

CogNIAC : High Precision Coreference with Limited Knowledge and Linguistic Resources

Breck Baldwin

3401 Walnut St.
IRCS Suite 400-a
Philadelphia, PA, USA 19104
breck@linc.cis.upenn.edu

Abstract

This paper presents a high precision pronoun resolution system that is capable of greater than 90% precision with 60% and better recall for some pronouns. It is suggested that the system is resolving a sub-set of anaphors that do not require general world knowledge or sophisticated linguistic processing for successful resolution. The system does this by being very sensitive to ambiguity, and only resolving pronouns when very high confidence rules have been satisfied. The system is capable of 'noticing' ambiguity because it requires that there be a unique antecedent within a salience ranking, and the salience rankings are not total orders, i.e. two or more antecedents can be equally salient. Given the nature of the systems rules, it is very likely that they are largely domain independent and that they reflect processing strategies used by humans for general language comprehension. The system has been evaluated in two distinct experiments which support the overall validity of the approach.

1 Introduction:

Pronoun resolution is one of the 'classic' computational linguistics problems. It is also widely considered to be inherently an 'A.I. complete' task--meaning that resolution of pronouns requires full world knowledge and inference. CogNIAC is a pronoun resolution engine designed around the assumption that there is a sub-class of anaphora that does not require general purpose reasoning. The kinds of information CogNIAC does require includes: sentence detection, part-of-speech tagging, simple noun phrase recognition, basic semantic category information like, gender, number, and in one configuration, partial parse trees.

What distinguishes CogNIAC from algorithms that use similar sorts of information is that it will not resolve a pronoun in circumstances of ambiguity. Crucially, ambiguity is a function of how much knowledge an understander has. Since CogNIAC does not have as rich a representation of world knowledge as

humans, it finds much more ambiguity in texts than humans do.

2 A path to high precision pronominal resolution-- avoid guesswork in ambiguous contexts:

It is probably safe to say that few referring pronouns are conveyed without the speaker/writer having an antecedent in mind. Ambiguity occurs when the perceiver cannot recover from the context what conveyer has in mind. I have found myself uttering pronouns which the hearer has no chance of recovering the antecedent to because they are not attending to the same part of the external environment, "He sure looks familiar", or in text I am so focused on the context of what I am writing that use a pronoun to refer to a highly salient concept for me, but the antecedent may completely evade a reader without my familiarity with the topic. Of course it is possible to explicitly leave the reader hanging as in, "Earl and Dave were working together when suddenly he fell into the threshing machine."

Humans, unlike most coreference algorithms, notice such cases of ambiguity and can then ask for clarification or at least grumble about how we cannot climb into the writers head to figure out what they meant. But in that grumble we have articulated the essence of the problem--we don't have sufficient knowledge to satisfy ourselves that an antecedent has been found.

Pronoun resolution systems have extremely limited knowledge sources, they cannot access a fraction of human common sense knowledge. To appreciate this consider the following text with grammatical tags replacing words with pronouns and names left in place:

The city council VERBGROUP the women NP CC they VB NN

Mariana VBD PP Sarah TO VB herself PP DT AJD NN

Without lexical knowledge a human attempting to resolve the pronouns is in much the knowledge impoverished position of the typical coreference

algorithm. It is no surprise that texts with so little information provided in them tend to be more ambiguous than the texts in fleshed out form. The conclusion to draw from this example is that the limiting factor in CogNIAC is knowledge sources, not an artificial restriction on domains or kinds of coreference. This point will be resumed in the discussion section when what the consequences of fuller knowledge sources would be on CogNIAC.

2.1 Using limited world knowledge to find possible antecedents:

For noun phrase anaphora, gathering semantically possible antecedents amounts to running all the noun phrases in a text through various databases for number and gender, and perhaps then a classifier that determines whether a noun phrase is a company, person or place¹. This set of candidate antecedents rarely has more than 5 members when some reasonable locality constraints are adhered to, and this set almost always contains the actual antecedent. The remainder of the coreference resolution process amounts to picking the right entity from this set.

For the kinds of data considered here (narratives and newspaper articles) there is a rarely a need for general world knowledge in assembling the initial set of possible antecedents for pronouns. This does not address the issue of inferred antecedents, event reference, discourse deixis and many other sorts of referring phenomenon which clearly require the use of world knowledge but are beyond the scope of this work. As it happens, recognizing the possible antecedents of these pronouns is within the capabilities of current knowledge sources.

