Improving anaphora resolution by identifying animate entities in texts

Richard Evans, Constantin Orasan

School of Humanities, Languages and Social Sciences,
University of Wolverhampton, Stafford Street,
Wolverhampton, UK.
{in6087, in6093}@wlv.ac.uk

Abstract

Some references to human beings can be identified in English texts using named entity (Chinchor, 1997) and pronoun recognition but in some genres this still leaves a large number of references to people unidentified. The remaining noun phrases have no overt marking as to their *animacy* and clues as to the appropriate classification of a NP as *animate* or *inanimate* are scarce in the surrounding textual material. In English anaphora resolution, recognition of the animacy of NPs improves the accuracy with which gender agreement restrictions can be enforced between pronouns and candidates.

In this work, the animacy of English NPs is identified using a combination of a number of tactics. The main one is the use of WordNet. Three noun hierarchies were identified as being indicative of *animate entities*. The remaining hierarchies are taken to indicate nouns referring to *inanimate entities*. Four hierarchies of verb senses were identified as containing verbs expected to require an animate subject. By examining the distribution of textual head nouns and main verbs in WordNet, the animacy of the NP or subject NP is assessed. A number of heuristics are used to reinforce or undermine the system's confidence as to the animacy of a NP. The new system is incorporated into an existing pronominal anaphora resolution system and tested on a text with a high proportion of gender-marked pronouns. The performance of the system using the new method for animate entity recognition is compared with that of the original.

1. Defining the problem

Most approaches to pronominal anaphora resolution rely on compatibility of the agreement features between pronouns and antecedents. Although, as noted in (Barlow, 1998), this assumption does not always apply, it is reliable in enough cases to be of great practical value in anaphora resolution systems. Such systems rely on knowledge about the number and gender of noun phrase (NP) candidates in order to check the compatibility between pronouns and candidates (e.g. Hobbs, 1976; Lappin and Leass, 1994; Nasukawa, 1994; Kennedy and Boguraev, 1996; Mitkov, 1998). None of the algorithms proposed by these researchers include a method for actually identifying the instantiated values of the NP agreement features. Information as to the number feature of a NP's referent is usually available to systems as a result of the pre-processing phase. Identification of NP gender is not so trivial, though numerous researchers (Hale and Charniak, 1998; Denber, 1998; Cardie and Wagstaff, 1999) have proposed automatic methods for identifying the potential gender of NPs' referents. An additional paper reporting the use of the system developed in (Hale and Charniak, 1998) was (Ge et al., 1998).

In the cases where an algorithm applies preferences to a set of candidates for antecedent of a pronoun, minimizing the size of the set makes a great contribution to the effectiveness of a system. As noted above, when dealing with pronominal anaphora in English, the identification of number information is easy because the majority of pre-processing programs (part of speech taggers and parsers) associate number information with the nouns that are tagged. The reason for this is that enforcing number agreement between constituents increases the accuracy of these tools. By contrast, most English sentence analysis methods gain nothing in performance by incorporating gender information into the algorithm. In fact, it has been observed that incorporation of gender information into the tag set reduces the accuracy of part of speech taggers for English.

Given that manual annotation is an expensive procedure, and different researchers consider that a corpus containing gender information brings no additional benefit, most of the corpora and resulting pre-processing software for English do not associate gender features with nominal expressions. This omission is seen to be of little consequence in the syntactic analysis of sentences, but in the field of pronominal anaphora resolution, the enforcement of number agreement alone is not sufficient as a means of producing suitable sets of candidates for pronouns. This means that other sources must be used in order to derive the information required for gender agreement.

In (Cobuild, 1995), it is written that "*Something that is **animate** has life…*" We therefore use the term *animate* to describe entities in texts that may also be referred to using gender-marked pronouns. This set of entities includes people and animals. One step toward the identification of NP gender is the identification of reference to animate entities by NPs in a text. There are a number of well-established tactics that may be applied in order to do this. The most reliable clue that an animate entity is being referred to is the use of gender-marked pronouns such as *he*, *him*, *his*, *himself*, *she*, *her*, *hers,* and *herself.*

Another tactic is the performance of *named entity recognition*, which has been tackled by numerous researchers, including (Wakao et al., 1996; Mikheev et al., 1999) in the *Message Understanding Conferences* (MUC) (Chinchor, 1997). The aim of named entity recognition is to correctly classify sequences of capitalised words or potentially mixed sequences of capitalised and non-capitalised words as organisations, locations, or persons. Evaluation has shown that the accuracy with which systems can perform named entity recognition is above the level of 90% for precision and recall (Mikheev et al., 1999), but only for restricted domains. However, sole reliance on such clues and systems may still, in some genres leave a large proportion of references to animate entities unidentified.

