Can Anaphoric Definite Descriptions be Replaced by Pronouns?

Judita Preiss, Caroline Gasperin, Ted Briscoe

University of Cambridge, Computer Laboratory
JJ Thomson Avenue, Cambridge CB3 0FD, UK
Judita.Preiss@cl.cam.ac.uk, Caroline.Gasperin@cl.cam.ac.uk, Ted.Briscoe@cl.cam.ac.uk

Abstract

We investigate the hypothesis that a pronoun is used in discourse, when its antecedent is in the focus of the discourse. We create a corpus of 'replaced anaphoric definite descriptions', where occurrences of definite descriptions are replaced with a corresponding pronoun. We use the Lappin and Leass (1994) anaphora resolution algorithm on the new corpus, and obtain a much lower performance than when the corpus only contains genuine pronouns, thus supporting the hypothesis.

1. Introduction

We investigate the hypothesis that a pronoun is used in discourse, when its antecedent is in the focus (Sidner, 1979) of the discourse. It is generally believed that to refer to an entity which is not in focus, anaphoric definite descriptions (Hawkins, 1978) are used, which override the most salient entity.

Anaphora is a form of presupposition, pointing back to some previous item (Halliday and Hasan, 1979). The following is an example of pronominal anaphora, where *it* refers to *the final total*:

> Whether the final total is £5 or £5000, it is all very much needed.

Another example of anaphora are definite descriptions. In this work, we create a corpus of 'replaced anaphoric definite descriptions'; we replace occurrences of definite descriptions with a corresponding pronoun. For example:

> ACET will shortly be opening a new office in London. Nurse Kay Hopps will have responsibility for running the office.

This becomes:

> ACET will shortly be opening a new office in London. Nurse Kay Hopps will have responsibility for running it.

This corpus is used with a known anaphora resolution algorithm (Lappin and Leass, 1994) to resolve the new pronouns. A high performance on this corpus would imply that the hypothesis is false, and that anaphoric definite descriptions could be solved without using lexical-semantic knowledge. We could then automate our method of replacing anaphoric definite descriptions (for example, using the RASP parser (Briscoe and Carroll, 2002)), and use the pronoun resolution algorithm to identify the original entities.[1] A low performance on replaced anaphoric definite descriptions would indicate that there is indeed extra information in the definite description which is not immediately recoverable from the pronoun.

Section 2. introduces our evaluation corpus. In Section 3. we describe the Lappin and Leass (1994) anaphora resolution algorithm, including the modifications made to allow more accurate resolution of definite descriptions. Our results are presented in Section 4., and they are analyzed in Section 5. We draw our conclusions in Section 6., and put our work into context of current research in Section 7.

2. Evaluation Corpus

We manually marked up anaphoric definite descriptions in the first 750 sentences of the BNC, which we have previously annotated with pronoun–antecedent relations (Preiss, 2000). We considered as an anaphoric definite description any noun phrase (NP) starting with a definite article (such as *the*) which co-refers with some other NP previously appearing in the text. We didn't mark up the definite descriptions whose antecedent were not an NP.[2]

We replaced by the appropriate pronoun 84 anaphoric definite descriptions in the initial segment of the BNC, of which 51 were direct and the remainder were bridging definite descriptions (Vieira and Poesio, 2000). Direct anaphora are the cases where the anaphoric definite description has the same head noun as its antecedent, as in the following example from the corpus:

> ACET provides a dedicated Home Care service using a team of doctors and nurses ... The service includes pain and symptom control ...

On the other hand, bridging anaphora are the cases where the head noun of the anaphoric description is different of its antecedent head noun, as in the following example.

> There is a simple covenant form attached to this leaflet. All you have to do is to fill in the details, and sign and date the document in front of a witness.

Usually, direct anaphora are solved by comparing the anaphor head noun with the candidates to antecedent head

[1] Of course, the performance of the pronoun resolution algorithm would probably decrease overall, as replacing definite descriptions could result in a loss of semantic 'agreement' information.

[2] In our experiment, we do not deal with discourse new definite descriptions, as for instance in Vieira and Poesio (2000), since we have chosen by hand the anaphoric definite descriptions to be replaced by the pronouns.

nouns, and if they are equal the anaphora is resolved (if there is no tie). But resolving bridging anaphora is more complicated, since it is necessary to acquire extra information about the (semantic) relation between to different head nouns.

It is interesting to note that, from the point of view of our replacement by pronouns experiment, both direct and bridging anaphora require the same information to be resolved. By replacing any anaphoric definite description by a pronoun, the lexical information about the head noun and the distinction of being direct or bridging is lost. If we successfully resolve the replaced pronouns, we could use an automatic replacement algorithm to resolve the harder bridging descriptions in general.