Better knowledge sources could be used to reduce the space of possible antecedents. For example the well known [Winograd 1972] alternation:

The city council refused to give the women a permit because they {feared/advocated} violence.

There are two semantically possible antecedents to *they*: *The city council*, and *the women*. The problem is picking the correct one. Dependent on verb choice, they strongly prefers one antecedent to the other. Capturing this generalization requires a sophisticated theory of verb meaning as relates to pronoun resolution. Speaking anecdotally, these kinds of resolutions happen quite often in text. CogNIAC recognizes knowledge intensive coreference and does not attempt to resolve such instances.

2.2 Using limited linguistic resources to find coreference:

¹ The named entity task at MUC-6 used a similar classification task and the best system performance was 96% precision/97% recall.

Fortunately not all instances of pronominal anaphora require world knowledge for successful resolution. In lieu of full world knowledge, CogNIAC uses regularities of English usage in an attempt to mimic strategies used by humans when resolving pronouns. For example, the syntax of a sentence highly constrains a reflexive pronoun's antecedent. Also if there is just one possible antecedent in entire the prior discourse, then that entity is nearly always the correct antecedent. CogNIAC consists of a set of such observations implemented in Perl.

CogNIAC has been used with a range of linguistic resources, ranging from scenarios where almost no linguistic processing of the text is done at all to partial parse trees being provided. At the very least, there must be sufficient linguistic resources to recognize pronouns in the text and the space of candidate antecedents must be identified. For the first experiment the text has been part of speech tagged and basal noun phrases have been identified with '[']' (i.e. noun phrases that have no nested noun phrases) as shown below:

```
[ Mariana/NNP ] motioned/VBD for/IN [
Sarah/NNP ] to/TO seat/VB [ herself/PRP ]
on/IN [ a/DT two-seater/NN lounge/NN ]
```

In addition, finite clauses were identified (by hand for experiment 1) and various regular expressions are used to identify subjects, objects and what verbs take as arguments for the purposes of coreference restrictions. With this level of linguistic annotation, nearly all the parts of CogNIAC can be used to resolve pronouns.

The core rules of CogNIAC are given below, with their performance on training data provided (200 pronouns of narrative text). In addition, examples where the rules successfully apply have been provided for most of the rules with the relevant anaphors and antecedents in boldface. The term 'possible antecedents' refers to the set of entities from the discourse that are compatible with an anaphor's gender, number and coreference restrictions (i.e. non-reflexive pronouns cannot corefer with the other arguments of its verb/preposition etc.).

- 1) **Unique in Discourse**: If there is a single possible antecedent *i* in the read-in portion of the entire discourse, then pick *i* as the antecedent: 8 correct, and 0 incorrect.
- 2) **Reflexive**: Pick nearest possible antecedent in read-in portion of current sentence if the anaphor is a reflexive pronoun: 16 correct, and 1 incorrect.

Mariana motioned for **Sarah** to seat **herself** on a two-seater lounge.

- 3) **Unique in Current + Prior**: If there is a single possible antecedent *i* in the prior sentence and the read-in portion of the current

sentence, then pick *i* as the antecedent: 114 correct, and 2 incorrect.

Rupert Murdoch's News Corp. confirmed his interest in buying back the ailing New York Post. But analysts said that if **he** winds up bidding for the paper,....

4) **Possessive Pro:** If the anaphor is a possessive pronoun and there is a single exact string match *i* of the possessive in the prior sentence, then pick *i* as the antecedent: 4 correct, and 1 incorrect.

After he was dry, Joe carefully laid out the damp towel in front of **his locker**. Travis went over to **his locker**, took out a towel and started to dry off.

5) **Unique Current Sentence:** If there is a single possible antecedent in the read-in portion of the current sentence, then pick *i* as the antecedent: 21 correct, and 1 incorrect.

Like a large bear, he sat motionlessly in the lounge in one of the faded armchairs, watching Constantin. After a week **Constantin** tired of reading the old novels in the bottom shelf of the bookcase--somewhere among the gray well thumbed pages he had hoped to find a message from one of **his** predecessors.....

6) **Unique Subject/ Subject Pronoun:** If the subject of the prior sentence contains a single possible antecedent *i*, and the anaphor is the subject of the current sentence, then pick *i* as the antecedent: 11 correct, and 0 incorrect.

