Coreference Resolution with Syntactico-Semantic Rules and Corpus Statistics

Sven Hartrumpf
Applied Computer Science VII
University of Hagen
58084 Hagen, Germany
Sven.Hartrumpf@fernuni-hagen.de

Abstract

A new hybrid approach to the coreference resolution problem is presented. The CORUDIS system (COreference RUles with DIsambiguation Statistics) combines syntactico-semantic rules with statistics derived from an annotated corpus. First, the rules and corpus annotations are described and exemplified. Then, the coreference resolution algorithm and the involved statistics are explained. Finally, the proposed method is evaluated against a baseline model and some directions for further research are indicated.

1 Introduction

Coreference resolution is a central problem in natural language understanding since coreference links play an important role for text coherence.[1] In sentence (1) for instance, one wants to know what the German personal pronouns *sie* and *ihr* refer to. Both can refer to *Mädchen* or *Zeitung* because grammatical gender agreement in German can be overruled by natural gender agreement in certain cases.

(1)

[Das	*Mädchen]$_i$*	*liest*	*[die*
The	girl+NEUT	reads	the
Zeitung]$_j$;	*danach*	*geht*	
newspaper+FEM;	afterwards	goes	
sie$_i$	*mit*	*ihr$_j$*	*[ins*
she+FEM	with	her+FEM	in the
Büro]$_k$.			
office.			

'The girl reads the newspaper; afterwards she goes to the office with it.'

The task in this paper is similar to the MUC coreference task (Hirschman and Chinchor, 1997)[2]:

- only **identity coreference** is treated (and not part-whole or other complex semantic relationships);
- only noun phrases (NPs) are considered as **markables** for coreference (and not situations expressed by clauses etc.).

This kind of coreference is an equivalence relation so that coreference resolution comes down to finding the correct partition[3] of markables. If there exists a genuine ambiguity for human readers (and not just a spurious one for computers), several partitions of markables would be the correct answer to the coreference problem. But since such ambiguities are rare the disambiguation method described in this paper always delivers only one partition.

In this paper, the full MUC coreference task is tackled with a new hybrid approach combining syntactico-semantic rules with rule statistics derived from an annotated corpus. Two question might arise. Why not a purely statistical approach: first, because why throw away traditional linguistic knowledge, and second, because statistics on rules reduce the sparse data problem since the applicability of one rule classifies combinations of many relevant features into one feature value. Why not a purely rule-based approach: because it would leave too many alternatives and would not indicate which to choose.

[1] I would like to thank Hermann Helbig, Rainer Osswald, and the anonymous reviewers for their helpful comments and suggestions.

[2] Some problems of this task definition are discussed by van Deemter and Kibble (2000).

[3] A partition of a set S is a set of pairwise disjoint subsets of S (the partition elements) that cover S.

2 Data

Two kinds of data are required for the coreference resolution method described in section 3: hand-crafted rules defining whether two markables can corefer or not and a corpus annotated with coreference information. The rules license possible coreferences; the corpus is used for scoring alternative coreference partitions with estimated probabilities.

2.1 Coreference rules

The coreference rules are designed to license *possible* coreference relations among two markables. Some rules are language-dependent, some are universal; in this paper, the rules (and the corpus) are for German, but the approach suits other languages as well. Each rule consists of a unique name, a premise, and a conclusion. For development and maintenance reasons, a rule is accompanied by a description, some positive example texts, and some negative example texts. A positive example shows that the rule premise is satisfied and the conclusion that the two markables at hand are coreferential would be correct, whereas a negative example shows that the rule premise is not satisfied and the conclusion would indeed be incorrect for the example.

The rule premise is a conjunction of (possibly negated) *constraints*; these can be constituent constraints (**c-constraints**) referring to feature values of one markable and interconstituent constraints (**ic-constraints**) referring to feature values of both markables that are to be tested for coreference. Both types of constraints can be attribute-value equations. The features used in coreference rules are listed in Table 1; the feature values for markables stem from a parser using a semantically oriented lexicon currently containing 14000 German lexemes (HaGenLex). A feature value can be a single type or a disjunction of types. Furthermore, one can construct constraints with predicates. The most important predicates are given in Table 2: they realize concepts from Dependency Grammar (depend/2) and Government and Binding Theory (c-command/2) or define simple relationships between constituents (e. g. compatible-gend-n-gend/2).

