Mining a Corpus to Support Associative Anaphora Resolution

Josef Meyer and Robert Dale
Centre for Language Technology
Macquarie University
Sydney, Australia
{jmeyer|rdale}@ics.mq.edu.au

Abstract

While work on pronominal anaphora resolution is well established, and there has been considerable work on definite noun phrase anaphora, associative anaphora—a phenomenon whereby an entity can be referred to by a definite referring expression without previously being mentioned in the text—has been much less investigated. In this paper, we describe work that aims to determine the appropriate antecedents for such referring expressions by mining a corpus to build **associative axioms**, and we present preliminary results of some experiments that use WordNet as a source of generalisations to overcome data sparseness in the derivation of these axioms.

1 Introduction

Different aspects of the task of anaphor resolution have been explored to differing degrees in the literature. While computational work on pronominal anaphora resolution is very well established, and there has been a considerable amount of research on definite noun phrase anaphora, work on associative anaphora is much less in evidence. By associative anaphora we mean here the phenomenon in discourse whereby an entity can be referred to by a definite referring expression without prior mention in the text. A typical example from the literature is the use of the definite noun phrase reference in the second sentence in example (1):[1]

(1) A bus came around the corner.
The driver had a mean look in her eye.

The usual explanation offered for the felicity of such examples is that either the context or general world knowledge makes available for reference entities that are associated in some way with an explicitly mentioned discourse referent; here, the referent of *the driver* is associated with the previously mentioned bus. For our purposes, we consider *a bus* to be the antecedent in a case like this, and so the process of resolution requires identification of this antecedent.

From a computational point of view, these anaphoric forms are problematic because their resolution would seem to require the encoding of substantial amounts of world knowledge. In this paper, we explore how evidence derived from a corpus might be combined with a semantic hierarchy such as WordNet to assist in the resolution of these anaphoric forms.

Section 2 provides some background context and presents our perspective on the problem. In Section 3, we describe the corpus we are using, and the techniques we have been exploring. Section 4 describes the current results of this exploration, and Section 5 draws some conclusions and points to a number of directions for future work.

2 The Problem

The phenomenon of associative anaphora as introduced above has been widely discussed in the linguistics literature: see, for example, (Hawkins, 1978; Clark and Marshall, 1981; Prince, 1981; Heim, 1982). However, computational approaches to the resolution of this form of anaphora are much less common.[2] This is not surprising: the trend over the last decade has been towards shallow processing approaches to anaphora resolution, but the absence of surface level cues makes associative anaphora difficult to handle using such techniques. On the other hand, using knowledge-based approaches of the kind that were commonly discussed in the literature in

[1] In these examples, italics are used to indicate anaphors.

[2] A notable exception here is the work of (Poesio et al., 1997) and (Vieira, 1998).

earlier decades (see, for example, (Grosz, 1977; Sidner, 1979)) is clearly problematic, especially given the almost limitless bounds on what can be associated with an already mentioned entity. The evidence would seem to suggest that a hearer can accommodate a posited associative relationship in a very wide range of circumstances; consequently, developing a knowledge-based approach to this problem is far from trivial, and probably unrealistic for practical broad coverage natural language processing tasks.

In processing a text, there are three possibilities we need to consider whenever we find a definite noun phrase. First, the definite noun phrase may, of course, be an anaphoric reference to an entity mentioned elsewhere in the text, where the antecedent reference may or may not share lexical content with the anaphor; such uses do not constitute associative anaphora and so are outside the scope of interest of this paper.

Second, the definite noun phrase may be a reference to an entity that is not explicitly mentioned in the text, but whose existence can be inferred on the basis of its association with some entity that is referred to elsewhere in the text; example (1) above demonstrates this phenomenon. We will refer to these uses as **textually-licenced associative anaphors**.

