# A corpus based investigation of morphological disagreement in anaphoric relations

## Cătălina Barbu, Richard Evans and Ruslan Mitkov

School of Humanities, Languages and Social Sciences
University of Wolverhampton
Stafford Street, Wolverhampton, WV1 1SB, United Kingdom
{C.Barbu, R.J.Evans, R.Mitkov}@wlv.ac.uk

### Abstract

This paper investigates the causes of the comparatively low success rates in finding the antecedents of plural pronouns as compared to finding antecedents of singular pronouns. We are trying to show experimentally that considering morphological agreement as a strong constraint in pronoun resolution results in the erroneous interpretation of almost a quarter of the plural pronouns. The work is based on analysing sample texts from the British National Corpus and online technical manuals.

## 1.    Introduction

In anaphora resolution, morphological agreement is often considered a hard constraint, i.e., noun phrases have to obey number and gender agreement with the anaphor if they are to be considered as possible antecedents [1]. Some previous works (Barlow, 1998) warned about the danger of restricting too much the searching space for antecedents by considering morphological agreement as a hard constraint, thus introducing errors in the resolution process. The problem of gender agreement has been practically addressed more extensively (Hale and Charniak, 1998; Orasan and Evans, 2001), but, although the problem of some cases of number disagreement have been previously identified and discussed (Barlow, 1998; Denber, 1998; Mitkov, 2002), no large scale corpus-based investigation has been conducted to show the extent to which plural pronouns can refer to other constituents than simple plural noun phrases. Identifying the possible causes of number disagreement allows us to design methods for tackling these pronouns in an automatic anaphora resolution system.

## 2.    Morphological agreement in anaphora resolution

A glance at the state-of-the-art literature in automatic anaphora resolution shows that most methods assume the necessity of number and gender agreement between the anaphor and the antecedent. This strong constraint is based on the intuition that anaphorically linked elements must share similar morphological and semantic features. Morphological agreement is used as a mechanism for reducing ambiguity by eliminating those noun phrases that do not agree with a pronoun from the list of possible antecedents of that pronoun.

This is the approach used by Lappin&Leass's RAP (Lappin and Leass, 1994), by Kennedy&Boguraev's parser-free resolution method (Kennedy and Boguraev, 1996) and by Mitkov's knowldege-poor algorithm (Mitkov, 1998). To our knowledge, there has been no study indicating what

percent of the errors in the resolution of pronouns is due to the elimination of the correct antecedent as a result of gender and number disagreement.

The importance of morphological agreement has been proved by several psycholinguistic studies, that put in evidence the fact that it influences the speed of interpretation in human anaphora resolution. Several psycholinguistic studies investigated the role of gender cues in the identification of antecedents. (Garnham, 1992) and subsequently other works showed that gender agreement does not usually speed up the reading of a text under normal circumstances; however, readers do react to flagrant gender mismatches. Cacciari et al (1997) analysed the role of gender cues for languages with arbitrary grammatical gender (Italian and French). Their investigation targeted nouns that have a fixed syntactic gender but can be referred to by both masculine and feminine pronouns (the so-called *epicenes*). They showed that pronouns, proper names and clitics were interpreted quicker if they matched the epicene in gender. A number of other works described the role of gender stereotype

The role of number agreement in the interpretation of pronouns has not benefitted of such an extensive psycholinguistic analysis. Garrod and Stanford (1982) investigated cases where pronouns followed a conjoined NP in subject position in the previous sentence. In a similar direction of research, Sanford and Lockhart (1990) discovered a small preference for plural subject pronouns over singular subject pronouns following a sentence containing a conjoint NP. Clifton and Ferreira (1987) showed that a sentence containing a plural pronoun was read as quickly when referring to a conjoint NP as when referring to a split antecedent. Their interpretation of the phenomena is that plural pronouns draw their interpretation from a discourse representation, not a surface one. This interpretation is backed up by another experiment (Carreiras, 1997) that showed that plural pronouns were more easily interpreted when their antecedents were spatially close, rather then when their antecedents were split.