Named entity recognition must be coupled with identification of animate reference using NPs headed by a common noun[1] in order to provide sufficient recall in the recognition of references to animate entities. In English, NPs and their head nouns often have no overt gender marking and clues as to their animacy are scarce in the surrounding textual material. The difficulty of the animate entity recognition task with respect to common NPs has led to the proposal of a new identification system in this paper.

1.1. Previous work

The automatic recognition of noun phrase gender on the basis of statistical information has been attempted before by (Hale and Charniak, 1998). The method operates by counting the frequency with which a NP is identified as the antecedent of a gender-marked pronoun by a simplistic pronoun resolution system. In that approach, "the probability that a referent is a particular gender class is just the relative frequency with which that referent is referred-to by a pronoun which is part of that gender class." It is reported that by using the syntactic Hobbs algorithm (Hobbs, 1976) for pronoun resolution, the method was able to assign the correct gender to proper names in a text with 68.15% precision, though the method was not evaluated with respect to the recognition of gender in common NPs. Further, in relation to the task of identifying the gender of named entities, it is apparent that the method will be ineffective unless operating over texts of a considerable size, containing a relatively large number of pronouns belonging to different gender classes. Generally, it is noted that the number of pronouns found in texts is highly variable from text to text and genre to genre and therefore, the method may not be usable in particular genres or texts.

For many purposes, clues restricted only to the proximity of pronouns will not offer sufficient coverage. In addition, general NPs referring to persons or groups of persons are not always found in the proximity of gender-marked pronouns, as in

there are ***two AI delegates*** *and one of* ***them****…*

once ***a person*** *registers for an interview* ***they*** *are removed from the unemployment figures*… (Barlow, 1998),

where the pronouns lack specific information as to the gender of their referents.

Automatic pronoun resolution systems can also be somewhat unreliable and there must be doubts about the validity of systems that are based on statistics derived from sources as unreliable as they are. In our system, we intend to use the recognition of NP gender or animacy in order to make anaphora resolution more effective. By contrast, the approach of (Hale and Charniak, 1998) seems to make the basic assumption that automatic pronoun resolution is already highly effective, that there is a high level of confidence in statistics derived from these systems, and the system can be used to recognise NP gender. The method may be quite suitable in the field of anaphora resolution but it does not provide a general solution to the problem of recognizing references to animate entities in texts. It is regarded by the present authors as more useful to have a system capable of operation over any text, regardless of the amount of pronoun usage in it. To this end, we formulated a general-purpose method that will operate equally effectively, regardless of the size of and distribution of pronouns in a text. In fact, we chose to test the system on a text containing a relatively high proportion of gender-marked pronouns, but this was only in order to provide a good basis for evaluation and to investigate the contribution of animate entity recognition to the process of anaphora resolution.

In the paper by (Denber, 1998), WordNet was used to determine the animacy of nouns and associate them with gender-marked pronouns. The details presented in (Denber, 1998) are sparse and it is difficult to judge the effectiveness with which WordNet was used in that system. It seems that there, nouns were the only elements of interest and WordNet was being used as a simple gazetteer in that method. In the paper, Denber notes that the large number of senses associated with each noun in the WordNet hierarchy means that it is problematic to treat it as a simple gazetteer. In the report, (Denber, 1998) does not evaluate the accuracy with which his system recognises animate entities, nor does he assess the contribution brought by animate entity recognition to anaphora resolution in English. (Cardie and Wagstaff, 1999) combined the use of WordNet with proper name gazetteers in order to obtain information on the compatibility of coreferential NPs in their clustering algorithm. As in the work of (Denber, 1998), the accuracy with which the method was able to identify NP animacy was not assessed.

In most cases, systems have consulted domain specific gazetteers or databases in order to obtain gender information about NPs in texts. This solution suffers from the drawbacks that the construction of such resources is labour-intensive, may require expert domain knowledge, and must be maintained in order to prevent the information from becoming outdated. For example, as times change, *bin men* become *refuse operatives* and *debt collectors* become *debt factors*. In general, it is desirable to minimize or even eliminate the amount of human intervention necessary to bring a system to an effective operating capacity.

In this paper, a method is proposed that is based on the combination of Conexor's FDG Parser (Tapanainen, 1997); WordNet (Fellbaum, 1998); a first name gazetteer; and a small set of heuristic rules for identifying the animate entities in English texts. As will be discussed in Section 2, it is not considered necessary at this point to perform specific gender identification (i.e. masculine, feminine, neuter) for NPs using this method. Section 3 reports empirical evidence that shows the contribution of animate entity recognition and agreement constraints in pronominal anaphora resolution. In Section 4, WordNet is

[1] Hereafter referred to as common NPs

presented as a resource for recognition of references to animate entities in texts. In Section 5, the algorithm for animate entity recognition is proposed. In Section 6, it is shown that the automatic identification of animate entities in texts is sufficient to induce increased performance in pronoun resolution. In Section 7, conclusions are drawn and future work considered.

