

Rules for Pronominalization

Franz Guentner, Hubert Lehmann
IBM Deutschland GmbH
Heidelberg Science Center
Tiergartenstr. 15, D-6900 Heidelberg, FRG

Abstract

Rigorous interpretation of pronouns is possible when syntax, semantics, and pragmatics of a discourse can be reasonably controlled. Interaction with a database provides such an environment. In the framework of the User Specialty Languages system and Discourse Representation Theory, we formulate strict and preferential rules for pronominalization and outline a procedure to find proper assignments of referents to pronouns.

1 Overview: Relation to previous work

One of the main obstacles of the automated processing of natural language sentences (and a fortiori texts) is the proper treatment of anaphoric relations. Even though there is a plethora of research attempting to specify (both on the theoretical level as well as in connection with implementations) "strategies" for "pronoun resolution", it is fair to say

- a) that no uniform and comprehensive treatment of anaphora has yet been attained
- b) that surprisingly little effort has been spent in applying the results of research in linguistics and formal semantics in actual implemented systems.

A quick glance at Hirst (1981) will confirm that there is a large gap between the kinds of theoretical issues and puzzling cases that have been considered on the one hand in the setting of computational linguistics and on the other in recent semantically oriented approaches to the formal analysis of natural languages.

One of the main aims of this paper is to bridge this gap by combining recent efforts forthcoming in formal semantics (based on Montague grammar and Discourse Representation Theory) with existing and relatively comprehensive grammars of German and English constructed in connection with the User Specialty Languages (USL) system, a natural language database query system briefly described below.

We have drawn extensively -- as far as insights, examples, puzzles and adequacy conditions are concerned -- on the various "variable binding" approaches to pronouns (e.g. work in the Montague tradition, the illuminating discussion by Evans (1980) and Webber (1978), as well as recent transformational accounts). Our approach has

however been most deeply influenced by those who have (like Smaby (1979), (1981) and Kamp (1981)) advocated dispensing with pronoun indexing on the one hand and by those (like Chastain (1973), Evans (1980), and Kamp (1981)) who have emphasized the "referential" function of certain uses of indefinite noun phrases.

2 Background

Contrary to what is assumed in most theories of pronominalization (namely that the most propitious way of dealing with pronouns is to consider them as a kind of indexed variable), we agree with Kamp (1981) and Smaby (1979) in treating pronouns as bona fide lexical elements at the level of syntactic representation.

Treatments of anaphora have taken place within two quite distinct settings, so it seems. On the one hand, linguists have primarily been concerned with the specification of mainly syntactic criteria in determining the proper "binding" and "disjointness" criteria (cf. below), whereas computational linguists have in general paid more attention to anaphoric relations in texts, where semantic and pragmatic features play a much greater role. In trying to relate the two approaches one should be aware that in the absence of any serious theory of text understanding, any attempt to deal with anaphora in unrestricted domains (even if they are simple enough as for instance children's stories), will encounter so many diverse problems which, even when they influence anaphoric relations, are completely beyond the scope of a systematic treatment at the present moment. We have thought it to be important therefore to impose some constraints right from the start on the type of discourse with respect to which our treatment of anaphora is to be validated (or falsified). Of course, what we are going to say should in principle be extendible to more complex types of discourse in the future.

The context of the present inquiry is the querying of relational databases (as opposed to say general discourse analysis). The type of discourse we are interested in are thus dialogues in the setting of a relational database (which may be said to represent both the context of queries and answers as well as the "world"). It should be clear that a wide variety of anaphoric expressions is available in this kind of interaction; on the other hand, the relevant knowledge we assume in resolving pronominal relations must come from the information

specified in the database (in the relations, in the various dependencies and integrity constraints) and in the rules governing the language.