At the same time, psycholinguistic studies showed that number disagreement is a frequent phenomenon that is part

---

[1]We are discussing nominal anaphora, the most extensively researched and best understood in NLP

of the natural language. Gernsbacher (1991) and Oakhill et al (1992) examined what they called *conceptual anaphors*, i.e. cases of number disagreement where *they* is used to denote a class or collection of items inferred on the basis of a singular NP: "I need a plate. Where do you keep them?". Their experiments show that a plural pronoun used in these cases makes the text more coherent than the same text using a singular referent.

In this work, we have only concentrated on the number disagreement phenomenon. This decision was based on the fact that gender (dis)agreement is not a real problem in automatic pronoun resolution for English. As English does not feature grammatical gender, the gender agreement problem is reduced to identifying animate and inanimate references, therefore entities that can be referred to by both neutral (*it*) and animated, masculine or feminine (*he/she*) pronouns. This information can be used for improving the resolution of singular pronouns and it has been tackled in (Hale and Charniak, 1998), (Orasan and Evans, 2001).

## 3. Corpus-based investigation

The corpus investigation had three main goals: to identify which are the factors that lay behind number disagreement, to assess how frequent this phenomena is in real texts and to assess the distribution of different cases of disagreement across several text genres. As a direct application of the corpus results, we tried to see which of these cases can be treated automatically and which can be categorised as more complex cases.

### 3.1. Corpus

The investigation was based on four types of text: technical manuals (Linux HOW-To documents available online), narrative texts, health promotion and medical information leaflets and diverse newspapers articles (the last three being extracted from the British National Corpus (Burnard, 1995)). The narrative texts were extracts from 4 novels, totalling approximately 38000 words, the health promotion material was extracted from three documents describing the activity of the ACET AIDS organisation, while the newspapers extracts contained politics and sports information. The technical texts were Linux manuals and were included in the analysis in spite of the low number of plurals, due to the fact that they were annotated for coreference, therefore allowing us to perform an automatic analysis of the plural disagreement. We preferred extracts from multiple texts, instead of contiguous files in order to avoid certain usages of plurals that hold to the style of individual authors. The total number of plural pronouns was about 2500. Table 1 describes the content of the corpus and the proportion of plural pronouns.

| | Technical | narrative | medical | news |
|---|---|---|---|---|
| #files | 5 | 3 | 3 | 2 |
| #words | 30000 | 38000 | 39500 | 36600 |
| #plural pronouns | 102 | 994 | 724 | 730 |

Table 1: Corpus

### 3.2. Methodology

The analysis consisted of two stages: identifying the situations of number disagreement and analysing their distribution in the corpus. The initial assumption was that the basic case of number agreement is represented by a plural pronoun referring to a plural noun phrase; everything else was considered as an exception. Following this basic definition of an "exception", we collected approximately 300 pairs of plural pronouns and antecedents and classified them further into finer categories. The classification was done with respect to the automatic treatment of plurals. For example, there is no semantic difference between an antecedent consisting in a sequence of coordinated noun phrases and a split antecedent (consisting in a number of noun phrases further apart in the text). However, the automatic identification of a split antecedent requires a far greater amount of computation, therefore the two cases were classified in different categories. After the initial classification, we obtained ten cases of number disagreement. Nevertheless, the distinction between some of the categories was sometimes too fine; this resulted in problems of classification and, from a more pragmatic point of view, did not bring any contribution to the automatic treatment of plural pronouns. Therefore, some of the less frequent and more similar categories were collapsed, resulting in the final classification described below. After agreeing on the classification, we analysed the remaining corpus and classified the instances of number mismatches. When analysing the texts, we had to consider only those cases that displayed genuine disagreement. Consequently, we made two assumptions:

- first, the number of the pronoun has to be compared with the number of the last full noun phrase in its coreferential chain

- second, if a coreferential chain contained more than one occurrence of number mismatch, only the first occurrence was taken into account.