The conclusion of a rule expresses a coreference relation with a semantic network (based on the MultiNet formalism defined by Helbig (2001) which has been applied in several other projects, see (Hartrumpf, 1999; Knoll et al., 1998)). For identity coreference, a relation named EQU (equivalence) leading from the anaphor (called c2 in rules)[4] to the antecedent (called c1 in rules) suffices.

Seven rules from the eighteen rules currently used are given in Figure 1. The rule *ident.gend_conflict* would license a link between *das Mädchen* and *sie* in sentence (1). The premise and conclusion can also be viewed as one attribute value matrix employing structure sharing for expressing ic-constraints.

2.2 Annotated corpus

A corpus (a collection of German newspaper articles from the *Süddeutsche Zeitung*) is annotated for coreference according to the guidelines for the MUC coreference task adapted from English to German. The annotations are inserted as SGML tags into the corpus, which is already marked up according to the Corpus Encoding Standard (Ide et al., 1996). The annotation for sentence (1) is given as (2):

(2) ⟨s⟩⟨coref id="125t129"⟩⟨w⟩Das⟨/w⟩ ⟨w⟩Mädchen⟨/w⟩⟨/coref⟩ ⟨w⟩liest⟨/w⟩ ⟨coref id="143t147"⟩⟨w⟩die⟨/w⟩ ⟨w⟩Zeitung⟨/w⟩⟨/coref⟩ ⟨w⟩;⟨/w⟩ ⟨w⟩danach⟨/w⟩ ⟨w⟩geht⟨/w⟩ ⟨coref ref="125t129" type="ident"⟩⟨w⟩sie⟨/w⟩⟨/coref⟩ ⟨w⟩mit⟨/w⟩ ⟨coref ref="143t147" type="ident"⟩⟨w⟩ihr⟨/w⟩⟨/coref⟩ ⟨w⟩ins⟨/w⟩ ⟨w⟩Büro⟨/w⟩ ⟨w⟩.⟨/w⟩⟨/s⟩

3 Coreference resolution

3.1 Algorithm overview

To resolve coreference ambiguities, one must find the partition of markables that corresponds to the correct coreference equivalence relation. The search space is huge since the number of different partitions for n markables is equal to the Bell number $B(n)$. These numbers are also called *Exponential numbers*, see (Bell, 1934); some example values are: $B(1) = 1$, $B(2) = 2$, $B(3) = 5$, $B(4) = 15$, $B(5) = 52$, $B(10) =$

[4]Rules for cataphora are also among the coreference rules. In such rules, c1 corresponds to the cataphor and c2 to the postcedent.

feature name	use*	description
CAT	c	syntactic category (*n* (noun), *perspro* (personal pronoun), *possdet* (possessive determiner), *reflpro* (reflexive pronoun), etc.)
ENTITY	ic	semantic classification comprising the semantic sort (feature SORT) and semantic Boolean features (currently 16, all defined for the MultiNet (multilayered extended semantic network) formalism by Helbig (2001))
ETYPE	ic	extension type (0 (an individual), 1 (a set), 2 (a set of sets), etc.), part of the complex feature LAY containing other extensional and intensional layer features like CARD (cardinality)
GEND	ic	gender (syntactic; masculine, feminine, and neuter in German)
NUM	c, ic	number (syntactic; singular and plural in German)
PERS	c, ic	person (only tested jointly with the other agreement features GEND and NUM)
PROPER	c	proper noun (Boolean feature)
REFER	c, ic	reference type (determinate, indeterminate; based on article choice)
SENTENCE-ID	ic	sentence number in text
SORT	c	semantic sort (45 hierarchically ordered values, 15 of them for nominal concepts)

*c means: feature is used in c-constraints; ic means: feature is used in ic-constraints.