We are making the following assumptions for database querying. A query dialogue is a sequence of pairs <query,answer>. For the sake of simplicity we assume that the possible answers are of the form

- yes/no answer
- singleton answer
(e.g. Spain, to a query like "Who borders Portugal?")
- set answer
({France, Portugal} to a query like "Who borders Spain?")
- multiple answer
({<France,Spain>, ... } to a query like "Who borders who?")
and
- refusal
(when a pronoun cannot receive a proper interpretation)

2.1 The User Specialty Languages system

The USL system (Lehmann (1978), Ott and Zoeppritz (1979), Lehmann (1980)) provides an interface to a relational data base management system for data entry, query, and manipulation via restricted natural language. The USL System translates input queries expressed in a natural language (currently German (Zoeppritz (1983), English, and Spanish (Sopeña (1982))) into expressions in the SQL query language, and evaluates those expressions through the use of System R (Astrahan & al (1976)). The prototype built has been validated with real applications and thus shown its usability. The system consists of (1) a language processing component (ULG), (2) grammars for German, English, and Spanish, (3) a set of 75 interpretation routines, (4) a code generator for SQL, and (5) the data base management system System R. USL runs under VM/CMS in a virtual machine of 7 MBytes, working set size is 1.8 MBytes. ULG, interpretation routines, and code generator comprise approximately 40,000 lines of PL/I code.

Syntactic analysis

The syntax component of USL uses the User Language Generator (ULG) which originates from the Paris Scientific Center of IBM France and has been described by Bertrand & al (1976). ULG consists of a parser, a semantic executor, the grammar META, and META interpretation routines. META is used to process the grammar of a language. ULG accepts general phrase structure grammars written in a modified Backus-Naur-Form. With any rule it allows the specification of arbitrary routines to control its application or to perform arbitrary actions, and it allows sophisticated checking and setting of syntactic features. Grammars for German, English, and Spanish have been described in a form accepted by ULG. The grammars provide rules for those fragments of the languages relevant

for communicating with a database. The USL grammars have been constructed in such a way that constituents correspond as closely as possible to semantic relationships in the sentence, and that parsing is made as efficient as possible. Where a true representation of the semantic relationships in the parse tree could not be achieved, the burden was put on the interpretation routines to remedy the situation.

Interpretation

The approach to interpretation in the USL system builds on the ideas of model theoretic semantics. This implies that the meaning of structure words and syntactic constructions is interpreted systematically and independent of the contents of a given database. Furthermore, since a relational database can be regarded as a (partial) model in the sense of model theory, the interpretation of natural language concepts in terms of relations is quite natural. (A more detailed discussion can be found in Lehmann (1978).)

In the USL system, extensions of concepts are represented as virtual relations of a relational database which are defined on physically stored relations (base relations). The set of virtual relations represents the conceptual knowledge about the data and is directly linked to natural language words and phrases. This approach has the advantage that extensions of concepts can relatively easily be related to objects of conventional databases.

For illustration of the connection between virtual relations and words, consider the following example. Suppose that for a geographical application someone has arranged the data in the form of the relation

CO (COUNTRY,CAPITAL,AREA,POPULATION)

Now virtual relations such as the following which correspond to concepts can be formed by simply projecting out the appropriate columns of CO:

CAPITAL (NOM_CAPITAL,OF_COUNTRY)

Standard role names (OF, NOM, ...) establish the connection between syntactic constructions and columns of virtual relations and enable answering questions such as

(1) What is Austria's capital?

in a straightforward and simple way. Standard role names are surface oriented because this makes it possible for a user not trained in linguistics to define his own words and relations. (For a complete list of standard role names see e.g. Zoeppritz (1983).)

We are currently working on the integration of the concepts underlying the USL system with Discourse Representation Theory which is described in the next section. We have already implemented a procedure which generates Discourse Representation Structures from USL's semantic trees and

which covers the entire fragment of language described in Kamp (1981).

2.2 Discourse Representation Theory (DRT)

In this section we give a brief description of Kamp's Discourse Representation Theory (DRT) in as much as it relates to our concerns with pronominalization. For a more detailed discussion of this theory and its general ramifications for natural language processing, cf. the papers by Kamp (1981) and Guentzner (1983a, 1983b).

