Resolving Anaphoric References on Deficient Syntactic Descriptions

Roland Stuckardt

Im Mellsig 25

D-60433 Frankfurt am Main, Germany

stuckardt@compuserve.com

Abstract

Syntactic coindexing restrictions are by now known to be of central importance to practical anaphor resolution approaches. Since, in particular due to structural ambiguity, the assumption of the availability of a unique syntactic reading proves to be unrealistic, robust anaphor resolution relies on techniques to overcome this deficiency. In this paper, two approaches are presented which generalize the verification of coindexing constraints to deficient descriptions. At first, a partly heuristic method is described, which has been implemented. Secondly, a provable complete method is specified. It provides the means to exploit the results of anaphor resolution for a further structural disambiguation. By rendering possible a parallel processing model, this method exhibits, in a general sense, a higher degree of robustness. As a practically optimal solution, a combination of the two approaches is suggested.

1 Introduction

The interpretation of anaphoric expressions is known to be a difficult problem. In principle, a variety of constraints and preference heuristics, including factors which rely on semantic, pragmatic, and world knowledge, contribute to this task (Carbonell and Brown, 1988). Operational approaches to anaphor resolution on unrestricted discourse, however, are confined to strategies exploiting globally available evidence like morphosyntactic, syntactic, and surface information.

Among the most promising practical work are approaches relying on the availability of syntactic surface structure by employing coindexing restrictions, salience criteria, and parallelism heuristics (Lappin and Leass, 1994; Stuckardt, 1996b). However, even the assumption of the availability of a unique syntactic description is unrealistic since, in general, parsing involves the solution of difficult problems like attachment ambiguities, role uncertainty, and the instantiation of empty categories. Based on this observation, Kennedy and Boguraev suggest an adaptation of the Lappin and Leass approach to the analysis frontend of English Constraint Grammar (Karlsson et al., 1995), which provides a part-of-speech tagging comprising an assignment of syntactic function but *no constituent structure*. This information deficiency is partially overcome by the application of a regular filter which heuristically reconstructs constituent structure (Kennedy and Boguraev, 1996).

The approach of Kennedy and Boguraev resorts to shallow input and heuristic reconstruction of surface structure in general, thus leaving open the question what may be gained by relying on the possibly partial, but potentially more reliable output of a conventional parser. This question is dealt with in the present paper. An operational approach to anaphor resolution is advocated which achieves robustness by a generalization to deficient syntactic descriptions rather than by resorting to shallow input. In section 2, notions of robustness are defined according to which different methods may be classified. Section 3 develops the perspective of fragmentary syntax and identifies the coindexing restrictions of binding theory as an important anaphor resolution strategy which is in particular affected by this loss of configurational evidence. In section 4, a solution is presented which accomplishes robustness against syntactic deficiency by a partly heuristic verification of coindexing constraints on fragmentary syntax. Finally, in section 5, a non-heuristic algorithm is specified which works on the standardized representation of ambiguous syntactic description by packed, shared parse forests. It achieves a higher degree of robustness by making available referential evidence

for a further disambiguation of syntactic structure. A combination of the two approaches is suggested as the practically optimal solution.

2 Notions of Robustness

2.1 Robustness in Natural Language Processing

In natural language processing in general, the robustness issue comprises the ability of a software system to cope with input that gives rise to *deficient descriptions* at some descriptional layer.[1] More or less implicit is the assumption that the system exhibits some kind of *monotonic behaviour*: the less deficient the description, the higher the quality of the output (Menzel, 1995).

Following Menzel further, this intuitive characterization may be refined. Processing should exhibit *autonomy* in the sense that complete failures at one stage of analysis should not cause complete failures at other stages of analysis or even a failure of the overall processing. Moreover, the processing model should ideally employ some kind of *interaction* between different stages of analysis: deficiency at one stage of analysis should be compensated by the information gained at other stages.

