environment [cf. 8]. Recently the resolution module has also been put to the test with dialogues of the Sundial system [cf. 7], however, in the work here we will concentrate on the SPICOS II system.
In SPICOS II 5 types of referring expressions [cf. 6, 11] are dealt with which is an extension to the work described in [5 . These types contain the full range of personal and possessive pronouns, demonstrative NP's (cf. *dieser Brief*), proforms for PP's (cf. *davon,dazu*) which are very common in German, as well as definite descriptions. Sentential anaphora and ellipsis are not taken into consideration in the system, but the algorithm is designed to capture these cases as well. Consequently NP's and PP's may be marked as possible antecedents for the anaphoric expressions. Only intersentential anaphoric expressions are taken into account in the system.
For deictic personal pronouns and deictic temporal and local expressions the referents are inserted during an introduction phase where the user is asked for name and company affiliation and place and time.
The module consists of 4 parts:

- the collection of anaphoric expressions
- collection and marking of antecedent candidates
- a reference grammar
- and the communication with the dialoge module and the formal semantic module.

In the following paragraphs the different steps to achieve the resolution are described.
Whether an expression is to be interpreted as anaphoric depends on the syntactic category of this element on the one hand. Thus pronouns and proforms as well as demonstratives have to be marked as anaphoric due to their syntactic category. The marking of the definite descriptions (as well as some indefinite complex NP's), however, depends on how they have been introduced into the dialogue.
In order to avoid that all definite NP's have to be marked as anaphoric we made some constraints: all definite NP's except those which are further specified by an attribute or PP were marked as anaphoric assuming that the latter are already sufficiently specified by the attribution. Most often such NP's are not anaphoric , but introduce a new entity. An exception

10.21437/ICSLP.1992-390

to this form the class of complex NP's in which a part-whole relation or a possessive relation between anaphoric expression and antecedent is established. For example a complex NP like *das Treffen in Hamburg* is not marked as being anaphoric whereas an NP like *Aufsatz ueber Akustik* is marked as anaphoric assuming that the speaker is referring to a special entity out of the set of all essays. This procedure is quite a simplification but it worked well for the majority of definite descriptions to be resolved.

The collection of antecedent candidates was restricted to NP's and PP's because only the treatment of those was required by the application. Of course all anaphoric expressions can in turn act as antecedent candidates.

Which features are to be collected is encoded in the SPICOS APSG grammar. Two methods for collecting the features were at hand: first one could traverse the sentence tree after parsing and collect all features needed for the resolution at the sentence node. A disadvantage of this procedure was that often not all features required by the anaphor resolver could be found at the sentence node, and the whole resolution becomes dependent on the syntactic grammar which is used. The alternative was to incorporate the collecting of the features into the grammar itself. This method was chosen for SPICOS II and SUNDIAL. The anaphor resolution becomes thus more or less independent of the syntactic grammar. Corresponding to each relevant grammar rule there exists a so-called anaphor part consisting of a context-free part of the rule which is shared with the syntactic grammar, a feature test part, where the right-hand side elements of the rules can be tested for special resolution purposes, and a feature percolation part. In the latter the features which are to be collected or have to be newly set are defined. In some cases features provided for by the syntactic grammar are used, often special features out of the lexicon or additional features are newly set separatly at different stages of the grammar. For example in the following part of a dialogue

(1) *Hat M. den Aufsatz ueber Akustik erhalten?*

(2) *Ist er am Projekt SPICOS beteiligt?*

the morphological information will be used from the syntactic grammar whereas information of reference of the complex NP (*den Aufsatz ueber Akustik*) will be set up in the rule where NP and PP are joined. If all conditions of a rule are fulfilled the features are percolated up the parse tree. Thus the sentence node contains the list of all anaphors and antecedent candidates plus the relevant feature lists as they have been provided for by the above described application of grammar rules.