According to DRT, each natural language sentence (or discourse) is associated with a so-called Discourse Representation Structure (DRS) on the basis of a set of DRS formation rules. These rules are sensitive to both the syntactic structure of the sentences in question as well as to the DRS context in which in the sentence occurs. In the formulation of Kamp (1981) the latter is really of importance only in connection with the proper analysis of pronouns. We feel on the other hand that the DRS environment of a sentence to be processed should determine much more than just the anaphoric assignments. We shall discuss this issue - in particular as it relates to problems of ambiguity and vagueness - in more depth in a forthcoming paper.

A DRS K for a discourse has the general form

$$K = \langle U, \text{Con} \rangle$$

where U is a set of "discourse referents" for K and Con a set of "conditions" on these individuals. Conditions can be either atomic or complex. An atomic condition has the form

$$P(t_1, \dots, t_n)$$

or

$$t_1 = c$$

where t_i is a discourse referent and c a proper name and P an n -place predicate.

The only complex condition we shall discuss here is the one representing universally quantified noun phrases or conditional sentences. Both are treated in much the same way. Let us call these "implicational" conditions:

$$K1 \text{ IMP } K2$$

where $K1$ and $K2$ are also DRSs. With a discourse D is thus associated a Discourse Representation structure which represents D in a quantifier-free "clausal" form, and which captures the propositional import of the discourse by - among other things, establishing the correct pronominal connections.

What is important for the treatment of anaphora in the present context is the following:

a) Given a discourse with a principal DRS K_0 and a set of non-principal DRSs (or conditions) K_i among its conditions all discourse referents of K_0 are admissible referents for pronouns in sentences or

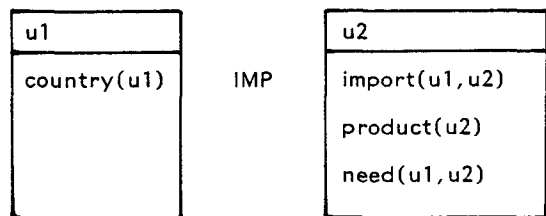
(phrases) giving rise to the various embedded K_i 's. In particular, all occurrences of proper names in a discourse will always be associated with discourse referents of the principal DRS K_0 . (This is on the (admittedly unrealistic) assumption that proper names refer uniquely.)

b) Given an implicational DRS of the form $K1 \text{ IMP } K2$ occurring in a DRS K , a relation of relative accessibility between DRSs is defined as follows:

$K1$ is accessible from $K2$ and all K' accessible from $K1$ are also accessible from $K2$.

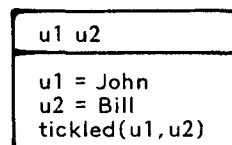
In particular, the principal DRS K_0 is accessible from its subordinate DRSs (for a precise definition cf. Kamp (1981)). The import of this definition for anaphora is simply that if a pronoun is being resolved (i.e. interpreted) in the context of a DRS K' from which a set K of DRSs is accessible, then the union of all the sets of discourse referents associated with every K_i in K is the set of admissible candidates for the interpretation of the pronoun. The following illustrations will make this clear:

$K(\text{Every country imports a product it needs})$



This sentence (as well as its interrogative version) allows only one interpretation of the pronoun *it* according to DRT. It does not introduce any discourse referent available for pronominalization in later sentences (or queries). But in a DRS like the following, DRT does not - as it stands - account for pronoun resolution:

$K(\text{John tickled Bill. He squirmed})$



At this point, the pronoun *he* has to be interpreted. There are two admissible candidates, $u1$ and $u2$, but DRT does not choose between them. So the DRS could be continued with either

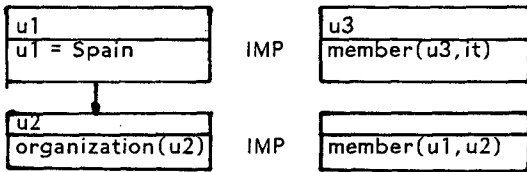
squirm(u1)

or

squirm(u2)

Similarly, in the following DRS

K (If Spain is a member of every organization, it has a member)



the pronoun *it* could only refer to Spain (on configurational grounds), and would have to be assigned that object if no other criteria are assumed. Obviously, as far as this sentence and the intended database is concerned, we should want to rule out such an assignment. (This can be done via rule S1 discussed below.)