2.2 Robustness and Anaphor Resolution

In the light of the above description, the robustness requirement for the anaphor resolution task may be rendered more precisely. In the aforesaid operational approaches, a *sequential processing model* is followed according to which anaphor resolution is performed by referring to the result of an *already completed* syntactic analysis. This architecture, however, tacitly ignores evidence for structural disambiguation that may be contributed by strong expectations at the referential layer (Stuckardt, 1996a). In terms of the general goals of robust processing, this means that, since there is no interaction, the robustness requirement merely shows up in form of the monotonicity and autonomy demands: the anaphor resolution module has to cope with deficient or shallow syntactic information. Besides the trivial way to achieve this kind of robustness by simply not exploiting deficient syntactic descriptions, the following two models may be followed:

- *the shallow description model:* by exploiting heuristic rules to reconstruct syntactic description, the anaphor resolution strategies are

adapted to *shallow input data* which are never defective.[2]

- *the deficient description model:* by extending anaphor resolution strategies to work on a possibly ambiguous or incomplete description, *syntactic evidence is exploited as far as available.*

In contrast to the approach of Kennedy and Boguraev, which is based on the shallow description model, the subsequent sections develop two methods that follow the deficient description model. At first, a new partly heuristic approach will be described. Secondly, a non-heuristic algorithm will be specified which establishes the conceptually superior degree of robustness through interaction: it makes available the results of anaphor resolution for syntactic disambiguation.

3 Fragmentary Syntax

3.1 Phenomena

The main phenomena which give rise to *structural ambiguity* of syntactic descriptions are *uncertainty of syntactic function* (involving subject and direct object) and *attachment ambiguities* of prepositional phrases, relative clauses, and adverbial clauses. In the example

Peter observes the owner of the telescope with it.

depending on the availability of disambiguating information, it may be uncertain whether the underlined prepositional phrase *with it* should be interpreted adverbially or attributively. From the configurational perspective, these ambiguities give rise to *fragmentary syntactic descriptions* which consist of several tree-shaped connected components. With the exception of the topmost tree fragment, all components correspond to a syntagma of type PP, S, or NP whose attachment or role assignment failed.

In addition, *cases in which no reading exists* give rise to fragmentary syntactic descriptions comprising the constituents whose combination failed due to constraint violation.

3.2 Fragmentary Syntax and Anaphor Resolution

Among the anaphor resolution strategies potentially affected by fragmentary syntax are heuristics as well as constraints. Preference criteria like salience factors and syntactic parallelism are not affected by

---

[1]The deficiency may result either because the input itself is deficient, or due to shortcomings of the processing resources, e.g. lexicon, grammar/parser, or semantic/pragmatic disambiguation.

[2]Here, the monotonicity demand of intuitive robustness virtually vanishes, since there is no longer a syntactic input prone to deficiency.

all types of syntactic defects. Moreover, there is a plethora of heuristics which do not rely on syntactic function or structure. Structural coindexing constraints, however, may lose evidence in all above cases of fragmentary syntax. Since they are known to be of central importance to the antecedent filtering phase of operational anaphor resolution approaches, the subsequent discussion focuses on the impact of deficient surface structure description to this class of restrictions.

According to the Government and Binding Theory of Chomsky, the core of the syntactic coindexing restrictions is stated as follows

Definition 1 (binding principles)

(A) A reflexive or reciprocal is bound in its binding category.

(B) A pronominal is free (i.e. not bound) in its binding category.

(C) A referring expression is free in any domain.

where *binding category* denotes the next dominator containing some kind of subject (Chomsky, 1981), and *binding* is defined as *coindexed and c-commanding*:

Definition 2 (the c-command relation)
Surface structure node X c-commands node Y if and only if the next "branching node" which dominates X also dominates Y and neither X dominates Y, Y dominates X nor X = Y.

A further structural well-formedness condition, commonly named *i-within-i filter*, rules out "referentially circular" coindexings, i.e. configurations matching the pattern $[\alpha \ldots [\beta \ldots]_i]_i$.

In the above example, the latter restriction comes to an application, licensing a coindexing of *telescope* and *it* only if the PP containing *it* is not interpreted as an attribute to *telescope* - otherwise, in contradiction to the i-within-i condition, the pronoun would be contained in the NP of the tentative antecedent. Hence, if the PP attachment ambiguity has not been resolved prior to anaphor resolution, the fragmentary syntactic description does not contribute the configurational evidence which is necessary for definitely confirming antecedent candidate *telescope*.

