Eliminating Non-Referring Noun Phrases from Coreference Resolution

Donna K. Byron and Whitney Gegg-Harrison

Department of Computer Science and Engineering
The Ohio State University
2015 Neil Ave.
Columbus, Ohio 43210 USA
dbyron/geggharr@cis.ohio-state.edu

Abstract

Indefinite noun phrases in certain contexts are unable to support anaphoric coreference to an individual entity, and therefore should be ignored when searching for coreferent antecedents of anaphoric pronouns. However, many algorithms for anaphora resolution utilize noun phrase chunking or shallow parsing, and therefore do not make the needed distinctions to avoid this type of spurious antecedent. This study used simple syntactic criteria to remove indefinite phrases from consideration as antecedents to evaluate the effect of their removal on pronoun resolution. Pronoun resolution performance improved only marginally, revealing some interesting properties of current pronoun resolution algorithms.

1. Introduction

Creating automated techniques to pair coreferent anaphoric pronouns with their antecedent remains a challenging research problem. In the computational linguistics literature, the task is typically modeled as a search process that selects between noun phrases to the left of the pronoun in the discourse, in order to determine which of them is most likely to be the pronoun's antecedent. For any particular pronoun, there will typically be many candidate NPs to consider. Improvements to the resolution process can be made either by developing more precise ways to order the search, or by creating better methods for eliminating inappropriate candidates from the search space.

A wide variety of different algorithms have been developed to order the search space using different criteria. For example, some algorithms use structural properties to order the search (e.g. Hobbs, 1986; Tetreault, 2001), some use information structure (e.g. Strube, 1998), and others use a combination of salience and syntactic features (Lappin & Leass, 1994, Ge et al. 1998). Rather then ordering the candidate NPs in a novel way, this study instead focuses on reducing the search space by eliminating spurious candidates, under the assumption that removing invalid candidates will result in a higher probability that the correct antecedent can be found.

The technique we employ is to use surface syntactic features to identify indefinite NPs that do not license anaphoric coreference, and to remove them from consideration as antecedents for anaphoric pronouns. This technique exploits the well-known fact that some items which analyze as noun phrases are non-referential, and therefore do not support anaphoric re-mention. Karttunen's paper *Discourse Referents* (Karttunen, 1976) identifies a variety of constructions in which indefinite noun phrases do not carry a presupposition of existence of an individual referent, and therefore do not support anaphoric re-mention of that entity outside of the local scope. An example is ``Pat is a great teacher'', in which the NP *Pat* introduces a discourse referent, but the NP *a great teacher* does not. No matter what method is used to gather candidate antecedents for subsequent anaphoric pronouns, the phrase *a great teacher* should not create an element of that set. Although Kartunnen's paper has been widely cited, its main message, that discourse referents *not* be instantiated for certain kinds of noun phrases, has not received much attention. Even coreference annotation guidelines, such as the MUC coreference task annotation instructions (SAIC, 1997), have failed to adequately make this distinction when describing which NPs are capable of participating in coreference relations. For a detailed discussion of this issue, see (van Deemter & Kibble, 1999).

In this study, we tested several pronoun resolution methods on the same dataset and compared their results when non-referring indefinite NPs were suppressed as candidate antecedents to the condition in which they were not suppressed. Our findings indicate that removing these indefinite expressions from the antecedent candidate set creates an insignificant gain in performance for pronoun resolution algorithms that take sentence structure into account, but provides a larger increase in techniques that do not. In addition, because a large number of candidate antecedents are eliminated, this technique results in a reduction in run-time for generate-and-test style pronoun resolution algorithms.

A discussion of categories of non-licensing indefinite NPs follows in Section 2, followed by the details of the algorithm we created for automatically identifying a subset of these noun phrases. Section 4 then briefly describes the pronoun resolution algorithms used in our experiment. Section 5 then describes our results utilizing this filter to remove non-referential NPs in a corpus of Wall Street Journal articles from the Penn Treebank.

2. Categories of non-licensing NPs

Karttunen's paper specifies the contexts in which indefinites do not support anaphoric coreference to a specific item (though they do license other anaphoric possibilities, such as reference to the kind). The guiding principle is that nonspecific indefinites cannot support anaphoric re-mention unless "the proposition represented by the sentence is asserted, implied, or presupposed by the speaker to be true" (page 371). When the proposition containing the indefinite is false, re-mention is not felicitous. The specific constructions identified by Karttunen in which the indefinites do not support re-mention are:

- Predicate Nominals: "Bill is *a linguist*"
- Negation: such as "Bill does not have *a car*", except in the case of factive verbs such as *know*
- Modal verbs: "Bill can make *a kite*"
- Quasi-modals: try, intend, plan, etc. in which the complement clause represents a yet-untrue proposition
- Negative-implicature verbs: "John forgot to write *a term paper*"
- Commands and yes/no questions

Karttunen points out that referents are sometimes available only within a local scope, for example in "I want to catch a fish and eat it" vs. "I want to catch a fish. *I want to eat it." Several of the constructions described by Karttunen can only be identified using semantic properties of the verb phrase. For example, a negative implication verb such as *fail* in ``Pat failed to write a term paper'' implies the non-existence of the entity described in object position, and therefore does not create a discourse referent that can be re-mentioned.

