NPs for Events: Experiments in Coreference Annotation

Laura Hasler[*], Constantin Orasan[*], Karin Naumann[†]

[*]Research Group in Computational Linguistics, SHLSS, University of Wolverhampton
Stafford Street, Wolverhampton, WV1 1SB, UK
{L.Hasler, C.Orasan}@wlv.ac.uk
[†]Seminar für Sprachwissenschaft, Universität Tübingen, Abt. Computerlinguistik
Wilhelmstraße 19, 72074 Tübingen, Deutschland
knaumann@sfs.uni-tuebingen.de

Abstract

This paper describes a pilot project which developed a methodology for NP and event coreference annotation consisting of detailed annotation schemes and guidelines. In order to develop this, a small sample annotated corpus in the domain of terrorism/security was built. The methodology developed can be used as a basis for large-scale annotation to produce much-needed resources. In contrast to related projects, ours focused almost exclusively on the development of annotation guidelines and schemes, to ensure that future annotations based on this methodology capture the phenomena both reliably and in detail. The project also involved extensive discussions in order to redraft the guidelines, as well as major extensions to PALinkA, our existing annotation tool, to accommodate event as well as NP coreference annotation.

1. Introduction

The computational treatment of coreference has recently become an important topic in Natural Language Processing (NLP). A wide range of applications including question answering, information extraction and multi-document summarisation benefit from coreference information. Progress in the interpretation of coreference of noun phrases (NPs) and events depends on the availability of suitable annotated corpora. In order to build such corpora, appropriate guidelines and schemes which allow the annotation of data need to be formulated.

To date, there exist several resources related to noun phrase and event coreference, but these are still relatively few. There are a number of small corpora annotated for within-document NP coreference (e.g. Ge, 1998; Mitkov et al., 2000). Other resources related to coreference and event annotation do not concentrate solely on these annotations, and include the TimeBank corpus (Pustejovsky et al., 2003b) and the corpus developed in the ACE program. Several annotation schemes have been developed to annotate existing resources, but an investigation showed that none of these were completely appropriate for our task.

This paper reports the efforts of a pilot project which investigated NP and event coreference[1]. The main objective of this project was to develop a methodology, consisting of detailed annotation schemes and guidelines, for the marking of NP and event coreference within documents. In order to develop the guidelines and schemes, a sample annotated corpus in the domain of terrorism/security was built. This methodology can be used as a basis for large-scale annotation to produce much-needed resources in the future. In contrast to other annotation projects, this project focused almost exclusively on the development of guidelines and schemes for the annotation of NP and event coreference, which should ensure that future annotations based on this methodology capture the phenomena both reliably and in detail. The project involved extensive discussions in order to redraft and improve the guidelines, as well as major changes to enable our existing annotation tool PALinkA (Orasan, 2003), to accommodate events as well as NPs.

The rest of this paper is structured as follows: Section 2 details existing work related to NP coreference and event annotation. A brief overview of the project is given in Section 3. A more detailed discussion of our annotation schemes and guidelines, along with issues arising during the annotation of NP coreference is presented in Section 4. Section 5 deals with the same aspects of our event annotation. Finally, Section 6 presents our conclusions.

2. Related Work in Coreference and Events

As mentioned above, there are a limited number of existing annotation schemes and guidelines both within and related to the field of coreference and events. There are currently more of these resources for NP coreference than there are for event coreference.

2.1. NP Coreference

In recent years, a number of annotation schemes for marking up anaphora and within-document coreference have been proposed. The most well-known schemes are the UCREL anaphora annotation scheme applied to newswire texts (Fligelstone, 1992; Garside, Fligelstone and Botley, 1997) and the MUC annotation scheme used in the MUC-7 coreference task (Hirschman, 1997). Other well known schemes include de Rocha's (1997) scheme for spoken Portuguese, Botley's (1999) scheme for demonstrative pronouns, Bruneseaux and Romary's (1997) scheme, DRAMA (Passonneau and Litman, 1997), Poesio and Vieira's (1998) scheme for definite noun phrases, the MATE scheme (Davies et al., 1998) for coreference in dialogues, and a MUC-based annotation scheme for technical manuals (Mitkov et al., 2000).