4 Checking Binding Constraints on Fragmentary Syntax

4.1 Basic Observations

The first step towards the verification of binding constraints on fragmentary syntax is suggested by the following observation:

If the anaphor as well as the antecedent candidate are contained in the same connected component of the fragmentary syntactic description, no (direct) binding theoretic evidence is lost.

In this case, the verification of the binding restrictions of anaphor and antecedent will be possible in a non-heuristic manner, since the necessary positive ($\rightarrow$ binding principle A) and negative ($\rightarrow$ binding principles B, C) syntactic-configurational evidence is entirely available.[3] If, however, the two occurrences belong to different fragments, relevant information may be lost.

These considerations give rise to a first solution: to be able to detect which one of the two cases holds, the descriptions of the discourse referent occurrences are supplemented with an attribute which uniquely identifies of the syntactic fragment to which the corresponding NP[4] belongs (e.g. a pointer to the root of the fragment, or a natural number). For a given pair of anaphor α and antecedent candidate γ, the following procedure is applied:

If anaphor α and candidate γ occur in the same fragment, verify binding restrictions as in case of unique syntactic description. If they occur in different fragments, consider γ a configurationally acceptable candidate for α, but reduce the plausibility score associated with the pair (α, γ).

Consequently, in certain, recognizable cases, the robust binding constraint verification merely yields a *heuristic* approval of coindexing. The strategy loses a part of its former strictness because configurational evidence is only partially available.

4.2 Rule Patterns

Even in the disadvantageous case, a closer look at the tree fragments of anaphor and antecedent candidate may reveal additional information. Figure 1 shows rule patterns which exploit this evidence.[5]

---

[3]This statement, however, solely applies to the direct comparison of the involved occurrences, since in case of further, transitive coindexings, negative evidence stemming from decision interdependency (cf. (Stuckardt, 1996a)) may get lost.

[4]This slightly sloppy assumption of a bijection between NP nodes of syntactic structure and discourse referent occurrences does not affect the validity of the subsequent discussion.

[5]The following notational conventions are used: round brackets delimit constituents; square brackets emphasize fragment boundaries; $bc(X)$ denotes the binding category of surface structure node X; $bn(X)$ denotes

$$\text{[F1]} \quad \checkmark \, \{ \, \ldots \, F_i = [\, \ldots \, bc(\gamma)(\ldots \gamma_{type\,B} \ldots) \, \ldots] \, , \, \ldots \, , \, F_j = [\, \ldots \, bc(\alpha)(\ldots \alpha_{type\,B} \ldots) \, \ldots] \, \ldots \}$$

$$\text{[F2]} \quad * \, \{ \, \ldots \, F_i = [\, \ldots \, bn(\gamma)(\ldots \gamma_{type\,B/C} \ldots) \, \ldots] \, , \, \ldots \, , \, F_j = [\, \ldots \, bc(\alpha)(\ldots \alpha_{type\,A} \ldots) \, \ldots] \, \ldots \}$$

$$\text{[FE1]} \quad \checkmark \, \{ \, \ldots \, F_d = [\, \ldots \, \gamma_{type\,B/C} \ldots] \, , \, \ldots \, , \, F_e = [\, \ldots \, bc(\alpha)(\ldots \alpha_{type\,B} \ldots) \, \ldots] \, \ldots \}$$

$$\text{[FE2]} \quad * \, \{ \, \ldots \, F_d = [\, \ldots \, \gamma_{type\,B/C} \ldots] \, , \, \ldots \, , \, F_e = [\, \ldots \, bc(\alpha)(\ldots \alpha_{type\,A} \ldots) \, \ldots] \, \ldots \}$$

$$\text{[FE3]} \quad * \, \{ \, \ldots \, F_d = [\, \ldots \, \gamma_{type\,B/C} \ldots] \, , \, \ldots \, , \, F_e = [\, \ldots \, \alpha_{type\,C} \ldots] \, \ldots \}, \quad \text{if } \gamma \text{ c-commands } \alpha$$
independently of the attachment choice

$$\text{[FE4]} \quad * \, \{ \, \ldots \, F_d = [\, \ldots \, \alpha_{type\,A} \ldots] \, , \, \ldots \, , \, F_e = [\, \ldots \, \gamma_{type\,B/C} \ldots] \, \ldots \}, \quad \text{if } root(F_e) \neq \gamma$$

Figure 1: rule patterns for binding constraint verification on fragmentary syntax

In fragmentations matching pattern [F1], both fragments are constituents which contain the binding categories $bc(\alpha)$ and $bc(\gamma)$ of the respective occurrences α and γ of type B. In particular, this implies that the fragments may not be attached in a way that one occurrence locally c-commands the other. Therefore, in both cases, binding principle B is respected, and the coindexing of α and γ is *non-heuristically* approved.