For example the anaphor lists for the sentence nodes in (1) and (2) would contain:

baseform	*Aufsatz über Akustik*
category	NP
case	accusative
refer	yes
number	singular
semcat	document
baseform	*er*
number	singular
person	third
semcat	individual asf.

where 'semcat' means the semantic category which is assigned from the semantic network of the system [cf. 8], 'refer' means the referential aspect of an NP.

To speed up the runtime of the system it is possible to separate out the anaphor part of the grammar and thus to activate only those anaphor rules after parsing which correspond to successfully applied syntactic rules.

After the anaphor lists are handed over to the resolution module additional features are explicitly set for the referents and anaphors. Among these there are features like the sentence number, later used in the reference grammar for information about recency, the chartnumber and the so-called focus feature . The completed feature lists of anaphoric elements and antecedent candidates are stored in a work space called "working area", a list where all candidates with their features remain accessible during all phases of a dialogue. Thus the "working-area" contains the dialogue history of all processed elements. For example the pronoun *er* in the above mentioned dialogue will get associated to its antecedent, it will be assigned the semantic category, (which is the one of its antecedent in the model described here), the focus feature will be updated asf.. All the collected information is stored with the pronoun and can in turn be used for further testing. This method is more or less common to all systems we know of.

The reference grammar finally contains as a set of rules the conditions under which anaphoric references can be established. After all candidates have been put into the "working-area", the rules of the reference grammar are applied to each of the anaphoric expressions of the last sentence. According to the 5 different types of anaphoric expressions there exist 5 different groups of rules. Depending on the type of an anaphoric expression different criteria are used to find the right antecedent. For pronouns agreement in number and person is demanded whereas proforms for PP's often require the same preposition with the potential antecedent. Common to all groups are factors like recency e.g.. The factors themselves in turn had to be adapted to the different groups.

For example in most cases the antecedent for the pronouns was found in the immediately preceding dialogue step (for the same results cf. [5], whereas the candidates for demonstratives were often found not in the immediately preceding dialogue but further down the history.

Compatible semantic relationship, however, is demanded for all types of referential expressions. Semantic information was

taken either from the lexical semantic types or from inferences on the semantic network (for details see [8]).

Further criteria like information about topicalization, the fact of being mentioned repeatedly, syntactic parallelism [c.f. 3, 9] are also used as additional conditions for resolution. The latter proved to be of less importance for the kind of dialogue in SPICOS II, because equality in the verbal case frames was rarely given in our domain. The fact of being mentioned repeatedly or "being in focus", (i.e. the theme of the dialogue), was a "hard" condition at least within the class of proforms and definite descriptions.

The rules themselves now are applied as follows: the potential candidates are marked as such in the "working-area" with a special initial feature test. For most cases in our domain potential candidates were those which passed the test for third person, namely NP's and PP's. The category of the anaphoric expression was questioned after the initial testing. Then the rules are applied in order of ascending restrictions on the antecedent candidates. This ensures that from competing candidates those are dropped early which are possible but less likely. The whole resolution process ends if only one candidate remains as a possible antecedent, if no candidate can be found at all or if all possible rules are applied and more than one candidate remains.

The results of the resolution are handed over to the dialogue module. In case of multiple links for one anaphor the dialogue module disambiguates the reference by means of a clarification dialogue with the user. The result of the disambiguation is returned to the anaphor resolution module where the links between anaphor and its antecedent are then updated consequently. These now unambiguous links are handed over to the formal semantic module where the proper formal semantic representation of the antecedent is inserted in place of the anaphoric expression. (for further details see [8]).

EVALUATION

The evaluation of the system took place in two steps. First the module was tested on 50 SPICOS dialogues which were set up by linguists.

Within the 50 dialogues there was an error rate of 6 % for all anaphoric expressions. One source of errors was the semantic network where at the time of evaluation some of the possible relations (like e. g. the part-whole relation) among elements were not implemented yet.