The UCREL annotation scheme is one of the most comprehensive schemes used for NP coreference, allowing annotators to mark a variety of phenomena, including the direction (anaphoric or cataphoric) and type of relation, various semantic features between referential expressions and the annotator's certainty. The main drawback of the UCREL scheme is that it did not use

[1] All the resources developed in the project can be found on the project web page at http://clg.wlv.ac.uk/projects/NP4E

SGML or XML for encoding, making it difficult for other researchers to use.

The majority of other annotation schemes use SGML or XML for encoding, and are based on the MUC-7 coreference task annotation scheme (Hirschman, 1997), focusing only on the relation of identity between NPs and ignoring other relations such as part-of and set-membership. The scheme was designed for developing annotated corpora for the automatic evaluation of coreference resolvers, and therefore has features specific to this aim, including the MIN attribute which indicates the minimum element to be matched for a correct resolution (often the head). Annotators can also mark optional elements in the chain. Despite some criticism (van Deemter and Kibble, 1999), the MUC scheme has proved a useful starting point for the standardisation of different annotation schemes.

2.2. Event Annotation

Whilst there is a limited amount of research related to event annotation and event coreference, there are not many annotation schemes or sets of annotation guidelines which are dedicated exclusively to events. Bagga and Baldwin (1999) report preliminary experiments for cross-document event coreference, but the focus is not on the development of annotation schemes and guidelines. Setzer and Gaizauskas (2000; 2002) are concerned with accurately positioning events in time and describe a scheme for the temporal annotation of events. However, they focus on temporal annotation and their event categories are not narrow enough for our annotation domain. Closely related to this is TimeML (Pustejovsky et al., 2003a), a specification language developed to capture temporal aspects of events. Guidelines are available which describe how to annotate text according to the TimeML language, and the TimeBank corpus (Pustejovsky et al., 2003b) was annotated using an early version of TimeML, which has since been extended. Similar to the work by Setzer and Gaizauskas (2000; 2002), TimeML concentrates on temporal information using fairly general event categories, making it unsuitable for our annotations.

The existing resources most closely related to our project have been produced in the ACE program, the aim of which is to develop automatic content extraction technology to support the automatic processing of texts. The corpora developed contains annotations of events, values, relations and entities for the evaluation of systems which recognise these phenomena for extraction. ACE annotates eight types of events (LIFE, MOVEMENT, TRANSACTION, BUSINESS, CONFLICT, CONTACT, PERSONNEL, JUSTICE), each with their own sub-types, totalling 33. For each sub-type, the trigger, the polarity, modality, genericity and tense of each event trigger, arguments for each trigger and coreference between triggers are marked. Arguments are participants and attributes associated with a particular event trigger, which differ from event to event. They can only be taken from the same sentence as the event trigger, which means that the amount of information captured for an event mention can be restricted. Coreference is mentioned in the ACE event annotation guidelines, but only briefly, annotators being instructed to mark coreference between two definite mentions of an event which refer to the same event.

3. Description of Project

In contrast to existing work, the project described here focused almost exclusively on the development of appropriate, detailed annotation schemes and guidelines for NP and event coreference, annotating a small sample corpus and refining the guidelines and schemes after a number of attempts at annotation in order to achieve this. Some aspects of our annotations, for example the idea of event categories and arguments, were similar to those in the ACE program, although our number of categories was smaller due to the sample nature of our corpus. We also dealt with certain aspects of events, such as modality and pronominal triggers, differently. Our treatment of NP coreference also differed from the ACE annotations, as we adapted the guidelines proposed by Mitkov et al. (2000).

3.1. Corpus

The goal of the project was to develop a set of annotation guidelines for NP and event coreference for newswire texts in the domain of terrorism/security. To this end, a set of documents were selected by indexing part of the Reuters corpus (Rose, Stevenson and Whitehead, 2002) using the ht://dig search engine. We first selected possible topics for clusters by retrieving documents containing words such as *bomb, explosion* and *kidnap*. Once several topics had been identified, ht://dig was used to retrieve more documents related to these topics. The reason for selecting clusters was the possible extension of this research to cross-document coreference, which would need sets of related documents. Five of the clusters selected were used: Bukavu bombing, Peru hostages, Tajikistan hostages, Israel suicide bomb and China-Taiwan hijack. The sample annotated corpus contains approximately 50,000 words and was built in order to redraft and improve our guidelines.