2. Specific gender vs. animate entity recognition

It is noted that there are very few surface clues as to the gender of NPs in English with the exception of third person pronouns, which are exceptional. In the past, it may have been possible to associate masculine and feminine gender with stereotypical occupations and the suffixes that have appeared in words like *manageress*, or *fireman*. These terms are no longer reliable in present usage as the current trend is to avoid using the feminine suffix *–ess* in most occupation names and the suffix *–man* is often used rather than *–woman* or *-person* regardless of the gender of the person being described, as for example, in *chairman*.

Although it should be noted that the binding of noun phrases by gender marked reflexives and the proximity of gender marked possessive pronouns are reliable clues as to the specific gender of noun phrases in the text, these clues are relatively infrequent. In addition, there are a variety of anaphoric phenomena that motivate the development of a system that can recognise the general animacy of NP referents rather than associating them with specific gender information (masculine, feminine, neuter).

Anaphorically linked entities are usually considered coreferential and coreference has been defined as a transitive relation (Vilain et al., 1995), but this view is weakened slightly by constructions such as

the user of the machine *should select* ***his*** *or* ***her*** *own settings*

in which two pronouns with different gender information are both anaphorically linked to the same NP and are therefore considered coreferential. In this case, if a system identifies the NP *the user of the machine* as a person and goes on to assign specific gender information to it, enforcement of gender agreement constraints will mean that one of the pronouns *his* or *her* cannot be resolved correctly. A more suitable alternative is to assign both masculine (*masc*) and feminine (*fem*) gender information to the agreement features of the NP and thus ensure its compatibility with all pronouns that may refer to people. These constructs motivate the distinction between animate and inanimate, but not between neuter, masculine and feminine.

The solution presented here addresses a relatively frequent phenomenon in which an expression is capable of describing people in general, regardless of their specific gender (as in the previous example). It also addresses some of the difficulties mentioned previously with respect to identifying the gender of referents of general occupation names. Incidentally, (Hirschman et al., 1998) circumnavigated the problem caused by '*his* / *her*' constructions by proposing that elements representing intensional mentions are *grounded* by elements reflecting extensional mentions, cutting the transitivity of the relation at the grounding point. In these constructions, each intensional pronoun is grounded by the NP antecedent and forms a different coreferential chain.

The foregoing argument reflects the view that a method based on surface clues and word definitions will be incapable of satisfactorily disambiguating the specific gender of animate entities in texts. Despite this, the recognition of references to animate entities is less problematic and, as shown in Sections 3 and 6, is useful in coreference and anaphora resolution.

It should be noted however that there are a number of genres and registers in which the recognition of references to animate entities is of limited use and will not assist in the task of anaphora resolution. For instance, texts from certain domains, such as scientific papers or technical documents, might not refer to entities with natural gender. (e.g. *"when the cashiers need to get their sub-totals, they can exchange information …"* from a technical manual describing Beowulf supercomputers in which the *cashiers* are parts of the computer system).

Alternatively, when there are a large proportion of entities with natural gender in a text and very few neuter pronouns, the binary distinction between animate and inanimate NPs will not be useful. In those cases, finer granularity would be required. Further, the recognition of either general or specific gender in NPs only assists in the resolution of pronouns, possessives and reflexives with singular number.

3. Identification of animate references makes antecedent selection easier

Given that many anaphora resolution systems, including our own, operate by assessing the likelihood that each of a set of candidate NPs are the antecedent of the current pronoun, it is in line with intuition that as long as the correct antecedent remains in the set of candidates, the smaller the set, the more likely a system is to identify the correct antecedent from it. Application of the gender agreement constraint is an effective way to minimize the cardinality of the candidate set.

In the case of preferential anaphora resolution where for each pronoun, a system must select the antecedent from a set of candidates, the enforcement of agreement constraints between pronouns and antecedents helps to maximise the ratio of antecedents to candidates for each pronoun. This process increases the algorithm's chances of making a correct selection. In processing the English language, the agreement constraint is normally enforced along two axes, number and gender. For the purpose of developing and evaluating a system for recognition of animate references, we used a text with the characteristics shown in Table 1.

Characteristic	Number
#Words	15839
#Pronouns	409
#Gender-marked Pronouns	199
#Noun phrases	5447
%Noun phrases referring to animate entities	30.1

Table 1: Characteristics of the text

The material itself was taken from the BNC corpus (Burnard et al., 1995), consisting of reports by the human-rights organisation, Amnesty International. The register is somewhat rhetorical/political. 69.9% of the noun phrases in the corpus refer to inanimate entities with only 30.1% referring to animate ones. By contrast, 199 (48.65%) pronouns require animate reference antecedents whereas 210 (51.35%) pronouns do not. It is therefore clear that we can improve the chances of an anaphora resolution system based on random chance and compatibility between pronoun and candidate agreement features if noun phrases are classified according to their animacy.