Besides, if **he** provoked Malek, uncertainties were introduced, of which there were already far too many. **He** noticed the supervisor enter the lounge ...

The method of resolving pronouns within CogNIAC works as follows: Pronouns are resolved left-to-right in the text. For each pronoun, the rules are applied in the presented order. For a given rule, if an antecedent is found, then the appropriate annotations are made to the text and no more rules are tried for that pronoun, otherwise the next rule is tried. If no rules resolve the pronoun, then it is left unresolved.

These rules are individually are high precision rules, and collectively they add up to reasonable recall. The precision is 97% (121/125) and the recall is 60% (121/201) for 198 pronouns of training data.

3 Evaluation:

3.1 Comparison to Hobbs' Naive Algorithm:

The Naive Algorithm [Hobbs 1976] works by specifying a total order on noun phrases in the prior discourse and comparing each noun phrase against the selectional restrictions (i.e. gender, number) of the anaphor, and taking the antecedent to be the first one to satisfy them. The specification of the ordering constitutes a traversal order of the syntax tree of the anaphors clause and from there to embedding clauses and prior clauses.

The Winograd sentences, with either verb, would yield the following ordering of possible antecedents:

The city council > the women

The algorithm would resolve *they* to *The city council*. This is incorrect on one choice of verb, but the algorithm does not integrate the verb information into the salience ranking.

In comparison, none of the six rules of CogNIAC would resolve the pronoun. Rules have been tried that resolved a subject pronoun of a nested clause with the subject of the dominating clause, but no configuration has been found that yielded sufficient precision². Consequently, *they* is not resolved.

The naive algorithm has some interesting properties. First it models relative salience as relative depth in a search space. For two candidate antecedents *a* and *b*, if *a* is encountered before *b* in the search space, then *a* is more salient than *b*. Second, the relative saliency of all candidate antecedents is totally ordered, that is, for any two candidate antecedents *a* and *b*, *a* is more salient than *b* xor *b* is more salient than *a*.

² In experiment 2, discussed below, the rule 'subject same clause' would resolve *they* to *the city council*, but it was added to the MUC-6 system without testing, and has shown itself to not be a high precision rule.

CogNIAC shares several features of the Naive Algorithm:

- Both use basic selectional restrictions to find semantically acceptable potential antecedents.
- Both use highly syntactic generalizations to resolve anaphors, and do not attempt to do more sophisticated semantic processing.

But they also differ in significant ways:

- CogNIAC is not committed to totally ordering all potential antecedents, the Naive Algorithm is.
- CogNIAC is sensitive to ambiguity, i.e. circumstances of many possible antecedents, and will not resolve pronouns in such cases. The Naive Algorithm has no means of noting ambiguity and will resolve a pronoun as long as there is at least one possible antecedent.

Perhaps the most convincing reason to endorse partially ordered salience rankings is that salience distinctions fade as the discourse moves on.

Earl was working with Ted the other day. He fell into the threshing machine.

Earl was working with Ted the other day. All of the sudden, the cows started making a ruckus. The noise was unbelievable. He fell into the threshing machine.

In the first example 'He' takes 'Earl' as antecedent, which is what rule 6, Unique Subject/Subject Pronoun, would resolve the pronoun to. However in the second example, the use of 'He' is ambiguous--a distinction that existed before is now gone. The Naive Algorithm would still maintain a salience distinction between 'Earl' and 'Ted', where CogNIAC has no rule that makes a salience distinction between subject and object of a sentence which has two intervening sentences. The closest rule would be Unique in Discourse, rule 1, which does not yield a unique antecedent.

3.2 Performance:

CogNIAC has been evaluated in two different contexts. The goal of the first experiment was to establish relative performance of CogNIAC to Hobbs' Naive Algorithm--a convenient benchmark that allows indirect comparison to other algorithms. The second experiment reports results on Wall Street Journal data.

3.2.1 Experiment 1:

The chosen domain for comparison with Hobbs' Naive Algorithm was narrative texts about two persons of the same gender told from a third person perspective. The motivation for this data was that we wanted to maximize the ambiguity of resolving pronouns. Only singular third person pronouns were considered. The text was pre-processed with a part-of-speech tagger over which basal noun phrases were delimited and finite clauses and their relative nesting were identified by

machine. This pre-processing was subjected to hand correction in order to make comparison with Hobbs' as fair as possible since that was an entirely hand executed algorithm, but CogNIAC was otherwise machine run and scored. Errors were not chained, i.e. in left-to-right processing of the text, earlier mistakes were corrected before processing the next noun phrase.