3.2. Annotations

In the longer-term, the resources produced in this project go some way towards developing corpora which can be used by researchers to train and test automatic methods for event processing. This meant that we needed to strike a balance between useful annotations and practical issues relating to the performance of computers in identifying certain phenomena. We therefore annotated not only information about events, but also coreferential links between NPs. This is because event arguments tend to be NPs forming part of a longer coreferential chain, and so if NP coreference information is available, it indicates that not only the marked NP can fill an argument slot, but also other elements from that coreferential chain.

3.3. PALinkA

The annotation was carried out using the multipurpose annotation tool PALinkA (Orasan, 2003), which has previously been employed in discourse annotation tasks such as marking NP coreferential chains, making it highly suited to the NP coreference task. PALinkA had to be significantly extended in order to accommodate the event annotation, the main extensions being related to the way argument slots of events are filled. It was previously only possible to annotate one link between two markables, but for events there often needed to be more than one.

A newly added plug-in support facilitated the annotation by allowing users to run programs which helped them in the annotation process, such as a plug-in to query WordNet (Fellbaum, 1998) about the relationship between two concepts (see Section 4.1). A plugable previewer for the annotation was another change made to PALinkA, due to the fact that the original way of displaying the markables and relations between them (a tree) was appropriate for NP but not event coreference.

4. Annotation of NP Coreference

The first stage of the project was to formulate an annotation scheme and guidelines for NP coreference. These were adapted from existing resources developed at the University of Wolverhampton (Mitkov et al., 2000). The progress of the NP annotation was important as certain aspects of the event annotation depended on it.

4.1. What Was Annotated and How

The aim of the first stage of annotation was to mark up coreferential links between noun phrases. Annotators first had to identify all the markables (NPs) in a text, regardless of whether they were coreferential or not, because the NPs marked in this phase would be used in the next stage of event annotation, where they would fill slots for arguments associated with events. As a coreferential relation is not a prerequisite for the assignment of an NP as an argument of an event trigger, it was important to mark all NPs. For the annotation of coreference, we offered the annotator the option to use the **coref** or the **ucoref** tag. The **coref** tag is for use where there is no doubt that one entity corefers with another, whilst **ucoref** should be used when the annotator is relatively sure of coreference but there is an element of uncertainty. This is useful because it gives a more fine-grained distinction than just one tag. For example, in the following sentence the verb *argue* may add uncertainty to the "objectivity" of the NP *the masterminds of the bombing plot*:

[*The government*] *will* ***argue*** *that…* [[***McVeigh***] ***and*** [***Nichols***]] *were* [***the masterminds of*** [***the bombing plot***]]

In addition to tagging markables and any coreferential relations between them, we also wanted to capture more detailed information within the coreferential links. PALinkA was adapted so we could annotate the type of relation between an NP and the antecedent with which it corefers. The list of possible relations comprised **identity**, **synonymy**, **generalisation**, **specialisation** and **other**. In order to ease this task for the annotators, a plug-in was developed which allows WordNet (Fellbaum, 1998) to be consulted. We point out that the terms generalisation and specialisation as used here are concerned with lexical choice and detail rather than concept. This is an important distinction as generalisation and specialisation of concept (e.g. *the house…the door*) are used in indirect anaphora, which we did not consider coreferential for our purposes. We use the terms to denote the level of detail present in one NP in relation to another with which it corefers.

A tag to encode the type of coreference was also added, as we wanted to distinguish the different forms of coreferential relations in the corpus. The list of possible types included **NP**, **copular**, **apposition**, **bracketed text**, **speech pronoun** and **other**. The relation between two NPs and the type of coreference was marked as attributes of the **coref** and **ucoref** tags. There was also the option for the annotator to comment on the annotation to note any problems, uncertainties or general observations.

4.2. Annotation Guidelines

We took the set of guidelines already developed at the University of Wolverhampton (Mitkov et al., 2000) as our starting point, and adapted them as a result of a manual analysis of several texts from our corpus. Two annotators performed the annotation, which involved frequent discussions and revisions to improve the guidelines.

