What is coreference, and what should coreference annotation be?

Kees van Deemter and Rodger Kibble
ITRI, University of Brighton
Lewes Road
Brighton BN2 4GJ
United Kingdom
Kees.van.Deemter@itri.brighton.ac.uk
Rodger.Kibble@itri.brighton.ac.uk

Abstract

In this paper, it is argued that 'coreference annotation', as currently performed in the MUC community, goes well beyond annotation of the relation of *coreference* as it is commonly understood. As a result, it is not always clear what semantic relation these annotations are actually encoding. The paper discusses a number of interrelated problems with coreference annotation and concludes that rethinking of the coreference task is needed before the task can be expanded (e.g., to cover part/whole relations) as has recently been advocated. As a step towards solution of the problems with coreference annotation, one possible simplification of the annotation task is suggested. This strategy can be summed up by the phrase "Coreference annotation should annotate coreference relations, and coreference relations only".

1 Introduction: Coreference Annotation

Various practical tasks requiring language technology including, for example, *information extraction* and *text summarization*, can be done more reliably if it is possible to automatically find parts of the text containing information about a given topic. For example, if a text summarizer has to select the most important information, in a given text, about the 1984 Wall Street crash, then the summarization task is greatly helped if a program can automatically spot all the clauses in the text that contain information about this crash. To 'train' a program of this kind, extensive language corpora have been prepared in which human readers have annotated what has been called the *coreference* relation. These annotated corpora are then used as a 'gold standard' against which the program's achievements can be compared.

The relation of coreference has been defined as holding between two noun phrases[1] if they 'refer to the same entity' (Hirschman et al. 1997). More precisely, let us assume that α_1 and α_2 are occurrences of noun phrases (NPs) and let us assume that both have a unique reference in the context in which they occur (i.e., their context in the corpus makes them unambiguous). Under these assumptions we can use a functional notation, e.g. Reference(α), as short for 'the entity referred to by α' and define (suppressing the role of context):

> **Definition**: α_1 and α_2 *corefer* if and only if Reference(α_1) = Reference(α_2).

Coreference annotation has been one focus of the 6th and 7th Message Understanding Conference (MUC-6, MUC-7) and various other annotation exercises (e.g. Davies et al. 1998), and it has been the topic of a number of separate workshops. We will limit the discussion to coreference annotations for information extraction. Because the MUC project is the best-known example of this type of coreference annotation, and because of the public availability of the MUC Task Definition (TD, MUC 1997), we will focus on coreference annotations in MUC.

It is clear that *anaphoric* relations are also of potential relevance for any task that requires text interpretation. It follows from the definition of coreference, however, that anaphora does not equal coreference. Coreference, for example, is a symmetrical and transitive relation, leading to a

[1] In some cases, a restriction to noun phrases *and nouns* is advocated (e.g. Hirschman et al. 1997) but it seems that, in practice, the annotation can be viewed as limited to noun phrases. If all common nouns (e.g. 'person', 'share price', etc.) were included, the notion of coreference would become even more difficult.

simple partitioning of a set of NPs.[2] Anaphora, by contrast, is a nonsymmetrical and nontransitive relation: if NP_1 is anaphoric to NP_2 then, usually, NP_2 is not anaphoric to NP_1, for example. Secondly, anaphora involves *context-sensitivity* of interpretation (an anaphoric pronoun, for example, cannot be interpreted without information about where it occurs), whereas a name (*W.J.Clinton*) and a description (*Mrs. H.Clinton's husband*) can corefer without any of the two depending on the other for its interpretation. Anaphoric and coreferential relations can coincide, as is the case with 'pragmatic' pronouns such as *he* in *W.J.Clinton took the oath; then he took a deep breath.* The point is just that not all coreferential relations are anaphoric, nor are all anaphoric relations coreferential.