Another source was the concept of focus we used. It is similar to that of Sidners "confirmed focus" [9 . The dialogue focus is not predicted by the system but an element is considered as being in focus if it has acted as an antecedent in a previous sentence. As mentioned above the fact of being in focus is a "hard" condition for some groups of referring expressions. This means that those antecedent candidates are preferred that pass the test of being in focus. This concept is suitable for short dialogues but doesn't work well with longer dialogues where changes in focus often occur. However, these changes can hardly be recognized without information about intonational structure. The above results were somewhat deceptive though: the number of dialogues was too small to prove the proper applicability of the concept, and moreover the test dialogues were used for setting up the reference grammar as well. In a next step we corrected the errors of the module as far as this was possible.

For testing the improved anaphor module (and the syntactic module as well) another 250 dialogues allowing questions concerning an office environment were built, 80 of them were written from non-linguists. These dialogues weren't seen by the system before the evaluation tests. The so called control dialogues contained 5-7 question-answer steps on average, some of them even up to 14 dialogue steps. All types of the above described referring expressions were allowed. The questions contained 2 types of anaphoric expressions on average. Within 1132 correctly parsed sentences an antecedent had to be found for 892 anaphoric elements. Out of these the algorithm found the correct antecedent in 89% of all cases (for similar results see [5]).

Whereas the above-mentioned semantic restrictions were the reason for failures of the algorithm in the 'SPICOS dialogues' the focus concept and the treatment of definite descriptions proved to be of more importance within the 250 control dialogues. As the latter usually contained more dialogue steps a more refined focus concept which would allow predictions of the dialogue focus would have been very helpful.

For the definite descriptions the most frequent error was that we classified them as anaphoric though they introduced a new topic in the dialogue. This could have been avoided in some cases if within the anaphor resolution module the more fine-grained formal semantic information would have been taken into account [cf.11 . Nevertheless it could be shown that without any specific knowledge of the discourse structure remarkable results can be achieved in resolving anaphoric references which may be driven well into the 90 % region , if the above mentioned measures were taken.

*The work described in this paper was carried out in collaboration with a former collegue, G. Dreckschmidt. Only the author is responsible for the contents of the paper. I'm grateful to two anonymous COLING writers for comments on an earlier version of this paper. The SPICOS project was supported by the German Ministry of Research and Technology under the number ITM 8801B9. The SUNDIAL project is supported by the Commission for the European Communities ESPRIT programme as project number 2218.

[1] S. Carberry. "A pragmatics-based approach to ellipsis resolution." *Computational Linguistics 15:2*, pp. 75-96, 1989.

[2] K. van Deemter. *On the composition of meaning*. Masch. phil. Diss. Amsterdam, 1991.

[3] B. Grosz and C.L. Sidner. "Attentions, intentions and the structure of discourse." *Computational Linguistics 12:3*, pp. 175-203, 1986.

[4] G. Hirst. Discourse-oriented anaphora resolution in natural language understanding: a review. *Computational Linguistics 7:2*, pp. 85-98, 1981.

[5] J. R. Hobbs. "Resolving pronoun references." B. Grosz et al. (eds.) *Readings in Natural Language Processing*, pp.339-352, 1986.

[6] H. Kamp. "A theory of truth and semantic interpretation." J.A.G. Groenendijk et al.(eds.) *Formal methods in the study of language*, vol.2, pp.277-322, 1981.

[7] G. Niedermair. "Linguistic modelling in the context of an oral dialogue." to appear: *Proc. of the ICSLP*, 1992.

[8] G. Niedermair, H. Tropf and M. Streit. "Linguistic processing related to speech understanding in SPICOS II". *Speech Communication*, vol.9, pp.565-585, 1990.

[9] P. Pause. "The interpretation of anaphoric expressions in Con3tra." *Proc. of the 6th European Conference on Artificial Intelligence*, pp.232-243, 1984.

[10] C.L.Sidner. "Focusing in the comprehension of definite anaphora." M. Brady and R.Berwick (eds.) *Computational models of discourse*, pp.267-330, 1983.

[11] B.L.Webber. "So what can we talk about now?" M. Brady and R.Berwick (eds.) *Computational models of discourse*, pp.331-371, 1983.