The MATE/GNOME Proposals for Anaphoric Annotation, Revisited

Massimo Poesio
University of Essex
Department of Computer Science and Centre for Cognitive Science
United Kingdom

Abstract

In the five years since it was proposed, the MATE scheme for anaphoric annotation has been used in a variety of annotation projects, and the resulting corpora have been used to study both anaphora resolution and NL generation. Annotation tools inspired by the proposals have been used in some of these projects. In this paper we discuss these first experiences with the scheme, some lessons that have been learned, and suggest a few modifications.

1 Introduction

The MATE 'meta-scheme' for anaphora annotation (Poesio et al., 1999) is one of the annotation schemes developed as part of the MATE project (McKelvie et al., 2001), whose goal was to develop annotation tools suitable for different types of dialogue annotation. The scheme has served as the basis for a number of annotation projects, such as the development of the GNOME corpus (Poesio, 2000a) and, more recently, of the VENEX corpus of anaphora in Italian spoken dialogue and text (Poesio et al., 2004a). The GNOME corpus has been used to study salience, particularly as formalized in Centering theory (Poesio et al., 2004c), to develop statistical models of natural language generation (e.g., (Poesio, 2000a; Henschel et al., 2000; Cheng et al., 2001; Cheng, 2001; Karamanis, 2003)) and to evaluate anaphora resolution systems, with a special focus on the resolution of bridging references (Poesio, 2003; Poesio and Alexandrov-Kabadjov, 2004; Poesio et al., 2004b). Aspects of the scheme have been implemented in annotation tools including MMAX (Müller and Strube, 2003) and the Annotator tool developed by ILSP. As a result of this work, many aspects of the proposals concerning anaphoric annotation made in MATE and GNOME have been subjected to a thorough test. In this paper we discuss some of the lessons learned through this work, some issues that have been raised, and how they have been or could be addressed.

2 The MATE Proposals

The design of an annotation scheme involves a number of decisions: what has to be annotated, how, and how the annotation should be recorded (the markup scheme). One of the most important motivations behind the design of the MATE proposals for anaphoric annotation is the belief that given the variety of phenomena that go under the name of anaphora, and the variety of possible applications, there can be no such thing as a general-purpose anaphoric annotation instructions. On the other hand, we also believed that it *is* possible to design a general purpose markup scheme (and therefore, general-purpose tools) that could then be used in different ways for different projects. The approach taken in MATE was then to design a general markup scheme (the 'meta-scheme') and then to show its basic building blocks could be used to implement different types of anaphoric annotation, including some of the most popular schemes for 'coreference annotation,' such as the MUC scheme (MUCCS) (Hirschman, 1998), Passonneau's DRAMA scheme (1997) , and the scheme used for annotation of references to landmarks in the MapTask corpus. In this section we summarize the most distinctive features of the proposals resulting from this basic assumption. The full description of the MATE scheme is available from the MATE project pages at `http://mate.nis.sdu.dk/`.

2.1 Coreference, Anaphora and Discourse Modeling

The MATE scheme differs from the best-known scheme for annotating 'coreference,' MUCCS (Hirschman, 1998) both in the conceptualization underlying the annotation (i.e., what type of information should be annotated) and in the way this information is marked up. MUCCS was designed to encode information deemed useful for a sub-

task of information extraction, and the instructions provided to annotators were meant to ensure that all information provided by a text about a certain entity would be marked using a single device, the IDENT relation. As van Deemter and Kibble (2000) point out, however, the result is rather *ad hoc*; the IDENT relation as defined by the instructions doesn't capture any coherent definition of 'coreference'. (In fact, the very notion of 'reference' is rather difficult to formalize precisely.)