In order to assess the effectiveness of the system for animate entity recognition, it was used as a pre-processing step in the operation of MARS, the anaphora resolution system developed at the University of Wolverhampton. MARS is based on the proposal of (Mitkov, 1998), but differs in certain respects from the original algorithm. In both cases, each pronoun in a text is associated with a set of candidate NPs that are required to agree with the pronoun in terms of number and gender. Candidates are extracted from the sentence in which the pronoun itself appears, from the two previous sentences up to the limit of a paragraph boundary, and also from the current section heading. The algorithm then applies preferential and non-preferential indicators that assign positive and negative scores to a pronoun's candidates. In each case, the highest scoring candidate is selected as the antecedent of the pronoun. The algorithm is fully described in (Mitkov, 1998) and the differences between MARS and the original statement of the algorithm are described in (Orasan et al., 2000).

In this system, a pronoun is compatible with a candidate if its agreement feature information matches or partially matches the agreement feature information of the candidate. In this way, a pronoun marked with masculine (*masc*) gender is compatible with candidates that contain this information (*masc*) or more general information such as both masculine and feminine (*mascfem*), masculine and neuter (*mascneut*), or masculine, feminine and neuter (*mascfemneut*).

Table 2 shows how agreement constraints facilitate MARS's selection of antecedents in the selected text.

As stated in (Mitkov, 1998) and reinforced by the evaluation in (Orasan and Evans, 2000), the pronoun resolution algorithm was developed on texts from the domain of technical manuals and its effectiveness was shown to be somewhat genre dependent. The text used to develop our system for recognition of animate references belongs to the political or rhetorical domain, rather than the domain of technical manuals and MARS's effectiveness is relatively poor over this text. The

MARS	Default (Gazetteer)	No Number Agreement	No Gender Agreement	No Number or Gender Agreement	With Automatic Animate Entity Recognition	With Perfect Animate Entity Recognition
#Candidates	3228	3394	3328	4873	2196	2154
#Antecedents	708	706	708	737	677	703
Candidates / Pronoun	7.89	8.3	8.14	11.91	5.37	5.27
Antecedents / Pronoun	1.73	1.73	1.73	1.8	1.65	1.72
Antecedents / Candidates	0.22	0.21	0.21	0.15	0.31	0.33
Antecedents / Candidates per Pronoun	0.3	0.29	0.29	0.19	0.4	0.42
#Pronouns Correctly Resolved	203	190	196	172	224	224
#Gender-Marked Pronouns Correctly Resolved	115	110	108	98	125	133

Table 2: Statistics showing how agreement constraints facilitate antecedent selection

algorithm used by MARS applies preferential indicators that exploit textual features common in the domain of technical manuals. Over the text used here, the indicators are not found to be so reliable. In its default mode, using only a first name gazetteer as the source of information to enforce gender agreement, MARS successfully resolves 203 pronouns, a success rate of 49.63%.

The different columns in Table 2 show how MARS's performance varies according to the degree to which agreement constraints are enforced. The information in the rows of the table is as follows:

- *#Candidates* shows the number of NPs considered as possible candidates for pronoun antecedents in the text.
- *#Antecedents* gives the number of NPs that can be considered valid antecedents in the sets of pronoun candidates once the relevant agreement restrictions have applied. The apparent anomaly in the table, whereby a lower number of antecedents is available *With Perfect Animate Entity Recognition* than for other cases, results from the occurrence of NP antecedents like *the government* which can be viewed as a collection of animate entities, but are referred to using neuter pronouns.
- *Candidates/Pronoun* is the mean number of candidates for the pronouns in the text once agreement constraints have applied. As mentioned earlier, the goal is to reduce this value as much as possible in order to obtain better performance from the anaphora resolution system.
- *Antecedents/Pronoun* is the mean number of valid antecedents remaining in pronouns' sets of candidates once agreement restrictions have applied. This measure can be used to assess how correctly the methods eliminate unsuitable candidates.
- *Antecedents/Candidates* is the ratio of valid antecedents to candidates available to MARS once agreement restrictions have applied. This can be regarded as the probability that random selection would choose the correct antecedents for pronouns.
- *Antecedents/Candidates per Pronoun* is the mean ratio per pronoun of valid antecedents to candidates in the sets of pronouns' candidates after agreement restrictions have applied. This is a more refined version of the previous ratio.
- *#Pronouns Correctly Resolved* shows the total number of pronouns in the text for which MARS is able to select a valid antecedent once agreement restrictions have applied. For some of these correctly resolved pronouns, the gender agreement constraint has no effect (as in the case of plural pronouns), therefore this value is not a wholly accurate measure of how useful it is to identify NP animacy in anaphora resolution.
- *#Gender-Marked Pronouns Correctly Resolved* is the total number of gender-marked pronouns for which MARS is able to select a valid antecedent once agreement restrictions have applied. This value indicates how useful it is to identify NP gender for resolution of gender-marked pronouns.