Conversely, if pattern [F2] is matched, γ is definitively ruled out as the local antecedent prescribed for type A anaphor α: it is impossible to connect the two fragments in a way that γ locally c-commands α. Here, the fragment of the candidate is only required to contain the branching node of γ, because this suffices to preclude that γ c-commands α if F_i is embedded in F_j.

For certain successive fragment pairs, the parsing result comprises additional information about immediate or transitive *embedding*. Based on this evidence, further non-heuristic rules ([FE1],[FE2], [FE3], and [FE4]) become applicable (F_d = dominating fragment, F_e = embedded fragment).

This list may be supplemented by rules which are based on more subtle configurational case distinctions, and, moreover, by *heuristic rules* which employ standardized assumptions about typical decision patterns of structural disambiguation. With each of the latter rules, an *individual* decision plausibility weight may be associated.

In an application context, the extension of heuristic rules should be limited to configurations which are known to be of practical relevance. Based on a suitable corpus annotated with syntactic and referential information, it should be possible to determine prob-

abilities for different coindexing configurations by a statistical distribution analysis. These probabilities may then be used to derive promising plausibility weighted rules.

4.3 An example

The following example illustrates the application of some of the above rules:[6]

> *Der Mann hat den Präsidenten besucht,*
> *der ihn von sich überzeugte.*

> *"The man has the president visited,*
> *who him from himself convinced."*

Because of the intervening past participle, the relative clause may be interpreted as an attribute to either *Mann* or *Präsidenten*. Hence, syntactic ambiguity arises, yielding a surface structure description which consists of the following two fragments

> *(S Mann*
> *(VP Präsident))*

> *(S der*
> *(VP ihn*
> *(VP (PP sich))))*

In addition, it is known that the second fragment is embedded in the first. There are three pronominal anaphors to be resolved: the reflexive pronoun *sich* of type A, the nonreflexive pronoun *ihn* of type B, and the relative pronoun *der* of type B.

For the reflexive pronoun *sich*, the syntactic restrictions may be applied nonheuristically. Candidates *der* and *ihn* are contained in the same surface structure fragment. Consequently, binding theoretic evidence is completely available. Since the candidates locally c-command *sich*, they are both determined to be possible antecedents. The two candidates *Mann* and *Präsident*, however, occur in the other fragment. Hence, it is attempted to apply one of the above rule

---

the branching node dominating X according to the c-command definition; the subscript of $X_{type\,Y}$ denotes that the binding theoretic class of the occurrence contributed by X is $Y \in \{A, B, C\}$, e.g. $P_{type\,B}$ is a pronominal. $\checkmark/*$ indicates that a rule pattern admits/forbids coindexing.

[6]The example is given in German, because the structural ambiguity comes out more strikingly.

patterns. Since the reflexive pronoun is of binding theoretic type A, and the fragment in which it occurs contains its binding category (the S node of the relative clause), the F_e fragment pattern of rule [FE2] is matched; analogously, the (dominating) fragment containing the type C candidates matches the F_d fragment pattern. Hence, rule [FE2] applies: it non-heuristically rules out *Mann* as well as *Präsident*. Similarly, for the pronouns *ihn* and *der*, type C candidates *Mann* and *Präsident* are definitively confirmed. Since these anaphors are of type B, the F_e fragment pattern of rule [FE1] is matched. Moreover, the (dominating) antecedent candidate fragment matches pattern F_d. Consequently, [FE1] applies and predicts the admissibility of the candidates.

4.4 An Implementation

The above technique for achieving robustness according to the deficient description model has been integrated into an anaphor resolution system for German text (Stuckardt, 1996b).[7] At present, only nonheuristic rules are employed.

