Using the WordNet Hierarchy for Associative Anaphora Resolution

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

Abstract

In this paper, we explore how the taxonomic inheritance hierarchy in a semantic net can contribute to the resolution of associative anaphoric expressions. We present the results of some preliminary experiments and discuss both their implications and the scope for improvements to the technique.

1 Introduction

Anaphor resolution is widely recognised as a key problem in natural language processing, and has correspondingly received a significant amount of attention in the literature. However, from a computational perspective, the primary focus of this work is the resolution of pronominal anaphora. There is significantly less work on full definite NP anaphora, and less still on what we will term here **associative anaphora**: that is, the phenomonen in which a definite referring expression is used to refer to an entity not previously mentioned in a text, but the existence of which can be inferred by virtue of some previously mentioned entity. Although these referring expressions have been widely discussed in the linguistics, psychology and philosophy literature, computational approaches are relatively rare (with a few notable exceptions, such as the work of (Poesio et al., 1997) and (Vieira, 1998).

A typical example from the literature is the use of the definite noun phrase reference in the second sentence in example (1):[1]

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

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

Here, the hearer is likely to infer that the driver referred to in the second sentence belongs to the bus mentioned in the first sentence. For our purposes, we consider *the driver* to be the **textual antecedent** of the anaphor, and the relationship between the referents of the anaphor and antecedent to be a part-of relationship. 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 process. Effectively, our goal is to extend the semantic network with information about pairs of senses that are 'associated' in a way that licenses possible associative anaphoric references. Our technique using involves unsupervised learning from a parsed corpus.

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, as noted above, computational approaches to resolving such anaphora are much less common. This is hardly surprising: given the almost limitless bounds on what can be associated with a previously mentioned entity, using knowledge-based approaches of the kind that were commonly discussed in earlier literature (see, for example, (Grosz, 1977; Sidner, 1979)) is a major undertaking, and probably unrealistic for practical broad coverage NLP tasks. On the other hand, the absence of surface level cues makes associative anaphora difficult to handle using the sort of shallow processing techniques that have become dominant over the last decade.

Our focus on the present paper is on those associative anaphors where there is a textual antecedent. The linguistic context provides us with a set of candidate antecedents, and our goal, for a given associative anaphor, is to identify the correct antecedent. Several antecedents may refer to the same entity; given an appropriate coreference resolution mechanism, this is non-problematic. Also, we are not concerned here with with determining the precise nature of the relationship that holds between the associative anaphor and its antecedent, although in most cases we consider this will be one of meronymy. All we require is the ability to be able to establish a connection between the entities mentioned in a text, effectively knitting the semantic fabric underlying the discourse.

As a way of moving towards this result, our motivating observation is a simple one, and one that has been explored in other areas (see, for example, (Hearst, 1992; Knott and Dale, 1992)): that semantic relationships which are left implicit for a reader to infer in some contexts may also occur explicitly in others, 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 the work described here, the corpus we are using consists of just over 2000 encyclopaedia articles drawn from the electronic versions of Grolier's Encyclopaedia and Microsoft's Encarta. All the articles used 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: for our purposes, it is sufficient to know that an association exists. As indicated, the possibility of such a relationship can be derived from a corpus. Our approach, then, is to mine a corpus for explicit **statements of association**, and to use this evidence

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 the entity in the preceding text with which this is associated. Suppose *the stingray* is one of a number of candidate antecedent NPs in the context. If the corpus contains expressions such as those italicised in (5a) and (5b), then we have prima facie evidence that the antecedent might be *the stingray*.

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. This could, for example, be used to infer that the expression *the tuna* is a possible antecedent for an associative anaphor *the gills*, as in example (6).

(6) *The tuna* has no respiratory mechanism to ensure the flow of water over *the gills*.

Our goal 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). 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; we find 17164 instances of the ⟨*NP of NP*⟩ construction over 11322 types, and 5662 instances of the ⟨*Genitive N*⟩ construction over 2133 types. 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:

(7) 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:[2]

(8) 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 Word-Net 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.

The second problem is that it is not always valid to assume that a property (or relationship) holds for all subtypes of a given type of entity just because it holds for a few; for example, although we know that okapis have legs, and okapis are a type of living organism, it would be incorrect to assume that trees (which are also living organisms) or snakes (which are also animals) have legs.

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: *body*, *color*, *head* and *tip*, as in the following examples:

(9) 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.[3] 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

---

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

[3]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 adjusted.

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 elements that are in the target set.

One measure of the effectiveness of the filters is 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 mean reduction factor provides a crude measure of the usefulness of the filter, since it reduces the search space for later processing stages.

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 the intersection of the target set and the selected set (henceforth 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 1 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 that are not in the target set 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 second set of measures that we used is 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 2 shows the mean of the single-case precision and recall values, along with the combined F-measure, 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. Although the F-measure indicates generally poor performance relative to the baseline, this is largely due to low precision, which would be improved by combining the semantic filter with other selection mechanisms, such as salience-based selection; this is the focus of current work.

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*.

Anaphor	measure	Level of generalisation						Baseline
		None	1	2	3	4	5	
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 1: Variation of reduction factor and accuracy with an increasing level of generalisation in the associative axioms used for filtering.

Anaphor	stat	initial	Level of generalisation						Baseline
			0	1	2	3	4	5	
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 2: Variation of precision and recall with an increasing level of generalisation in the associative axioms used for filtering.

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 in themselves 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 our current work, we have not explored the possibility of using information about associations that is explicitly encoded in existing machine-acessible ontologies. WordNet, for example, actually encodes meronym relationships. Our reason for not relying on this information in the first place was the limited set of relationships that were encoded, and the fact that associative relationships were encoded far less reliably than the hypernym relationship. However, it would be interesting to compare the results

that could be obtained by using the ontology as a source for associative axioms with those that could be achieved by automatically deriving axioms from the data.

Another direction we have not explored is the complementary information about anaphora resolution that derives from explicit statements of association: in line with the Gricean maxims, the author's decision to use an expression such as *the leg of the okapi* may constitute evidence that there is more than one previously mentioned entity in the context that may have legs. This information might be used, for example, to rule out an otherwise most likely antecedent.

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 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.*

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. 1992. Using linguistic phenomena to motivate a set of rhetorical relations. Technical Report HCRC/RP-39, Human Communication Research Centre, University of Edinburgh.

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.