The problems that will be pointed out in Section 2 suggest that coreference and anaphora are not properly distinguished in MUC and that this has led to a TD that is difficult to understand and apply. Four criteria are listed (MUC 97) for the MUC TD, in order of priority:

1. The MUC task should be supported by the annotations
2. Good (defined as 95%) inter-annotator agreement should be achievable
3. It should be possible to annotate texts quickly and cheaply
4. A corpus should be created that can be used as a tool for linguists not working on the MUC information extraction task

The TD makes it clear that the annotation task has been simplified in a number of ways. In particular, only Noun Phrases were annotated (thereby circumventing problems of null anaphora, summation, abstraction, etc., see e.g. Kamp & Reyle 1993). Such eminently sensible simplifications notwithstanding, we will argue that the above-mentioned criteria are extremely difficult to achieve. We shall argue that this is due to fundamental unclarities in the TD and we will suggest that a rethinking of the coreference annotation enterprise is in order before it ventures into new domains involving speech, noisy data, etc., (see e.g. Bagga et al. 1999), or before it extends the relation of coreference to cover whole/part and class/instance relations (e.g. Popescu-Belis 1998, MUC 1997), as has been proposed recently.

[2]The somewhat confusing use of the REF feature, in SGML-based MUC annotations, which records the 'antecedent' of a 'referring expression' (MUC 1997) could be taken to imply that the notion of coreference relevant for MUC is nonsymmetrical, but the explanations elsewhere (see e.g. Hirschman et al. 1997, and MUC 1997, Section 8) make it clear that an equivalence relation is intended.

2 Problems

In this section, we will discuss what we view as some of the most fundamental obstacles for coreference annotation. We will explore the implications of the observation that many NPs do not refer (Section 2.1), after which we will move on to problems of *intensionality* (Section 2.2) and the issue of determining the 'markables' in a corpus (Section 2.3). Some conclusions will be drawn in the final section (Section 2.4).

2.1 Non-referring NPs

When a speaker/writer uses an NP to refer to an entity (i.e., either an object of some sort or a set of objects), he or she tries to single out the entity uniquely. Thus, when someone says *The owner of this wig is bald*, the speaker uses the NP *The owner of this wig* to enable his or her audience to determine what person, say Mr. X, they are ascribing the property of baldness to. Like everything in language, the notion of referring is not *entirely* unproblematic. For example, the speaker's belief that Mr. X owns the wig may be mistaken; worse even, nobody might own the wig. But, as is recognized in virtually any semantic theory (for elaboration, see e.g. Gamut 1982, Chapter 1), as well as in the MUC TD itself, reference is a relatively clear notion. Especially in the very factual text *genres* targeted in Information Extraction (see the Appendix of the present paper for an example), few problems are likely to occur. In an annotation exercise that has been carried out separate from MUC and that will be reported on elsewhere (e.g. Poesio et al. 1999 for a preliminary report), it has been found that the question whether an NP refers (based on definitions in Lyons 1977) can be answered by annotators with very high inter-annotator agreement.

One thing that is clear about reference is that many NPs do not refer. When someone says

1a. *No solution emerged from our conversations*, or
1b. *A few solutions may emerge from our conversations*

the subject NPs do not refer to any single solution, nor to any definite set of solutions. They have no reference.[3] As a result, the *co*reference relation as defined in Section 1 is inapplicable.

Nonreferring NPs can stand in various semantic relations to each other including anaphoric relations. For example, the NP *a few solutions* can be embedded in a conditional, saying *Whenever a few solutions emerged, we embraced them.* The anaphoric relation between *a few solutions* and *them* cannot be modeled by a theory of reference. Instead, a variable binding account may be employed to reflect that two sets of entities must *co-vary*: the set of any solutions that emerged at a given moment and the set of any solutions that we embraced at that moment. Of course, it would be possible to ask annotators to annotate anaphoric relations, in which case one would need to explain what anaphora is. This would be a substantial task which would require the writing of a new TD[4].
For the reasons sketched above, NPs of the following types do not refer:

- Quantifying NPs (e.g. *'Every man'*, *'Most computational linguists'* (MUC 97))
- Most occurrences of Indefinite NPs (e.g., *'I don't have a machete'* (MUC 97), *'Do you own a machete?'*)
- Predicative NPs (*'... became (president of Dreamy Detergents)'*, (MUC 97, see also Section 2.2))

A 'substitution' test can be used to confirm that NPs that stand in anaphoric relations to NPs of these types do not corefer with them. For instance, one may observe that *Every man loves his mother* does not mean the same as *Every man loves every man's mother*, contrasting with referring NPs, which do allow such substitutions (e.g., *John loves his mother* equals *John loves John's mother*).[5]

So, substantial classes of NPs do not refer, and consequently they cannot *co*refer. Yet, MUC's annotators have been asked to mark NPs of each of the above-mentioned categories and to let them 'corefer' with other NPs. So clearly, the relation annotated in MUC – let's call it the IDENT relation, following (MUC 97) – differs sharply from the coreference relation. The TD admits that certain instructions may be incompatible with the definition of coreference but no reason is given for these incompatibilities and no intuitive motivation for the relation IDENT is offered. The annotator is left with a long series of instructions which fail to be held together by a common rationale.

2.2 Intensionality (and text-asserted identity)

The coreference annotation community is well aware of some of the problems with the TD. The problem that has received most of their attention is the problem of intensionality (Hirschman et al. 1997). This awareness has led to considerable complexities in the relevant parts of the TD. For example, in Section 1.3 of MUC (1997), where the implications of 'change over time' are considered, where the example *the stock price fell from $4.02 to $3.85* is discussed, the instructions tell annotators to consider *the stock price* as standing in the IDENT relation with $3.85 but not with $4.02, for the reason that $3.85 is 'the more recent value'. Quite reasonably, $4.02 is not considered to stand in the IDENT relation with *the stock price* because transitivity would lead to the conclusion that $4.02 and $3.85 are equal. The first question this raises is, what

[3] Of course, an NP like 'no solution' has a *meaning*, but 'having the same meaning' is different from coreference. For example, in *Mary is married to a nice man and Sue is also married to a nice man*, both occurrences of *a nice man* have the same meaning, but one would expect them to refer to different individuals.

[4] Sometimes the term 'co-specification' has been used to replace coreference by a wider notion which subsumes at least some types of anaphora including, specifically, the use of pragmatic pronouns (e.g. Sidner 1983). Co-specification, however, is not an intuitively clear notion either – what does it mean for an expression to 'specify' something? – and no definition of it that would be useful to annotators is known to us. In particular, it is unclear whether a bound anaphor and its antecedent co-specify, or how the notion should be applied to intensional constructions (see Section 2.2).

[5] Tests of this kind could be offered to annotators to simplify their task. Space does not allow their exact formulation, since qualifications are needed to account for NPs in attitude contexts and for specifically used indefinites.

happens if the next sentence asserts that, later on, the price fell even lower?

> 2. *(a) The stock price fell from $4.02 to $3.85*; (b) Later that day, it fell to an even lower value, at $3.82.

Does the annotator have to go back to (a), deciding that $3.82 is an even more recent value and *the stock price* does not stand in the IDENT relation with $3.85 after all?

Later parts of the TD contradict what is said in Section 1.3. Section 6.4 tells annotators that 'Two markables should be recorded as coreferential if the text asserts them to be coreferential at *any time*'. Accordingly, in

> 3. *Henry Higgins, who was formerly sales director of Sudsy Soaps, became president of Dreamy Detergents,*

annotators are asked to mark (1) Henry Higgins, (2) sales director of Sudsy Soaps, and (3) president of Dreamy Detergents as standing in the IDENT relation. But, by the same reasoning as above, this implies that Henry Higgins is presently the sales director of Sudsy Soaps as well as the president of Dreamy Detergents, which is not what the text asserts. Clearly, this is not a sensible instruction either.

