Corpus Annotation and Reference Resolution

A. McEnery, I. Tanaka & S. Botley,
Department of Linguistics,
Lancaster University,
Bailrigg,
Lancaster,
LA1 4YT

email: mcenery@comp.lancs.ac.uk

Abstract

A variety of approaches to annotating reference in corpora have been adopted. This paper reviews four approaches to the annotation of reference in corpora. Following this we present a variety of results from one annotated corpus, the UCREL anaphoric treebank, relevant to automated reference resolution.

Introduction

The application of corpora to the problems of pronoun resolution is a rapidly growing area of corpus linguistics. Work by Dagan and Itai (1990) and Mitkov (1994, 1995, 1996; Mitkov, Choi and Sharp 1995) are good examples of this growth. However, the application of suitably annotated corpora to the problem of pronoun resolution has been largely hampered to date by a lack of availability of suitable corpus resources. This paper is going to review what work has been undertaken in the production of corpora including discourse annotations. We will then show what quantitative data is available from such corpora which can be of use in the construction of robust pronoun resolution systems.

Corpus Annotation

While an increasingly wide range of linguistic analyses (both automatically and manually produced) are becoming available as annotations in corpora, morphosyntactically annotated corpora have long been available, and syntactically annotated corpora are now becoming more readily available too. Examples include the parsed LOB corpus, the Susanne corpus (Sampson, 1995) and the Penn Treebanks. While it is widely perceived that appropriately annotated corpus data is of importance in the study of reference resolution, corpora which include appropriate discourse annotations have not become more readily available in the public domain, however. Evidence for the growing appreciation of the importance of anaphorically annotated corpora can be seen in the slow but sure growth of a range of corpus annotation systems for reference annotation in the 1990s - Fligelstone (1992), Aone and Bennett (1994), Botley (1996), de Rocha (1997) and Gaizauskas and Humphries (1997). Yet while the proposals for an appropriately annotated corpus are growing, there is little corpus data available in English[1]. The only corpus that is available, developed by Aone and Bennett, has a variety of shortcomings - it covers only one genre of written language (newspaper articles), it deals only with anaphora, it is a corpus of Japanese and Spanish[2], and its annotations were not produced to meet the need of a wide range of end-users, only participants in the fifth message understanding competition. Hence the work which has been undertaken with corpora in the field of reference resolution has not been able to exploit and evaluate the type of reliable quantitative data that an anaphorically annotated corpus could yield.

Our aim at Lancaster over the past three years has been to develop a series of tools to retrieve quantitative data on a range of reference features in text. We have done this on the basis of one corpus which was developed in collaboration with IBM Yorktown Heights, and a second which we have developed in-house. Neither of these corpora are available for general release because of restrictions placed upon us by the providers of the corpus text. What we can release, however, are the results

[1] It should be noted that other languages have already started to generate such resources - Aone and Bennett (1994) have been working on such a corpus for Japanese and Spanish.

[2] Aone and Bennett (1994:74) appreciate the importance of extending this work to English.

of our mining of the corpus for quantitative data. This is the first publication aimed at the release of such information.

The uses to which such information may be put in anaphor resolution are obvious. Using quantitative data of this sort may be one means of providing "knowledge poor" reference resolution. Also, as McEnery (1995) suggests, such quantitative data may be used to provide a more restricted search area for knowledge intensive anaphor resolution systems to work in. We believe that the data we present later in this paper shows clearly, for one text type at least, that the use of quantitative data to limit the search space of a reference resolution algorithm is a possibility.

Before we present any such data, however, we will review the limited body of work that exists in the
field of reference resolution-oriented annotation in corpus linguistics. In doing so we will cover two schemes developed at Lancaster University (Botley, 1996, Fligelstone, 1992), and two further schemes developed at Sheffield and Sussex Universities (Gaizauskas and Humphries, 1997, de Rocha, 1997). This type of review is of interest because it shows the range of features that may be annotated in such a corpus, and consequently gives a sense of the type of quantitative data we may hope to extract from an appropriately annotated corpus.