Third, the definite noun phrase may be a reference to an entity that is not explicitly mentioned in the text, but whose existence can be assumed on the basis of world knowledge. For our purposes, this case covers both reference to entities present in the physical environment and those whose existence can simply be taken for granted; we will refer to these as **contextually-licenced associative anaphors**.

There are essentially two related questions we want to be able to answer: Given a definite NP, is it a textually-licenced associative anaphor? And if so, how can we determine its antecedent? The linguistic context provides us with a set of candidate antecedents: we are not concerned in the present paper with how this set of candidate antecedents is derived or represented, although our current work uses an approach similar in spirit to that of (Lappin and Leass, 1994; Boguraev and Kennedy, 1996). Neither are we concerned in the current paper with determining the precise nature of the semantic or real-world relationship between the associative anaphor and its antecedent. We focus here on the second question above: if we assume that the anaphor is a textually-licenced associative anaphor, how do we assess the likelihood of each candidate being its antecedent?

Our motivating observation is a simple one, and one that has been explored in other areas (see, for example, (Hearst, 1992; Knott and Dale, 1995)): that semantic relationships which are left implicit for a reader to infer in some contexts may also occur explicitly in other contexts. Specifically, the kinds of entities that figure in associative anaphoric relationships are also often referred to in contexts where the relationship between the two entities is made explicit, as in example (2):

(2) A bus nearly collided with a car.
The driver of the bus had a mean look in her eye.

Here, we have prima facie evidence of the existence of a relationship between drivers and buses. Our goal is to see whether this kind of evidence can be gathered from a corpus and then used in cases where the association between the two entities is not made explicit.

3 Extracting Evidence from a Corpus

3.1 The Corpus

For our experiments, we have been working with a corpus of some 2000 encyclopaedia articles drawn from the electronic versions of Grolier's Encyclopaedia and Microsoft's Encarta. All the articles we are using are descriptions of animals, with 1289 from Grolier's and 932 from Encarta. Manual analysis of portions of the corpus suggests that it contains a significant number of instances of associative anaphora. Some interesting examples are presented below:

(3) The head of a ground beetle is narrower than its body; long, thin, threadlike antennae jut out from *the sides of the head*.
The mouthparts are adapted for crushing and eating insects, worms, and snails.

(4) Beetles undergo complete metamorphosis.
The larvae are cylindrical grubs, with three pairs of legs on *the thorax*; *the pupae* are usually encased in a thin, light-colored skin with *the legs* free; *the adults* have biting mouth parts, in some cases enormously developed.

These examples should make it clear that identifying the antecedent is already a difficult enough problem; identifying the nature of the relationship between the entities referred to is significantly more complicated, and often requires quite sophisticated semantic notions.

3.2 Our Approach

If we were pursuing this work from a knowledge-based perspective, we might expect to have available a collection of axioms that could be used in resolving associative anaphoric expressions. So, for example, we might have an axiom that states that buses have drivers; this axiom, and many others like it, would then be brought to bear in identifying an appropriate antecedent.

As noted earlier, we are not concerned in the present paper with the precise nature of the association: often such relationships are meronymic, but this is clearly not always the case. For our purposes, it is sufficient to know that an association exists. As indicated above, the possibility of such a relationship can be derived from a corpus; in effect, the corpus provides us with existence proofs.

Our approach, then, is to mine a corpus for explicit statements of association, and to use the evidence thus garnered as a source for constructing what we will call **associative axioms**; these axioms can then be used as one component in an anaphor resolution process.

Statements of association take a number of different forms, and one issue we face is that these are of varying reliability, a point we will return to in Section 5. In the present work we focus on two forms of statements of association that we suspect are of quite high reliability: genitive constructions and *of NP* constructions, as in examples (5a) and (5b) below.

(5) a. *The stingray's head* is not well defined, and there is no dorsal or caudal fin.

b. *The head of the stingray* is not well defined, and there is no dorsal or caudal fin.