Since the Naive Algorithm resolves all pronouns, two lower precision rules were added to rules 1-6) for comparisons sake. The rules are:

7) **Cb-Picking**³: If there is a Cb *i* in the current finite clause that is also a candidate antecedent, then pick *i* as the antecedent.

8) **Pick Most Recent**: Pick the most recent potential antecedent in the text.

The last two rules are lower precision than the first six, but perform well enough to merit their inclusion in a 'resolve all pronouns' configuration. Rule 7 performed reasonably well with 77% precision in training (10/13 correct for 201 pronouns), and rule 8 performed with 65% precision in training (44/63 correct). The first six rules each had a precision of greater than 90% for the training data with the exception of rule 4 which had a precision of 80% for 5 resolutions. The summary performance of the Naive Algorithm and CogNIAC (including all 8 rules) for the first 100 or so pronouns in three narrative texts are:

Naive Alg.	CogNIAC resolve all	CogNIAC High Prec.
235 (78.8%)	232 (77.9%)	190/206 (92%) P 190/298 (64%) R

Results for 298 third person pronouns in text about two same gender people.

Since both the Naive Algorithm and the resolve all pronouns configuration of CogNIAC are required to resolve all pronouns, precision and recall figures are not appropriate. Instead % correct figures are given. The high precision version of CogNIAC is reported with recall (number correct/number of instances of coreference) and precision (number correct/number of guesses) measures.

The conclusion to draw from these results is: if forced to commit to all anaphors, CogNIAC performs comparably to the Naive Algorithm. Lappin and Leas

³ Rule 7 is based on the primitives of Centering Theory (Grosz, Joshi and Weinstein '86). The Cb of an utterance is the highest ranked NP (Ranking being: Subject > All other NPs) from the prior finite clause realized anaphorically in the current finite clause. Please see Baldwin '95 for a full discussion of the details of the rule.

1994 correctly resolved 86% of 360 pronouns in computer manuals. Lapin and Leass run Hobbs' algorithm on the their data and the Naive Algorithm is correct 82% of the time--4% worse. This allows indirect comparison with CogNIAC, with the suggestive conclusion that the resolve all pronouns configuration of CogNIAC, like the Naive Algorithm, is at least in the ballpark of more modern approaches⁴.

The breakdown of the individual rules is as follows:

Rule	Recall	Precision
1)Uniq in Discourse	11% (32/298)	100% (32/32)
2) Reflexive	3% (10/298)	100% (10/10)
3) Uniq Curr + Prior	35% (104/298)	96% (104/110)
4) Possessive Pro	1% (2/298)	100% (2/2)
5) Uniq Curr Sentence	6% (18/298)	81% (18/22)
6) Uniq Subj/ Subj Pro	8% (24/298)	80% (24/30)
7) Cb-Picking	4% (13/298)	42% (13/31)
8) Pick Most Recent	10%(29/298)	48% (29/61)

Performance of individual rules in Experiment 1. Note the high precision of rules 1 - 6). Recall = #correct/#actual, Precision = #correct/#guessed

Far more interesting to consider is the performance of the high precision rules 1 through 6. The first four rules perform quite well at 96% precision (148/154) and 50% recall (148/298). Adding in rules 5 and 6 resolves a total of 190 pronouns correctly, with only 16 mistakes, a precision of 92% and recall of 64%. This contrasts strongly with the resolve-all-pronouns results of 78%. The last two rules, 7 and 8 performed quite badly on the test data. Despite their poor performance, CogNIAC still remained comparable to the Naive Algorithm.

3.2.2 Experiment 2-- All pronouns in MUC-6 evaluation:

CogNIAC was used as the pronoun component in the University Pennsylvania's coreference entry⁵ in the MUC-6 evaluation. Pronominal anaphora constitutes 17% of coreference annotations in the evaluation data used. The remaining instances of anaphora included common noun anaphora and coreferent instances of proper nouns. As a result being part of a larger system, changes were made to CogNIAC to make it fit in better with the other components of the overall system in addition to adding rules that were specialized for the new kinds of pronominal anaphora. These changes include:

⁴ This is not to say that RAP was not an advancement of the state of the art. A significant aspect of that research is that both RAP and the Naive Algorithm were machine executed--the Naive Algorithm was not machine executed in either the Hobbs 76 paper or in the evaluation in this work.