The first step was to annotate all suitable NPs as markables. Individual elements within coordinated NPs were treated as markables, as were numerals, dates and quantified NPs. Possessive pronouns and other possessors, interrogative pronouns functioning as possessives but not as relative pronouns, reciprocal pronouns and gerunds which are true nominalisations of verbs were also considered markables. The annotators were instructed to mark NPs at all levels, both definite and indefinite, from base to complex and co-ordinated, including all modifiers and embedded NPs within a larger NP, and including all the noun phrases involved in events, regardless of whether or not they were coreferential. For example:

[*Three Israeli women killed by* [*a suicide bomb in* [*a Tel Aviv café*]] *on* [*Friday*]]…

We did not annotate relative pronouns or relative clauses as markables, but this could be addressed in future work (see section 4.3). Gerunds functioning as verbs were not annotated, along with NPs which are part of fixed expressions and other multi-word lexemes because these NPs depend on the whole expression for their meaning. We also stated explicitly that *here* and *there* should not be marked because they are not NPs and they appear relatively frequently to refer to the place of an attack or incident, especially in direct speech.

One important decision was to determine the definition of coreference to be used. Following van Deemter and Kibble (1999), reflected in Mitkov et al. (2000), we used a narrow definition to ensure higher quality and reliability of annotation. We did not consider indefinite NPs as coreferential with an antecedent. Nor did we consider identity-of-sense anaphora or indirect anaphora between anaphors and antecedents as coreferential. We did not annotate bound anaphors as coreferential, or relations that could be "potentially" regarded as coreferential as they are not truly coreferential at all times. The texts used to formulate our guidelines contained some instances of uncertainty as to the "objectivity" of relations, so it was important to mention this explicitly in our guidelines. This led to the addition of the **ucoref** relation (see section 4.1).

Different readings of the same NP occurred frequently, the most common case being illustrated by the use of the name of a country both as a geographical entity and a governmental/authoritative entity. This is exemplified by the name of a country often being used both to represent an authority issuing orders and commenting on situations, and a place where attacks/fighting happen. Therefore the guidelines include the instruction not to annotate different readings of an NP as coreferential. In the following example, the two mentions of *China* should not be marked as coreferential as they refer to two different entities:

[*A jobless Taiwanese journalist who commandeered* [*a Taiwan airliner*] *to* [***China***]]… [***China***] *ordered* [[*its*] *airports*] *to beef up* [*security*]…

When marking coreference, definite NPs referring to the same entity in the real world were the only possible candidates. The one exception to this rule is that we allowed indefinite NPs occurring at the beginning of a text which refer to an NP appearing in the headline to be marked as coreferential with that NP. The manual analysis highlighted this point, which is due to the newswire genre of our corpus. In the following example, the NPs *Blast*, *an explosion* and *the blast* would be marked as coreferential:

[***Blast***] *kills* [*man*] *in* [*Kinshasa airport*] (headline)

[***An explosion***] *killed* [*one man*]…[***the blast***]…

Annotators also needed to assign tags describing the relation and the type of text by which the coreferential relation is realised (see section 4.1). For most coreferential links, there were two options for the selection of the antecedent: the first mention or the nearest mention. The way the coreferential relation is realised determines which of the antecedents the anaphor is linked to. For NPs annotated with the tag NP (including non-speech pronouns), the anaphor should be linked back to the first mention of its antecedent in the text. For all the other tags: copular, apposition, bracketed text and speech pronouns (pronouns which occur in direct speech), the anaphor should be linked back to the nearest mention of the antecedent in the text. This decision was based on the manual test analysis, where it was discovered to be easier and more intuitive for the annotator to link them as such. There are no practical implications for this, as once an anaphor is linked to any mention of its antecedent, it will appear in the coreferential chain for that entity.

There were a number of instances of direct speech containing the personal pronouns *I*, *we* and *you*. We decided that these should be marked as coreferential with their antecedents otherwise information could be missed. These pronouns should not be marked as the first element in a coreferential chain, so this means that they can indicate a cataphoric reference, which currently does not have a separate tag. It was decided that the pronoun *we* should be taken to refer to the organisation/group that the person is speaking on behalf of, as it is the views of the organisation/group as opposed to the individual which tend to be expressed in these cases. For example, here, *We* should be annotated as coreferential with *IATA*:

[[***IATA***]*'s director of security services*] *said, "*[***We***] *consider that* [[*Aeroflot*]*'s air security measures*] *correspond to* [*international standards*]*."*

4.3. Issues Arising During NP Annotation

Most of the annotation issues arising were resolved during the annotation and are reflected in the final set of guidelines. However, there were a number of points which we found inappropriate to change during the annotation.