In general, then, given a sentence (or discourse) represented in a DRS there will be more candidates for admissible pronoun assignments as one should like to have available when a particular pronoun is to be interpreted. The rules described in Section 3 are meant to capture some of the regularities that arise in typical database querying interactions.

c) Finally, given a DRS for a discourse *D* we can say that a pronoun is properly referential iff it is represented by (i.e. eliminated in favor of) a discourse referent *u_i* occurring in the domain of the principal DRS representing *D*. (In the context of the constructions illustrated so far, this will be true in particular of proper names as well as of indefinite noun phrases not in the scope of a universal noun phrase or a conditional.)

The main problem then for the treatment of anaphora is to determine which possible discourse referents should be chosen when we come to the interpretation of a particular pronoun occurrence *p_i* in the formation of the extension of the DRS in which we are working.

We would like to suggest the following strategy as a starting point. Consider a query dialogue *Q* with an already established DRS *K* and the utterance of a query *S*, where *S* contains occurrences of personal pronouns. Suppose further that *A(S)* is the sole syntactic analysis available for *S*. Then we regard the construction of the extension of the DRS obtained on the basis of *S* and *K* as the value of a partial function *f* defined on *K* and *A(S)*. More generally still, as Kamp himself suggests, we can regard the "meaning" (or information content) of a sentence to be that partial function from DRSs to DRSs.

In a given dialogue both the queries and the answers will have the side effect of introducing new individuals and "preference" or "salience" orderings on these individuals, and we want to allow for pronominal reference to these much in the same way that in a text preceding sentences may have determined a set of possible antecedents for pronouns in the currently processed sentence. The DRS built up in the process of a querying session will constitute the "mutual knowledge" available to the user in specifying his further queries as well

as in his uses of pronouns. It is on the individuals introduced in the DRSs that the rules to be discussed below are intended to operate.

3 Interplay of syntax, semantics, and pragmatics in pronominalization

The process of pronominalization is governed by rules involving morphological, syntactic, semantic, and pragmatic criteria. These rules are discussed and illustrated with examples drawn from the context of querying a geographical database. Then a procedure is outlined which uses these rules and applies them in the following order:

- First morphological criteria are checked, if they fail no further tests are required.
- Then syntactic (or configurational) criteria are tested. Again, if they fail, no further tests are necessary.
- Next semantic criteria are applied, and if they do not fail,
- the pragmatic criteria have to be tested. If more than one candidate remains, the use of the pronoun was pragmatically inappropriate and must be noted as such.

3.1 Strict factors determining the admissibility of anaphora

3.1.1 Morphological criteria

Morphological criteria concern the agreement of gender and number. Complications come in, when coordinated noun phrases occur, e.g.

- (2) John and Bill went to Pisa. They delivered a paper.
- (3) *John and Bill went to Pisa. He delivered a paper.
- (4) John and Sue went to Pisa. He delivered a paper.
- (5) *John or Bill went to Pisa. They delivered a paper.
- (6) *John or Bill went to Pisa. He delivered a paper.
- (7) Neither John nor Bill went to Pisa. They went to Rome.
- (8) *Either John or Bill did not go to Pisa. He went to Rome.

The starred examples contain inappropriate uses of pronouns. With and-coordination, reference to the complete NP is possible with a plural pronoun. When the members of the coordination are distinct in gender and/or number, reference to them is possible with the corresponding pronouns. Clearly, the same observations hold for interrogative sentences.

3.1.2 Configurational criteria

Syntactic criteria operate only within the boundaries of a sentence, outside they are useless. The configurational criteria stemming from DRT however work independent of sentence boundaries.