Table 1: Features in coreference rules

predicate name/arity	description
=/2	The values are unifiable.
c-command/2	The first argument (a constituent) c-commands the second.
compatible-gend-n-gend/2	The grammatical gender value at the first argument position is compatible with the natural gender value at the second argument position.
compatible-possdet/2	The first argument (a possessive determiner) can refer to the second argument (a constituent).
copula-related/2	The arguments (two constituents) are related by a copula.
depend/2	The first argument (a constituent) depends on the second.
difference>/3	Numerical difference between two feature values is greater than a third value.
matching-names/2	Two constituents containing (possibly complex) names match.
maximal/1	The argument (a feature value) is maximal, i. e., a leaf node in the type hierarchy.
similar-nouns/2	One argument (a constituent) is a compound suffix of the other argument (a constituent) or both arguments have the same nominal head.

Table 2: Predicates in coreference rules

115975, $B(15) \approx 1.38 \times 10^9$, $B(20) \approx 5.17 \times 10^{13}$, $B(25) \approx 4.64 \times 10^{18}$.

The evaluated algorithm for coreference resolution is implemented as the CORUDIS system (COreference RUles with DIsambiguation Statistics) and works as follows:

1. The markables in a given text are identified. For this task and for gaining the syntactico-semantic feature values to be accessed by rules in step 2, each sentence in the text is parsed independently. If a sentence parse fails, a chunk parse is gener-

id	ident.n_perspro
pre.	(C1 CAT) *n*
	(C2 CAT) *perspro*
	(= (C1 NUM) (C2 NUM))
	(= (C1 PERS) (C2 PERS))
	(= (C1 GEND) (C2 GEND))
	(= (C1 ENTITY) (C2 ENTITY))
	(not (c-command C1 C2))
	(not (c-command C2 C1))
desc.	same gender - anaphoric
exam.	Der Mann liest [das Buch]$_i$. Er versteht [es]$_i$ nicht.

id	ident.perspro_n
pre.	(C1 CAT) *perspro*
	(C2 CAT) *n*
	(= (C1 NUM) (C2 NUM))
	(= (C1 PERS) (C2 PERS))
	(= (C1 GEND) (C2 GEND))
	(= (C1 ENTITY) (C2 ENTITY))
	(not (c-command C1 C2))
	(not (c-command C2 C1))
	(difference> (C1 SENTENCE-ID) (C2 SENTENCE-ID) 0)
desc.	personal pronoun - cataphoric
exam.	[Sie]$_i$ will die Welt ändern; und [die Wissenschaftlerin]$_i$ macht sich frisch ans Werk.

id	ident.perspro_perspro
pre.	(C1 CAT) *perspro*
	(C2 CAT) *perspro*
	(= (C1 NUM) (C2 NUM))
	(= (C1 PERS) (C2 PERS))
	(= (C1 GEND) (C2 GEND))
	(= (C1 ENTITY) (C2 ENTITY))
	(not (c-command C1 C2))
	(not (c-command C2 C1))
desc.	same gender - anaphoric
exam.	[Sie]$_i$ schreiben viel. Und [sie]$_i$ lesen viel.

id	ident.gend_conflict
pre.	(C1 CAT) *n*
	(C2 CAT) *perspro*
	(= (C1 NUM) (C2 NUM))
	(= (C1 PERS) (C2 PERS))
	(not (= (C1 GEND) (C2 GEND)))
	(compatible-gend-n-gend (C2 GEND) (C1 N-GEND))
	(not (c-command C1 C2))
	(not (c-command C2 C1))
desc.	A personal pronoun refers to an NP with a nominal head and conflicting grammatical gender.
exam.	[Das Mädchen]$_i$ lacht. [Sie]$_i$ war stets so.

id	ident.num_conflict
pre.	(C1 CAT) *n*
	(C1 PROPER) *noproper*
	(C2 CAT) *n*
	(C2 PROPER) *noproper*
	(not (= (C1 NUM) (C2 NUM)))
	(= (C1 ETYPE) (C2 ETYPE))
	(= (C1 ENTITY) (C2 ENTITY))
	(not (c-command C1 C2))
	(not (c-command C2 C1))
desc.	different number values (one aggregate and one nonaggregate but equal etype values)
exam.	[Der Vorstand]$_i$ entschied über die Entlassungen. [Diese Männer]$_i$ hatten keine Skrupel.