Given a unmodified NP like *the head*, we want to identify which entity in the preceding text this is associated with. Suppose there are a number of candidate antecedent NPs in the context, and *the stingray* is one. If we find instances in the corpus of expressions like those italicised in (5a) and (5b), then we have prima facie evidence that the antecedent might be *the stingray*: at the very least, we have evidence that stingrays have heads.

Of course, such an approach is prone to the problems of data sparseness. The chance of finding such explicit evidence elsewhere in a corpus is low, unless the corpus is very large indeed. Our response to this is, again, similar to the solution taken by other tasks that face this problem: we try to find useful generalisations that allow us to overcome the data sparseness problem. The source for our generalisations is WordNet (Fellbaum, 1998), although it could in principle be any available taxonomic or ontological knowledge source.

WordNet tells us that heads are body parts, and that stingrays are fish; thus, the appearance of examples like (5a) and (5b) above could be considered as evidence that fish have body parts. The data also supports a host of other generalisations, some more useful than others: for example, we have some evidence that fish have heads, and that stingrays have body parts; we can also, of course, climb higher up the taxonomic hierarchy, although clearly the higher we go, the less useful or informative the resulting relationships are: the more we generalise, the more we risk overgeneralising. Ultimately, of course, we will determine that somethings have somethings, but this not useful information.

Our goal, then, is to see what useful relationships we might be able to mine from explicit statements in a corpus, and then to use these relationships as a factor in determining antecedents of associative anaphora. The key problem we face is in determining the appropriateness or reliability of the generalisations we extract.

4 An Experiment

4.1 Associative Constructions

To support the generalisations that we wish to extract from the corpus, we need to identify cases where the anaphoric element appears in a syntactic configuration that makes the presence of an associative relationship explicit; we refer to these syntactic configurations as **associative constructions**. Examples of such associative constructions are the forms ⟨*NP of NP*⟩ and ⟨*Genitive NP*⟩ as in example (5) above. In these constructions, we will refer to the head of the first NP in the case of the pattern ⟨*NP of NP*⟩, and the N in the case of the pattern ⟨*Genitive N*⟩, as the **head** of the associative construction, and to the other head noun in each case as the **modifier** of the associative construction; thus, in the example under discussion, the head is *head* and the modifier is *stingray*.

To identify associative constructions, we first process our texts using Conexor's FDG parser (Tapanainen and Jarvinen, 1997). The results of this analysis for example (5a) are shown below:

```
1   The      the      det:>2   @DN> %>N DET
2   head     head     subj:>6  @SUBJ %NH N NOM
3   of       of       mod:>2   @<NOM-OF %N< PREP
4   the      the      det:>5   @DN> %>N DET
5   stingray stingray pcomp:>3 @<P %NH N NOM SG
6   is       be       main:>0  @+FMAINV %VA V PRES SG3
7   not      not      neg:>6   @ADVL %EH NEG-PART
8   well     well     man:>6   @ADVL %EH ADV
9   defined  defined  comp:>6  @PCOMPL-S %NH A ABS
```

We then use a collection of regular expression matching procedures to identify the NPs in the text. A further filter over the extracted NPs identifies the expressions that meet the patterns described above, producing the results shown in Table 1. The ratios of tokens to types here may

Associative Construction	Types	Tokens
⟨*NP of NP*⟩	11322	17164
⟨*Genitive N*⟩	2133	5662

Table 1: Associative constructions in the corpus

seem surprisingly high; however, the data is of course fairly skewed. For example, the statement of association *member of family* occurs 193 times in the corpus, and *bird of prey* occurs 25 times. It is clear from a rudimentary analysis of this data that many of the high frequency forms are of a semantic type other than that which we are interested in. Also, not all expressions which match our patterns for associative constructions actually express associative constructions. Some of these can be filtered out using simple heuristics and stop word lists; for example, we know that the relationship expressed by the *of* in *number of N* is not of interest to us. Other candidates that can be ignored are terms like *north of*, *south of*, and so on.