As in the case of non-referring NPs (Section 2.1), the root of the trouble lies in the fact that the relatively clear (but limited) notion of coreference is extended to one that aims to be applicable in a wider class of cases, but which is no longer clear. On linguistic grounds, Two strategies could be used to solve the problem. One would be to exclude predicatively used NPs from entering coreference relations and to leave their analysis to other MUC tasks. The other, more sophisticated strategy, consistent with Dowty et al. (1981), would be to say that, in cases like this, *The stock price* refers, not to a number (such as the number $3.85) but to a Montague-type individual concept (Dowty et al. 1981), that is, a function from times to numbers. It would have followed that *The stock price* does not corefer with either $4.02 or $3.85 and no problem would have arisen. Analogously, *president of Dreamy Detergents*, in the context cited above, would denote an individual concept rather than an individual. If the next sentence goes on the say *He died within a week*, *he* would be marked as coreferential with Henry Higgins; if, instead, the text proceeds saying *This is an influential position, but the pay is lousy*, then *this* would be marked as coreferential with *president of Dreamy Detergents.* It is possible that this second strategy would be asking rather too much from annotators, in which case the first strategy would be preferable.

2.3 Markables

Experience with the coreference task has shown that it is surprisingly difficult, and this has been tackled by breaking it down into more manageable subtasks. The emerging practice (recommended by Hirschman et al. 1997) is to separate annotation into a two-stage process: annotation of markables is to be carried out before linking coreferring elements. This means that the coreference task becomes a matter of partitioning the set of markables into equivalence classes, which may be interpreted as corresponding to 'discourse referents' (cf. Popescu-Belis and Robba 1998). It turns out, however, that the distinction between marking up and linking is not strictly followed even in the MUC-7 specification. Certain elements are only marked up if they corefer with an existing markable: these include conjuncts and prenominal modifiers. In the following example, the first occurrence of *aluminum* is markable as it 'corefers' with the occurrence of this noun as a bare NP in the second clause.

> 4. *The price of* aluminum *siding has steadily increased, as the market for* aluminum *reacts to the strike in Chile.*

Bare nouns in modifier position are not said to be markable unless there is a coreference relation of this type.

There are various ways one could address these difficulties. One possibility is to explicitly separate out the task of marking up all elements which might participate in coreference. However, this approach will increase the complexity of the task, which is likely to become unmanageable if the scheme is extended to cover "discontinuous elements, including conjoined elements" as suggested in Section 1.4. Consider the example (emphasis added):

5. Oestradiol *is a form of oestrogen;* norethisterone acetate *is a progestogen.*
They *belong to a group of medicines known as Hormone Replacement Therapy (HRT).* (ABPI 1997)

The problem here is that the antecedent of *They* is the conjunction of *Oestradiol* and *norethisterone acetate*, which doesn't appear as a contiguous sequence in the text. This relation can be annotated by adding new tags for composite referring expressions, but it is obviously undesirable to encode these tags in advance for every possible combination of referents in a text, since the number would increase exponentially with the number of "basic" referring expressions. An extreme alternative is to have a first pass where only referring expressions which look like anaphors are marked up, such as pronouns, definite NPs and reduced forms of proper names. Subsequent passes would look for antecedents for these expressions and link coreferring elements. An intermediate approach would be to mark up a "core" set of referring expressions on the first pass, allowing for further referring expressions to be identified on subsequent passes if this is necessary to resolve coreference. The extent to which each of these strategies would contribute to accuracy and speed of annotation remains to be determined, but it seems unrealistic to expect great benefits from any of them.

Beyond the noun phrase. It has been suggested above that the scope of the coreference task might better be restricted to cases of 'strict' coreference involving NPs. This would be compatible in the longer term with extending the domain of the task to cover abstract objects such as events, when they are not described using an NP. When analysing naturally occurring text one often finds pronouns and full NPs which refer back in some way to the content of a clause, as in the following example:

6. *Bates had crashed an F-14 into the Pacific during a routine training flight in April. Navy officials blamed him for causing* the accident...