id	ident.similar_sem
pre.	(C1 CAT) *n*
	(C1 SORT) *co*
	(C2 CAT) *n*
	(C2 REFER) *det*
	(C2 PROPER) *noproper*
	(= (C1 NUM) (C2 NUM))
	(similar-nouns C1 C2)
	(difference> (C1 SENTENCE-ID) (C2 SENTENCE-ID) 0)
desc.	two semantically similar NPs. Cases contained in similar-nouns: compound and base noun; synonyms.
exam.	[der Buchautor]$_i$ ... [der Autor]$_i$
	[die Großstädte]$_i$ ... [die Städte]$_i$
	[Krankenhaus]$_i$ ... [Klinik]$_i$

id	ident.compatible_sem
pre.	(C1 CAT) *n*
	(C1 SORT) *co*
	(C1 PROPER) *noproper*
	(C2 CAT) *n*
	(C2 REFER) *det*
	(C2 PROPER) *noproper*
	(= (C1 NUM) (C2 NUM))
	(= (C1 ETYPE) (C2 ETYPE))
	(= (C1 ENTITY) (C2 ENTITY))
	(not (similar-nouns C1 C2))
	(maximal (C1 ENTITY))
	(maximal (C2 ENTITY))
	(difference> (C1 SENTENCE-ID) (C2 SENTENCE-ID) 0)
desc.	two semantically compatible NPs.
exam.	[die Täter]$_i$ ... [die Männer]$_i$
	[einer holländischen Familie]$_i$ ... [die entführte Deutsche]$_j$

Figure 1: Example coreference rules

ated. (In such cases, constraints in rule premises that involve predicates requiring full parses (e. g. c-command) are ignored in step 2.) For details on the parser, see (Helbig and Hartrumpf, 1997).

2. All possible coreference rule activations that link an anaphor to an antecedent candidate are collected. This is done by testing rule premises on all markable pairs (constituent c1 must precede constituent c2). For two markables, one rule (at most) is activated since the rules have disjoint premises for real text purposes.

3. For each anaphor, one antecedent candidate is selected. This decision is based on rule statistics gained from the annotated training corpus. The sparse data problem is alleviated by backed-off estimation (see for example (Katz, 1987; Collins and Brooks, 1995)).

The algorithm deals with three sets of objects: first, the possible anaphors (all identified markables); second, the candidate antecedents for each possible anaphor (all preceding markables and the artificial **nonreferable markable** explained below); third, the coreference rules. The nonreferable markable is used as the artificial anaphor of a nonreferring markable in order to represent all alternative references for a possible anaphor as a pair. For *first-mentions*, the disambiguation algorithm should select a coreference with the nonreferable markable as antecedent. Currently, one rule licenses the nonreferable markable as antecedent. But it might be useful to apply more finely grained rules and not just one rough licensing rule, as indicated by promising research results for definite descriptions referring to discourse-new entities (see (Vieira and Poesio, 2000)).

3.2 Disambiguating between antecedent candidates

Step 3 of the algorithm given in section 3.1 is the most interesting one and needs some explanation. Leaving the issue of search algorithms aside for a moment, all possible and licensed partitions of identified markables are generated, filtered, and finally scored using estimated probabilities.

The partitions are generated incrementally starting with the first possible anaphor in a singleton partition element. For each antecedent candidate licensed by a coreference rule in step 2, an extended partition with this antecedent in the same partition element as the anaphor in question is introduced. This process is iterated until all possible anaphors have been investigated.

Partitions are filtered out if they violate one of the following **distance and compatibility constraints**:

sentence distance The distance between the anaphor and the antecedent measured in sentences must be below the limit for the linking coreference rule. These limits have been learned from the training corpus.

paragraph distance The distance between the linked markables measured in paragraphs must be below the limit learned for the licensing coreference rule. Typically, pronominal anaphoras can span only two paragraphs, while for example coreferences between named entities can span arbitrary distances.

semantic compatibility All markables in a partition element must bear compatible semantics (unifiable ENTITY and LAY feature values, see Table 1).

Because of the huge search space (see section 3.1), the generation of partitions and the filtering is intertwined in a heuristic search algorithm so that impossible alternatives in the search tree are pruned early. Also the scoring described below is done during the search so that alternatives with low (bad) scores can be delayed and possibly discarded early by the search algorithm.

