Learning selectional preferences for use in resolving associative anaphora

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

Abstract

Anaphora resolution is well recognised as one of the more important — and most difficult — tasks in natural language processing. It has applications in a wide variety of areas machine translation, text summarisation, information extraction, and information retrieval. Most work in the area to date has focussed on cases involving pronouns (and certain full noun phrases) that are coreferent with other noun phrases in a text. This paper presents some preliminary results from experiments aimed at extending the coverage of an anaphora resolution system to deal with certain cases of *associative anaphora* — where the anaphor and the expression that allows it to be interpreted (the *antecedent*) do not refer to the same thing. The technique that has been used involves automatically acquiring semantic relationships from a parsed corpus, using the WordNet thesaurus as a resource to overcome the problem of data sparseness. The results in this paper extend previous work (Meyer and Dale, 2002) by examining the effect of applying word sense disambiguation when deriving the rules used to filter the set of candidate antecedents.

Keywords: associative anaphora, anaphora resolution, WordNet, selectional preferences

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 the phenomenon in which a definite noun phrase is used to refer to something that has not been previously mentioned in a text, but the existence of which can be inferred from that of some previously mentioned entity. 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 the context makes available for reference entities 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, and so the process of resolution involves identifying 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

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

paper, we explore how evidence derived from a corpus might be combined with a semantic hierarchy such as WordNet to assist in the resolution.

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 5 describes the current results of this exploration, and section 6 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 (Hawkins, 1978; Clark and Marshall, 1981; Prince, 1981; Heim, 1982, for example). However, computational approaches to resolving such anaphora are much less common.[2] This is hardly surprising: 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. On the other hand, using knowledge-based approaches of the kind that were commonly discussed in earlier literature (Grosz, 1977; Sidner, 1979, for example) is clearly problematic, given the almost limitless bounds on what can be associated with a previously mentioned entity. 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 NLP 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 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. 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. These are what we refer to as ASSOCIATIVE ANAPHORS. And third, the definite noun phrase may be exophoric: 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.

There are essentially two related questions we want to be able to answer: given a definite NP, is it an 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). Nor are we concerned in the current paper with determining the precise nature of the relationship between the associative anaphor and its antecedent. We focus instead on the second question above: given an 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 (Knott and Dale, 1992, for example): 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.

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

3 Acquiring Axioms

3.1 The Corpus

For our experiments, we have been working on a corpus of some 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. This corpus has some interesting characteristics which are not explored in the present work; for example, for many animals we have two different (and not always compatible) descriptions. Here, however, we focus at a lower level, examining the particular anaphoric relationships exhibited in the corpus. 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) These beetles are most often unmarked black or brown; several species have wing cases that are striped or bordered with metallic blue, green, or bronze. 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 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, 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 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 6. 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 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 (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 Method

The set of associative constructions that we used in deriving our set of associative axioms consisted of all genitive and of-NP constructions in which the head noun was included amongst those used in evaluation: namely, *head*, *body*, *color*, and *tip*. In the first experiment, reported in (Meyer and Dale, 2002), we simply used the base forms of words that the parser identified as the head of each associative construction and the head of the modifier as the elements in the lexical (or level-0) axioms that were entered into the knowledge base:

(7) has(0, head , stingray)

(8) has(0, head , bulldog)

(9) has(0, head , viper)

(10) has(0, head , snake)

The first level of DERIVED AXIOMS (level-1 axioms) were obtained by looking up the concepts (or *synsets*) associated with the modifier elements (or ANTECEDENT TEMPLATE) in each axiom:

(11) has(1, head , ⟨STINGRAY⟩)

(12) has(1, head , ⟨BULLDOG⟩)

(13) has(1, head , ⟨VIPER⟩)

(14) has(1, head , ⟨SNAKE_1, ...⟩)

(15) * has(1, head , ⟨SNAKE_2, ...⟩ [person])

(16) * has(1, head , ⟨SNAKE_3, ...⟩ [river])

(17) * has(1, head , ⟨SNAKE_4, ...⟩ [constellation])

(18) * has(1, head , ⟨SNAKE_5, ...⟩ [tool])

This first stage represents a generalisation when different words are used to refer to the same concept (such as *English bulldog* and *bulldog*). In cases where the element being generalised had more than one sense (as with *snake*), all senses were used. The starred examples represent inappropriate generalisations generated using this approach.

Further levels of derived axioms were obtained using the hypernym relationships encoded in WordNet [3]:

(19) has(2, head , ⟨RAY⟩)

(20) has(2, head , ⟨WORKING DOG⟩)

(21) has(2, head , ⟨DIAPSID REPTILE⟩)

(22) * has(2, head , ⟨BAD PERSON⟩)

(23) * has(2, head , ⟨RIVER⟩)

(24) * has(2, head , ⟨CONSTELLATION⟩)

(25) * has(2, head , ⟨HAND TOOL⟩)

The generalisation for 13 is 14, and so no generalisation for this axiom is added at level 2.