The first assumption was made the view of the efficiency of the analysis in mind. As the corpus was not previously annotated with coreferential links, and all the analysis was performed manually, it seemed extremely time consuming to try and identify the head of the coreferential chain containing a case of pronoun disagreement.

The second assumptions envisaged cases such as the sentence below, where the plural pronoun "they" appears twice, referring to the indefinite pronoun "everybody":

"She had soon learned that almost *everybody* has something *they* want to hide, and something *they*'re eager to share."

However, only the first instance of the pronoun was taken into account as a genuine case of disagreement.

This is consistent with the intuition that subsequent references can be interpreted with respect to the last element in the chain, independently of the head of the chain. The transitivity of the coreference relation insures the fact that the mental representation constructed for an element in the chain gathers the semantic information of all the previous references.

|         | technical |        | narrative |        | medical |        | newspapers |        | Total | %      |
|---------|-----------|--------|-----------|--------|---------|--------|------------|--------|-------|--------|
| Class 1 | 6         | 17.6%  | 56        | 21.2%  | 55      | 27.7%  | 28         | 24.3%  | 145   | 23.7%  |
| Class 2 | 2         | 5.8%   | 49        | 18.6%  | 26      | 13.3%  | 14         | 12.1%  | 91    | 14.8%  |
| Class 3 | 13        | 38.2%  | 44        | 16.7%  | 41      | 20.7%  | 22         | 19.1%  | 120   | 19.6%  |
| Class 4 | 5         | 14.7%  | 15        | 5.70%  | 38      | 19.1%  | 10         | 8.69%  | 68    | 11.1%  |
| Class 5 | 7         | 20.5%  | 11        | 4.18%  | 19      | 9.5%   | 19         | 16.5%  | 56    | 9.1%   |
| Class 6 | 5         | 14.7%  | 12        | 4.56%  | 16      | 8.08%  | 17         | 14.7%  | 50    | 8.1%   |
| Class 7 | 0         | 0%     | 54        | 20.5%  | 1       | 0.5%   | 1          | 0.8%   | 56    | 9.1%   |
| Class 8 | 0         | 0%     | 22        | 8.36%  | 2       | 0.1%   | 4          | 3.4%   | 28    | 4.5%   |
| Total   | 34        |        | 263       |        | 198     |        | 115        |        | 612   |        |

Table 2: Distribution of plural pronouns

### 3.3. Cases of number disagreement

The final classification consisted of the following eight categories:

1. The antecedent is a **conjunction/disjunction of plural or singular NPs**:
"You know *the Daily Mirror, and the Sun, and ITV, and the Unions*, what are *they* telling people to do?" 2. The pronoun has a **split antecedent**
"Only when *they* hang up did *Jay* realise that *she* hadn't given her a date."

3. **Collective nouns**
This category includes both collective noun such as *police*, *government*, *army*, singular nouns denoting more than one person, such as *a group of people*, *a number of people*, names of companies or associations.
"It belongs to *the Customs and Excise mob. They*'re not using it any more.

4. **Class representation**
Singular noun phrases that stand for a class of entities can be referred to by plural pronouns in English, as in:
"*Mackerel*! Terrific idea. I'll bet *they*'re really fresh."

5. **Gender underspecification**
In a context where the antecedent is a person, but does not feature grammatical gender, and the speaker is not aware of the gender of the person, it can be referred to by a plural pronoun, as an alternative to "he or she":
"You were called on the 30th of April at 21:38 hours. *The caller* withheld *their* number" (BT standard message)

6. Plural pronouns that refer to a **quantified noun or indefinite pronoun** ("someone", "every person")
"*Someone* will remember to wake me up early in the morning, won't *they*?"