The columns of the same table indicate the following:

- *Default (Gazetteer)*, MARS is run in its original configuration where the only source of gender information comes from first name gazetteers for masculine and feminine names. NPs starting with such names are assigned the appropriate gender. The presence of a name word in both masculine and feminine gazetteers means that the NP is assigned the gender *mascfem* and becomes a suitable candidate for all but neuter pronouns.
- *No Number Agreement* shows the performance of MARS when perfect person recognition is applied, but the number agreement restriction between pronouns and NPs is dropped, so that for example singular pronouns may have plural NP candidates.
- *No Gender Agreement* shows MARS's performance when no gender restriction of any kind applies to pronouns and candidates. In this mode, MARS does not even have access to information from the gazetteers mentioned earlier. In this mode, the average ratio across pronouns of valid antecedents to candidates is at the level of 0.29, identical to the ratio when *No Number Agreement* is applied.
- *No Number or Gender Agreement* shows MARS's performance when no agreement restrictions of any kind are applied to pronouns and their candidates. Here, it can be seen that the average ratio across pronouns of suitable antecedents to candidates is at its lowest level.
- *With Automatic Animate Entity Recognition* shows MARS's performance when the system based on the extended algorithm described in Section 5 is incorporated to allow recognition of animate reference in the text. In this mode, NPs determined by the algorithm to be referring to animate entities are assigned the gender *mascfem*.
- *With Perfect Animate Entity Recognition* shows MARS's performance when the system has perfect knowledge from the key about which NPs are referring to animate entities in the text. Here, the average ratio across pronouns of valid antecedents to candidates is at its highest level, 0.42. Inspection of Table 2 shows that this ratio is correlated with the resulting effectiveness of the pronoun resolution system.

This data provides empirical evidence that the enforcement of agreement constraints increases a system's chances of selecting an antecedent from a set of candidates. The effectiveness of the implemented system for animate entity recognition is further described in Section 6.

4. WordNet as a resource for recognition of animate entities

WordNet (Fellbaum, 1998) is an electronic lexical resource organized hierarchically by relations between sets of synonyms or near-synonyms called synsets. Each of the four primary classes of content-words, nouns, verbs, adjectives and adverbs are arranged under a small set of so-called *unique beginners*. In the case of nouns and verbs, which are the concern of the present paper, the unique beginners are the most general concepts under which the entire set of entries is organized on the basis of hyponymy and entailment.

It was noticed that several hierarchies were of interest with respect to the aim of identifying the animate entities in texts. Firstly, in the case of nouns, three of the hierarchies are expected to contain nouns that refer to entities that may also be referred to using gender-marked pronouns. The unique beginners in these cases were *animal*, reference number (05), *person* (18), and *relation*, (24). Secondly, there are four verb sense hierarchies that allow the inference to be made that their subject NPs must be referring to people. The unique beginners in these cases are *cognition* (31), *communication* (32), *emotion* (37) and *social* (41). It was clear that a lexical resource arranged in such a hierarchical fashion could be exploited in order to associate the heads of noun phrases with a numerical measure of confidence that the associated NP has either an animate or inanimate referent.

Further motivation for the use of WordNet arises from several criteria. Firstly, as mentioned previously, knowledge as to the animacy of common NPs cannot be readily computed from explicit features of the text. Unlike the situation with proper names and their associated clues such as titles and initial capitalisation, knowledge as to the animacy of common NPs appears to be purely implicit. Recognition of animate entities must, at some point be grounded in world-knowledge. Embodying this knowledge to a certain extent, WordNet allows a system to recognise references to animate entities that are not proper names. Incorporating information from verbs has the advantage that whereas a text is quite likely to contain many NP references in which the head noun does not appear in the WordNet ontology, this situation is less likely to arise in the case of verbs. By associating subject NPs that contain an unknown head with the verbs of which they are subject, the system is more likely to correctly identify them as references to animate or inanimate entities.

In the present system, the text is analysed using Conexor's FDG Parser (Tapanainen and Jarvinen, 1997). This parser provides information on the parts of speech of and the functional dependency relations between words in a text. Using this information, it is possible to identify complex NPs by grouping words dependent on noun heads and to associate all NPs with the appropriate head. The FDG Parser also returns the syntactic roles, such as subject or object, of the analysed words, as well as their lemmas. By tracing the dependency relations between nouns and verbs it is possible to obtain the subject of a given verb automatically.

In the next section, the algorithm is described, as is the method of synthesising a decision as to the animacy of NPs from this information.

5. Combining information from WordNet with heuristics for recognition of references to people

Having obtained the global set of NPs in a text, the algorithm for animate entity recognition is put into effect. In the first step of the algorithm, some surface clues pointing to the likely animacy of a NP are obtained. During this phase, the lemma of the head of the NP is determined and in the case of subject NPs, the lemma of the verb that they are subject of is extracted.