The first issue discussed was the marking of relative pronouns. Initially we decided not to mark these because they could refer to more than just an NP. However, after discussions and the annotation of a number of texts, it became clear that these pronouns should be annotated. This should be addressed in future work. Another issue was which antecedent the pronoun *we* in direct speech should be linked to: the individual speaker, the group being represented by the speaker and *we*, or nothing. This needs further discussion and should be determined before any future annotations. There were also discussions about general concepts such as *violence*, *terror*, *terrorism*, *police*, *rebels*, *militants* etc. In several texts, these are used in a general sense and it is difficult to decide whether they should be annotated as coreferential and if so, with what and how. The annotators agreed that there should be some recognition that if a general concept is mentioned throughout the text then there should be some way of encoding that the same concept is under consideration, but this is difficult because coreference is much more specific.

There were several cases in the corpus where there were two correct possible antecedents for a coreferential anaphor in the text. Two indefinite NPs were followed by a definite NP, but all three referred to the same entity. It needs to be decided which antecedent is most appropriate to mark: the first, because it is the first mention, or the second because it is the nearest mention. Or should we ignore these factors and take into account the level of detail or amount of information offered by the possible antecedents? This issue remains to be decided, but however this is dealt with, the coreferential chain will lack one NP because both are indefinite. For example:

…the man detained for hijacking [***a Taiwanese airliner***]*… Liu forced* [***a Far East Air Transport domestic plane***]*… Beijing returned* [***the Boeing 757***]…

5. Annotation of Event Coreference

The annotation of event coreference followed the NP annotation, using markables identified in that stage. The most suitable existing guidelines related to our project were from the ACE program (see section 2.2), which were consulted during the development of our guidelines.

5.1. What Was Annotated and How

The aim of this second stage of annotation was to annotate coreference between events. We annotated a set of events falling into five categories representative of our domain of terrorism/security. An annotation scheme to encode all possible information about an event was developed, so that we did not just annotate the word best expressing the event (the event **trigger**) and omit vital information which can help identify coreference between event mentions.

Our event annotation scheme encodes information about the event **trigger**, including the trigger **type** (verb, noun, adjective, pronoun), **polarity, tense** (past, present, future, unspecified) and **modality**, including the **modality indicator**. Polarity was set by default to YES, and annotators needed to specify if this was different. Modality was set to NO, with annotators specifying if it was YES. If annotators selected YES for modality, they also needed to insert the word which indicated modality.

As well as this more general information about the trigger, the scheme also captures information about the event **category** (ATTACK, DEFEND, INJURE, DIE, CONTACT) and its **arguments**. Arguments vary with each event category, with the exceptions of TIME and PLACE, and include, amongst others, ATTACKER, MEANS, VICTIM, CAUSE, AGENT, TOPIC and MEDIUM, each containing more specific labels[2].

As in the NP annotation, the scheme also encodes coreferential links between event triggers referring to the

[2] Due to space restrictions we cannot describe all our arguments and their sub-parts here. They are fully detailed in our annotation guidelines, available at http://clg.wlv.ac.uk/projects/NP4E

same event occurring in the real world and offers annotators the option to comment on the annotation.

5.2. Annotation Guidelines

The ACE program has developed a set of guidelines for the annotation of events in their corpora. However, due to differences between their annotations and ours, they are not completely suitable for our annotation. Some aspects of the ACE guidelines proved useful for this project, so we adapted these for our purposes. To formulate our guidelines, as in the NP annotation stage, there was a manual analysis of texts from the corpus. Two annotators performed the annotation, and the process included discussions and revisions to improve the guidelines.

The first task was to decide on our definition of **event**. In our annotation, we consider an event to be something that happens or takes place; a single specific occurrence, either instantaneous or ongoing, that is unique and can be anchored at a point in time. This definition does not include repeated or generic events triggered by words such as *terrorism*. In the following example, we do not consider any of the underlined words a markable event:

Police attempts at a crackdown have only worsened the violence, in which gangs attack and burn property…

We did not mark all events appearing in the texts, just five categories relevant to our domain of terrorism/ security. ATTACK events are physical actions aiming to cause harm/damage to things/people. In DEFEND events, people/organisations defend something, often against someone/something else. They include self-defence and escape. INJURE events involve people experiencing physical harm. DIE events signal the end of a person's life. CONTACT events occur when parties communicate in order to negotiate, resolve something, reach an agreement, better relations, etc. They also include threats, demands and promises made during negotiations.