7. **Generic plurals**
Sometimes plural pronouns are used with an impersonal sense, therefore there is no antecedent in the text and it cannot be inferred from any other entity in the text, as in:
"*They* have lessons in everything these days, don't *they*?"

8. **Indirect anaphora**
This is the case of plural pronouns whose interpretation is triggered by another entity in the text (possibly of singular number), as in:
"My sister's wedding was beautiful. *They* were the happiest couple",
where "they" is interpreted as "my sister and her husband"; this relation is established through a chain of mental inferences that links the textual elements "wedding" and "couple" to the inferred antecedent.
This last case is not a genuine case of disagreement, as we cannot talk about morphological agreement unless in the context of coreference. However, acknowledging and identifying such cases helps filtering unlikely candidates.

### 3.4. Results

Table 1 shows the distribution of the eight types of number mismatches across the four genres of texts considered. For each class and each type of text we provide the number of occurrences and the percentage reported to the total number of exceptions in that file. The last column displays the percentage of each type of exception with respect to the total number of exceptions in all documents.

### 3.5. Interpretation

A first look at the results shows that out of the 2500 pronouns inspected, 612 were exceptions, meaning that almost a quarter of the plural pronouns could not be interpreted as referring to a plural noun phrase. The largest percentage of exceptions was displayed by the technical manuals, with 36 out of 102 plurals (35.29%) constituting exceptions. The widest variety of disagreement cases was displayed by the narrative texts, which contained a significant number of pronouns in each category.
Overall, the most common case of disagreement seems to be represented by references to coordinated noun phrases, while the least common was the indirect anaphora. In fact, only narrative texts contained a significant number of indirect anaphora cases, in the other types of text the frequency of these cases being far below the frequency of the other cases of disagreement. The percentage of plurals referring to quantified nouns was, as predicted, proportional with the frequency of quantified nouns in the different types of text.
Among the disagreements present in technical manuals, the largest category was represented by references to collective nouns. A closer look showed that most of them were names of companies. The second most important category were references to a class representative, while generic plurals and indirect anaphora never appeared in our texts. Pronouns with split antecedents were also poorly represented.
The medical information documents contained

approximately the same percentage of pronouns referring to collective nouns as pronouns referring to a class representative. Most collective nouns were names of organisations, companies, hospitals, as well as nouns expressing the idea of collectivity (*a number, folk, mob*). Narrative texts contained an unexpectedly high number of generic plurals, which was only exceeded by the number of plurals with coordinated antecedent. It was also remarkable the low number of disagreement cases due to gender under-specification.

We do not believe that the data analysed is sufficient for us to comment on the linguistic and stylistic reasons and implications of the distribution of pronouns, and this was beyond the aim of our work. However, this investigation gives a good starting point for identifying those cases that are frequent enough to deserve a separate computational treatment in an automatic anaphora resolver.

## 4. A practical approach

### 4.1. Experiments

Table 3 presents the results reported by three anaphora resolvers in the resolution of plural pronouns. The evaluation was performed on the same set of technical manuals, which were manually annotated for coreferential links. The evaluation measure used was precision, defined as:

$$Precision = \frac{number\ of\ correctly\ resolved\ anaphors}{number\ of\ anaphors\ attempted\ to\ be\ resolved}$$
(1)

We are presenting here the results obtained by Mitkov's knowledge-poor approach (Mitkov, 1998), Kennedy&Boguraev's parser-free method (Kennedy and Boguraev, 1996) and Baldwin's Cogniac (Baldwin, 1997). All the implementations run in a fully automatic mode, using the same pre-processing tools and employing the same evaluation strategy[2].

| | #plurals | Mitkov | K&B | Cogniac |
|------|----------|--------|--------|---------|
| WIN | 6 | 0% | 50.0% | 50.0% |
| ACC | 39 | 48.7% | 48.7% | 43.5% |
| CDR | 24 | 29.1% | 30.0% | 29.1% |
| BEO | 17 | 35.2% | 41.17% | 35.2% |
| Total | 86 | 37.2% | 43.0% | 38.37% |