In a quantitative evaluation on a corpus of architect biographies (Lampugnani, 1983), the algorithm correctly resolved about 82 per cent of type B pronouns (including possessives). In an idealized test scenario in which correct syntactic readings were manually provided, a precision of 90 per cent was obtained. Hence, on fragmentary syntax, the result quality only decreases by 8 points of percentage. Compared with the 75 per cent achieved by the shallow description approach of Kennedy and Boguraev, this indicates that approaches to robust anaphor resolution which follow the deficient description model may achieve a higher precision. A principled, instructive comparison based on a broader set of text genres has to confirm this improvement.

5 A Complete Algorithm

By the nonheuristic rules of the above method, only those parts of surface structure description are exploited which are valid independently of further disambiguation. It is, however, possible to follow a more principled approach which utilizes configurational evidence that is confined to certain readings. As it will be shown in the following, this may be achieved by tracing the reading dependency of antecedent decisions relying on particular configurations. Through this technique, the results of referential disambiguation can be utilized as evidence for further narrowing structural ambiguity.

5.1 Dominance Relations in Packed Shared Forests

Following a standardized framework, ambiguous parsing results are henceforth assumed to be represented as *packed shared forests* (PSFs) (Tomita, 1985). In this representation, structural ambiguity is encoded by *packing* different derivation variants of input substrings into single interior nodes. Moreover, subtrees common to different readings are allowed to be *shared*. Formally, such a parse forest can be described as a *directed acyclic graph* (DAG) with a distinguished topmost element, and leaves corresponding to input words.

Given a PSF T with nodes $V = \{v_1, \ldots, v_k\}$, let $P(v_1), \ldots, P(v_k)$ be the respective derivation variants according to packing. Hence,

$$n := \prod_{i=1}^{k} |P(v_i)|$$

denotes the maximum number of readings (parse trees) represented by T. Consequently, sets of readings may be specified by bit vectors of length n.

The application of binding principles crucially rests on the availability of information about dominance relations between parse tree nodes. In ambiguous structure, configurational relations may be confined to certain subsets of readings. The idea now is to qualify dominance information by bit vectors of length n. For each pair (v_i, v_o) consisting of an interior node v_i and an occurrence contributing node v_o (usually preterminal), vectors $\vec{ad}(v_i, v_o)$ and $\vec{ld}(v_i, v_o)$ are introduced. Vectors $\vec{ad}(v_i, v_o)$ and $\vec{ld}(v_i, v_o)$ characterize the readings in which v_i arbitrarily dominates, or (in the sense of binding theory) locally dominates v_o. Based on these vectors, it will be possible to apply the binding restrictions in a reading sensitive way.

By generalizing a technique for unambiguous syntax (Correa, 1988), the reading-qualified dominance information may be precomputed as follows.[8] The tree traversal process starts at the preterminal nodes which are assumed to be shared among all readings. (If this condition is not satisfied, a preprocessing is performed which, by topdown propagation, determines reading characterization vectors which are then taken for a *qualified* initialization of the vectors $\vec{ld}$ and $\vec{ad}$.) Each preterminal node v_p is assigned the following vectors:

$$\vec{ad}(v_p, v_o) = \vec{ld}(v_p, v_o) := \begin{cases} (1, \ldots, 1), & v_p = v_o \\ (0, \ldots, 0), & v_p \neq v_o \end{cases}$$

---

[7] In its principal layout, the algorithm coincides with the one that will be described in section 5.3.

[8] Alternatively, this information may be determined on demand of the anaphor resolution task.

34

$$\vec{ad}(v_i, v_o) := \bigvee_{1 \leq m \leq |P(v_i)|} (\vec{p}(P(v_i), m) \bigwedge (\bigvee_{v_d \in D(P(v_i), m)} \vec{ad}(v_d, v_o)))$$

$$\vec{ld}(v_i, v_o) := \begin{cases} \bigvee_{1 \leq m \leq |P(v_i)|} (\vec{p}(P(v_i), m) \bigwedge (\bigvee_{v_d \in D(P(v_i), m)} \vec{ld}(v_d, v_o))), & \text{if } \neg bcateg(v_i) \\ (0, \ldots, 0), & \text{if } bcateg(v_i) \end{cases}$$

Figure 2: bottom-up computation of dominance vectors

The computation proceeds bottom-up as follows. Let v_i be an interior node for which all descendants of all derivation variants in $P(v_i)$ have already been processed. By taking into consideration whether the node delimits a local domain of binding, the vectors are assigned as shown in figure 2 (the operators $\bigwedge$ and $\bigvee$ denote bitwise conjunction and disjunction of vectors, respectively).