In the next step, WordNet is consulted in order to obtain more information about the NPs in the text. With respect to the lemmas of the head noun, WordNet's entries for the relevant noun are examined and counts are obtained of the number of senses of the noun that belong to any of the three hierarchies mentioned in Section 4 (#NAS) and the number of senses that belong to other hierarchies (#NNAS). In the cases of subject NPs, similar counts are obtained with respect to the verbs whose subject they are. The number of senses of the verb that belong under the unique beginners mentioned in Section 4 and taken to require an animate subject (#VAS) is obtained, as is the number of senses of the verb that belong under other unique beginners (#VNAS).

Two ratios are then computed that will be referred to as *noun animacy* (NA) and *noun non-animacy* (NN). These are defined in the equations below.

$$NA = \frac{\#NAS}{\#NAS + \#NNAS}$$

$$NN = \frac{\#NNAS}{\#NAS + \#NNAS}$$

Similar counts and ratios are then computed for the verbs of NPs identified as subject NPs. These measures will be referred to as *verb subject animacy* (VSA) and *verb subject non-animacy* (VSN).

In the next step, some additional low-level information is obtained. Namely, the questions of whether or not the NP contains a reflexive, as in *the Secretary General himself*, whether or not it contains the complementiser *who*, as in *a prosecutor who may well specialize in capital law*, and whether or not the NP is itself a gender-marked pronoun.

The final decision as to the personhood of a NP is then made according to one of two methods, the first, *basic algorithm* yields high precision but moderate recall in its performance. The second, *extended algorithm* tends to equalise precision and recall but gives a higher F-measure than the basic algorithm on the text used for testing the system.

The basic algorithm is presented in Figure 1. The values of the thresholds t_1, t_2, t_3 were determined to be 0.71, 0.92, and 0.9 respectively on the basis of a small number of discrete experiments.

The extended algorithm is identical to the basic one, but additional steps have been added in order to increase the recall of the system. The additional steps can be viewed as simple heuristics applying in order to address the system's inability to perform named entity recognition. Some tests are applied to the NPs remaining after the basic algorithm has applied and would ordinarily be classed as inanimate. Firstly, any NP matching the names of months or a sequence of abbreviated month names and slashes are classified as inanimate. All remaining capitalised NPs are classified as animate. This particular heuristic was markedly effective in the present case due to the large number of proper names being used to refer to people in the text that were not present in the gazetteers consulted by the system. The remaining NPs are assigned to the set of inanimate references.

6. How well does it work?

In order to assess the usefulness of the animacy recognition system for anaphora resolution, MARS was run over the text described in Section 3 and the candidates of pronouns were filtered using different methods. The performance of the system for identifying the animate entities in texts is presented in Table 3. The columns assess two classification tasks for NPs, one for NPs as

```
NP is the noun phrase for which animacy has to be determined

Compute NA, NN, VSA, VSN for NP

1. If NA > t1 then NP is animate; Stop; Endif
2. If NN > t2 then NP is inanimate; Stop; Endif
3. If NA>NN and VSA>VSN then NP is animate; Stop; Endif
4. If (NP contains the complementiser who or a reflexive) or
      (VSA > t3) then
       NP is animate; Stop
   Endif
5. If unique_beginner(NP) = 15 or
      Unique_beginner(NP) = 17 then
      NP is inanimate; Stop;
      // the unique beginners 15 and 17 from WordNet refer
      // to objects and locations
   Endif
6. If head(NP) does not contain vowels then
     NP is inanimate; Stop;
     // this rule eliminates acronyms
   Endif
7. If head(NP) is in a first name gazetteer then
     NP is animate; Stop;
   Endif
8. NP is inanimate
```

Figure 1: The basic algorithm

items referring to entities that can also be referred to using gender-marked pronouns and one for NPs as items referring to entities that can also be referred to using neuter pronouns. The row *WordNet Alone* reports the performance of the system when only steps 1 and 2 of the algorithm are used. NPs with head lemmas not present in WordNet are determined to refer to inanimate entities. Operating in this mode, precision is relatively high but as expected, many references to people are missed by the system. There is great disparity between precision and recall in this system. The *Basic Algorithm* gives increased levels of precision and recall, again with quite a disparity between those measures. The *Extended Algorithm* has performance that tends to equalise precision and recall over the text. Additionally, the F-measure is greater for the *Extended Algorithm* than for either of the other two methods.

Note that in the extended algorithm, just considering noun heads gives precision 77.32% and recall 75.76%. If we also take information about the verb for which the NP is a subject, precision drops slightly to 77.25% whereas recall is increased to 75.88%, demonstrating a slight gain in F-measure.

One crucial aspect of the evaluation of the animate entity recognition system is the assessment of its contribution to the performance of a system for pronominal anaphora resolution. Indeed, this is the primary motivation for the current line of research. To this end, the method for animate entity recognition was applied to MARS.