In this corpus, the average number of sentences between an anaphoric definite description and its antecedent was found to be 1.7, whereas the average number of sentences between a pronoun and its antecedent is 0.5. So pronouns usually have antecedents in the same sentence, whereas anaphoric definite descriptions tend to have antecedents about 2 sentences away.

3. Anaphora Resolution Algorithm

In this work we use the Lappin and Leass (1994) anaphora resolution system, reimplemented as described in the work of Preiss and Briscoe (2003) to use the grammatical relation (GR) output from the RASP system (Briscoe and Carroll, 2002).

3.1. Original Algorithm

For each pronoun, this algorithm uses syntactic criteria to rule out noun phrases that cannot possibly corefer with it. An antecedent is then chosen according to a ranking based on salience weights.

For all types of pronoun, noun phrases are ruled out if they have incompatible agreement features. Pronouns are split into two classes: lexical (reflexives and reciprocals) and non-lexical anaphors. There are syntactic filters for the two types of anaphors.

Candidates which remain after filtering are ranked according to their salience. A salience value corresponding to a weighted sum of the relevant feature weights (summarized in Table 1) is computed. If we consider the sentence *John walks*, the salience of *John* will be:

$$\begin{aligned} \mathrm{sal}(J) &= w_{\text{sent}} + w_{\text{subj}} + w_{\text{head}} + w_{\text{non-adv}} \\ &= 100 + 80 + 80 + 50 \\ &= 310 \end{aligned}$$

The weights are scaled by a factor of $\left(\frac{1}{2}\right)^s$ where s is the distance (number of sentences) of the candidate from the pronoun.

The candidate with the highest salience is proposed as the antecedent.

3.2. Modified Algorithm

To investigate the performance of the Lappin and Leass algorithm on definite descriptions, genuine pronouns are placed in their antecedent's equivalence class according to the manual annotation. The salience of each equivalence class is therefore initialized to a correct value, and the algorithm doesn't start with any chaining errors due to wrong resolution of genuine pronouns.

Factor	Weight
Sentence recency	100
Subject emphasis	80
Existential emphasis	70
Accusative emphasis	50
Indirect object/oblique	40
Head noun emphasis	80
Non-adverbial emphasis	50
Parallelism	35
Cataphora	175

Table 1: LL Salience Weights

An important change is also made to the salience weighting scheme: as noted in Section 2., the average number of sentences between a definite description pronoun and the antecedent definite description is larger than the number of sentences between a genuine pronoun and its antecedent. The Lappin and Leass algorithms scales salience values using the following scaling function f_{ll}:

$$f_{ll}(s) = \left(\frac{1}{2}\right)^s$$

(where s is the number of sentences between the pronoun and its antecedent). This means that the salience of noun phrases (NPs) in the same sentence as the pronoun is not reduced at all, the salience of NPs in the previous sentence halves and so on. Therefore a noun phrase in the same sentence as the pronoun is very likely to be selected as the antecedent. This is clearly not desired for a definite description pronoun; in this case, noun phrases in the previous sentence should be more likely.

We assume that a Geometric distribution (as used by Lappin and Leass) is optimal for scaling, and preserve the $\frac{1.7}{0.5}$ ratio of the number of sentences between pronoun and antecedent as follows:

$$\begin{aligned} \mathcal{P}(X = n) &= (1-r)r^n && (1) \\ \mathcal{E}(X) &= \sum_{n \in N} n(1-r)r^n && (2) \end{aligned}$$

Equation 1 describes the Geometric distribution with parameter r, and equation 2 gives the mean of this distribution. Assuming that $\frac{1}{2}$ is the optimal value for the genuine pronoun resolution, we preserve the ratio in the means as shown in the Table below:

No. of sents	$\mathcal{E}(X)$	r
0.5	1	$1/2$
1.7	3.4	$17/22$

After shifting the distribution to start at one, and selecting the optimal value for same sentence weighting, we obtain the following scaling function f_{dd}:

$$f_{dd}(s) = \begin{cases} \frac{1}{2} & \text{if } s = 0 \\ \left(\frac{17}{22}\right)^{s-1} & \text{if } s > 0 \end{cases}$$

The new function f_{dd} decreases the salience of entities in the current sentence, boosting saliences in the previous sentence (as per the average distances above).

4. Results

The performance of the original Lappin and Leass pronoun resolution algorithm was found to be 23% on the replaced anaphoric definite descriptions, whereas it achieved an accuracy of 62% on genuine inter-sentential pronouns (the performance was even higher for genuine intra-sentential pronouns). The Lappin and Leass pronoun resolution algorithm was also run with the new weighting function described in Section 3.2.; the performance on the replaced anaphoric definite descriptions rises to 29%.

The distribution of correctly resolved anaphora according to the distance (number of sentences) between the anaphor and the antecedent is presented in Table 2. The results using the original distance weighting in the Lappin and Leass algorithm are shown in the first line (original), whereas the results with the optimized distance weighting are in the second line (optimized).