The computation of the vectors $\vec{ad}(v_i, v_o)$, which characterize the readings in which v_i arbitrarily dominates v_o, denotes the basic case. The outermost vector disjunction sums over the derivation variants $P(v_i)$ which, due to packing, exist for the interior node v_i. Bit vector $\vec{p}(P(v_i), m)$ acts as a filter characterizing the subset of readings in which the mth derivation variant of v_i, $1 \leq m \leq |P(v_i)|$, is valid.[9] For each derivation variant, there exists a set $D(P(v_i), m)$ of descendants v_d which correspond to nonterminals on the right-hand side of the respective rewriting rule. The nodes that are dominated by these nonterminals are transitively dominated by v_i. Hence, the overall result is obtained by recursively summing up the individual contributions of the descendants and qualifying them by conjoining them with $\vec{p}(P(v_i), m)$.

Vectors $\vec{ld}(v_i, v_o)$, which characterize the readings in which v_i locally dominates v_o, are computed similarly. The only difference arises if v_i delimits a local domain of binding. In this case, the dominance relation computation starts from scratch, i.e. with the zero vector.

5.2 Binding Principle Verification on Packed Shared Forests

The vectors $\vec{ad}(v_i, v_o)$ and $\vec{ld}(v_i, v_o)$ are used to perform referential disambiguation which is sensitive to structural ambiguity. For this purpose, during anaphor resolution, each pair of anaphor α and antecedent candidate γ (identified with the corresponding preterminal nodes) is assigned a vector $\vec{r}(\alpha, \gamma)$ characterizing the readings under which a

[9]Upon fixation of a particular encoding scheme for reading characterization, vectors $\vec{p}(P(v_i), m)$ may be computed according to a simple formula.

coindexing of α and γ is configurationally admissible. By taking into account the respective binding principles, $\vec{r}(\alpha, \gamma)$ may be determined as follows:

$$\vec{r}(\alpha, \gamma) := (\bigvee_{v \in bn(\gamma)} (\vec{p}(P(v), m) \bigwedge \vec{bp_s}(v, \alpha)))$$
$$\bigwedge (\bigvee_{v \in bn(\alpha)} (\vec{p}(P(v), m) \bigwedge \vec{bp_w}(v, \gamma)))$$

where $bn(x)$ represents the set of dominators of a node x which are branching nodes for x in the sense of the c-command definition. The first conjunct specifies the bitvector characterizing the subset of readings under which the binding principle of the anaphor α is (constructively) satisfied. Analogously, the second conjunct describes the (unconstructive) binding principle verification for the antecedent candidate γ. In both cases, branching nodes v have to be considered because they determine the starting point for the application of the dominance information. Since the property of being a branching node relatively to another node may in general depend on packing variants, reading dependency arises. Again, this subtlety is modeled by adding up a set of disjuncts which are qualified by vectors $\vec{p}(P(v), m)$, $v \in bn(x)$. Here, these vectors characterize the subset of readings in which the property of being a branching node relatively to node x holds for v.

The strong (constructive) and weak (unconstructive) verification of binding principles is accomplished by a conjunction with vectors $\vec{bp_s}(v, \alpha)$ and $\vec{bp_w}(v, \gamma)$, respectively, which, depending on the applicable binding principle, exploit the reading-qualified domination information:

$$\vec{bp_s}(v_1, v_2) := \begin{cases} \vec{ld}(v_1, v_2), & \text{if } bttype(v_2) = A \\ \overline{\vec{ld}(v_1, v_2)}, & \text{if } bttype(v_2) = B \\ \overline{\vec{ad}(v_1, v_2)}, & \text{if } bttype(v_2) = C \end{cases}$$

$$\vec{bp_w}(v_1, v_2) := \begin{cases} (1, \ldots, 1), & \text{if } bttype(v_2) = A \\ \overline{\vec{ld}(v_1, v_2)}, & \text{if } bttype(v_2) = B \\ \overline{\vec{ad}(v_1, v_2)}, & \text{if } bttype(v_2) = C \end{cases}$$

The sole difference between the strong and the weak

1. For each anaphoric NP α, determine the set of admissible antecedents γ:
   (a) Verify morphosyntactic or lexical agreement with γ
   (b) If the antecedent candidate γ is intrasentential: by checking that $\vec{r}(\alpha, \gamma) \neq (0, \ldots, 0)$, verify that
      i. the binding restriction of α is constructively satisfied,
      ii. the binding restriction of γ is not violated.
   (c) If α is a type B pronoun, antecedent candidate γ is intrasentential, and, according to surface order, γ follows α, verify that γ is definite.