Work to Date

Having established that a range of annotation schemes are being developed to encode anaphoric reference resolution in corpora, we need to review this work here. We need to review it because in doing so one gains a flavour of the types of quantitative data that may become available in the near future from the fruit of such efforts. When we have completed our review of four important annotation schemes, we will go in and look in some detail at the type of data forthcoming from corpora annotated by one scheme, the Lancaster scheme, and assess its potential impact upon practical, robust knowledge poor anaphor resolution.

The UCREL Anaphoric Treebank

This treebank consists of 100,000 words of morphosyntactically-annotated Associated Press Newswire stories, and was developed as part of a collaborative project between UCREL and IBM Yorktown Heights (Garside, 1993). The treebank is marked up using the IBM/UCREL discourse annotation scheme (Fligelstone, 1992). This scheme encodes a wide range of anaphoric and cohesive features based on the typology of Halliday and Hasan (1976). Each feature is associated with annotation symbols which encode the type of relationship involved, and the direction of reference, where relevant.

In addition to these features, it is possible to mark uncertainty (of direction of reference, or of the antecedent), multiple antecedents, and semantic values on second and third-person pronouns. For a more detailed treatment of the annotation scheme, see Fligelstone (1992).

The UCREL discourse annotation scheme was applied to corpus texts using a tailor-built editing tool, called XANADU (Garside, 1993). XANADU is an interactive tool which allows an analyst to rapidly introduce cohesion annotations into corpus texts. The tool was developed as part of the above-mentioned collaborative project between UCREL and IBM Yorktown Heights (Garside, 1993: 5-27).

Evaluation of the UCREL annotation scheme

The UCREL annotation scheme scores very highly in terms of granularity of analysis - it is possible to mark a wide range of cohesive phenomena using the scheme. This means that it is possible to provide a corpus resource which is very rich in data that could be useful to an algorithm for resolving anaphora. However, one area of weakness is that it only works effectively when marking antecedents that are surface linguistic strings, such as noun phrases and clauses. It has been found, especially by Francis (Francis, 1994 as well as Botley (Botley, 1996) that some antecedents are indirectly related to their anaphors. One example of this is where demonstrative anaphors function to encapsulate or label (Francis, 1994) a large stretch of preceding discourse. Using the existing UCREL annotation scheme, the only way of marking such a feature would be to place antecedent annotations around the entire previous text. While it is possible to do this for situations where an antecedent is a single sentence, it is far from certain how to mark an antecedent that

is not a clearly identifiable surface element of the text. Therefore, this class of 'indirect anaphors' are not easily markable using the UCREL scheme. Despite this limitation, it is still eminently feasible to mark surface antecedents.

De Rocha's Work.

In a notation scheme developed by Marco de Rocha (de Rocha, 1997), spoken corpus texts [3] in English and Portuguese are segmented and annotated according to the topic structure of the texts analysed. This approach reflects the widely-accepted view in discourse analysis and text linguistics that the topic of the discourse [4] tends to be the preferred antecedent for a given anaphoric expression. Therefore, de Rocha's annotation is aimed at exploring the complex relationships between anaphora and discourse topic.

Firstly, de Rocha establishes, for each discourse fragment under analysis, a global topic, or **discourse topic**. The discourse topic can be valid throughout a whole text, or may change at different points, in which case, a new discourse topic will be established and annotated. The discourse topic is annotated above the text fragment as a noun phrase within asterisks. The next step is to divide the text into discourse segments according to local topic continuity. This is done by assigning a **segment topic**, which is only valid throughout a given segment of the discourse. Whenever the local topic changes, a new segment topic is assigned, and appropriate annotation is inserted manually into the text. Segment topics are annotated using the letter **s**, followed by an index number, similar to those assigned using the UCREL scheme.