Given these analyses as evidence of associations, we then refer to any ⟨head, modifier⟩ pair for which we have evidence as a **lexical associative axiom**. From example (5) we thus have the following lexical associative axiom:

(6) have(*stingray*, *head*)

The 'have' predicate effectively encodes what we might think of as 'unspecified association'.

4.2 Generalising Associative Axioms

There are 1092 ⟨*NP of NP*⟩ forms that appear twice in the corpus, and 9391 that appear only once; and it is these low frequency constructions that appear more relevant to our purpose. Given the low frequencies, we therefore want to generalise the lexical associative axioms we can derive directly from the text. WordNet's hypernymic relationships give us an easy way to do this. Thus, an expression like *the leg of the okapi* supports a number of associative axioms, including the following:[3]

(7) have(*okapi*, *leg*)
have(*okapi*, LIMB)
have(GIRAFFE, *leg*)
have(GIRAFFE, LIMB)
...
have(LIVING THING, BODY PART)

Of course, there are two notable problems with this that lead to inappropriate generalisations.

First, since many or most lexical items in WordNet have multiple senses, we will produce incorrect generalisations: the above is fine for the sense of leg as 'a structure in animals that is similar to a human leg and used for locomotion' (sense 2), but there are eight other senses in WordNet, including such things as 'a section or portion of a journey or course' (sense 9). Generalisations derived from these senses will clearly be in error. This could be addressed, of course, by first applying a word sense disambiguation process to the source texts.

A second problem is that, to use the example above, just because the okapi has a leg does not mean that the giraffe, or the referents of any superordinate terms, also have legs. This overgeneralisation is simply a reflection of the fact that default inheritance may not hold.

Notwithstanding these problems, for each generalisation we make, we take the view that we have some evidence. If we measure this as the number of instances that support the generalisation, then, as we go higher up the WordNet taxonomy, our putative evidence for a generalisation will increase. At the same time, however, as the generality increases, the less potentially useful the generalisations are likely to be in anaphora resolution.

We refer to each generalisation step as an **expansion** of the axiom, and to the result as a **derived associative axiom**. We would like to have some indication, therefore, of how useful a given degree of expansion is, so that we are in a better position to decide on the appropriate trade off between the increased evidence and decreased utility of a given generalisation.

4.3 Evaluating the Axioms

For an evaluation of the effectiveness of our associative axioms, we focussed on four particular heads that appeared in our extracted statements of association: *body*, *color*, *head* and *tip*, as in the following examples:

(8) a. *its head*, *the snake's head*, *the head of the stingray*
b. *its color*, *the snake's color*, *color of the skin*, *color of its coat*
c. *its body*, *the female's body*, *the bird's body*
d. *its tip*, *the tip of the island*, *the tip of the beak*

For each of these heads, we automatically extracted all the **contexts of occurrence** from the corpus: we defined a context of occurrence to be an occurrence of the head without a modifier (thus, a suspected associative anaphor) plus its two preceding sentences. [4] Omitting those cases where the antecedent was not present in the context, this delivered 230 contexts for *body*, 19 for *color*, 189 for *head*, and 33 for *tip*. Then, we automatically identified all the NPs in each context; these constitute the candidate antecedent sets for the associative anaphors, referred to here as the **initial candidate sets**. We then manually annotated each instance in this test set to indicate the true antecedents of the associative anaphor; since the antecedent entity may be referred to more than once in the context, for each anaphor this gives us a target antecedent set (henceforth the **target set**).

To test the utility of our axioms, we then used the lexical and derived axioms to filter the initial candidate set, varying the number of generalisation steps from zero (i.e., using only lexical associative axioms) to five (i.e., using derived axioms generated by synset lookup followed by four levels of hypernym lookup): at each step, those candidates for which we do not have evidence of association are removed, with the remaining elements being referred to as the **selected set**. Ideally, of course, the axioms should reduce the candidate set without removing

[3] Small caps are used here to indicate generalised terms.