Disjoint reference

The rule of "disjoint reference" according to Reinhart (1983) goes back to Chomsky and has been refined by Lasnik (1976) and Reinhart (1983). It is able to handle a variety of well-known cases, such as

- (9) When did it join the UN?
- (10) Which countries that import it, produce petrol?
- (11) *Does it entertain diplomatic relations with Spain's neighbor?

(In the starred example, the use of "it" is inappropriate, if it is to be coreferential with "Spain".)

Rather than using c-command to formulate this criterion, which is elegant but too strict in some cases (as noted by Reinhart herself and Bolinger (1979), we have chosen an admittedly less elegant, but hopefully reliable, approach to disjoint reference, in that we specify the concrete syntactic configurations where disjoint reference holds. We do not rely here on the syntactic framework of USL grammar, but use more or less traditionally known terminology for expressing our rules. We need the terms "clause", "phrase", "matrix", "embedding", and "level". These can be made explicit, when a suitable syntactic framework is chosen.

Now we can formulate our disjoint reference rule and some of its less obvious consequences.

C1. The referent of a personal pronoun can never be within the same clause at the same phrase level. (Note that this rule does not hold for possessive pronouns.)

C1 has a number of consequences which we now list:

C1a. The (implicit) subject of an infinitive clause can never be referent of a personal pronoun in that clause

- (12) Does the EC want to dissolve it?

C1b. Nouns common to coordinate clauses cannot be referred to from within these coordinate clauses

- (13) Which country borders it and Spain?

C1c. Noun complements of nouns in the same clause can never be referred to.

- (14) Does it border Spain's neighbors?

The following rules have to do with phrases and clauses modifying a noun. They too can be regarded as consequences of C1.

C2. Head noun of a phrase or clause can never be referent of a personal pronoun in that phrase or clause

C2a. Head noun of participial phrase

- (15) a country exporting petrol to it

C2b. Head noun of that-clause

- (16) the truth is that it follows from A.

C2c. Head noun of relative clause

- (17) the country it exports petrol to

The following two rules deal with kataphoric pronominalization (sometimes called backward pronominalization).

C3a. Kataphora into a more deeply embedded clause is impossible

- (18) Did it export a product that Spain produces?

C3b. Kataphora into a succeeding coordinate clause is impossible

- (19) Who did not belong to it but left the UN?

The accessibility relation on DRSs

C4. Only those discourse referents in the accessibility relation defined in sec. 2.2 are available as referents to a pronoun.

3.1.3 Semantic criteria

Widely used is the criterion of semantic compatibility. It is usually implemented via "semantic features". In the USL framework we can derive this information from relation schemata. We state the criterion as follows:

S1. If *s* is a sentence containing a pronoun *p* and *c* a full noun phrase in the context of *p*. If *p* is substituted by *c* in *s* to yield *s'* and *s'* is not semantically anomalous, i.e. does not imply a contradiction, then *c* is semantically compatible with *s* and is hence a semantically possible candidate for the reference of *p*.

- (20) What is the capital of Austria? - Vienna. What does it export?

If it is assumed that only countries but not capitals export goods, then the only semantically possible referent for "it" is Austria.

S2. Non-referentially introduced nouns cannot be antecedents of pronouns.

- (21) Which countries does Italy have trade with? How large is it?

Since "trade" is used non-referentially, it cannot be antecedent of "it". Unfortunately, in many cases where this criterion could apply, there is an ambiguity between referential and non-referential use.

Apart from the type of semantic compatibility covered by rule S1, more complex semantic properties are used to determine the referent of a pronoun. The "task structures" described by Grosz (1977) illustrate this fact. We hence formulate the rule

S3. The properties of and relationships between predicates determine pronominalizability.

For an illustration of its effect, consider the following query:

(22) What country is its neighbor?

The irreflexivity of the neighbor-relation entails that "its" cannot be bound by "what country" in this case, but has to refer to something mentioned in the previous context.

Given a subject domain, one can analyze the properties of the relations and the relationships between them and so build a basis for deciding pronoun reference on semantic grounds. In the framework of the USL system, information on the properties of relations is available in terms of "functional dependencies" given in the database schema or as integrity constraints.