⁵ Please see Baldwin et al '96 for performance statistics and a bit more detail about the entire system.

- Processing quoted speech in a limited fashion (**Quoted Speech**).
- Addition of a rule that searched back for a unique antecedent through the text at first 3 sentences back, 8 sentences back, 12 sentences back and so on (**Search Back**).
- Addition of a partial parser [Collins 1996] to determine what a finite clause is.
- A new pattern was added which selected the subject of the immediately surrounding clause (**Subject Same Clause**).
- Addition of a pleonastic-it detector which filtered uses of *it* that were not pronominal.
- Disabling of several rules because they did not appear to be appropriate for the domain; 4, 7 and 8.

A total of thirty articles were used in the formal evaluation, of which I chose the first fifteen for closer analysis. The remaining fifteen were retained for future evaluations. The performance of CogNIAC was as follows:

	Recall (for pros)	Precision
All Pronouns	75% (85/114)	(73%) 85/116

The precision (73%) is quite a bit worse than that encountered in the narrative. The performance of the individual rules was quite different from the narrative texts, as shown in the table below:

Rule	Recall (pros)	Precision
Quoted Speech	11% (13/114)	(87%) 13/15
1) Uniq in Discourse	4% (5/114)	(100%) 5/5
3) Uniq Curr + Prior	50% (57/114)	(72%) 57/79
Search Back	1% (1/114)	(33%) 1/3
2) Reflexive	0% (0/114)	0/0
5) Uniq Curr Sentence	4% (5/114)	(70%) 5/7
Subject Same Clause	4% (4/114)	(57%) 4/7

The results for CogNIAC for all pronouns in the first 15 articles of the MUC-6 evaluation.

Upon closer examination approximately 75% of the errors were due to factors outside the scope of the CogNIAC pronominal resolution component. Software problems accounted for 20% of the incorrect cases, another 30% were due to semantic errors like misclassification of a noun phrase into person or company, singular/plural etc. The remaining errors were due to incorrect noun phrase identification, failure to recognize pleonastic-it or other cases where there is no instance of an antecedent. However, 25% of the errors were due directly to the rules of CogNIAC being plain wrong.

4 Discussion:

CogNIAC is both an engineering effort and a different approach to information processing in variable knowledge contexts. Each point is addressed in turn.

4.1 The utility of high precision coreference:

A question raised by a reviewer asked whether there was any use for high precision coreference given that it is not resolving as much coreference as other methods. In the first experiment, the high precision version of CogNIAC correctly resolved 62% of the pronouns as compared to the resolve all pronouns version which resolved 79% of them--a 27% loss of overall recall.

The answer to this question quite naturally depends on the application coreference is being used in. Some examples follow.

Information Retrieval

Information retrieval is characterized as a process by which a query is used to retrieve relevant documents from a text database. Queries are typically natural language based or Boolean expressions. Documents are retrieved and ranked for relevance using various string matching techniques with query terms in a document and the highest scoring documents are presented to the user first.

The role that coreference resolution might play in information retrieval is that retrieval algorithms that a) count the number of matches to a query term in a document, or b) count the proximity of matches to query terms, would benefit by noticing alternative realizations of the terms like 'he' in place 'George Bush'.

In such an application, high precision coreference would be more useful than high recall coreference if the information retrieval engine was returning too many irrelevant documents but getting a reasonable number of relevant documents. The coreference would only help the scores of presumably relevant documents, but at the expense of missing some relevant documents. A higher recall, lower precision algorithm would potentially add more irrelevant documents.

Coherence Checking

A direct application of the "ambiguity noticing" ability of CogNIAC is in checking the coherence of pronoun use in text for children and English as a second language learners. Ambiguous pronoun use is a substantial problem for beginning writers and language learners. CogNIAC could scan texts as they are being written and evaluate whether there was sufficient syntactic support from the context to resolve the pronoun--if not, then the user could be notified of a potentially ambiguous use. It is not clear that CogNIAC's current levels of performance could support such an application, but it is a promising application.