The first experiment yielded relatively poor results, that represented no significant improvement over a simple baseline that selected as an

[3] The level associated with a given axiom was defined to be the *least* number of generalisation steps required to derive the axiom from a lexical axiom

antecedent the most recent sentential subject. This is clearly shown by the comparisons in section 5; in particular, due to a significant loss of precision after the synset lookup stage, the best results were obtained when no generalisation was performed. Since data sparseness is an issue for our application, this poses a problem.

We identified lexical ambiguity as one of the most likely source of error. Many words occurring in the associative constructions that we extracted from the corpus have alternative senses relating to humans or human activity. For example, as well as its primary sense, *snake* has senses describing a type of person, a tool used in plumbing, a river, and a constellation (the latter two being proper nouns). If we generalise these additional senses, our filter will allow tools and people as things that could have heads[4]. The problem with this is not principally that it allows people and tools through the filter, but that other words have similar metaphorically extended senses. For instance, *plant* also has a sense that denotes a type of person, and so *plants* would be allowed as a candidate antecedent for *the head*. To overcome the problems caused by the assignment of inappropriate senses to elements in associative axioms, we tried two things: (1) allowing collocations such as *carpet snake* that occur in WordNet to appear in lexical axioms, and (2) manually disambiguating the axioms by editing the set of level-1 axioms before further generalisation.

5 Results

Table 1 compares the results with and without word sense disambiguation across a number of different levels of generalisation. As can be seen, the results obtained using the disambiguated axioms show a relatively consistent improvement over those obtained without disambiguation. The improvement is particularly notable at around two levels of generalisation, where, for all words but *tip*, the F-measure actually improves.

Table 2 shows the increase in the number of relevant derived axioms at each level of expansion. The number of "explicit" axioms corresponds to the total number of lexical axioms, plus the number of axioms that can be derived through n steps of generalisation; the "implied" figure is the total number of axioms that can be derived by taking the closure of the explicit set under hyponym lookup. The latter figure provides a true indication of the number of word-senses that will be allowed through the filter.'

Given that there are 66,025 synsets in the version of WordNet that we have been using, the derived axioms with 5 levels of expansion permit on average 62% of the senses in WordNet for the disambiguated case; this includes all of the descendants of ⟨ENTITY, SOMETHING⟩ , a top-level synset, but not ⟨ABSTRACTION⟩ . This explains why the filter becomes ineffective after the first two levels of generalisation, even with disambiguation.

6 Discussion

Although the results from our first experiment in using axioms automatically derived from a parsed corpus to filter the set of candidate antecedents when resolving an associative anaphor were disappointing, disambiguating the axioms and taking account of collocations result in a major improvement. With these minor modification the technique improves on the simple baseline in which the antecedent is taken to be the preceding sentential subject, and generalising the associative axioms leads, as expected, to an improvement in performance.

Our results support the impression that the technique is more effective in cases where there is a stronger association between the class of the anaphor and that of the antecedent, as is the case with body parts (which are generally associated with animals), but not properties such as colour or more abstract parts such as tip.

However, it should be stressed that the results are based on a quite limited set of examples. This is due to the relatively small amount of the corpus that has been annotated so far. We are currently working on increasing the size of the annotated portion of our corpus so that we can

[4] Most people and some tools do have heads, but similar overlaps occur with other body parts like *wings*

Table 1: The change in precision (P), recall (R), and F-measure (F) with increasing levels generalisation, both with and without WSD (base = previous subject)

				Level of generalisation						
Anaph.	WSD	stat	rough	0	1	2	3	4	5	base
body	N	P	0.10	0.37	0.32	0.12	0.11	0.11	0.11	0.47
		R	1.00	0.44	0.46	0.71	0.83	0.87	0.87	0.33
		F	0.18	0.40	0.38	0.21	0.19	0.20	0.20	0.39
body	Y	P	0.10	0.40	0.40	0.28	0.13	0.13	0.13	0.47
		R	1.00	0.43	0.43	0.91	0.94	0.95	0.95	0.33
		F	0.18	0.41	0.41	0.43	0.23	0.23	0.23	0.39
head	N	P	0.10	0.31	0.29	0.11	0.10	0.10	0.10	0.51
		R	1.00	0.39	0.39	0.58	0.79	0.84	0.85	0.39
		F	0.18	0.35	0.33	0.18	0.18	0.18	0.18	0.44
head	Y	P	0.10	0.38	0.38	0.38	0.28	0.13	0.13	0.51
		R	1.00	0.38	0.38	0.88	0.91	0.94	0.96	0.39
		F	0.18	0.38	0.38	0.53	0.43	0.23	0.23	0.44
color	N	P	0.10	0.45	0.45	0.16	0.10	0.10	0.10	0.37
		R	1.00	0.56	0.56	0.59	0.69	0.79	0.79	0.31
		F	0.18	0.50	0.50	0.25	0.17	0.18	0.18	0.34
color	Y	P	0.10	0.41	0.41	0.32	0.17	0.10	0.10	0.37
		R	1.00	0.40	0.40	0.58	0.66	0.71	0.71	0.31
		F	0.18	0.40	0.40	0.41	0.27	0.18	0.18	0.34
tip	N	P	0.07	0.37	0.33	0.18	0.09	0.08	0.08	0.56
		R	1.00	0.64	0.64	0.85	0.85	0.88	0.91	0.55
		F	0.13	0.47	0.44	0.30	0.16	0.15	0.15	0.55
tip	Y	P	0.07	0.37	0.37	0.13	0.11	0.09	0.09	0.56
		R	1.00	0.64	0.64	0.85	0.85	0.85	0.85	0.55
		F	0.13	0.47	0.47	0.23	0.19	0.16	0.16	0.55

