# Learning to Disambiguate Relative Pronouns


**Claire Cardie**

Department of Computer Science
University of Massachusetts
Amherst, MA 01003
(cardie@cs.umass.edu)



### Abstract

In this paper we show how a natural language system can learn to find the antecedents of relative pronouns. We use a well-known conceptual clustering system to create a case-based memory that predicts the antecedent of a wh-word given a description of the clause that precedes it. Our automated approach duplicates the performance of hand-coded rules. In addition, it requires only minimal syntactic parsing capabilities and a very general semantic feature set for describing nouns. Human intervention is needed only during the training phase. Thus, it is possible to compile relative pronoun disambiguation heuristics tuned to the syntactic and semantic preferences of a new domain with relative ease. Moreover, we believe that the technique provides a general approach for the automated acquisition of additional disambiguation heuristics for natural language systems, especially for problems that require the assimilation of syntactic and semantic knowledge.


## Introduction

Relative clauses consistently create problems for language processing systems. Consider, for example, the sentence in Figure 1. A correct semantic interpretation should include the fact that "the boy" is the actor of "won" even though

Tony saw **the boy** who won the award.

***Figure 1*** **: Understanding Relative Clauses**

the phrase does not appear in the embedded clause. The interpretation of a relative clause, however, depends on the accurate resolution of two ambiguities, each of which must be performed over a potentially unbounded distance. The system has to 1) find the antecedent of the relative pronoun and 2) determine the antecedent's implicit position in the embedded clause. The work we describe here focuses on (1): locating the antecedent of the relative pronoun.Indeed, although relative pronoun disambiguation seems a simple enough task, there are many factors that make it difficult[1]:

**The head of the antecedent of a relative pronoun does not appear in a consistent position or syntactic constituent.** In both S1 and S2 of Figure 2, for example, the antecedent is "the boy." In S1, however, "the boy" is the direct object of the preceding clause, while in S2 it appears as the subject of the preceding clause. On the other hand, the head of the antecedent is the phrase that immediately precedes "who" in both cases. S3, however, shows that this is not always the case. In fact, the antecedent head may be very distant from its coreferent wh-word[2] (e.g., S4).

**S1.** Tony saw *the boy* who won the award.
**S2.** *The boy* who gave me the book had red hair.
**S3.** Tony ate dinner with *the men* from Detroit who sold □□ computers.
**S4.** I spoke to *the woman* with the black shirt and green hat over in the far corner of the room who wanted a □ second interview.
**S5.** I'd like to thank *Jim, Terry, and Shawn*, who provided the desserts.
**S6.** I'd like to thank *our sponsors, GE and NSF*, who provide financial support.
**S7.** We wondered who stole the watch.
**S8.** We talked with *the woman and the man* who were/was dancing.
**S9.** We talked with *the woman and the man* who danced.
**S10.** *The woman* from Philadelphia who played soccer was my sister.
**S11.** The awards for *the children* who pass the test are in the drawer.

***Figure 2*** **: Relative Pronoun Antecedents**

[1] Locating the gap is a separate, but equally difficult problem because the gap may appear in a variety of positions in the embedded clause: the subject, direct object, indirect object, or object of a preposition. For a simple solution to the gap-finding problem that is consistent with the work presented here, see (Cardie & Lehnert, 1991).

[2] Relative pronouns like *who*, *whom*, *which*, *that*, *where*, etc. are often referred to as wh-words.

**The antecedent is not always a single noun phrase.** In S5, for example, the antecedent of "who" is a conjunction of three phrases and and in S6, either "our sponsors" or its appositive "GE and NSF" is a semantically valid antecedent.

**Sometimes there is no apparent antecedent.** (e.g., S7).

**Disambiguation of the relative pronoun may depend on information in the embedded clause.** In S8, for example, the antecedent of "who" is either "the man" or "the woman and the man," depending on the number of the embedded clause verb.

**Sometimes, the antecedent is truly ambiguous.** For sentences like S9, the real antecedent depends on the surrounding context.

**Locating the antecedent requires the assimilation of both syntactic and semantic knowledge.** The syntactic structure of the clause preceding "who" in sentences S10 and S11, for example, is identical (NP-PP). The antecedent in each case is different, however. In S10, the antecedent is the subject, "the woman;" in S11, the antecedent is the prepositional phrase modifier, "the children."