The column *#Pronouns Successfully Resolved* in Table 3 indicates the performance of MARS when it has access to the information about which NPs refer to animate entities and which ones do not, as computed by the different automatic methods for animate entity recognition. It is apparent that MARS's performance is relatively poor[2] when no heuristics are included in the system for animate entity recognition. The low level of recall obtained under that method means that a large number of antecedents are excluded from the sets of candidates in the case of gender-marked pronouns. In fact, the application of information from *WordNet Alone* induces a worse success rate than when MARS is run in its default mode with 185 rather than 203 pronouns successfully resolved.

When the *Basic Algorithm*, with high precision and moderate recall, is applied, the resulting gain in performance is of just one more successful resolution than the original system, 204 rather than 203 pronouns successfully resolved. Again, it is suspected that the moderate level of recall obtained with respect to recognition of animate entities by the *Basic Algorithm* means that many antecedents are erroneously being excluded from the candidate sets of gender-marked pronouns.

The state of affairs is much improved for the text used in this work when the *Extended Algorithm* is applied. Here, the closeness of the levels of precision and recall

[2] It should be reiterated that MARS was not originally intended to process texts in this genre

Animate Entity Recognition	Precision (Animate Entity References)	Recall (Animate Entity References)	Precision (Non-Animate Entity References)	Recall (Non-Animate Entity References)	#Pronouns Successfully Resolved
WordNet Alone	0.9119	0.2845	0.7623	0.9882	185
Basic Algorithm	0.9363	0.5055	0.8223	0.9853	204
Extended Algorithm	0.7725	0.7588	0.8969	0.9038	224

Table 3: Evaluation of the different algorithms for recognition of animate entities

and the small increase in F-measure means that the effectiveness of the system is improved. The sizeable reduction in precision is counter-balanced by the positive effects of increased recall. When applied, the *Extended Algorithm* induces a level of performance in MARS that is identical to the level that would be obtained if references to animate entities could be recognised with perfect accuracy. This is a marked improvement over the original version of MARS, with 224 rather than 203 pronouns resolved successfully. It now correctly resolves 125 gender-marked pronouns. Prior to incorporation of the person recognition system, it only resolved 115 of them.

Unfortunately, there is some doubt about the confidence with which the extended algorithm can be applied to texts from other domains where there are fewer named entity references to people. To illustrate, when the extended algorithm was applied to MARS and run on different texts from the domain of software instruction manuals, the performance of the anaphora resolution system deteriorated to a level below the default (73.68% vs. 82.89% in one case). It is clear that the lack of precision caused by the heuristics simulating named entity recognition was detrimental to the system's performance. It has been noted in that domain that capitalised sequences refer to products, software and system components far more frequently than to people. By contrast, when the basic algorithm was applied, MARS's performance was unchanged in most cases and positively improved in others.

7. Future work

In this paper, the recognition of NP reference to animate entities has been motivated with respect to the process of pronominal anaphora resolution and the difficulty of that task has been described. A method addressing the identification of common NP reference to animate entities, based on the combination of the WordNet ontology and some heuristic rules, was presented and evaluated.

The results so far have been promising, but more serious testing of the system is required on texts from a variety of genres. There are a number of reliable clues to the animacy of entities referred to using NPs that were not implemented in the current system because they did not occur in the text used to develop the system. One example is the phenomenon of reflexive binding. Principle A of the binding theory states that reflexive pronouns must be bound by an antecedent in the same governing category. Binding occurs when one element c-commands another and the elements are coreferential (Haegeman, 1992). For current purposes, the required relation of c-command can be captured in a simplistic way using dependency and syntactic role information from the FDG Parser (Tapanainen and Jarvinen, 1997). It may then be predicted that any NP dependent on the same main verb as a reflexive but having the syntactic role of subject in the case of an object reflexive, shares the gender of the reflexive. Fortunately, reflexives are some of the few elements in English that are explicitly gender-marked. This clue was not incorporated into the current system due to a complete lack of positive examples in the text used to develop it. In general, it is recognised that this indicator can apply with strong reliability.

It would also be interesting to examine the contribution that relations between gender-marked possessive pronouns and their antecedents may make. It seems that the relative positions of possessive pronouns and their antecedents is constrained to a much greater extent than is the case for other pronouns, but to a lesser extent than is the case with reflexives.

Readers will note that the basic algorithm described here is dependent on the setting of various thresholds with respect to the measures NA, NN, and VSA. In the system presented here, those thresholds were set manually on the basis of a relatively small number of discrete experiments. In combination with a larger training corpus, it would be interesting to try neural network, algebraic or genetic optimisation techniques for obtaining the optimal values for these thresholds.