Table 2: The change in the total number of axioms with increasing generalisation after WSD has been applied

WSD		Lex.	L1	%	L2	%	L3	%	L4	%	L5	%
N	Explicit	161	624		1,033		1,364		1,571		1,665	
N	Implied	161	26,802		123,812		180,337		216,969		230,958	
Y	Explicit	161	294	47	343	33	366	23	370	22	371	
Y	Implied	161	28,350	100	97,499	88	160,104	74	162,725	70	162,775	

perform more comprehensive evaluations in the future. We also intend to determine whether it is possible for us to make use of the annotated corpora used in some recently published work on associative anaphora (**?**).

The ultimate purpose for the filtering mechanism described in this paper is to improve the performance of an anaphora resolution system by reducing the set of possible candidates. We have implemented such a system based on (Lappin and Leass, 1994), which is currently restricted to resolving pronouns that are coreferent with NPs in the same or preceding sentences, and detecting repeated mention for certain classes of full NP. We intend to extend this to handle associative anaphors. The filtering mechanism described in this paper will determine possible associative relationships involving noun phrases that are identified by a set of domain-independent heuristics as possible associative anaphors. The use of a salience ranking is expected to improve precision considerably. It is also our intention to investigate the effects of taking associative reference into account when determining the salience of a discourse entity.

We have been considering a number of possible methods for improving the accuracy and coverage of our filtering technique, including the following:

- Applying automatic WSD to both associative constructions and candidate antecedents
- Automatically determining how far to generalise lexical axioms
- Making use of other sources of information in generating associative axioms
- Resolving pronominal anaphors before constructing associative axioms

A variety of standard techniques could be adapted to our purposes in the first two cases.

Manual or assisted manual disambiguation of the associative axioms may be feasible, but it still represents a significant amount of human effort when porting an anaphora resolution system to a new domain. This is something that we would prefer to avoid, as the basis of our approach so far has been to take advantage of methods that require minimal human intervention. So an obvious extension of the work discussed in this paper is to automate the process of disambiguating the lexical axioms. Automatic WSD applied to the candidates could also be of benefit, and standard WSD techniques should prove as applicable to these as to the elements in associative axioms. Take the noun *gray* as an example; the WordNet synsets corresponding to this word include (1) the colour grey, (2) grey horses, (3) grey clothing, and (4) organisations whose members conventionally dress in grey. The first sense is the most common in the corpus, although the second also occurs in the entry describing horses. However, taking all possible senses for a word results in *gray* being taken as a possible antecedent for any anaphoric body part expression in the corpus, because it can also refer to a type of animal. In most contexts this is inappropriate. Whether automatic WSD will perform well enough to improve results remains in question, but it is certainly something that is worth testing.

The approach of generalising to a fixed number of steps from lexical axioms is flawed. There are two main reasons for this: (1) lexical items do not all map onto the same level in the WordNet ontology, and (2) the number of levels between two concepts in the hierarchy is not necessarily indicative of the degree of generalisation. There is a significant body of work on learning selectional preferences for verbs using WordNet and similar resources (Resnik, 1993; Abe and Li, 1998; Clark and Weir, 2001, for example)

The use of pronouns in lexical axioms is something that works reasonably well for our corpus, but probably would not work as effectively with texts from other domains. What we are doing when we allow lexical axioms including pronouns is effectively trading on the fact that in a particular context there are a limited number of types of entity that are likely to be pronominalised. When describing the physical appearance or physiology of animals, it is likely that a singular third person pronoun will refer to some kind of animal. It would be worthwhile investigating whether having pronouns in lexical axioms is beneficial in other types of corpus, and also how resolving pronouns prior to constructing axioms affects the performance of the filter in the domain of animal descriptions.

References

N. Abe and H. Li. 1998. Generalizing case frames using a thesaurus and the mdl principle. *Computational Linguistics*, 24(2):217–244.

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*, pages 10–63. Cambridge University Press, New York.

S. Clark and D. Weir. 2001. Class-based probability estimation using a semantic hierarchy. In *Proceedings of the 2nd Conference of the North American Chapter of the ACL.*

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.

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

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

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

J. Meyer and R. Dale. 2002. Mining a corpus to support associative anaphora resolution. In *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium*, September.

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.

P. Resnik. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships.* Ph.D. thesis, University of Pennsylvania.

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