Segments where further local topic shift occurs are further subdivided into subsegments, with their own appropriate annotation, consisting of the string **ss**, followed by an index number as with segment annotations. Also, topics which have been dropped, but have been re-introduced in the conversation are also marked by adding the letter **r** to the **s** or **ss** annotations for discourse segments.

As well as the above segment and subsegment annotations, de Rocha's scheme allows for discourse segments to be annotated according to the discourse function they serve, for instance 'introduce the discourse topic' is annotated using the string **intro_dt**. Also, each annotation string contains a short phrase describing the current topic for the segment or subsegment under analysis.

The final stage in de Rocha's analytical framework is to annotate each case of anaphora that take place within the discourse segments identified. This is done by specifying four properties of anaphora:

1. **type of** anaphora, such as 'subject pronoun' or 'full noun phrase', with each type having its own tag,
2. **type of antecedent**, defined as either implicit or explicit, each of which is tagged separately in the annotation,
3. **topicality status of the antecedent**, in other words whether the antecedent is the discourse topic, segment topic or subsegment topic.
4. **processing slot**, by which anaphora cases can be classified according to the type of knowledge used in processing them, such as syntactic, collocational or discourse knowledge.

Evaluation of de Rocha's annotation scheme.

De Rocha's scheme has a number of innovations. Primarily, it goes beyond annotating anaphoric cases in texts, and attempts to encode information about the relationship between anaphora and topicality in discourse, which goes a long way towards providing annotated corpora that can be used in studies of discourse structure and anaphora. Secondly, rather than simply identifying anaphors and antecedents, it classifies them according to some rigorous criteria which are more detailed than the framework laid down by Halliday and Hasan, which was at the core of the UCREL scheme. Thirdly, de Rocha's scheme is developed for use with spoken dialogues in more than one language, which introduces extra analytical dimensions to the corpus-based analysis of anaphora. And finally, de Rocha introduces information concerning the kind of knowledge used in processing anaphors, which is not included in other schemes, but would be very useful in any research that marries corpus-based description with a knowledge-based approach to anaphor resolution.

[3] de Rocha used extracts from the London-Lund Corpus for his English data.

[4] Topic is known by various terms in the literature, for instance focus (Sidner, 1986) or center (Mitkov, 1994b).

The main disadvantage to de Rocha's scheme is that it does not use a widely-accepted text encoding format in its annotation symbols, a requirement that is becoming increasingly important in modern corpus-based research. The next anaphoric annotation system to be described here does do this, however.

Gaizauskas and Humphries Scheme

Gaizauskas and Humphries (1997) use SGML (Standard Generalised Markup Language) tags to annotate anaphoric expressions in texts used in a coreference resolution task. SGML is becoming a widely-recognised standard for encoding electronic texts for interchange between different computer systems in natural language engineering research.

Evaluation of Gaizauskas and Humphries' annotation

This system has the main advantage of being in a widely-recognised text interchange format. However, it only allows a small subset of anaphoric relations to be marked, in this case, reference involving 'it'. Also, the scheme was developed for use in a rigidly restricted automatic resolution task where the success of each annotation had to be measured. It was not developed for use on a large corpus-based project, as with other annotation schemes described in this chapter. Despite this, however, the SGML framework does provide a useful starting point by which other schemes may be converted to SGML in the future.

Botley's Annotation Scheme

The final annotation scheme to be described here was developed by Simon Botley (Botley, 1996), and, like that of de Rocha, attempts to classify anaphoric expressions according to various external criteria. Botley's scheme was developed to describe the different ways in which demonstrative expressions function anaphorically in written and spoken corpus texts. Essentially, Botley classifies demonstrative anaphors according to five distinctive features, each of which can have one of a series of values: **Recoverability of Antecedent** (the extent to which the antecedent is a recoverable surface string), **Direction of Reference, Phoric Type** (derived from Halliday and Hasan's framework), **Syntactic Function** and **Antecedent Type**.

Each case of demonstrative anaphora in a 300,000 word corpus [5] was annotated with a five-character tag which encoded each of the above values for each of the five features identified.