[4] An informal analysis suggests that the antecedent of an associative anaphor generally occurs no further back than the two previous sentences. Of course, this parameter can be modified.

elements that are in the target set: in other words, the **overlap set** (the intersection of the target set and the selected set) should be non-empty.

Our first pass at determining the effectiveness of the filters was to measure the extent to which they reduce the candidate sets: so, for example, if the context in a test instance contains four possible antecedents, and the filter only permits one of these and rejects the other three, we have reduced the candidate set to 25% of its original size. We will call this the **reduction factor** of the filter for that instance. The reduction factor can be viewed as a measure of how much the filter has reduced the search space for later processing stages. The mean reduction factor thus provides a crude measure of the usefulness of the filter.

Reducing the size of the search space is, of course, only useful if the search space ends up containing the correct result. Since the target set is defined as a set of coreferent elements, we hold that the search space contains the correct result provided it contains at least one element in the target set. So another useful measure in evaluating the effectiveness of a filter is the ratio of the number of cases in which the overlap set was non-empty to the total number of cases considered. We refer to this as the **overall accuracy** of the filter.

Table 2 summarises the overall accuracy and mean reduction factor for each of the four anaphoric heads we considered in this evaluation, measured at each level of generalisation of the associative axioms extracted from the corpus. What we would like our filtering to achieve is a low reduction factor (i.e., the selected set should be small) but a high overall accuracy (the filter should rarely remove an actual antecedent). As a baseline to evaluate against, we set the selected set to consist of the subjects of the previous sentences in the context, since these would seem to constitute reasonable guesses at the likely antecedent.

As can be seen, the synset lookup step (generalisation level 1) does not have a significant effect for any of the words. For all of the words there is a significant worsening in the reduction ratio after a single hypernym lookup: not surprisingly, as we generalise the axioms, their ability to filter out candidates decreases. This is accompanied by an increase in accuracy over the next two steps, indicating that the more specific axioms have a tendency to rule out the correct antecedents. This clearly highlights the trade-off between the two measures.

The last set of measures that we used are based on the precision and recall figures for each application of a filter to a set of candidate antecedents. The **single-case recall** is the ratio of the size of the overlap set to the size of the target set (i.e, how many real antecedents remain after filtering), while the **single-case precision** is the ratio of the size of the overlap set to the size of the selected set (i.e., what proportion of the selected set are real antecedents).

Table 3 shows the mean of the single-case precision and recall values, taken over all of the cases to which the filters were applied. As might be expected from the previous results, there is an obvious trade-off between precision and recall, with precision dropping sharply after a single level of hypernym lookup, and recall beginning to increase after one or two levels.

It is worth noting that with both sets of figures, there are substantial differences between the scores for each of the words. The filter performed best on *tip*, reasonably on *head* and *body*, and fairly poorly on *color*.

5 Conclusions and Further Work

Our intention in this paper has been to explore how we might automatically derive from a corpus a set of axioms that can be used in conjunction with an existing anaphor resolution mechanism; in particular, it is likely that in conjunction with an approach based on saliency, the axioms could serve as one additional factor to be included in computing the relative likelihood of competing antecedents.

The preliminary results presented above do not yet make a strong case for the usefulness of the technique presented in this paper. However, they do suggest a number of possibilities for further work. In particular, we have begun to consider the following.

First, we can make use of word sense disambiguation to reduce the negative consequences of generalising to synsets. Second, we intend to explore whether it is possible to determine an appropriate level of generalisation based on the class of the anaphor and antecedent. Third, there is scope for building on existing work on learning selectional preferences for WSD and the resolution of syntactic ambiguity; we suspect that, in particular, the work on learning class-to-class selectional preferences by (Agirre and Martinez, 2001) may be useful here.

We are also looking for better ways to assess the results of using the axioms. Two directions here are clear. First, so far we have only a relatively small number of hand-annotated examples, from a single source. Increasing the number of examples will let us investigate questions like whether different choices of parameters are appropriate to different classes of anaphor. Second, it should be possible to refine the evaluation metrics: it is likely that even without looking at the effect of different filters in the context of a particular anaphora resolution system, we could provide a more meaningful analysis of their probable impact.