2. Scoring and sorting:
   (a) For each remaining anaphor-candidate pair (α_i, γ_j), determine, according to salience and parallelism heuristics, the numerical plausibility score $v(\alpha_i, \gamma_j)$.
   (b) For each anaphor α: sort candidates γ_j according to decreasing plausibility $v(\alpha, \gamma_j)$.
   (c) Sort the anaphors α according to decreasing plausibility of their individual best antecedent candidate.

3. Antecedent Selection: Initialization $\vec{r}^* := (1, \ldots, 1)$. Consider anaphors α in the order determined in step 2c. Suggest antecedent candidates $\gamma(\alpha)$ in the order determined in step 2b. Select $\gamma(\alpha)$ as candidate if there is no interdependency, i.e. if
   (a) the morphosyntactic features of α and $\gamma(\alpha)$ are still compatible,
   (b) for each NP δ whose coindexing with $\gamma(\alpha)$ has been determined in the *current* invocation of the algorithm: the coindexing of α and δ which results as a side effect when chosing $\gamma(\alpha)$ as antecedent for α does *not violate* the binding principles, i.e. $\vec{r}(\alpha, \delta) \neq (0, \ldots, 0)$. (Here, for both occurrences, the weak predicate bp_w applies.)
   (c) $\vec{r}^* := \vec{r}^* \wedge (\bigwedge_{\delta, \alpha \; as \; above} \vec{r}(\delta, \alpha)) \wedge \vec{r}(\alpha, \gamma)$ does not become $(0, \ldots, 0)$.

Figure 3: robust anaphor resolution on PSFs

version holds for type A occurrences. While binding principle A constructively demands the *existence* of a local binder, it does not preclude further nonlocal coindexings. This prediction is in accordance with the following, intuitively acceptable example:

The barber$_i$ admits that he$_i$ shaves himself$_i$.

During the *candidate filtering* phase of anaphor resolution, compliance with the binding theoretic disjoint reference rules is now verified by computing vectors $\vec{r}(\alpha, \gamma)$. γ is considered a suitable candidate for α only if $\vec{r}(\alpha, \gamma)$ is not completely zero. In the course of the *antecedent selection* phase, the vectors of the individual decisions as well as the vectors pertaining to dynamically resulting transitive coindexings are conjoined to form a vector $\vec{r}^*$ which characterizes the overall reading dependency:

$$\vec{r}^* = \bigwedge_{X \; and \; Y \; coindexed} \vec{r}(X, Y)$$

This gives rise to a further restriction which has to be checked during the decision interdependency test step: to assure the existence of an overall reading, only choices may be combined for which $\vec{r}^*$ does not become the zero vector.

5.3 An Algorithm

Figure 3 shows an anaphor resolution algorithm which employs the above specified method. In step 1, restrictions are applied. By determining vectors $\vec{r}(\alpha, \gamma)$, the binding constraints are verified. In step 2, numerical preference scoring and sorting is performed.[10] Finally, in step 3, antecedent selection takes place. Only decisions are combined which do not interdepend. The reading compatibility is verified by the stepwise computation of vector $\vec{r}^*$.[11]

A theoretical analysis shows that, under the assumption of a clever organization of the computation, the number of bitvector conjunctions and disjunctions (including dominance vector determination) is bounded by $O(bq^2 + s)$, where q is the number of occurrence contributing NPs, s denotes the size of the PSF, and b stands for the maximal degree of branching due to packing and sharing. For natural language grammars, it is justified to assume that b is a (small) constant.[12] The complexity of sorting in step 2 is $O(q^2 \log(q))$. Hence, the overall time complexity of the approach amounts to $O(q^2(n + \log(q)) + s)$. The practical contribution of n, however, is reduced: in a reasonable implementation, the conjunction or disjunction of w bits (w = processor word length)

---

[10]The optimal choice of preference factors and weights remains to be investigated; at least some of the criteria which have been investigated by (Lappin and Leass, 1994) do not immediately generalize to deficient syntax.