Evaluation of Botley's annotation scheme.

Like de Rocha, Botley's scheme has the advantage of being able to mark a great deal more information about anaphoric phenomena in the text than the UCREL scheme at present can. Also, it is relatively straightforward to derive statistics concerning frequency of occurrence of particular demonstrative features using the Botley scheme, from which sophisticated statistical modelling can be carried out. Also, the Indirectly Recoverable value allows analysts to home in on areas of demonstrative anaphora which are worthy of further study. It was mentioned above that the UCREL scheme cannot provide much information about those cases of anaphora where the antecedent is not an identifiable surface noun phrase. However, schemes which classify antecedents according to directness or indirectness of recoverability (Botley) or explicitness versus implicitness (de Rocha) are highly valuable and sensitive tools which can help analysts to derive richer descriptions of particular anaphoric features in a corpus.

Findings to Date

Having reviewed the work undertaken on reference oriented corpus annotation to date, we can now present a few examples of the type of data that we have extracted from the anaphoric treebanks held at Lancaster. It should be emphasised that the data we are presenting here is but a sample of the data we have [6], which will be presented fully in Tanaka (1998) and Botley (forthcoming). [NOTE: need to say how many anaphors were detected here] The data all refers to the genre of newswire reporting, using the anaphoric treebank described above. Yet in presenting

[5] Consisting of 3x100-word samples from the Associated Press Treebank, the Canadian Hansard and the American Printing House for the Blind Corpus.

[6] Coming, as it does, from the 100,000 word Anaphoric Treebank.

these samples of data we believe that we are showing at least two things. Firstly, that existing quantitatively oriented studies of reference in English are generally supportable, as far as they go, by reference to corpus data. Secondly, and more importantly, it is possible to go beyond the bounds of existing studies, and provide relevant quantitative data that is not currently available.

Table Two: Distance between referent and antecedent measured by number of intervening sentences.

A	B	C	D	E
0	2073	44.29 %	2073	44.29 %
1	1449	30.95 %	3522	75.24 %
2	487	10.40 %	4009	85.64 %
3	205	4.38 %	4214	90.02 %
4	141	3.01 %	4355	93.04 %
5	88	1.88 %	4443	94.92 %
6	64	1.37 %	4507	96.28 %
7	34	0.73 %	4541	97.01 %
8	37	0.79 %	4578	97.80 %
9	21	0.45 %	4599	98.25 %
10	20	0.43 %	4619	98.68 %
11	19	0.41 %	4638	99.08 %
12	10	0.21 %	4648	99.30 %
13	9	0.19 %	4657	99.49 %
14	7	0.15 %	4664	99.64 %
15	6	0.13 %	4670	99.77 %
16	2	0.04 %	4672	99.81 %
17	5	0.11 %	4677	99.91 %
18	1	0.02 %	4678	99.94 %
19	2	0.04 %	4680	99.98 %
21	1	0.02 %	4681	100.00 %

Key to table: **A** is numbers of intervening sentence boundaries, **B** is number of occurrences, **C** is rate of occurrences, **D** is the sum of occurrences to that point and **E** is rate of sum.

Let us begin with the first case. Ariel (1988) found a relatively normal distribution for distance between anaphor and antecedent in the data she observed. We have tried a variety of distance measures (intervening NPs, intervening words, intervening sentences) and found a very similar distribution in all cases. Table Two and Figure Five below show, in detail and graphically, how sentence distance shows the behaviour of anaphors.