The score for a partition is constructed as the sum of estimated probabilities for adding the possible anaphor m currently under investigation to one of the antecedent candidates $C = \langle c_1, c_2, \ldots, c_k \rangle$. The candidates are ordered by distance; each c_i is a feature structure representing the parse result from algorithm step 1 for the corresponding markable. Each coreference between m and c_i is licensed by a coreference rule r_i so that this **coreference alternative** can be represented as the triple (m, c_i, r_i).

In order to generalize from the token-based representation (m, c_i, r_i) and to make useful statistics from an annotated corpus, an **abstraction function** a is applied that abstracts from the given anaphor, antecedent candidate, and linking coreference rule to a type-based representation. The abstraction function in equation (3) turned out to be a good compromise between limited sparseness of statistical matrices and distinctiveness for disambiguation purposes: It reduces a coreference alternative (m, c_i, r_i) to the candidate antecedent position i and the licensing coreference rule r_i:

$$a(m, c_i, r_i) \quad := \quad (i, r_i) \tag{3}$$

Let a_i be the **abstracted coreference alternative** $a(m, c_i, r_i)$ and A be the list $\langle a_1, a_2, \ldots, a_k \rangle$ of abstracted coreference alternatives for the possible anaphor m. Then, the probability that a_i corresponds to the *closest* correct antecedent for m is estimated as the relative frequency $rf(i, A)$:

$$rf(i, A) \quad := \quad \frac{f(i, A)}{\sum_{l=1}^{k} f(l, A)} \tag{4}$$

The equation uses the statistical values $f(i, A)$, which count how many times in the annotated training corpus the abstracted coreference alternative a_i wins as the one with the closest correct antecedent in the context of abstracted coreference alternatives A.

Further experiments have shown that looking at more than 5 antecedent candidates does not improve disambiguation results. Therefore, k is reduced to 5 if necessary.

Backed-off estimation can alleviate sparse data problems. The basic idea is that if for a context A no statistical values are known, they are estimated by looking at increasingly smaller parts of A until statistical values are found. One might call such a backed-off estimation **backed-off estimation over alternatives**. Backed-off estimation as defined by equations (5) to (7) is applied in the coreference resolution method when all counts $f(i, A)$ are zero and $f^j(i, A)$ is calculated for $j = 1$.

$$f^0(i, A) := f(i, A) \quad (5)$$

$$f^j(i, A) := \sum_{a_i \in A' \subset A, |A'| = k-j} f^{j-1}(i, A') \quad \text{for} \quad 1 \leq j < k-1 \quad (6)$$

$$rf^j(i, A) := \frac{f^j(i, A)}{\sum_{l=1}^{k} f^j(l, A)} \quad \text{for} \quad 0 \leq j < k-1 \quad (7)$$

The parameter j is increased by one until one of the $f^j(i, A)$ becomes positive (then, the $rf^j(i, A)$ are used as scores for the antecedent candidates) or j reaches $k - 1$ (in this case, all candidates receive equal scores). If the back-off process stops at $j = b$, the relative frequencies $rf^b(i, A)$ are used as estimates for the conditional probabilities $p(i|C)$ that c_i is the closest correct antecedent given antecedent candidates C:

$$p(i|C) \approx rf^b(i, A) \quad (8)$$

One could add other scores to those based on estimated probabilities. In the literature, syntactic parallelism between anaphor and antecedent (based on syntactic case), semantic parallelism (based on semantic roles), and maximality of antecedent NPs are proposed among others. In several experiments, such additional scores have been applied for certain rules (e. g. rules involving pronouns). Small improvements have been achieved, but this topic has not been investigated completely yet.

4 Evaluation

Evaluation results from 12-fold cross-validation for 502 anaphors[5] are listed in Table 3. The standard definitions for recall and precision used in information retrieval are as follows:

$$r := \frac{\#\text{true positives}}{\#\text{true positives} + \#\text{false negatives}} \quad (9)$$

$$p := \frac{\#\text{true positives}}{\#\text{true positives} + \#\text{false positives}} \quad (10)$$

For coreference resolution, true positives are correct coreference links found, false negatives are correct coreference links not reported, and false positives are incorrect coreference links reported. Vilain et al. (1995) illustrate that these definitions sometimes yield counter-intuitive results for coreference evaluations and propose model-theoretic definitions of recall and precision. The values in Table 3 are calculated with these modified definitions.