Weighting	Sentences					Total
	0	1	2	3	≥ 4	
Original	54%	26%	23%	8%	0%	22%
Optimized	46%	42%	23%	8%	6%	28%

Table 2: Distribution of correctly resolved anaphora according to anaphor-antecedent distance (in sentences)

The distribution in Table 2 is more flat when the scaling factors are optimized. The decrease in the performance on intra-sentential definite descriptions is to be expected, as we have lowered the significance of the saliences for the current sentence. However, in so doing we have greatly increased the performance on the definite descriptions with antecedent in the previous sentence (of which there are 31 whereas intra-sentential definite descriptions only occur 13 times).

5. Discussion

5.1. Number Disagreement

In the case of bridging anaphora, the anaphor and its antecedent do not have to agree in number. So, when the anaphor is replaced by a pronoun containing number information, it is possible that the correct antecedent is filtered out due to gender agreement in the Lappin and Leass algorithm. For example, the coreference between *churches* and *the congregation* in the following discourse will not be found when *the congregation* is replaced by *it*:

> Most churches are completely unprepared for the shock of finding an established member of the congregation is infected with HIV or dying with AIDS . . .

Similarly, the correct antecedent *illnesses* will be ruled out as the antecedent of *the cause of death* in the following:

> The biggest changes are in the length of time people ill with the disease are now surviving and in the nature of the illnesses themselves. (. . .) The "cause of death" figures are also changing beyond all recognition.

5.2. Baseline Comparison

For the results obtained to have meaning, we need to consider them in context of a baseline system; that is an unsophisticated system which carries out the task of resolving definite descriptions to their antecedent. An example of such a system is a simple string match – in this system, a definite description is linked to its antecedent only if they share the same word:

> After every client visit you are asked to call the office so that you can report how the visit went.

The baseline system run over the same 84 definite descriptions achieves a performance of 62%. Given that the performance of the anaphora resolution system on the definite descriptions was only 28%, these two results present strong support for the statement that definite descriptions are used whenever a pronoun would not provide sufficient information.

6. Conclusion

We have shown that the replaced definite descriptions are not immediately resolvable by an anaphora resolution system. A comparison with a baseline system which uses a basic string match, supports our hypothesis that definite descriptions are used under different circumstances than pronouns, and require extra semantic information.

Our experiment also shows that salience factors need to be optimised differently for pronouns and replaced definite descriptions.

For future work, we intend to test if our anaphoric definite descriptions replacement experiment can be more successful in domain-specific texts. For example, in biomedical texts the definite descriptions seem to have a more restricted behaviour: most of them are anaphoric and co-refer to precise named entities instead of larger pieces of text.

7. Related Work

According to Sidner (1986), a definite description is preferably used, instead of a pronoun, to refer back to a previous focus in the focus stack, overriding the current focus (which could be referred to by a pronoun). Sidner's algorithms rely on a semantic network encoding elements and their associations, providing links expressing their general class, and inheritance of associations.

Much of the earlier work in anaphora resolution exploited domain and linguistic knowledge which was difficult to acquire (considerable human input required), represent and process. More recently, work has been done on knowledge-poor anaphora resolution strategies, which require extensive annotated corpora (Mitkov, 2001). However, many anaphora resolution algorithms are evaluated on authors' own corpora (e.g., computer manuals for the Lappin and Leass anaphora resolution algorithm) possibly making the results unreasonably high.

There is movement on generating coreference annotated corpora, since annotating anaphora and coreference in general is a very difficult task. The majority of NLP work on anaphora and coreference resolution will be able to benefit from such corpora by using it for evaluation and training purposes.

Acknowledgements

The authors were supported by UK EPSRC project GR/N36462/93: 'Robust Accurate Statistical Parsing (RASP)', and the Brazilian Government - CAPES, Brazil.

8. References

E. J. Briscoe and J. Carroll. 2002. Robust accurate statistical annotation of general text. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, pages 1499–1504.

M. A. Halliday and R. Hasan. 1979. *Cohesion in English*. Longman English Language Series.

J. A. Hawkins. 1978. *Definiteness and Indefiniteness*. London: Croom Helm.

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

R. Mitkov. 2001. Outstanding issues in anaphora resolution. In Al. Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 110–125. Springer.

J. Preiss and E. Briscoe. 2003. Shallow or full parsing for anaphora resolution? An experiment with the Lappin and Leass algorithm. In *Proceedings of the Workshop on Anaphora Resolution*, pages 1–6.

J. Preiss. 2000. Word sense disambiguation using CIDE+. Master's thesis, University of Cambridge.

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

C. Sidner. 1986. Focusing in the comprehension of definite anaphora. In B. J. Grosz, K. Sparck Jones, and B. L. Webber, editors, *Readings in Natural Language Processing*, pages 363–394. Morgan Kaufmann Publishers Inc.

R. Vieira and M. Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–594.