One frequent reason for the system's inability to recognise some references to animate entities was caused by its inability to disambiguate the appropriate senses of words like *pupil*, *producer*, *viewer* and even *teacher* in a given context. It may therefore by beneficial to explore the possible application of tactics from the field of word sense disambiguation to the system. Given the complexity of the word sense disambiguation task, a more reasonable solution may be to prune the WordNet ontology in light of information about the domain from which the text comes.

As hinted at in Section 1, any system attempting to recognise references to animate entities in texts must also approach the problem of named entity recognition. The simple heuristics applied in the present system with respect to capitalised words, for all their problems, demonstrate that named entity recognition has much to contribute to the recall of the system. The challenge will be to keep precision at an effective level, too. This will require the implementation of a much more sophisticated named entity recognition component. In order to process texts from the domain of computer and software technical manuals it will be necessary to address the classification of capitalised words into one of four classes, *locations*, *organisations*, *people*, and *artifacts*. This latter class is not normally considered in current named entity recognition task guidelines, though recognition of artifacts in English forms part of the MUC-7 *Information Extraction Task* (Chinchor, 1998). An additional problem is that in the classes mentioned previously, examples can be found of named entities which are not initially capitalised (e.g. *eXcite* and other "dot.com" companies, and the computer program *ears*).

Another area of interest is the examination of *n-grams* at the character level in lemmas of NPs referring to people. It has been noted that certain sequences of characters appear frequently in the names of professions. Examples are the sequences '*cian*' as in *musician*, *statistician*, *technician*, or '*tor*' as in *doctor*, *operator*, or *benefactor*, '*ist*' as in *facist*, *dentist*, or *optometrist* as well as '*ess*' as in *actress*, *princess*, or *manageress*, etc. It may be possible to find indicative sequences of letters by comparing the frequency of character n-grams in the head lemmas of NPs referring to people with those in other NP head lemmas.

8. References

Barlow, M., 1998. Feature Mismatches and Anaphora Resolution. In *Proceedings of DAARC2*, University of Lancaster.

Burnard, L. et al. 1995 *Users Reference Guide British National Corpus Version 1.0*, Oxford University Computing Services.

Cardie, C. and Wagstaff, K. 1999. Noun Phrase Coreference as Clustering. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in NLP and Very Large Corpora*. ACL.

Chinchor, N. 1997. *MUC-7 Named Entity Task Definition*. Version 3.5.

Chinchor, N. 1998. *MUC-7 Information Extraction Task Definition*, Version 5.1.

Denber, M. 1998. *Automatic Resolution of Anaphora in English*. Eastman Kodak Co.

Fellbaum, C. (Ed.), 1998. *WordNet An Electronic Lexical Database*. London: The MIT Press.

Ge, N., Hale, J., and Charniak, E. 1998. A Statistical Approach to Anaphora Resolution. In *Proceedings of the Sixth Workshop on Very Large Corpora (COLING-ACL '98)*. Morgan Kaufmann.

Haegeman, L. 1992. *Introduction to Government and Binding Theory*. Blackwell.

Hale, J. and Charniak, E. 1998. Getting Useful Gender Statistics from English Text. *Tech Report CS-98-06*, Brown University.

Hirschman, L., Robinson, P., Burger, J., and Vilain, M. 1998. The Role of Annotated Training Data. In *Proceedings of the workshop on Linguistic Coreference*. ELDA

Hobbs, J. 1996. Pronoun Resolution. *Research report 76-1*. City University of New York.

Kennedy, C. and Boguraev, B. 1996. Anaphora for everyone: pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)*. Morgan Kaufmann.

Lappin, S. and Leass, H. J. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*. MIT Press.

Mikheev, A., Moens, M., and Grover, C. 1999. Named Entity recognition without gazetteers. In *Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics (EACL 1999)*. ACL.

Mitkov, R. 1998. Robust pronoun resolution with limited knowledge. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING'98 / ACL'98)*. Morgan Kaufmann.

Nasukawa, T. 1994. Robust Method of Pronoun Resolution Using Full-Text Information. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*. Morgan Kaufmann.

Orasan, C., Evans, R., and Mitkov, R. 2000. Enhancing Preference-Based Anaphora Resolution with Genetic Algorithms. In *Proceedings of Natural Language Processing – NLP 2000*. Springer-Verlag.

Sinclair, J. et al. 1995. *Cobuild English Dictionary*. HarperCollins.

Tapanainen, P. and Jarvinen, T. 1997. A non-projective dependency parser. In *Proceedings of the 5th Conference of Applied Natural Language Processing*. ACL.

Vilain, M., Burger, J., Aberdeen, J., Connolly, D., and Hirschman, L. 1995. A Model Theoretic Coreference Scoring Scheme. In *Proceedings of the 6th Message Understanding Conference*. Morgan Kaufmann.

Wakao, T., Gaizauskas, R., and Wilks, Y. 1996. Evaluation of an Algorithm for for the Recognition and Classification of Proper Names. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING96)*. Morgan Kaufmann.