In this paper we show how a natural language system can learn to disambiguate relative pronouns. We describe the use of an existing conceptual clustering system to create a case-based memory that predicts the antecedent of a wh-word given a description of the clause that precedes it. In addition,

- Our approach duplicates the performance of hand-coded rules.
- It assumes only minimal syntactic parsing capabilities and the existence of a very general semantic feature set for describing nouns.
- The technique requires human intervention only to choose the correct antecedent for the training instances.
- The resulting memory is automatically tuned to respond to the syntactic and semantic preferences of a particular domain.
- Acquiring relative pronoun disambiguation heuristics for a new domain requires little effort.

Furthermore, we believe that the technique may provide a general approach for the automated acquisition of additional disambiguation heuristics, especially for problems that require the assimilation of syntactic and semantic knowledge. In the next section, we describe current approaches to relative pronoun disambiguation. The remainder of the paper presents our automated approach and compares its performance to that of hand-coded rules.

## Finding Wh-Phrase Antecedents: Previous Approaches

Any natural language processing system that hopes to process real world texts requires a reliable mechanism for locating the antecedents of relative pronouns. Systems that use a formal syntactic grammar often directly encode information for relative pronoun disambiguation in the grammar. Other times, the grammar proposes a set of syntactically legal antecedents and this set is passed to a semantic component which determines the antecedent using inference or preference rules. (For examples of this approach, see (Correa, 1988; Hobbs, 1986; Ingria, & Stallard, 1989; Lappin, & McCord, 1990).) Alternatively, semantically-driven systems often employ disambiguation heuristics that rely for the most part on semantic knowledge but also include syntactic constraints (e.g., UMass/MUC-3 (Lehnert et. al., 1991)). Each approach, however, requires the manual encoding of relative pronoun disambiguation rules, either 1) as part of a formal grammar that must be designed to find relative pronoun antecedents for all possible syntactic contexts, or 2) as heuristics that include both syntactic and semantic constraints. Not surprisingly, NLP system builders have found the hand-generation of such rule sets to be both time consuming and error prone.

Furthermore, the resulting rule set is often fragile and generalizes poorly to new contexts. For example, the UMass/MUC-3[3] system began with 19 rules for finding the antecedents of relative pronouns. These rules were based on approximately 50 instances of relative pronouns that occurred in 10 newswire stories from the MUC-3 corpus. As counterexamples were identified, new rules were added (approximately 10) and existing rules changed. Over time, however, we became increasingly reluctant to modify the rule set because global effects of local rule changes were difficult to measure. In addition, most rules were based on a class of sentences that the UMass/MUC-3 system had found to contain important information. As a result, the hand-coded rules tended to work well for relative pronoun disambiguation in sentences of this class (93% correct for one set of 50 texts), but did not generalize to sentences outside of this class (78% correct for the same set of 50 texts).

In the next section, we present an automated approach for learning the antecedents of wh-words. This approach avoids the problems associated with the manual encoding of heuristics and grammars, and automatically tailors the disambiguation decisions to the syntactic and semantic profile of the domain.

## A Case-Based Approach

Our method for relative pronoun disambiguation consists of three steps:

1. Create a hierarchical memory of relative pronoun disambiguation cases.

---

[3]MUC-3 is the Third Message Understanding System Evaluation and Message Understanding Conference (Sundheim,1991). UMass/MUC-3 (Lehnert et. al., 1991) is a version of the CIRCUS parser (Lehnert, 1990) developed for the MUC-3 performance evaluation.

2. For each new occurrence of a wh-word, retrieve the most similar case from memory using a representation of the clause preceding the wh-word as the probe.

3. Use the antecedent of the retrieved case to guide selection of the antecedent for the probe.

These steps will be discussed in more detail in the sections that follow.

## Building the Case Base

We create a hierarchical memory of relative pronoun disambiguation cases using a conceptual clustering system called COBWEB (Fisher, 1987).[4] COBWEB takes as input a set of training instances described as a list of attribute-value pairs and incrementally discovers a classification hierarchy that covers the instances. The system tracks the distribution of individual values and favors the creation of classes that increase the number of values that can be correctly predicted given knowledge of class membership. While the details of COBWEB are not necessary for understanding the rest of the paper, it is useful to know that nodes in the hierarchy represent probabilistic concepts that increase in generality as they approach the root of the tree. Given a new instance to classify after training, COBWEB retrieves the most specific concept that adequately describes the instance.

We provide COBWEB with a set of training cases, each of which represents the disambiguation decision for a single occurrence of a relative pronoun.[5] We describe the decision in terms of four classes of attribute-value pairs: one or more *constituent* attribute-value pairs, a *cd-form* attribute-value pair, a *part-of-speech* attribute-value pair, and an *antecedent* attribute-value pair. The paragraphs below briefly describe each attribute type using the cases in Figure 3 as examples.