Information Extraction

Information extraction amounts to filling in template like data structures from free text. Typically

the patterns which are used to fill the templates are hand built. The latest MUC-6 evaluation involved management changes at companies. A major problem in information extraction is the fact that the desired information can be spread over many sentences in the text and coreference resolution is essential to relate relevant sentences to the correct individuals, companies etc. The MUC-6 coreference task was developed with the idea that it would aid information extraction technologies.

The consequences for an incorrectly resolved pronoun can be devastating to the final template filling task--one runs the risk of conflating information about one individual with another. High precision coreference appears to be a natural candidate for such applications.

4.2 The methodology behind CogNIAC

CogNIAC effectively circumscribes those cases where coreference can be done with high confidence and those cases that require greater world knowledge, but how might CogNIAC be a part of a more knowledge rich coreference application?

CogNIAC as a set of seven or so high precision rules would act as an effective filter on what a more knowledge rich application would have to resolve. But the essential component behind CogNIAC is not the rules themselves, but the control structure of behind its coreference resolution algorithm. This control structure could control general inference techniques as well.

An interesting way to look at CogNIAC is as a search procedure. The Naive Algorithm can be over simplified as depth first search over parse trees. Depth first search is also a perfectly reasonable control structure for an inference engine-- as it is with PROLOG. The search structure of CogNIAC could be characterized as parallel iterative deepening with solutions being accepted only if a unique solution is found to the depth of the parallel search. But there is not enough room in this paper to explore the general properties of CogNIAC's search and evaluation strategy.

Another angle on CogNIAC's role with more robust knowledge sources is to note that the recall limitations of CogNIAC for the class of pronouns/data considered are due to insufficient filtering mechanisms on candidate antecedents. There is not a need to expand the space of candidate antecedents with additional knowledge, but rather eliminate semantically plausible antecedents with constraints from verb knowledge and other sources of constraints currently not available to the system.

However, there are classes of coreference that require strong knowledge representation to assemble the initial set of candidate antecedents. This includes the realm of inferred definites "I went to the house and opened the door" and synonymy between definite common nouns as in "the tax" and 'the levy.

4.3 The possibility of perfect coreference

Hobbs 1976 ultimately rejects the Naive Algorithm as a stand-alone solution to the pronoun resolution problem. In that rejection he states:

The naive algorithm does not work. Anyone can think of examples where it fails. In these cases it not only fails; it gives no indication that it has failed and offers no help in finding the real antecedent.

Hobbs then articulates a vision of what the appropriate technology is, which entails inference over an encoding of world knowledge. But is world knowledge inherent in resolving all pronouns as Hobbs skepticism seems to convey?

It has not been clear up to this point whether any anaphora can be resolved with high confidence given that there are clear examples which can only be resolved with sophisticated world knowledge, e.g. the Winograd city council sentences. But the results from the first and second experiments demonstrate that it is possible to have respectable recall with very high precision (greater than 90%) for some kinds of pronominal resolution. However, good performance does not necessarily falsify Hobbs' skepticism.

The high precision component of CogNIAC still makes mistakes, 8-9% error for the first experiment--it is harder to evaluate the second experiment. If it were the case that integration of world knowledge would have prevented those errors, then Hobbs' skepticism still holds since CogNIAC has only minimized the role of world knowledge, not eliminated it. In looking at the mistakes made in the second experiment, there were no examples that appeared to be beyond the scope of further improving the syntactic rules or expanding the basic categorization of noun phrases into person, company or place. For the data considered so far, there does appear to be a class of anaphors that can be reliably recognized and resolved with non-knowledge intensive techniques. Whether this holds in general remains an open question, but it is a central design assumption behind the system.

A more satisfying answer to Hobbs' skepticism is contained in the earlier suggestive conjecture that world knowledge facilitates anaphora by eliminating ambiguity. This claim can be advanced to say that world knowledge comes into play in those cases of anaphora that do not fall under the purview of rules 1 through 7 and their refinements. If this is correct, then the introduction of better world knowledge sources will help in the recall of the system rather than the precision.

Ultimately, the utility of CogNIAC is a function of how it performs. The high precision rules of CogNIAC performed very well, greater than 90% precision with

good recall for the first experiment. In the second experiment, components other than the rules of CogNIAC began to degrade the performance of the system unduly. But there is promise in the high precision core of CogNIAC across varied domains.