3.2 Pragmatic criteria

The generation of discourse is controlled by two factors: communicative intentions and mutual knowledge. In the context of database interaction, we can assume that the communicative intentions of a user are simply to obtain factual answers to factual questions. His intentions are expressed either by single queries or by sequences of queries, depending on how complex these intentions are or how closely they correspond to the information in the database. As will be shown below, in many cases the system will not have a chance to determine whether a given query is a "one-shot query", or whether it is part of a sequence of queries with a common "theme". For the resolution of pronouns, this means that the system should rather ask the user back than make wild guesses on what might be the most "plausible" referent. This is of course not possible when running text is analyzed in a "batch mode", and no user is there to be asked for clarification.

Mutual knowledge (see e.g. Clark and Marshall (1981) for a discussion) determines the rules for introducing and referencing individuals in the discourse. In the context of database interaction we assume the mutual knowledge to consist initially of:

- the set of proper names in the database,
- the predicates whose extensions are in the database,
- the "common sense" relationships between and properties of these predicates.

It will be part of the design of a database to establish what these "common sense" relationships and properties are, e.g. whether it is generally known to the user community, whether "capital" expresses a one-one relation. Each question-answer pair occurring in the discourse is added to the stock of mutual knowledge.

It is a pragmatic principle of pronominalization that only mutual knowledge may be used to determine the referent of a pronoun on semantic grounds, and hence it may be legal to use the same sentence containing a pronoun where earlier in the

discourse it was illegal, because the mutual knowledge has increased in the meantime.

3.2.1 A first attempt using preference rules

What the topic of a discourse is, which of the entities mentioned in it are in focus, is reflected in the syntactic structure of sentences. This has been observed for a long time. It has also often been observed that discourse topic and focus have an effect on pronominalization where morphological, configurational, and semantic rules fail to determine a single candidate for reference. However, it has not been possible yet to formulate precise rules explaining this phenomenon. We have the impression that such rules cannot be absolutely strict rules, but are of a preferential nature. We have developed a set of such rules and tested them against a corpus of text containing some 600 pronoun occurrences, and have found them to work remarkably well. Similar tests (with a similar set of rules) have been conducted by Hofmann (1976).

In the sequel we formulate and discuss our list of rules. Their ordering corresponds to the order in which they have to be applied.

P1 (principle of proximity). Noun phrases within the sentence containing the pronoun are preferred over noun phrases in previous or succeeding sentences.

Consider the sequence

- (23) What country joined the EC after 1980? - Greece.
(24) What country consumes the wine it produces?

One could argue that "Greece" is just as probably the intended referent of "it" in this case as the bound interpretation and that hence the use of "it" should be rejected as inappropriate. However, there is no way to avoid the "it", if the bound variable interpretation is intended, and one can use this as a ground to rule out the interpretation where "it" refers to "Greece".

P1a. Noun phrases in sentences before the sentence containing the pronoun are preferred over noun phrases in more distant sentences.

This criterion is very important to limit the search for possible discourse referents.

P2. Pronouns are preferred over full noun phrases.

This rule is found in many systems dealing with anaphora. One can motivate it by saying that pronominalization establishes an entity as a theme which is then maintained until the chain of pronouns is broken by a sentence not containing a suitable pronoun. For an example consider:

- (25) What is the area of Austria?
(26) What is its capital?
(27) What is its population?

P3. Noun phrases in a matrix clause or phrase are preferred over noun phrases in embedded clauses or phrases.

P3a. Noun phrases in a matrix clause are preferred over noun phrases in embedded clauses.

Example:

- (28) What country imports a product that Spain produces? - Denmark.
(29) What does it export?

Here "it" has to refer to the individual satisfying "what country", not to "Spain" which occurs in an embedded clause.

P3b. Head nouns are preferred over noun complements.

Example:

- (30) What is the capital of Austria? - Vienna.
(31) What is its population?