Because the antecedent of a relative pronoun usually appears in the clause just preceding the relative pronoun, we include a *constituent* attribute-value pair for each phrase in that clause.[6] The attribute describes the syntactic class of the phrase as well as its position with respect to the relative pronoun. In S1, for example, "of the 76th court" is represented with the attribute *pp1* because it is a prepositional phrase and is in the first position to the left of "who." In S2, "Rogoberto Matute" and "Juan Bautista" are represented with the attributes *np1* and *np2*, respectively, because they are noun phrases one and two positions to the left of "who." The subject and verb

**S1:** [The judge] [of the 76th court] [,] who ...

**constituents:** *( s human ) ( pp1 physical-target ) ( v nil )*
**cd-form:** *( cd-form physical-target )*
**part-of-speech:** *( part-of-speech comma )*
**antecedent:** *( antecedent ((s)) )*

**S2:** [The police] [detained] [Juan Bautista] [and]

**constituents:** *( s human ) ( v t ) ( np2 proper-name )*
**cd-form:** *( cd-form proper-name )*
**part-of-speech:** *( part-of-speech comma )*
**antecedent:** *( antecedent ((np2 cd-form)) )*

**S3:** [Her DAS bodyguard] [,] [Dagoberto Rodriquez] [,] who...

**constituents:** *( s human ) ( np1 proper-name ) ( v nil )*
**cd-form:** *( cd-form proper-name )*
**part-of-speech:** *( part-of-speech comma )*
**antecedent:** *( antecedent ((cd-form) (s cd-form) (s)) )*

***Figure 3*: Relative Pronoun Disambiguation Cases**

constituents (e.g., "the judge" in S1 and "detained" in S2) retain their traditional *s* and *v* labels, however — no positional information is included for those attributes.

The value associated with a constituent attribute is simply the semantic classification that best describes the phrase's head noun. We use a set of seven general semantic features to categorize each noun in the lexicon: human, proper-name, location, entity, physical-target, organization, and weapon.[7] The subject constituents in Figure 3 receive the value *human*, for example, because their head nouns are marked with the "human" semantic feature in the lexicon. For verbs, we currently note only the presence or absence of a verb using the values *t* and *nil*, respectively.

In addition, every case includes a *cd-form*, *part-of-speech*, and *antecedent* attribute-value pair. The *cd-form* attribute-value pair represents the phrase last recognized by the syntactic analyzer.[8] Its value is the semantic feature associated with that constituent. In S1, for example, the constituent recognized just before "who" was "of the 76th court." Therefore, this phrase becomes part of the case representation as the attribute-value pair *(cd-form physical-target)* in addition to its participation as the constituent attribute-value pair *(pp1 physical-target)*. The *part-of-speech* attribute-value pair specifies the part of speech or item of punctuation that is active when the relative

[4] The experiments described here were run using a version of COBWEB developed by Robert Williams, now at Pace University.

[5] Our initial efforts have concentrated on the resolution of "who" because the hand-coded heuristics for that relative pronoun were the most complicated and difficult to modify. We expect that the resolution of other wh-words will fall out without difficulty.

[6] Note that we are using the term *constituent* rather loosely to refer to all noun phrases, prepositional phrases, and verb phrases **prior to any attachment decisions**.

[7] These features are specific to the MUC-3 domain.

[8] The name *cd-form* is an artifact of the CIRCUS parser.

pronoun is recognized. For all examples in Figure 3, the part of speech was *comma*.[9]

Finally, the position of the correct antecedent is included in the case representation as the *antecedent* attribute-value pair.[10] For UMass/MUC-3, the antecedent of a wh-word is the head of the relative pronoun coreferent – without any modifying prepositional phrases. Therefore, the value of this attribute is a list of the constituent and/or cd-form attributes that represent the location of the antecedent head or *(none)* if no antecedent can be found. In S1, for example, the antecedent of "who" is "the judge." Because this phrase occurs in the subject position, the value of the antecedent attribute is *(s)*.