Table 3: Resolution rates of pronouns in technical manuals

Two things can be noticed in the results presented in Table 3: first, the low success rate in the resolution of plurals, and second, the fact that the results obtained with the three methods are very close to each other. The first observation should be made in the context of the evaluation of the resolution on all types of pronouns. Evaluating the aforementioned methods on the same texts, (Barbu and Mitkov, 2001) report a precision ranging from 51% for Cogniac up to 71% for Kennedy&Boguraev, therefore sensibly higher than for the interpretation of plurals alone. The second observation leads us to believe that the failures

---

[2]For more details on the implementation, see (Barbu and Mitkov, 2001)

are not entirely due to the malfunctioning of the methods. The small number of plurals in these texts allowed us to perform a manual experiment that was aimed at finding the upper limit of the resolution rate. In order to do this, we have manually corrected the number of the antecedents of the plurals, so that there was no disagreement. For split antecedents, we have removed the number indication from each of the components, and considered a resolution correct if one of the components was identified as antecedent. No other features have been corrected from the output of the parser. By running the three systems on the corrected input, we have obtained the results described in Table 4. Of course, the resolution of plurals is not only dependent on the correct identification of the number of the antecedent, many other types of pre-processing errors being responsible for failures in pronoun resolution. These pre-processing errors further add to those induced by the malfunctioning of the anaphora resolvers themselves.

| | #plurals | Mitkov | K&B | Cogniac |
|------|----------|---------|--------|---------|
| WIN | 6 | 83.3%% | 100.0% | 75.0% |
| ACC | 39 | 76.9% | 82.0% | 61.5% |
| CDR | 24 | 66.6% | 70.8% | 58.3% |
| BEO | 17 | 64.7% | 76.4% | 52.9% |
| Total | 86 | 72.0% | 79.0% | 59.3% |

Table 4: Accuracy rates on corrected input

### 4.2. Tackling difficult cases

The corpus investigation showed that the most frequent types of number disagreement belonged to categories 1 (coordinated antecedents), 4 (class representation) and 3 (collective nouns). This gives us an indication as to the areas that can be improved in anaphora resolution for achieving better resolution accuracy.

We have tried to identify which of the cases described above can be easily tackled in an automatic anaphora resolver, and which can be at least identified as difficult. The three categories of disagreement that could be solved more easily are: reference to coordinated NPs, reference to collective nouns and references to indefinite pronouns and quantified nouns. Some basic rules for identifying references to a class representative could also be attempted, while solving indirect anaphora definitely requires the most amount of knowledge and the most complicated inferential process.

**Coordinated NPs**

This case is easily tackled automatically, since coordinated NPs can be identified with a high degree of accuracy using a small number or rules. The resulting noun phrase will be allowed to function as an antecedent candidate for a plural pronoun. At the same time, none of the constituents of such a composed noun phrase should be allowed as antecedent for a plural pronoun. For example, in "Tom, John and Mary went to the cinema. They saw a comedy", it is impossible for *they* to refer to any of the groups *Tom, John* and John and Mary.

**Collective nouns**

Collective nouns such as "government" or "police" constitute a restricted set and therefore can be singled out using a basic lexicon. The same applies to noun phrases pre-modified by a collective noun (*a number of*, *a group of*). A more difficult problem is posed by names of companies, which have to be identified as such by a named entity recogniser. A simple grammar consisting of a small number of rules has been implemented that identifies as a company name a capitalised string followed by one of the suffixes:*inc., co., ltd., lab., corp.* (or the full denomination *Corporation, Limited, Laboratories, Incorporated*). Experiments performed on technical manuals showed that when using this restricted grammar, the resolution of plural pronouns referring to companies improved by about 55%. In general, named entity recogniser perform extremely well, with an accuracy of classification approaching 100%.