"Vienna", not "Austria" becomes the referent of "its", and the argument is analogous to that for P3a.

P4. Subject noun phrases are preferred over non-subject noun phrases.

In declarative contexts, this rule works quite well. It corresponds essentially to the focus rule of Sidner (1981). In a question-answering situation it is hardly applicable, since especially in wh-questions subject position and word order, which both play a role, tend to interfere. We therefore tend to not use this rule, but rather to let the system ask back in cases where it would apply. For illustration consider the following examples:

- (32) Does Spain border Portugal? What is its population?
(33) Is Spain bordered by Portugal? What is its population?
(34) Which country borders Portugal? What is its population?
(35) Which country does Portugal border? What is its population?

P5. Accusative object noun phrases are preferred over other non-subject noun phrases.

P6. Noun phrases preceding the pronoun are preferred over noun phrases succeeding the pronoun (or: anaphora is preferred over kataphora).

3.3 Outline of a pronoun resolution procedure

We now outline a procedure for "resolving" pronouns in the framework of the USL system and DRT.

Let $M = \langle U, Con \rangle$ be the DRS representing the mutual knowledge, in particular the past discourse. Let $K(s)$ be the DRS representing the current sentence s and let p be a pronoun occurring in s for which an appropriate discourse referent has to be

found. Let $U_{a(p)}$ be the set of discourse referents accessible to p according to the accessibility relation given in sec. 2.2

Let further c be a function that applies to $U_{a(p)}$ all the morphological, syntactic, and semantic criteria given above and yields a set $U_{c(p)}$ as result. Now three cases have to be distinguished:

1. $U_{c(p)}$ is empty. In this case the use of p was inappropriate.
2. $Card(U_{c(p)})$ is 1. In this case a referent for p has been uniquely determined, p is replaced by it in the DRS, and the procedure is finished.
3. $Card(U_{c(p)})$ is greater than 1. In this case the preference rules are applied.

Let p be a function that applies to $U_{c(p)}$ if the cardinality of $U_{c(p)}$ is greater than 1 all the preference rules given above in the order indicated there yielding the result U_p . $Card(U_p)$ can never be 0, hence two cases are possible, either the cardinality is 1, then a referent has been uniquely determined and the pronoun p can be eliminated in K , or the cardinality is greater than 1, and then the use of p was inappropriate.

It can be inferred from the formulation of the pronominalization rules given above, what morphological and syntactic information has to be stored with the discourse referents in the DRSs, and what semantic information has to be accessible from the schema of the database to enable the application of the functions c and p . Hence, we will not spell out these details here.

4 Open questions and conclusions

Many well-known and puzzling cases have not been addressed here, among them plural anaphora, so-called pronouns of laziness, one pronominalization, to name just a few.

We have not said anything about phenomena such as discourse topic, focus, or coherence and their influence on anaphora. Their effects are captured in our preference rules to some degree, but no one can precisely say how. In spite of claims to the contrary, we believe that much work is still required, before these notions can be used effectively in natural language processing.

By limiting ourselves to the relatively well-defined communicative situation of database interaction, we have been able to state precisely, what rules are applicable in the fragment of language we are dealing with. We are currently working on the analysis of running texts, but again in a well-delineated domain, and we hope to be able to extend our theory on the basis of the experience gained.

We are convinced that serious progress in the understanding of anaphora and of discourse phenomena in general is only possible through a careful control of the environment, and on a solid syntactic and semantic foundation.