Sometimes, however, the antecedent is actually a conjunction of constituents. In these cases, we represent the antecedent as a list of the constituent attributes associated with each element of the conjunction. Look, for example, at sentence S2. Because "who" refers to the conjunction "Juan Bautista and Rogoberto Matute," the antecedent can be described as *(np2 cd-form)* or *(np2 np1)*. Although the lists represent equivalent surface forms, we choose the more general *(np2 cd-form)*.[11] S3 shows yet another variation of the antecedent attribute-value pair. In this example, an appositive creates three semantically equivalent antecedent values, all of which become part of the antecedent feature: 1) "Dagoberto Rodriguez" — *(cd-form)* , 2) "her DAS bodyguard"— *(s)*, and 3) "her DAS bodyguard, Dagoberto Rodriguez"— *(s cd-form)*.

The instance representation described above was based on a desire to use all relevant information provided by the CIRCUS parser as well as a desire to exploit the cognitive biases that affect human information processing (e.g., the tendency to rely on the most recent information). Although space limitations prevent further discussion of these representational issues here, they are discussed at length in (Cardie, 1992b).

It should be noted that UMass/MUC-3 automatically creates the training instances as a side effect of syntactic analysis. Only specification of the antecedent attribute-value pair requires human intervention via a menu-driven interface that displays the antecedent options. In addition, the parser need only recognize low-level constituents like noun phrases, prepositional phrases, and verb phrases because the case-based memory, not the syntactic analyzer, directly handles the conjunctions and appositives that comprise a relative pronoun antecedent. For a more detailed discussion of parser vs. learning component tradeoffs, see (Cardie, 1992a).

## Retrieval and Case Adaptation

As the training instances become available, they are passed to the clustering system which builds a case base of relative pronoun disambiguation decisions. After training, we use the resulting hierarchy to predict the antecedent of a wh-word in new contexts. For each novel sentence containing a wh-word, UMass/MUC-3 creates a probe case that represents the clause preceding the wh-word. The probe contains constituent, cd-form, and part-of-speech attribute-value pairs, but no antecedent feature. Given the probe, COBWEB retrieves the individual instance or abstract class in the tree that is most similar and the antecedent of the retrieved case guides selection of the antecedent for the novel case. For novel sentence S1 of Figure 4, for example, the retrieved case specifies *np2* as the location of the antecedent. Therefore, UMass/MUC-3 chooses the contents of the *np2* constituent — "the hardliners" — as the antecedent in S1.

Sometimes, however, the retrieved case lists more than one option as the antecedent. In these cases, we rely on the following case adaptation algorithm to choose an antecedent:

1. Choose the first option whose constituents are all present in the probe case.
2. Otherwise, choose the first option that contains at least one constituent present in the probe and ignore those constituents in the retrieved antecedent that are missing from the probe.
3. Otherwise, replace the *np* constituents in the

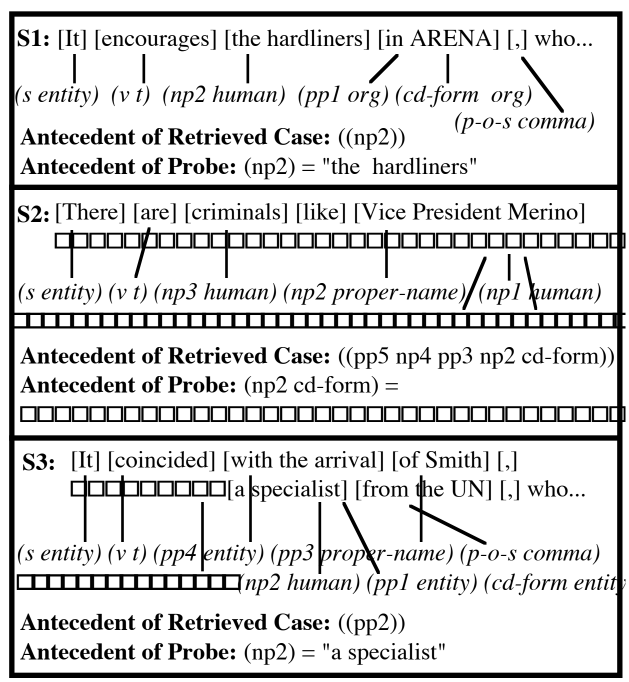


***Figure 4*: Case Retrieval and Adaptation**

[9]Again, the fact that the *part-of-speech* attribute may refer to an item of punctuation is just an artifact of the UMass/MUC-3 system.

[10]COBWEB is usually used to perform unsupervised learning. However, we use COBWEB for supervised learning (instead of a decision tree algorithm, for example) because we expect to employ the resulting case memory for predictive tasks other than relative pronoun disambiguation.