This is a clear case of coreference, where two expressions refer to a well-defined event. Cases like this are currently excluded from the MUC coreference task, which limits itself to relations between NPs. On the other hand, they are on the "wish list" in Section 1.4, "Future Directions" and, from the point of view of Information Extraction, it is obviously desirable to incorporate reference to events. If and when this is done, the problems that were noted above will be exacerbated. In particular, difficulties arise if the strategy of identifying markables in advance is maintained, since it is difficult to determine which types of elements can serve as antecedents (ABPI 1997):

7. *Be careful not to get the gel in your eyes*
If this *happens, rinse your eyes with clean water and tell your doctor.*

8. *The label will tell you how much to use and how often.*
This *will usually be two or three times a day.*

To sum up, the news from this quarter is both good and bad. There are clear cases of event coreference which can be incorporated into the coreference task. On the other hand, existing problems with annotation of NPs will be made worse since annotators will be confronted with some difficult problems both in identifying markables and deciding on coreference links.

3 Conclusion

Based on the above, we would like to argue that current 'coreference annotation' practice, as exemplified by MUC, has over-extended itself, mixing elements of coreference with elements of anaphora in unclear ways. As a result, the annotated corpus that is likely to emerge from MUC may not be very useful for the research community outside MUC (Criterion 4), the more so because generalization to other subject domains is likely to make problems worse. For example, in many domains, there are other sources of intensionality than just change over time.

Let us briefly return to the other success criteria mentioned in MUC (1997). It would seem that the current MUC Task Definition is already rather complex, to the point where it becomes doubtful that it can be applied quickly and cheaply (Criterion 3). Indeed, one has to ask whether it can be applied with a sufficient

degree of accuracy, even given plenty of time. Hirschberg et al. (1997), when discussing this question, note that inter-annotator agreement (Criterion 2) at the time of writing, was in the low eighties. The material in Section 2 suggests that this relative lack of success is no accident and that unclarities and internal inconsistencies stand in the way of success.

A separate issue that has been discussed in Section 2.3 is the identification of 'markables'. A clear separation between (1) the task of marking all the markables and (2) that of annotating coreference relations between markables appears to be difficult to uphold. Consequently, one has to ask whether the coreference task can be made easier by finding a more effective separation of subtasks. This becomes even more urgent if generalization to other phenomena than relations between NPs, as has recently been advocated, are contemplated.

Given this situation, we suggest that coreference annotation might do well to restrict itself to annotation of the coreference relation, as defined in MUC (1997) and Hirschman et al.(1997) (See our Section 1). Instead of the IDENT relation practiced in MUC, annotation of the *coreference* relation promises better chances for success. If this strategy were adopted, annotation would become a more modest enterprise, which would provide its consumers with information that is smaller in volume but more reliable in quality. (For an example, see Appendix.)

In conclusion, it appears that there is scope for new collaboration between the coreference annotation community and the computational semantics community. The present paper attempts to be a small step in this direction.

APPENDIX

To show that, in the text genre targeted by MUC, not much is lost if annotation is limited to coreference as defined in Section 1, we took an excerpt of a MUC-6 Wall Street Journal article, precisely as it was rendered and annotated in MUC (1997), Appendix A, where it was used as an extended 'sample annotation' of non-dialogue annotation. Reflecting the 'official' view (see Section 1), according to which only the partitioning into equivalence classes matters, we simplify the notation by printing a number following the NP that it annotates. Thus, NPs that are followed by the same number stand in the IDENT relation (Section 2.1):

> Ocean Drilling & Exploration Co.(1) will sell its(1) contract-drilling business(2), and took a $50.9 million loss from discontinued operations in the third quarter(3) because of the planned sale.
>
> The New Orleans oil and gas exploration and diving operations company(1) added that it(1) doesn't expect any further adverse financial impact from the restructuring.
>
> In the third quarter(3), the company(1), which is 61%-owned by Murphy Oil Corp. of Arkansas, had a net loss of(4) $46.9 million(4), or 91 cents a share(4).
>
> It has long been rumored that Ocean Drilling(1) would sell the unit(2) to concentrate on its(1) core oil and gas business.