The data in itself is quite remarkable, and shows a variety of points clearly. First, the preoccupation with intra-sentential anaphora in generative linguistics is shown to be unhealthy. Intra-sentential anaphora is shown to be the minority case in this data. Most anaphors in the data sample are inter-sentential. Any anaphor resolution system which dwells upon intra-sentential anaphora at the expense of inter-sentential anaphora is doomed to failure. Second, the behaviour of the anaphors is remarkably uniform. In practical terms, if you accepted that you were prepared to limit your search for an antecedent five sentences distant from an anaphor, then although you would be placing an upper limit on the accuracy of your algorithm of 94.92%, you would be receiving a bonus, in a reduction of some 75% of the potential relevant search space. If we assume that with an increased search space accuracy declines, then quantitatively motivated limitations such as that suggested may boost the success rate of a knowledge intensive system which suffers from declining accuracy and speed with an open ended search space.

Third, the majority of anaphors are either inter-sentential, or occur in the previous sentence. Admittedly, this observation only covers around 75% of cases, but nonetheless, it is indicative of the type of probabilistic information that may be incorporated into search algorithms to aid with selection of an antecedent - if there is no case to choose on rational grounds between an antecedent five sentences away and one in the previous sentence, the one in the previous sentence is fundamentally more likely to be the right one.

Let us now move to our second point - using annotated corpora to derive quantitative data not available currently. When we look at the analysis above, and studies such as Ariel (1988), there is a great deal that this simple distance oriented type of analysis does not show. Do all of the anaphors have the same pattern of distribution, for instance? It may be that this averaging of distances as shown in table one looks quite different when it is broken down by anaphor. Also, how do features other than distance influence reference resolution - for example, how do pronouns in direct quotations behave? Do they have antecedents beyond the scope of the quotation itself?

Table 3: Distance data for each pronoun in direct speech and non-quoted text

Intervening Sentences	0	1	2	3	4	5	6	7	8	9	10	11	12	13	1

															4
I_PPIS1	163	290	31	9	5	3	2	0	0	0	0	1	0	0	0
in DS	163	290	31	9	5	3	2	0	0	0	0	1	0	0	0
my_APP$	44	28	3	1	0	0	0	0	0	0	0	0	0	0	0
in DS	44	28	3	1	0	0	0	0	0	0	0	0	0	0	0
me_PPIO1	27	28	9	1	0	0	0	0	0	0	0	0	0	0	0
in DS	27	28	9	1	0	0	0	0	0	0	0	0	0	0	0
myself_PPX1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
in DS	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
we_PPIS2	80	123	30	8	8	3	5	0	0	0	0	0	0	0	0
in DS	80	123	29	8	8	3	5	0	0	0	0	0	0	0	0
our_APP$	34	18	6	2	1	0	0	0	1	0	0	0	0	0	1
in DS	34	18	6	2	1	0	0	0	1	0	0	0	0	0	1
us_PPIO2	9	17	2	1	3	0	2	0	0	0	0	0	0	0	0
in DS	9	16	2	1	3	0	2	0	0	0	0	0	0	0	0
ourselves_PPX2	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0
in DS	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0
you_PPY	4	10	0	1	1	0	1	0	0	0	0	0	0	0	0
in DS	4	10	0	1	1	0	1	0	0	0	0	0	0	0	0
your_APP$	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0
in DS	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0
yours_PP$	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
in DS	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
yourself_PPX1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
in DS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
yourselves_PPX2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
in DS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
he_PPHS1	494	485	28	7	3	0	1	1	1	0	0	0	0	0	0
in DS	37	75	8	2	0	0	0	0	0	0	0	0	0	0	0
his_APP$	489	69	5	2	0	4	1	0	0	0	0	0	0	0	0
in DS	40	5	0	1	0	2	1	0	0	0	0	0	0	0	0
his_PP$	15	8	0	0	1	0	0	0	0	0	0	0	0	0	0
in DS	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
him_PPHO1	90	25	4	0	0	0	0	0	0	0	0	0	0	0	0
in DS	10	11	3	0	0	0	0	0	0	0	0	0	0	0	0
himself_PPX1	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0
in DS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
she_PPHS1	81	89	5	0	1	0	0	0	0	0	0	0	0	0	0
in DS	2	10	2	0	0	0	0	0	0	0	0	0	0	0	0
her_APP$	85	23	3	0	0	0	0	0	0	0	0	0	0	0	0
in DS	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0
her_PPHO1	25	9	2	0	0	0	0	0	0	0	0	0	0	0	0
in DS	7	3	1	0	0	0	0	0	0	0	0	0	0	0	0
herself_PPX1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
in DS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The ways in which we have elaborated the analysis presented above are numerous. But for the purpose of showing briefly and succinctly the reasons why we may want to do a more detailed analysis of our data, see Table 2 below. In this table, we have analysed sentence distance for each pronoun, both within direct speech (line "in DS") and outside direct speech. For purposes of presentation we have limited the distance to 14 sentences, though as noted previously, a few antecedents lie beyond this range. In the following table, each pronoun has two entries. On the first line, the total number of occurrences of each pronoun are given in each of the distance categories represented. The second line gives the total number of occurrences of each pronoun appearing in direct speech in each distance category represented.