There are three different evaluation results. The first is the full coreference task. The second one could be called **markable-relative evaluation** since the numbers are calculated only for the markables that have been successfully identified (in some sense, this concentrates on the coreference *relation* aspect of the coreference task). And the final evaluation result comes from a baseline model: “always select the closest antecedent candidate that is licensed by a rule and fulfills the distance and compatibility constraints from section 3.2”.

5 Related Work

There are many recent approaches to this problem, e. g. syntax-based approaches (Lappin and Leass, 1994), cooccurrence-based approaches (Dagan and Itai, 1990), machine-learning approaches (Connolly et al., 1994; Aone and Bennett, 1996; Soon et al., 1999), uncertainty reasoning approaches (Mitkov, 1995; Mitkov, 1997), and robust knowledge-poor approaches (Kennedy and Boguarev, 1996; Baldwin, 1997;

[5] The number of markables that are coreferential with some other markable ranges from 28 to 63 for the folds because the texts in the evaluation corpus were not broken up for cross-validation in order to yield statistical data about *whole* texts. Therefore the training corpus size varied between 439 and 474 and the test corpus between 28 and 63 during the cross-validation.

method	evaluation results in percentage		
	recall	precision	F-measure
coreference (incl. markable identification)	55	82	66
markable-relative coreference evaluation	76	82	79
baseline: always closest candidate	46	42	44

F-measure is calculated with equal weight to recall r and precision p as $\frac{2rp}{r+p}$.

Table 3: Coreference resolution results

Mitkov, 1998b; Mitkov, 1999).[6] The following two systems tackle the MUC coreference task and bear some similarities to CORUDIS.

The system described by Cardie and Wagstaff (1999) resembles the presented system in that it views coreference resolution in a text as partitioning (or clustering). The difference in terms of clustering is that the first system uses greedy clustering while CORUDIS optimizes using global scores. The fundamental difference is that the first system partitions based on a similarity function over markable representations as attribute value pairs, while CORUDIS applies linguistic rules to license possible coreference links and applies corpus statistics to choose one link because typically alternatives exist.

The SWIZZLE system (Harabagiu and Maiorano, 2000) applies heuristics and heuristic ordering by bootstrapping to pick one antecedent per anaphor; in the CORUDIS system, rules license alternatives and one is selected based on a learned statistical model. CORUDIS uses sentence parsing, SWIZZLE as an intentionally knowledge-poor approach only *approximate phrasal parsing*.

6 Conclusion

I have presented a disambiguation method which combines traditional linguistically motivated rules and a backed-off statistical model derived form an annotated corpus in a powerful way. Comparison to other approaches is difficult since evaluation results for German are not available for the MUC coreference task. But the results presented seem to be competitive compared to the 60% F-measure results for English in MUC-7.

Additional filtering conditions, additional scores (preferences), and features from Centering Theory (Grosz et al., 1995) might improve the results reported in this paper significantly. The use of a large lexical-semantic network like GermaNet would solve some problematic coreference cases. More sophisticated evaluations centered around different error types as recommended by Mitkov (1998a) and larger data sets are planned for the future.

[6] The cited works deal only with pronominal anaphors, except the approaches by Aone and Bennett (1996), Baldwin (1997), Connolly et al. (1994), and Soon et al. (1999).

References

Chinatsu Aone and Scott William Bennett. 1996. Applying machine learning to anaphora resolution. In Stefan Wermter, Ellen Riloff, and Gabriele Scheler, editors, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, volume 1040 of *LNAI*, pages 302–314. Springer, Berlin.

Breck Baldwin. 1997. CogNIAC: High precision coreference with limited knowledge and linguistic resources. In *Proceedings of the ACL-EACL'97 workshop on operational factors in practical, robust anaphora resolution for unrestricted texts*, pages 38–45, Madrid, Spain.

Eric T. Bell. 1934. Exponential numbers. *American Mathematical Monthly*, 41:411–419.

Claire Cardie and Kiri Wagstaff. 1999. Noun phrase coreference as clustering. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-99)*, pages 82–89, College Park, Maryland.