[11]This form is more general because it represents both *(np2 np1)* and *(np2 pp1)*.

| Exp # | Training Sets (# of instances) | Test Set (# of instances) | Automated Approach | Adjusted Evaluation | Hand-Coded Rules | Baseline Strategy |
|---|---|---|---|---|---|---|
| 1 | set1 + set2 (170) | set3 (71) | 89% | 93% | 87% | 86% |
| 2 | set2 + set3 (159) | set1 (82) | 83% | 86% | 80% | 72% |
| 3 | set1 + set3 (153) | set2 (88) | 74% | 81% | 75% | 67% |

***Figure 5*** **: Results (% correct)**

retrieved antecedent that are missing from the probe with *pp* constituents (and vice versa) and try 1 and 2 again.

In general, the case adaptation algorithm tries to choose an antecedent that is consistent with the context of the probe or to generalize the retrieved antecedent so that it is applicable in the current context. S1 illustrates the first case adaptation filter. In S2, however, the retrieved case specifies an antecedent from five constituents, only two of which are actually represented in the probe. Therefore, we ignore the missing constituents *pp5*, *np4*, and *pp3* and look to just *np2* and *cd-form* for the antecedent. For S3, filters 1 and 2 fail because the probe contains no *pp2* constituent. However, if we replace *pp2* with *np2* in the retrieved antecedent, then filter 1 applies and “a specialist” is chosen as the antecedent. Note that, in this case, the case adaptation algorithm returns an antecedent that is just one of three valid antecedents (i.e., “Smith,” “a specialist,” and “Smith, a specialist”).

## Experiments and Results

To evaluate our automated approach, we extracted all sentences containing “who” from 3 sets of 50 texts in the MUC-3 corpus. In each of 3 experiments, 2 sets were used for training and the third reserved for testing. The results are shown in Figure 5. For each experiment, we compare our automated approach with the hand-coded heuristics of the UMass/MUC-3 system and a baseline strategy that simply chooses the most recent phrase as the antecedent. For the “adjusted” results, we discount errors in the automated approach that involve antecedent combinations never seen in any of the training cases. In these situations, the semantic and syntactic structure of the novel clause usually differs significantly from those in the hierarchy and we cannot expect accurate retrieval from the case base. In experiment 1, for example, 3 out of 8 errors fall into this category.

Based on these initial results, we conclude that our automated approach to relative pronoun disambiguation clearly surpasses the “most recent phrase” baseline heuristic and at least duplicates the performance of hand-coded rules. Furthermore, the kind of errors exhibited by the learned heuristics seem reasonable. In experiment 1, for example, of the 5 errors that did not specify previously unseen antecedents, 1 error involved a new syntactic context for “who”— “who” preceded by a preposition, i.e., “regardless of who.” The remaining 4 errors cited relative pronoun antecedents that are difficult even for people to specify. (In each case, the antecedent chosen by the automated approach is indicated in italics; the correct antecedent is shown in boldface type.)

1. “... 9 rebels died at the hands of **members** of *the civilian militia*, who resisted the attacks”
2. “... the government expelled **a group** of *foreign drug traffickers* who had established themselves in northern Chile”
3. “*...the number* of **people** who died in Bolivia...”
4. “...**the rest** of *the contra prisoners*, who are not on this list...”

## Conclusions

Developing state-of-the-art NLP systems or extending existing ones for new domains tends to be a long, labor-intensive project. Both the derivation of knowledge-based heuristics and the (re)design of the grammar to handle numerous classes of ambiguities consumes much of the development cycle. Recent work in statistically-based acquisition of syntactic and semantic knowledge (e.g., (Brent, 1991; Church, et. al., 1991; de Marcken, 1990; Hindle, 1990; Hindle, & Rooth, 1991; Magerman & Marcus, 1990)) attempts to ease this knowledge engineering bottleneck. However, statistically-based methods require very large corpora of on-line text.

In this paper, we present an approach for the automated acquisition of relative pronoun disambiguation heuristics that duplicates the performance of hand-coded rules, requires minimal syntactic parsing capabilities, and is unique in its reliance on relatively few training examples. We require a small training set because, unlike purely statistical methods, the training examples are not word-based, but are derived from higher level parser output. In addition, we save the entire training case so that it is available for generalization when the novel probe retrieves a poor match.

In spite of these features of the approach, the need for a small training set *may*, in fact, be problem-dependent.

Future work will address this issue by employing our case-based approach for a variety of language acquisition tasks. Further research on automating the selection of training instances, extending the approach for use with texts that span multiple domains, and deriving optimal case adaptation filters is also clearly needed. However, the success of the approach in our initial experiments, especially for finding antecedents that contain complex combinations of conjunctions and appositives, suggests that the technique may provide a general approach for the automated acquisition of additional disambiguation heuristics, particularly for traditionally difficult problems that require the assimilation of syntactic and semantic knowledge.