Following ACE, we annotated a **trigger**, and then added further information to this trigger by assigning event-dependent arguments taken from the surrounding context. Events can be lexicalised as verbs, nouns, adjectives and pronouns, and we took these as our set of **trigger types**. More than one trigger can appear in a sentence, and it is important to recognise this, especially where they trigger events that can be seen as related, for example, in ATTACK and DIE events. Triggers were generally taken to be one word, often the head of a verbal group or noun phrase. However, the annotators were instructed not to split fixed expressions and other multi-word lexemes, and also to include pre-modifiers in noun triggers. Examples of event categories and trigger types:

{*The blast*} **{*killed*}** *168 people…and* **{*injured*}** *hundreds more…* (ATTACK: noun, DIE: verb, INJURE: verb)

The **{*dead*}** *man was a retired employee of the state telecommunications company* (DIE: adjective)

*"…***{*it*}** *doesn't look like an accident."* (ATTACK: pronoun - previous mention *the blast*)

As **{*Friday's talks*}** *got underway…* (CONTACT: noun)

The military says it is **{*reinforcing*}** *Zaire's third city of Kisangani…* (DEFEND: verb)

Adjectives were included as triggers, because although they do not trigger an event in the strict sense, they do signal a state resulting from a past event, and can therefore be seen as a trigger. The guidelines state that plural NPs should not be marked as triggers, as they often signify repetition. The exception to this is in the CONTACT event category, where nouns such as *talks* can signal a single event. Pronoun triggers were treated differently to others because rather than triggering an event itself, they refer to an event trigger already present in the text. Pronoun triggers do not have argument slots, instead they are linked to their antecedent, meaning that the arguments for the antecedent are also available for the pronoun.

Event arguments are usually NPs marked in the previous stage of annotation and should preferably fall within the same sentence as the trigger, although this is not crucial, especially if an argument in the surrounding sentences adds more information about the event. The fact that we re-used the NP annotations as arguments for event triggers meant that the annotators could concentrate more on the events themselves. If, however, an NP functioning as an argument was marked incorrectly or not marked at all, the annotator could easily create another markable.

Arguments are split into two categories: **participants**, which take part in the event, and **attributes**, which are related to the event but are not strictly participants. We used the attributes TIME and PLACE for each event category, but participants and other attributes differ depending on the event category. It was possible for annotators to mark more than one argument per slot. Not all argument slots needed to be filled for each event trigger, as not all the necessary information always falls within a reasonable distance of the trigger.

The main issue relating to event coreference was whether one event was only part of another, or if it was the same event. This was especially noticeable in texts containing several CONTACT events, where it was sometimes difficult to establish whether one set of talks was coreferential with or part of another mention. Unlike NP coreference, we also allowed an indefinite mention of an event to be marked as coreferential with another mention of exactly the same event. The first example illustrates a coreferential relation, but the second does not:

On January 30, a gunman **{*shot*}** *and killed a bicycle vendor in* **{*an attack*}** *in San Sebastian…*

…a member of a mediating panel overseeing talks… a second round of crisis talks…

5.3. Issues Arising During Event Annotation

There were many discussions both before and during the annotation relating to what should constitute an event and an event trigger, as well as issues about event arguments regarding how best to label these to ensure we captured the maximum possible relevant information. There were also discussions on how restrictive our event categories should be, and which triggers in our corpus reflected what event category. The current annotation guidelines and scheme reflect most of our decisions. However, there are still several outstanding issues.

Two issues regard marking the tense of event triggers. We currently annotate the tenses of all event triggers according to the speaker/writer of the document, which takes into account the time of production of the article. However, many texts contain direct speech where events are reported in the present tense, although in relation to the production time of the article they are in the past. Discussions proved that it might be useful to indicate that

certain events are reported in the present tense because they happen at the time of speaking. Related to this is the use of the present tense to report past events in headlines of newswire texts. This is not as important because it is a matter of style rather than text production time.

There was much discussion about the inclusion of demands, threats and promises in our CONTACT category. Finally, it was decided that these should be included as events in their own right rather than just to signal modality, because omitting them would have also omitted many central events in the corpus. Related to this is the issue of including a speaker/hearer or producer/ recipient distinction when demands, threats, promises, etc. are made. Currently we only annotate the producer as an argument, because otherwise it is difficult to distinguish the producer from the receiver. In future it would be better to add to a slot to include this information.

Another issue arising during the annotation was how to mark participles functioning as triggers. There were several cases of these, mainly as INJURE and DIE triggers, such as *one man was* **{*killed*}**. These are currently annotated as a verb, but in future it may be feasible to add a category to incorporate this as a separate type of trigger.