Looking at this table, it is clear that all pronouns do not act alike, and further, that the determination of whether a pronoun is in direct speech or not can be of great practical relevance. Most[7] first and second person pronouns (*I, my, me, myself, we, our, us, ourselves, you, yours*) only ever occur within direct speech in this genre. On examination of the text, their antecedent almost invariably lies beyond the direct quotation itself. Consequently, it is noticeable that these pronouns tend to have longer distance ties than the third person pronouns, as the quotation itself has been included within the distance measure. If we look, for instance, at references occurring at a distance of six sentences, I, *we, us* and *you* occurring in direct speech account for ten of the thirteen cases (the remainder being one each of *he, his* and *they*, of which only his occurs in direct speech).

Overall, the predominant tendency is still towards very short range reference - with most antecedents being within one sentence of the anaphor. There are a very few cases of pronouns with a more substantial tail than others - 6.1% of the 504 cases of *I* observed had an antecedent at two sentences distance, compared, say to the 213 cases of *its*, which has no antecedents more than one sentence distant, and which only had 1.9% (4 cases) at 1 sentence distance..

[7] In our discussion here, we will use the raw freqeuncy data from the tables as the basis of our discussion. A more exacting statistical analysis of the type of data we present here is given in Tanaka(1998) and Botley (forthcoming).

While the overall characterisation of distance suggests that the most populous category of references is intra-sentential followed by references one sentence distant, this pattern is not true for all pronouns. While there is no pronoun observed that prefers antecedents at two or more sentences distant over those occurring within the same sentence, there are anaphors which prefer an antecedent in the previous sentence over an antecedent in the same sentence, adding further weight to our observation that ignoring inter-sentential anaphora is not an option. *I, me, we, us, you* and *she* prefer an antecedent at one sentence distance; intra-sentential anaphora does not seem to be the norm for these anaphors.

In other words, given different pronouns and different circumstances, there are variations in the behaviour of pronouns. Although we are only reporting on one genre here, the work of Botley (1996) looking at determiners in three genres, suggest that genre is another dimension of variation where we may see significant differences in the pattern of distributions of pronouns across various distance measures.

Conclusion

The point of our investigations are to illuminate a variety of features that corpus based pronoun resolution systems may benefit by, and which they must certainly be aware of:

1. Pronoun antecedents do exhibit clear quantitative patterns of distribution
2. Genre may influence those patterns
3. Direct speech is an important factor in explaining some of those patterns
4. Inter sentential pronoun resolution is not always the norm
5. Some patterns of pronoun antecedent distribution are prone to longer tails than others
6. Characterisation of pronoun distributions based on all pronouns distorts the picture which may be observed on the individual pronoun level

An obvious criticism of our work to date is that we have not based a pronoun resolution system upon the data that we have extracted from our corpus (although Tanaka, 1998 does include a report on such a system). The reason we have not done so, is that we are far from convinced that the right corpus resources and the right

type of quantitative data are currently available. As we produce more and more refined data from our corpus, we are seeing patterns of distribution which are masked in more general representations of the data, as we have exemplified. Also, we are only able to produce this data for a severely limited genre of written texts. What we need to do next is work towards a balanced corpus, including both written and spoken language, which would allow us to extract quantitative data similar to that shown in this paper, for a wide range of text types and for spoken language. Our experience to date indicates that while we may observe patterns of usage which are of use and of importance to robust pronoun resolution, that data should at least be extracted on a by genre basis. The compilation of such data is our next research aim.