## Acknowledgments

I thank Wendy Lehnert for many helpful discussions and for comments on earlier drafts. This research was supported by the Office of Naval Research, under a University Research Initiative Grant, Contract No. N00014-86-K-0764, and Contract No. N00014-92-J-1427, NSF Presidential Young Investigators Award NSFIST-8351863, and the Advanced Research Projects Agency of the Department of Defense monitored by the Air Force Office of Scientific Research under Contract No. F49620-88-C-0058.

## References

Brent, M. (1991). Automatic acquisition of subcategorization frames from untagged text. *Proceedings, 29th Annual Meeting of the Association for Computational Linguistics*. University of California, Berkeley. Association for Computational Linguistics.

Cardie, C. (1992a). Corpus-Based Acquisition of Relative Pronoun Disambiguation Heuristics. To appear in *Proceedings, 30th Annual Meeting of the Association for Computational Linguistics*. University of Delaware, Newark. Association for Computational Linguistics.

Cardie, C. (1992b). Exploiting Cognitive Biases in Instance Representations. Submitted to *The Fourteenth Annual Conference of the Cognitive Science Society*. University of Indiana, Bloomington.

Cardie, C., & Lehnert, W. (1991). A Cognitively Plausible Approach to Understanding Complex Syntax. *Proceedings, Eighth National Conference on Artificial Intelligence*. Anaheim, CA. AAAI Press / The MIT Press.

Church, K., Gale, W., Hanks, P., & Hindle, D. (1991). Using Statistics in Lexical Analysis. In U. Zernik (Ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon* (pp. 115-164). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Correa, N. (1988). A Binding Rule for Government-Binding Parsing. *Proceedings, COLING '88*. Budapest.

de Marcken, C. G. (1990). Parsing the LOB Corpus. *Proceedings, 28th Annual Meeting of the Association for Computational Linguistics*. University of Pittsburgh. Association for Computational Linguistics.

Fisher, D. H. (1987). Knowledge Acquisition Via Incremental Conceptual Clustering. *Machine Learning, 2,* p. 139-172.

Hindle, D. (1990). Noun classification from predicate-argument structures. *Proceedings, 28th Annual Meeting of the Association for Computational Linguistics*. University of Pittsburgh.Association for Computational Linguistics.

Hindle, D., & Rooth, M. (1991). Structural ambiguity and lexical relations. *Proceedings, 29th Annual Meeting of the Association for Computational Linguistics*. University of California, Berkeley. Association for Computational Linguistics.

Hobbs, J. (1986). Resolving Pronoun References. In B. J. Grosz, K. Sparck Jones, & B. L. Webber (Eds.), *Readings in Natural Language Processing* (pp. 339-352). Los Altos, CA: Morgan Kaufmann Publishers, Inc.

Ingria, R., & Stallard, D. (1989). A Computational Mechanism for Pronominal Reference. *Proceedings, 27th Annual Meeting of the Association for Computational Linguistics*. Vancouver. Association for Computational Linguistics.

Lappin, S., & McCord, M. (1990). A Syntactic Filter on Pronominal Anaphora for Slot Grammar. *Proceedings, 28th Annual Meeting of the Association for Computational Linguistics*. University of Pittsburgh.Association for Computational Linguistics.

Lehnert, W. (1990). Symbolic/Subsymbolic Sentence Analysis: Exploiting the Best of Two Worlds. In J. Barnden, & J. Pollack (Eds.), *Advances in Connectionist and Neural Computation Theory*. Norwood, NJ: Ablex Publishers.

Lehnert, W., Cardie, C., Fisher, D., Riloff, E., & Williams, R. (1991). University of Massachusetts: Description of the CIRCUS System as Used for MUC-3. *Proceedings, Third Message Understanding Conference (MUC-3)*. San Diego, CA.Morgan Kaufmann Publishers.

Magerman, D. M., & Marcus, M., P. (1990). Parsing a Natural Language Using Mutual Information Statistics. *Proceedings, Eighth National Conference on Artificial Intelligence*. Boston.AAAI Press/The MIT Press.

Sundheim, B. M. (May,1991).Overview of the Third Message Understanding Evaluation and Conference. *Proceedings, Third Message Understanding Conference (MUC-3)*. San Diego, CA.Morgan Kaufmann Publishers.