References

- Astrahan, M. M., M. W. Blasgen, D. D. Chamberlin, K. P. Eswaran, J. N. Gray, P. P. Griffiths, W. F. King, R. A. Lorie, P. R. McJones, J. W. Mehl, G. R. Putzolu, I. L. Traiger, B. W. Wade, V. Watson (1976): "System R: Relational Approach to Database Management", *ACM Transactions on Database Systems*, vol. 1, no. 2, June 1976, p. 97.
- Bertrand, O., J. J. Laudennarde, D. Starynkevich, A. Stenbock-Fermor (1976): "User Application Generator", *Proceedings of the IBM Technical Conference on Relational Data Base Systems*, Bari, Italy, p. 83.
- Bolinger, D. (1979): "Pronouns in Discourse", in: T. Givon (ed.): *Syntax and Semantics*, Vol. 12: *Discourse and Syntax*, Academic Press, New York, p. 289.
- Chastain, Ch. (1973): *Reference and Context*, Thesis, Princeton.
- Clark, H. H. and C. R. Marshall (1981): "Definite Reference and Mutual Knowledge", in: B. L. Webber, A. K. Joshi, and I. A. Sag (eds.): *Elements of Discourse Understanding*, Cambridge University Press, Cambridge, p. 10.
- Donnellan, K. S. (1978): "Speaker Reference, Descriptions and Anaphora", in P. Cole (ed.): *Syntax and Semantics*, Vol. 9: *Pragmatics*, Academic Press, New York, p. 47.
- Evans, G. (1980): "Pronouns", *Linguistic Inquiry*, vol. 11.
- Grosz, B. J. (1977): "The Representation and Use of Focus in Dialogue Understanding", *Technical Note 151*, SRI International, Menlo Park, California.
- Guenther, F. (1983a) "Discourse Representation Theory and Databases", forthcoming.
- Guenther, F. (1983b) "Representing Discourse Representation Theory in PROLOG", forthcoming.
- Hirst, G. (1981): *Anaphora in Natural Language Understanding: A Survey*, Springer, Heidelberg.
- Hofmann, J. (1976): "Satzexterne freie nicht-referentielle Verweisformen in juristischen Normtexten, unpublished dissertation, Univ. Regensburg.
- Kamp, H. (1981) "A Theory of Truth and Semantic Representation", in Groenendijk, J. et al. *Formal Methods in the Study of Language*. Amsterdam.
- Lasnik, H. (1976): "Remarks on Coreference", *Linguistic Analysis*, vol. 2, nr. 1.
- Lehmann, H. (1978): "Interpretation of Natural Language in an Information System", *IBM J. Res. Develop.* vol. 22, p. 533.
- Lehmann, H. (1980): "A System for Answering Questions in German", paper presented at the 6th International Symposium of the ALLC, Cambridge, England.
- Ott, N. and M. Zoeppritz (1979): "USL - an Experimental Information System based on Natural Language", in L. Bolc (ed): *Natural Language Based Computer Systems*, Hanser, Munich.
- Ott, N. and K. Horländer (1982): "Removing Redundant Join Operations in Queries Involving Views", TR 82.03.003, IBM Heidelberg Scientific Center.
- Reinhart, T. (1979): "Syntactic Domains for Semantic Rules", in F. Guenther and S. J. Schmidt (eds.): *Formal Semantics and Pragmatics for Natural Languages*, Reidel, Dordrecht.
- Reinhart, T. (1983): "Coreference and Bound Anaphora: A Restatement of the Anaphora Questions", *Linguistics and Philosophy*, vol. 6, p. 47.
- Sidner, C. L. (1981): "Focusing for Interpretation of Pronouns", *AJCL*, vol. 7, nr. 4, p. 217.
- Smaby, R. (1979): "Ambiguous Coreference with Quantifiers", in F. Guenther and S.J. Schmidt (eds.) *Formal Semantics and Pragmatics for Natural Languages*, Reidel, Dordrecht.
- Smaby, R. (1981): "Pronouns and Ambiguity", in U. Mönnich (ed.): *Aspects of Philosophical Logic*, Reidel, Dordrecht.
- de Sopeña Pastor, L. (1982): "Grammar of Spanish for User Specialty Languages", TR 82.05.004, IBM Heidelberg Scientific Center.
- Webber, B. L. (1978): "A Formal Approach to Discourse Anaphora", TR 3761, Bolt, Beranek & Newman, Cambridge, MA.
- Zoeppritz, M. (1983): *Syntax for German in the User Specialty Languages System*, Niemeyer, Tübingen.