5 The future of CogNIAC:

CogNIAC is currently the common noun and pronoun resolution component of the University of Pennsylvania's coreference resolution software and general NLP software (Camp). This paper does not address the common noun coreference aspects of the system but there are some interesting parallels with pronominal coreference. Some changes planned include the following sorts of coreference:

The processing of split antecedents,

John called Mary. They went to a movie.

This class of coreference is quite challenging because the plural anaphor 'they' must be able to collect a set of antecedents from the prior discourse--but how far should it look back, and once it has found two antecedents, should it continue to look for more?

Event reference is a class of coreference that will also prove to be quite challenging. For example:

The computer won the match. It was a great triumph.

The antecedent to 'It' could be any of 'The computer', 'the match' or the event of winning. The space of ambiguity will certainly grow substantially when events are considered as candidate antecedents.

Currently the system uses no verb semantics to try and constrain possible coreference. While the Winograd sentences are too difficult for current robust lexical semantic systems, simpler generalizations about what can fill an argument are possible, consider:

The price of aluminum rose today due to large purchases by ALCOA Inc. It claimed that it was not trying to corner the market.

Since 'It' is an argument to 'claimed', a verb that requires that its subject be animate, we can eliminate 'The price of aluminum' and 'today' from consideration, leaving 'ALCOA Inc.' as the sole singular antecedent from the prior sentence. Work has been done along these lines by Dagan '90.

6 Acknowledgments:

I would like to thank my advisors Ellen Prince and Aravind Joshi for their support. Also the comments of two anonymous reviewers proved quite helpful.

7 References:

- [Baldwin 1995]
Baldwin, B., "CogNIAC: A discourse processing engine", University of Pennsylvania, Department of Computer and Information Science, Ph.D. dissertation.
- [Baldwin et al, 1995]
Baldwin, B., Reynar, J., Collins, M., Eisner, J., Ratnaparkhi, A., Rosenzweig, J., Sarkar, A., Srinivas (1995), "University of Pennsylvania: Description of the University of Pennsylvania system used for MUC-6", Proceedings of Sixth Message Understanding Conference, November 1995.
- [Collins 1996]
Collins, M., "A New Statistical Parser based on Bigram Lexical Dependencies", Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, June, 1996.
- [Carter 1986]
Carter, D. (1986), "Common Sense Inference in a Focus-Guided Anaphor Resolver", Journal of Semantics, 4.
- [Chinchor and Sundheim 1994]
Chinchor, N., Sundheim, B. (1994) "Coreference Task Definition v1.1", supporting documentation for MUC-6 coreference task.
- [Dagan 1990]
Dagan, I., Itai, A., (1990) A Statistical Filter for Resolving Pronoun References, Proceedings of 7th Israeli Symposium on Artificial Intelligence and Computer Vision.
- [Grosz, Joshi and Weinstein 1986]
Grosz, B. J, Joshi, A. K., and Weinstein, S. (1986) "Towards a Computational Theory of Discourse Interpretation" Ms.
- [Hobbs 1976]
Hobbs, J. (1976) "Pronoun Resolution", Research Report #76-1, City College, City University of New York.
- [Hobbs 1977]
Hobbs, J. (1977) "38 Examples of Elusive Antecedents from Published Texts", Research Report #77-2, City College, City University of New York.
- [Kennedy 1996]
C. Kennedy, B. Boguraev, "Anaphora for everyone: pronominal anaphora resolution without a parser." Proceedings of the 16th International Conference on Computational Linguistics COLING'96, Copenhagen, Denmark, 5-9 August 1996
- [Lappin and Leass 1994]
Lappin, S., Leass, H. (1994) "An Algorithm for Pronominal Anaphora Resolution", Computational Linguistics.
- [Mitkov 1997]
Mitkov R., "Pronoun resolution: the practical alternative", In S. Botley, T. McEnery (Eds) "Discourse Anaphora and Anaphor Resolution". University College London Press, 1997
- [Sidner 1986]
Sidner, C. L. (1986) "Focusing in the Comprehension of Definite Anaphora", Readings in Natural Language Processing, eds Grosz, Jones, Webber. Morgan Kaufman, Los Altos CA.
- [Winograd 1972]
Winograd, T. 1972, "Understanding Natural Language", New York: Academic Press.