Bibliography

[Aone and Bennett 1994] C. Aone, S.W. Bennett - Discourse tagging and discoourse tagged multilingual corpora. Proceedings of the International Workshop on Sharable Natural Language Resources, Nara, Japan, 71-77.

[Ariel, 1988] M. Ariel. Referring and Accessibility, Journal of Linguistics 24, 65-87, 1988.

[Botley 1996] SP Botley - Comparing Demonstrative Features in Three Written English Genres. In S P Botley, J Glass, A M McEnery and A Wilson (eds), Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Resolution Colloquium (DAARC96). University Centre for Computer Corpus Research on Language Technical Papers 8 (special issue). pp. 86-105.

[Botley forthcoming] S. Botley. Corpora and Discourse Anaphora, PhD. Thesis, Lancaster University.

[Dagan and Itai 1990] I. Dagan, A. Itai - Automatic processing of large corpora for the resolution of anaphora references. Proceedings of the 13th International Conference on Computational Linguistics, COLING'90, Helsinki, 1990

[Fligelstone 1992] S. Fligelstone. Developing a scheme for annotating text to show anaphoric relations. In Leitner, G. (ed), New directions in English language corpora. Methodology, results, software developments. Berlin: Mouton de Gruyter, pp.153-70.

[Francis, 1994] G. Francis. Labelling discourse: an aspect of nominal-group lexical cohesion (1994) in Coulthard, M: Advances in Written Text Analysis, Routledge, 1994.

[Gaizauskas and Humphries 1997] R. Gaizauskas and K. Humpreys - Quantitative evaluation of coreference algorithms in an information extraction system. In S.P. Botley and A.M. McEnery (eds) Corpus-Based and Computational Approaches to Discourse Anaphora, UCL Press, (forthcoming)

[Garside 1993] R. Garside, The Marking of Cohesive Relationships: Tools for the Construction of a Large Bank of Anaphoric Data. ICAME Journal 17, 5-27.

[Halliday and Hasan, 1976] M. Halliday and R. Hasan. Cohesion in English, Longman 1976.

[McEnery 1995] A.M. McEnery, Computational Pragmatics, PhD Thesis, Lancaster University.

[Mitkov 1994] Mitkov R. - An integrated model for anaphora resolution. Proceedings of the 15th International Conference on Computational Linguistics COLING'94, Kyoto, Japan, 5-9 August 1994

[Mitkov 1995] R. Mitkov - An uncertainy reasoning approach to anaphora resolution. Proceedings of the Natural Language Pacific Rim Symposium, 4-7 December 1995, Seoul, Korea

[Mitkov 1996] R. Mitkov - Two engines are better than one: generating more power and confidence in the search for the antecedent. In R. Mitkov, N. Nicolov (Eds) Recent Advances in Natural Language Processing , John Benjamins (forthcoming)

[Mitkov, Choi and Sharp 1995] R. Mitkov R., S.K. Choi, R. Sharp - Anaphora resolution in Machine Translation. Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation, Leuven, Belgium, 5-7 July 1995

[de Rocha 1997] M. de Rocha - Corpus-Based Study of Anaphora in English and Portuguese. In S.P. Botley and A.M. McEnery (eds) Corpus-Based and Computational Approaches to Discourse Anaphora, UCL Press, (forthcoming)

[Sampson, 1995] G. Sampson. English for the Computer, Clarendon Press, Oxford, 1995.

[Tanaka forthcoming] I. Tanaka. Exploiting an Anaphoric Treebank, PhD. Thesis, Lancaster University.