In addition to the cases described by Karttunen, indefinite NPs in several other contexts do not license subsequent mention. For example, negated NPs such as "no man", the predication portion of an apposition, for example "a Manchester banker" in "Peter VanHausen, a Manchester banker", and noun phrase predication expressed as prepositional phrases, such as 'a nonexecutive director' in "Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29".

Our aim is to create a preprocessing step for pronoun resolution that automatically detects and removes these phrases from consideration as antecedents. The pronouns considered in this paper are only those pronouns that specify the identical entity as their antecedent phrase. So although the indefinite noun phrases that we are labelling as 'non-licensing phrases' could in fact license pronouns through other anaphoric relationships, such as one-anaphora referring to the kind (i.e. "Chris failed to find an apartment but I found one") or reference to a complement set (i.e. "No senator was available for comment, as they were all attending the hearing"), their inability to license a pronoun referring to the same individual as the indefinite NP is the property of interest in this study.

3. Automatic Classification of non-licensing NPs

This section describes the details of our algorithm to automatically tag non-referential NPs in the Penn Treebank corpus. The tagged NPs are filtered out in a preprocessing step so that they are not listed among the candidate antecedents for pronoun resolution. We chose not to incorporate semantic information in our labeling process, but to concentrated rather on items that could be easily identified using the Penn Treebank-3 tagset (with one exception – appositive phrases were hand-labeled to indicate the referential vs. the predication portion of the phrase). Inspired by Karttunen's analysis, but adding some rules of our own, we gathered the following set of attributes for each NP during pre-processing:

- **Determiner** The determiner used in this noun phrase, if there is one.
- **Predicate** Indicates that the noun phrase is in the predicate of a copular verb phrase (the –PRD phrase tag).
- **Negated** Indicates that the NP is dominated by a negated verb phrase.
- **Appositive** Indicates that the noun phrase is the predication in an apposition (these phrases were hand-labeled).
- **Modal** Indicates that the NP is dominated by a modal verb phrase from the set {'D CA CAN MIGHT SHOULD OUGHT COULD MAY WOULD}
- **Modifier** Indicates that the NP is part of an adjective phrase or predication adjunct (tagged with the phrase tag CLR in the Treebank)
- **Numeric-Value** Noun phrases that express a value are considered to be predication rather than referring expressions, for example ``4 Billion''

Using these flags, non-candidate NPs were classified using the rule shown in Figure 1. An NP was considered indefinite if its determiner was a member of the set:
{a an any some another either neither every each all}.

```
NON-REF if (appositive is true
            or determiner is 'no'
            or numeric-value is true
            or (the NP is indefinite
                 and (negated is true
                      or predicate is true
                      or modal is true
                      or modifier is true)))
```

Figure 1: Pseudocode for the non-referential filter

As mentioned above, it is possible for indefinite NPs to create discourse referents that are only accessible for re-mention within a particular scope, such as the fish in "I want to catch a fish and eat it." Also, in many of the contexts that Karttunen describes, a referent does exist if the modality of the sentence initially describing it is maintained (i e. "John wants to marry a millionaire as long as she is devout"). However, the process we created was not sophisticated enough to allow these short-term referents. Instead, the indefinite NPs labeled as non-ref in the filter were completely eliminated. An examination of the results indicates that this was not a source of error in the pronoun resolution results; in other words our test dataset did not happen to contain any of these short-term referents.

4. Pronoun Resolution Algorithms

To determine whether elimination of non-referential NPs would aid in pronoun resolution accuracy, we chose two leading pronoun resolution techniques to compare, as well as a very simplistic baseline algorithm. All three algorithms take advantage of agreement features, so antecedents must match the anaphoric pronoun in number, gender, and animacy. The Hobbs (Hobbs, 1986) tree-search algorithm utilizes a search order determined by the syntactic structure of each sentence. The algorithm starts from a pronoun to be resolved and probes each base noun phrase looking for an antecedent that matches agreement features for the pronoun. The probe order works leftward from the pronoun, examining disjoint noun phrases in the current sentence, then executes a breadth-first search on noun phrases from previous sentences, moving progressively further back from the pronoun.