6. Conclusions and Future Work

This paper reported a pilot project which developed detailed and appropriate sets of guidelines and schemes for the annotation of NP and event coreference, with a view to providing resources suitable for use in developing corpora on which to train and test automatic methods for event processing in the longer term. We have given an overview of the schemes and guidelines developed, as well as of the discussions and issues arising from the annotation, with reference to the texts in the small sample corpus developed in order to achieve the project goals. One practical aspect of our project was to minimise the annotation load by employing NPs annotated in the first phase as arguments for events annotated in the second phase. In addition, the schemes, guidelines and annotations can pave the way for the annotation of cross-document coreference, of both NPs and events. The project proved that the annotation of events, and even of within-document NP coreference, is not a trivial task.

7. Acknowledgements

The research reported in this paper was jointly funded by the British Academy through the project "Annotation of Cross-Document Coreference: A Pilot Study", and the British Council through the "SYNCAR" project.

8. References

ACE. http://www.itl.nist.gov/iaui/894.01/tests/ace/

Bagga, A., Baldwin, B. (1999). Cross-document event coreference: annotations, experiments and observations. In *Proceedings of the ACL'99 Workshop on Coreference and its Applications*. pp. 1-8.

Botley, S. (1999). Corpora and discourse anaphora: using corpus evidence to test theoretical claims. PhD Thesis. University of Lancaster.

Bruneseaux, F., Romary, L. (1997). Codage des références et coréférences dans les dialogues hommemachine. In *Proceedings of ACH-ALLC '97*. pp. 15-17.

Davies, S., Poesio, M., Bruneseaux, F., Romary, L. (1998). Annotating coreference in dialogues: proposal for a scheme for MATE. First draft. Available at http://www.hcrc.ed.ac.uk/~poesio/MATE/anno_manual.html

de Rocha, M. (1997). Supporting anaphor resolution with a corpus-based probabilistic model. In *Proceedings of the ACL'97/EACL'97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution*. pp. 54-61.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fligelstone, S. (1992). Developing a scheme for annotating text to show anaphoric relations. In G. Leitner (Ed.) *New Directions in English Language Corpora: Methodology, Results, Software Developments*. Berlin: Mouton de Gruyter, pp. 153-170.

Garside, R., Fligelstone, S., Botley, S. (1997) Discourse annotation: anaphoric relations in corpora. In R. Garside, G. Leech, A. McEnery (Eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman, pp. 66-84.

Ge, N. (1998). *Annotating the Penn Treebank with coreference information*. Internal report, Department of Computer Science, Brown University.

Hirschman, L. (1997). *MUC-7 coreference task definition*. Version 3.0

Mitkov, R., Evans, R., Orasan, C., Barbu, C., Jones, L., Sotirova, V. (2000). Coreference and anaphora: developing annotating tools, annotated resources and annotation strategies. In *Proceedings of DAARC2000*. pp. 49-58.

Orasan, C. (2003). PALinkA: A highly customisable tool for discourse annotation. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue, ACL'03*. pp. 39-43.

Passonneau, R. J., Litman, D. L. (1997). Discourse segmentation by human and automated means. *Computational Linguistics* 23(1), pp. 103-139.

Poesio, M., Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics* 24(2), pp. 183-216.

Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., Katz, G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of IWCS-5, Fifth International Workshop on Computational Semantics.*

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizuaskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., Lazo, M. (2003). The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics 2003*. pp. 647-656.

Rose, T.G., Stevenson, M., Whitehead, M. (2002). The Reuters Corpus Volume 1 - from Yesterday's News to Tomorrow's Language Resources. In *Proceedings of LREC2002*. pp. 827-833

Setzer, A., Gaizauskas, R. (2000). Annotating Events and Temporal Information in Newswire Texts. In *Proceedings of LREC2000*. pp. 1287-1293.

Setzer, A., Gaizauskas, R. (2002). On the Importance of Annotating Event-Event Temporal Relations in Text. In *Proceedings of the Workshop on Annotation Standards for Temporal Information in Natural Language, LREC2002*. pp. 52-60.

van Deemter, K., Kibble, R. (1999). What is coreference and what should coreference annotation be? In *Proceedings of the ACL'99 Workshop on Coreference and its Applications*. pp. 90-96.