Michael Collins and James Brooks. 1995. Prepositional phrase attachment through a backed-off model. In *Proceedings of the 3rd Workshop on Very Large Corpora (WVLC-3)*, pages 27–38, Cambridge, Massachusetts. http://xxx.lanl.gov/abs/cmp-lg/9506021.

Dennis Connolly, John D. Burger, and David S. Day. 1994. A machine learning approach to anaphoric reference. In *Proceedings of the International Conference on New Methods in Language Processing*, pages 255–261, Manchester, England.

Ido Dagan and Alon Itai. 1990. Automatic processsing of large corpora for the resolution of anaphora references. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING 90)*, volume 3, pages 330–332, Helsinki, Finland.

Barbara J. Grosz, Aaravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, June.

Sanda M. Harabagiu and Steven J. Maiorano. 2000. Multilingual coreference resolution. In *Proceedings of the Language Technology Joint Conference on Applied Natural Language Processing and the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL'2000)*, Seattle, Washington.

Sven Hartrumpf. 1999. Hybrid disambiguation of prepositional phrase attachment and interpretation. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-99)*, pages 111–120, College Park, Maryland.

Hermann Helbig and Sven Hartrumpf. 1997. Word class functions for syntactic-semantic analysis. In *Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing (RANLP'97)*, pages 312–317, Tzigov Chark, Bulgaria, September.

Hermann Helbig. 2001. *Die semantische Struktur natürlicher Sprache: Wissensrepräsentation mit MultiNet*. Springer, Berlin.

Lynette Hirschman and Nancy Chinchor. 1997. MUC-7 coreference task definition (version 3.0). In *Proceedings of the 7th Message Understanding Conference (MUC-7)*. http://www.itl.nist.gov/iaui/894.02/related_projects/muc/.

Nancy Ide, Greg Priest-Dorman, and Jean Véronis, 1996. *Corpus Encoding Standard*. http://www.cs.vassar.edu/CES/.

Slava M. Katz. 1987. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3):400–401, March.

Christopher Kennedy and Branimir Boguarev. 1996. Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING 96)*, pages 113–118, Copenhagen, Denmark.

A. Knoll, C. Altenschmidt, J. Biskup, H.-M. Blüthgen, I. Glöckner, S. Hartrumpf, H. Helbig, C. Henning, Y. Karabulut, R. Lüling, B. Monien, T. Noll, and N. Sensen. 1998. An integrated approach to semantic evaluation and content-based retrieval of multimedia documents. In C. Nikolaou and C. Stephanidis, editors, *Proceedings of the 2nd European Conference on Digital Libraries (ECDL'98)*, number 1513 in Lecture Notes in Computer Science, pages 409–428, Berlin. Springer.

Shalom Lappin and Herbert Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–561, December.

Ruslan Mitkov. 1995. An uncertainty reasoning approach for anaphora resolution. In *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLPRS'95)*, pages 149–154, Seoul, Korea.

Ruslan Mitkov. 1997. Two engines are better than one: Generating more power and confidence in the search for the antecedent. In Ruslan Mitkov and Nicolas Nicolov, editors, *Recent Advances in Natural Language Processing: Selected Papers from RANLP'95*, pages 225–234. John Benjamins, Amsterdam.

Ruslan Mitkov. 1998a. Evaluating anaphora resolution approaches. In *Proceedings of the Second Colloquium on Discourse Anaphora and Anaphor Resolution (DAARC 2)*, pages 164–177, Lancaster, England.

Ruslan Mitkov. 1998b. Robust pronoun resolution with limited knowledge. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL'98)*, pages 869–875, Montréal, Canada.

Ruslan Mitkov. 1999. Multilingual anaphora resolution. *Machine Translation*, 14:281–299.

Wee Meng Soon, Hwee Tou Ng, and Chung Yong Lim. 1999. Corpus-based learning for noun-phrase coreference resolution. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-99)*, pages 285–291, College Park, Maryland.

Kees van Deemter and Rodger Kibble. 2000. On coreferring: Coreference in MUC and related annotation schemes. *Computational Linguistics*, 26(4):629–637, December.

Renata Vieira and Massimo Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–593, December.

Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Proceedings of the 6th Message Understanding Conference (MUC-6)*, pages 45–52, San Mateo. Morgan Kaufmann.