The LRC algorithm, as described in (Tetreault, 2001), uses a breadth-first walk of the syntax tree of the sentence containing the pronoun to find intra-sentential antecedents, then utilizes the same breadth-first probe order as Hobbs to search for antecedents from previous sentences. For intra-sentential antecedents, there are two differences between Hobbs and LRC. LRC searches for intra-sentential antecedents by starting at the beginning of the sentence and working toward the pronoun, rather than starting from the pronoun and working leftwards. Also, LRC takes all nps from sentence-initial dependent clauses (adverbials appearing before the subject) and moves them to the end of the sentence. Therefore, spurious antecedents from supporting material at the beginning of the sentence are not visited before noun phrases in the subject. Notice that LRC always prefers antecedents that are objects of the main verb of the sentence over noun phrases deeper in the syntax tree, while Hobbs can resolve the pronoun to an antecedent deeper in the sentence structure, depending on the position of the pronoun.

We also experimented with a simplistic baseline algorithm, which we label 'most recent' in the results section. The 'most recent' algorithm does not use syntactic information or any other sophisticated source of evidence to chose an antecedent, but instead simply starts from the pronoun moving leftwards, testing each noun phrase encountered for agreement. The first noun phrase encountered that matches the pronoun's agreement features is selected as the antecedent. This method has an equal likelihood of resolving a pronoun to an antecedent in the predicate or the subject of a sentence, and does not differentiate based on the depth of the potential antecedent within the syntactic structure.

5. Results

The corpus used in this study consists of stories 0001 - 0199 of the Wall Street Journal portion of the Penn Treebank that had previously been annotated with correct pronominal referents for base-noun-phrase coreferent pronouns by (Ge et al., 1998). This corpus is the same dataset used in the evaluation of the LRC, Hobbs, and BFP algorithms reported in (Tetreault, 2001). The Treebank syntax trees are preprocessed by a noun-phrase chunker to extract a list of base noun phrases from each sentence. The noun phrase chunker was supplemented to collect the attributes described in Section 3 and to filter out the appropriate NPs using the non-referential filter described in Figure 1. The test corpus contains 25,249 noun phrases in 4,090 sentences, and contains 1,699 test tokens from the types {IT, ITS, HE, THEY, THEIR, THEM, HIS, SHE, THEMSELVES, ITSELF, HER, HIM, HIMSELF and HERSELF}. The non-referential filter identified a total of 3,156 noun phrases (12%), which were eliminated from consideration as antecedents for the pronouns. Each pronoun was considered correct if the resolution process paired it with any element of the correct coreference chain.

5.1. Pronoun Resolution Performance

Table 1 shows the small amount of improvement gained for each of three different pronoun resolution techniques when we applied the filter. The LRC algorithm, shown in row (a), when all NPs are candidate antecedents, finds the correct antecedent for 1366 of the pronouns, with Hobbs not far behind at 1339. Both Hobbs and LRC correctly resolve only a few additional pronouns when the filter is used. LRC shows a gain of only 7 pronouns (from 80.4% to 80.8% correct), and Hobbs picked up an additional 9 pronouns when the non-referential filter is employed (to improve from 78.8% to 79.3%). This small amount of improvement is quite remarkable in light of the fact that 12% of the noun phrases in the corpus have been removed. Both of these techniques utilize structural properties of the syntax tree to prefer NP heads over their modifiers, and both techniques prefer subjects of sentences over predicates for inter-sentential search. Based on this structural search order, it is apparent that the noun phrases eliminated by our filter were rarely creating problems for these algorithms.

Our filter labeled 3,156 noun phrases as non-referential. Had these noun phrases been left in the sentences as potential antecedents, only 163 of them would have been examined as potential antecedents by LRC, using its left-to-right search order, and only 86 would have been examined by Hobbs. This statistic reveals something interesting about the construction of these algorithms. Because they use search orders that prefer foregrounded over backgrounded material, the search naturally avoids the non-referring NPs without requiring their removal.

Table 1: Resolution accuracy for pronouns with coreference antecedents

	Overall		Breakdown by Position of Pronoun			
	Original	**With Filter**	**Original**		**With Filter**	
			In Subject	**In Predicate**	**In Subject**	**In Predicate**
Total Pronouns	1,699	1,699	1,430	268	1,430	268
(a) LRC	1,366	1,373	1,194	172	1,201	172
(b) Hobbs	1,339	1,348	1,175	164	1,181	167
(c)Most-Recent	1,057	1,088	906	153	924	164

As the table shows, a method with less structural sophistication, such as the ``Most Recent" technique shown in row (c), which simply binds the pronoun to its closest left-hand neighbor using a simple linear search, benefits more from the filter. Although the Hobbs and LRC algorithms improve by less than one percent using the filter, Most Recent picks up nearly a 2% improvement in accuracy (from 62.2% without the filter to 64% with).