**Quantified noun/Indefinite pronoun antecedent**

In most cases, such noun phrases are referred to by plural pronouns, therefore easy rules can be implemented that allow them as antecedents for plural pronouns. Moreover, singular pronouns should normally not be allowed to refer to quantified nouns or indefinite pronouns, unless in special circumstances.

**Class representation**

Identifying singular noun phrases that stand for a class of entities is not a trivial task. The only simple solution would be to allow non-definite noun phrases to function as antecedents for plural pronouns. This could help solving cases where there is no plural noun phrase in the text that can be antecedent. However, the method is likely to introduce errors in the resolution of plural pronouns, by making the search space too wide.

## 5. Conclusion & Further work

We have addressed in this paper one of the factors responsible for errors in anaphora resolution, mainly cases of number disagreement between a pronoun and its antecedent. Although the problem may be interesting from a linguistic point of view, our approach was rather oriented towards identifying cases that can be dealt with automatically. The amount of data analysed may not be sufficient for drawing a definite conclusion regarding the frequency of occurrence of each type of number disagreement, but it gives an overall view of the cases that need to be tackled in order to improve the performance of an anaphora resolver. Our further work will concentrate on implementing some of the ideas resulting from this analysis in an automatic anaphora resolver.

## 6. References

Breck Baldwin. 1997. Cogniac: High precision coreference with limited knowledge and linguistic resources. In R. Mitkov and B. Boguraev, editors, *Operational factors in practical, robust anaphora resolution for unrestricted texts*, pages 38 – 45.

Catalina Barbu and Ruslan Mitkov. 2001. Evaluation tool for rule-based anaphora resolution methods. In *Proceedings of ACL'01*, Toulouse, France.

Michael Barlow. 1998. Feature mismatches in anaphora resolution. In *Proceedings of DAARC'98*, Lancaster, UK.

Lou Burnard. 1995. *Users Reference Guide British National Corpus Version 1.0*. Oxford University Computing Services, UK.

M. Carreiras. 1997. Plural pronouns and the representation of their antecedents. *European Journal of Cognitive Psychology*, (9):53–87.

C. Clifton and F.Ferreira. 1987. Discourse structure and anaphora: some experimental results. In M. Coltheart, editor, *Attention and performance XII: The psychology of reading*, pages 635–654. Lawrence Erlbaum Associates Ltd.

M. Denber. 1998. Automatic resolution of anaphora in english. Technical report, Eastman Kodak Co, Imaging Science Division.

A. Garnham. 1992. Minimalism versus constructionism: A false dichotomy in theories of inference during reading. *Psycoloquy*, 3(63):Reading Inference, 1.1.

S.C. Garrod and A.J. Sanford. 1982. The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of semantics*, (1):21–41.

M. A. Gernsbacher. 1991. Comprehending conceptual anaphors. *Language and Cognitive Processes*, (6):81–105.

John Hale and Eugene Charniak. 1998. Getting useful gender statistics from english text. Technical Report CS-98-06.

Christopher Kennedy and Branimir Boguraev. 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, Copenhagen, Denmark.

Shalom Lappin and H.J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535 – 562.

Ruslan Mitkov. 1998. Robust pronoun resolution with limited knowledge. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING'98/ACL'98*, pages 867 – 875. Morgan Kaufmann.

R. Mitkov. 2002. *Anaphor resolution*. Studies in Language and Linguistics. Longman.

J. Oakhill, A. Garnham, M.A. Gernsbacher, and K. Cain. 1992. How natural are conceptual anaphors? *Language and Cognitive Processes*, (7):257–280.

C. Orasan and R. Evans. 2001. Learning to identify animate references. *Proceedings of the Workshop on Computational Natural Language Learning (CoNLL-2001), ACL2001*.

A.J. Sanford and F. Lockhart. 1990. Description types and methods of conjoining as factors influencing plural anaphora: A continuation study of focus. *Journal of Semantics*, (7):365–378.