In conclusion, we have shown in this paper how associative axioms can be derived automatically from a corpus, and we have explored how these axioms can be used to filter the set of candidate antecedents for instances of

		Level of generalisation						
Anaphor	measure	None	1	2	3	4	5	Baseline
color	reduction	0.15	0.15	0.42	0.64	0.71	0.74	0.08
	accuracy	0.63	0.63	0.63	0.74	0.79	0.79	0.37
body	reduction	0.14	0.17	0.63	0.76	0.79	0.79	0.07
	accuracy	0.57	0.58	0.79	0.88	0.91	0.91	0.45
head	reduction	0.14	0.15	0.54	0.72	0.80	0.80	0.07
	accuracy	0.49	0.49	0.66	0.84	0.88	0.89	0.49
tip	reduction	0.13	0.14	0.37	0.64	0.72	0.77	0.06
	accuracy	0.64	0.64	0.85	0.85	0.88	0.91	0.55

Table 2: Variation of reduction factor and accuracy with an increasing level of generalisation in the associative axioms used for filtering.

			Level of generalisation						
Anaphor	stat	initial	0	1	2	3	4	5	Baseline
color	precision	0.10	0.45	0.45	0.16	0.10	0.10	0.10	0.37
	recall	1.00	0.56	0.56	0.59	0.69	0.79	0.79	0.31
body	precision	0.10	0.37	0.32	0.12	0.11	0.11	0.11	0.47
	recall	1.00	0.44	0.46	0.71	0.83	0.87	0.87	0.33
head	precision	0.10	0.31	0.29	0.11	0.10	0.10	0.10	0.51
	recall	1.00	0.39	0.39	0.58	0.79	0.84	0.85	0.39
tip	precision	0.07	0.37	0.33	0.18	0.09	0.08	0.08	0.56
	recall	1.00	0.64	0.64	0.85	0.85	0.88	0.91	0.55

Table 3: Variation of precision and recall with an increasing level of generalisation in the associative axioms used for filtering.

associative anaphora. Our initial evaluation of the impact of using these filters suggests that they are of limited value; yet the intuition that generalisations of this kind should be useful remains strong, and so our next steps are to find ways of refining and improving the approach.

References

E. Agirre and D. Martinez. 2001. Learning class-to-class selectional preferences. In *Proceedings of the ACL CONLL Workshop. Toulouse, France.*

B. Boguraev and C. Kennedy. 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.

H. Clark and C. Marshall, 1981. *Definite reference and mutual knowledge*. Cambridge University Press, New York.

C. Fellbaum, editor. 1998. *WordNet*. MIT Press.

B. Grosz. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Ph.D. thesis, Stanford University.

J. Hawkins. 1978. *Definiteness and Indefiniteness: a study in reference and grammaticality prediction*. Croom Helm, London.

M. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*.

I. Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts at Amherst.

Alistair Knott and Robert Dale. 1995. Using linguistic phenomena to motivate a set of rhetorical relations. *Discourse Processes*, 18(1):35–62.

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

M. Poesio, R. Vieira, and S. Teufel. 1997. Resolving bridging references in unrestricted text. In *Proceedings of the ACL-97 Workshop on Operational Factors in Practical, Robust, Anaphora Resolution For Unrestricted Texts*.

E. Prince. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*, pages 223–256. Academic Press, New York.

C. Sidner. 1979. *Towards a computational theory of definite anaphora comprehension in English discourse*. Ph.D. thesis, MIT.

P. Tapanainen and T. Jarvinen. 1997. A non-projective dependency parser. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, pages 64–71. Association for Computational Linguistics.

R. Vieira. 1998. *Definite Description Processing in Unrestricted Text*. Ph.D. thesis, University of Edinburgh.