[11]In some cases, it may be necessary to retract decisions. Hence, step 3 has to be supplemented with a backtracking facility.

[12]In the general case, however, b may be $\Theta(|V|)$.

will be performed by an elementary operation.

5.4 Structural Disambiguation by Referential Evidence

Since vectors $\vec{r^*}$ describing the reading dependency are available, any set of anaphor resolution choices may now be referred to as further evidence for structural disambiguation. PSF trees which are from now on invalid can be eliminated by pruning all packing variants whose characterizing vectors are orthogonal to $\vec{r^*}$. By this means, based on the above described framework, it becomes possible to realize a parallel processing model which accomplishes the refined version of robustness by interaction.

6 Conclusion

Two approaches to robust anaphor resolution have been presented. The first one, which has been implemented, works on fragments of parses representing subtrees which, due to ambiguity or constraint violation, have not been conjoined. The second one has been formally specified. It exploits structural information as far as possible by taking into account dominance relations that are confined to certain readings. Moreover, by producing an exact description of the reading dependency of its decisions, anaphor resolution according to the latter model yields further evidence for structural disambiguation, thereby rendering possible a higher degree of robust processing by allowing structural and referential disambiguation to interact.

An implementation of the theoretically described approach has to show whether its practical behaviour, which depends on the maximum number of PSF readings n determining the length of the bitvectors, is acceptable. A further generalization is needed for the processing of (truely) fragmentary PSFs in case there is no reading at all. A practically feasible solution might be obtained by *combining the two approaches*: the theoretically complete method is applied to cope with ambiguity that is clearly confined to certain "local" fragments (e.g. PP attachment within clausal fragments), thereby keeping n small. The heuristic approach handles the remaining cases (e.g. unattached clausal fragments). While a systematic investigation and evaluation of the latter issue is pending, the first results for the practical method confirm that, in accordance with psycholinguistic evidence, high-quality anaphor resolution does not hinge on the availability of a unique syntactic description.

References

Jaime G. Carbonell and Ralf D. Brown. 1988. *Anaphora Resolution: A Multi-Strategy Approach*. In: *Proceedings of the 12th International Conference on Computational Linguistics, COLING*, Vol. I, 96-101, Budapest.

Noam Chomsky. 1981. *Lectures on Government and Binding*. Foris Publications, Dordrecht.

Nelson Correa. 1988. *A Binding Rule for Government-binding Parsing*. In: *Proceedings of the 12th International Conference on Computational Linguistics, COLING*, Vol. I, 123-129, Budapest.

Fred Karlsson, Atro Voutilainen, Juha Heikkila, and Arto Antilla. 1995. *Constraint Grammar: A language-independent system for parsing free text*. Mouton de Gruyter, Berlin/New York.

Christopher Kennedy and Branimir Boguraev. 1996. *Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser*. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Vol. I, August 1996, Kopenhagen, 113-118.

Vittorio M. Lampugnani (ed). 1983. *Lexikon der Architektur des 20. Jahrhunderts*. Hatje, Stuttgart.

Shalom Lappin and Herbert J. Leass. 1994. *An Algorithm for Pronominal Anaphora Resolution*. In: *Computational Linguistics*, 20 (4), 535-561.

Wolfgang Menzel. 1995. *Robust Processing of Natural Language*. Working report, Arbeitsbereich Natürlichsprachige Systeme (NaTS), Universität Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Deutschland.

Roland Stuckardt. 1996a. *An Interdependency-Sensitive Approach to Anaphor Resolution*. In: *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium (DAARC96)*, UCREL Technical Papers Series, Vol. 8, Lancaster University, July 1996. 400-413.

Roland Stuckardt. 1996b. *Anaphor Resolution and the Scope of Syntactic Constraints*. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Vol. II, August 1996, Kopenhagen, 932-937.

Masaru Tomita. 1985. *Efficient Parsing for Natural Language*. Kluwer Academic Publishers, Dordrecht, The Netherlands.