The annotation shows a partitioning into four non-singleton sets of NPs: 7 NPs in class **(1)** (*the company*), 2 in class **(2)** (*the contract-drilling business*, 2 in class **(3)** (*the third quarter*), and 3 in class **(4)** (*the loss*).

It is easy to see how the text would be annotated using the notion of ('strict') coherence advocated in Section 2. Each of the above-defined classes except **(4)** are coreference relations, and consequently they would be annotated in the exact same way if only coreference were annotated. The only difference is class **(4)**, and this class exemplifies the problems discussed in section 2.2. We conclude that, for this text, the only MUC annotations that are not inherited by 'strict' coreference are questionable.

In addition, it may be noted that the most important links that are missed by both annotation schemes concern the use of *which*, which refers to the company, and various references to the selling/restructuring of the division (e.g., *the planned sale*, *the restructuring*). Both are coreferential relationships that could be covered by extensions of an annotation scheme based on 'strict' coreference (see Section 2.3).

Acknowledgements

The authors wish to thank the following for helpful feedback on earlier drafts of this pa-

per: Adam Kilgarriff, Richard Power, Paul Piwek and colleagues in the GNOME project. Kibble's work was funded by the UK EPSRC as part of this project under grant reference GR/L51126.

Literature

ABPI (1997). *1996-1997* ABPI *Compendium of Patient Information Leaflets.* Association of the British Pharmaceutical Industry.

BAGGA A. ET AL. (1999). Coreference and its Applications. Call for papers for workshop associated with ACL'99. See `www.cs.duke.edu/~amit/acl99-wkshp.html`

MUC (1997). MUC-7 Coreference Task Definition, 13 July 1997. See `www.muc.saic.com`

DAVIES S. ET AL. (1998). Annotating Coreference in Dialogues: Proposal for a scheme for MATE. See `www.cogsci.ed.ac.uk/~poesio/MATE/anno_manual.html`

DOWTY D. ET AL. (1981). Dowty, D., Wall, R. & Peters, S. *Introduction to Montague Semantics.* Dordrecht: Kluwer.

GAMUT L. T. F. (1991). *Logic, Language and Meaning Volume 2.* Chicago and London: University of Chicago Press.

HIRSCHMAN L., ROBINSON P., BURGER J. & VILAIN M. (1997). Automating Coreference: The Role of Annotated Training Data. In *Proceedings of AAAI Spring Symposium on Applying Machine Learning to Discourse Processing.*

KAMP H. & REYLE U. (1993). *From Discourse to Logic.* Dordrecht: Kluwer Academic Publishers.

LYONS J. (1977). *Semantics.* Cambridge: Cambridge University Press.

POESIO, M., R. HENSCHEL, J. HITZEMAN, R. KIBBLE, S. MONTAGUE, AND K. VAN DEEMTER. (1999). Towards an Annotation Scheme for Noun Phrase Generation. To appear in H. Uszkoreit et al. (eds), *EACL-99 Workshop on Linguistically Interpreted Corpora*, Bergen.

POPESCU-BELIS A. (1998). How Corpora with Annotated Coreference Links Improve Reference Resolution. In R. ANTONIO ET AL., Eds., *First Int. Conf. on Language Resources and Evaluation*, p. 567-572. Granada: European Language Resources Association.

POPESCU-BELIS A. & ROBBA I. (1998). Three New Methods for Evaluating Reference Resolution. In *First International Conference on Language Resources & Evaluation: Workshop on Linguistic Coreference.* Granada: European Language Resources Association.

SIDNER C. (1983). Focusing in the comprehension of definite anaphora. InM.Brady and R.Berwick eds. *Computational Models of Discourse.* Cambridge, Mass:MIT Press.