Table 1 also breaks down the algorithms' performance based on the pronoun's position in its sentence. In error analysis undertaken as part of a previous study, we had noticed that both Hobbs and LRC perform worse on pronouns that are themselves part of a predicate than on pronouns in the subject of their sentence. We expected that the non-referential filter would improve the resolution of pronouns within the predicate, because non-referential NPs within the predicate are removed. As expected, Hobbs algorithm does show more improvement for predicates (1.3% improvement in predicate pronouns vs .4% improvement for subject pronouns), however the counts are so small here that no reliable conclusions can be drawn. The results for the Most Recent technique are more divergent, as the filter does result in a higher performance gain for pronouns in predicates (4% improvement) than for those in subjects (2% improvement).

5.2. Running Time Improvement

In addition to the gain in pronoun resolution accuracy introduced by the filter, removing these indefinite phrases can also improve the algorithm's running time, as fewer candidate antecedents are generated from each sentence. This would impact some algorithms more than others. For example, the BFP algorithm (Brennan et al, 1987) and its Optimality Theory equivalent COT (Beaver, 2004) both begin by building a list of all possible pairings of pronouns in a sentence with antecedents from the previous sentence. Generating these hypotheses is time consuming, and this has been a frequent criticism of these centering-style algorithms. Using our rule to filter out the non-licensing indefinite NPs resulted in a 15% reduction in running time for our implementation of COT.

5.3 Future Work

The coverage of the simplistic filter we describe in this paper can be extended in a number of ways. Additional categories of non-licensing phrases could be eliminated, and developing an automatic process to detect these additional categories is a point for future work. For example, in one sentence in our test corpus 'busloads' is selected as the antecedent for 'their' in the phrase 'busloads of executives and their wives'. As a quantifier phrase, the NP 'busloads' should not be considered a candidate antecedent, but quantifier phrases in the Treebank do not carry a special tag. Automatic detection of these phrases may depend on lexical resources or gazetteers. In addition, the coverage of the filter would be extended if the objects of certain verbs were eliminated, as described by Karttunen. For example, identifying negative implicature verbs using a lexical resource such as Wordnet is a point for future work.

6. Conclusions

This paper describes our initial attempt to improve pronoun resolution performance by eliminating certain types of indefinite noun phrases from consideration as antecedents. We found that the NPs eliminated by our filter tend not to occur in a prominent position in their sentence, so pronoun resolution algorithms that use breadth-first syntax tree traversal as a search heuristic avoid these phrases for the most part without removing them. Other techniques that do not consider syntax-tree depth are more likely to incorrectly select these phrases as pronoun antecedents.

Although the improvement to pronoun resolution accuracy is small using this filter, we feel that this technique could be part of a set of preprocessing steps that work together to increase pronoun resolution accuracy. The state of the art for pronoun resolution has plateaued at around 80%, and there may not be any single techniques left to push us through that last 15-20% improvement. Apart from developing deep analysis and true understanding, we feel that the remaining improvements will come one percent at a time, slogging through a series of small process improvements. The filter explored in this paper is an example of one such improvement.

Acknowledgements

The authors would like to thank Joel Tetreault for providing LISP code for the NP chunker and the LRC algorithm, Eugene Charniak, Niyu Ge, and John Hale for the coreference-annotated WSJ corpus, and the anonymous reviewers for many helpful remarks.

7. References

Beaver, David I. (2004). The optimization of discourse anaphora. *Linguistics and Philosophy,* 27(1):3-56.

Brennan, Susan, Friedman, Marilyn W. and Pollard, Carl J. (1987). A centering approach to pronouns. In *Proceedings of the 25th Meeting of the Association for Computational Linguistics (ACL '87),* pages 155-162.

Ge, Niyu, Hale, John, and Charniak, Eugene. (1998). A statistical approach to anaphora resolution. In *Proceedings of the Sixth Workshop on Very Large Corpora*. pages 161-170.

Hobbs, Jerry. (1986). Resolving pronoun reference. In *Readings in Natural Language Processing*, pages 339-352. Morgan Kaufmann.

Karttunen, Lauri. (1976). Discourse referents. In J. McKawley, editor, *Syntax and Semantics*, vol. 7, pages 361-385. Academic Press.

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

SAIC. (1997) The MUC-7 Coreference Task Definition. Science Applications International Corporation. Available from ftp://ftp.muc.saic.com/pub/MUC/MUC7-guidelines/

Strube, Michael. (1998). Never Look Back: An Alternative to Centering, In the *Proceedings of the 38th annual meeting of the Association for Computational Linguistics (ACL '98),* pages 1251-1257.

Tetreault, Joel R. (2001). A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, 27(4):507-520.

Van Deemter, Kees & Kibble, Roger. (1999). What is coreference, and what should coreference annotation be? In the Proceedings of the Workshop Coreference and its Applications (ACL '99), Amit Bagga, Breck Baldwin, and Sara Shelton, eds. pages 90-96.