The MATE proposals, by contrast, while still labeled as proposals for 'coreference annotation,' because the name has become a *de facto* standard as a result of the MUC initiative, are explicitly based on the DISCOURSE MODEL assumption adopted almost universally by linguists (computational and not) working on anaphora resolution and generation (Webber, 1979; Heim, 1982; Kamp and Reyle, 1993; Gundel et al., 1993). This is the hypothesis that interpreting a discourse involves building a shared discourse model containing DISCOURSE ENTITIES that may or may not 'refer' to specific objects in the world, as well as the relations between these entities. The type of annotation for which the MATE scheme was developed–and that we'll call here 'anaphoric annotation,'.[1] is meant as a partial representation of the discourse model evoked by a text (hence, for example, the tag used for nominal expressions denoting discourse entities, ⟨de⟩).[2]

2.2 The Markup Scheme

The design of the MATE workbench was strongly inspired by the concept of STANDOFF ANNOTATION developed for the reorganization of the MapTask. The main principle of standoff annotation is that each level of annotation–for example, syntactic annotation, dialogue act annotation, and anaphoric annotation–should be stored independently; in this way, annotators working on one level need not be concerned about the other levels of annotation, and can start immediately without having to wait for other annotation tasks to be completed. The separate levels of annotation are synchronized via a *base file*, to which the separate levels point using the HREF mechanism of XML.

The markup scheme for anaphoric relations is the core aspect of the MATE proposals and its most distinctive aspect. As in the MUC scheme, it is assumed that annotation of anaphoric information involves identifying MARKABLES (the text constituent that realize semantic objects that may enter in anaphoric relations), and marking up anaphoric relations between them. The main difference from MUCCS is that whereas in MUCCS anaphoric relations are annotated using an attribute of the markables, in the MATE markup scheme–following the recommendations of the Text Encoding Initiative (Burnard and Sperberg-McQueen, 2002), and of Bruneseaux and Romary (1998)–the distinction between these two steps of annotation is mirrored by a distinction between two XML elements: ⟨de⟩, used to indicate the markables, and ⟨link⟩, used to mark information about anaphoric relations (or any other semantic relation).[3] However, unlike in the TEI proposals, in the MATE markup scheme ⟨link⟩ elements are *structured* elements, containing one or more ⟨anchor⟩ element. The ⟨link⟩ element specifies the anaphoric expression (using XML's HREF mechanism) and the relation between the anaphoric expression and its antecedent; whereas the ⟨anchor⟩ element specifies the antecedent, as in (1) where, for example, the first ⟨link⟩ elements encodes the information that the discourse entities realized by the NPs *the engine E3* and *it* denote the same object.

(1)
```
coref.xml

<de ID="de_01">we</de>'re gonna take
<de ID="de_07"> the engine E3 </de>
and shove <de ID="de_08"> it </de> over
to <de ID="de_02">Corning</de>,
hook <de ID="de_09"> it </de> up to
<de ID="de_03">the tanker car</de>...
<link href="coref.xml#id(de_07)"
      type="ident">
  <anchor href="coref.xml#id(de_08)"/>
</link>
<link href="coref.xml#id(de_08)"
      type="ident">
  <anchor href="coref.xml#id(de_09)"/>
</link>
```

There were two main reasons for having ⟨link⟩ elements separated from the elements used to indicate markables. The first reason is that in this way ⟨link⟩ elements can be kept in a separate file from ⟨de⟩ elements, in keeping with the idea of standoff annotation. The second, and more important, reason is that in this way it is possible to annotate multiple anaphoric relations involving the same anaphoric expression without having multiple attributes for each markable.

The reason why ⟨link⟩ elements may have more than one ⟨anchor⟩ element is to allow for the possibility to annotate ambiguities. For some types of applications, it may be a good idea not to ask annotators to decide upon the interpretation of ambiguous anaphoric expressions.

[1] van Deemter and Kibble (2000) give a stricly textual definition of 'anaphora' which is very distant from the common use of the term 'anaphora resolution' in computational linguistics, typically used to indicate the interpretation of (parts of) the meaning of an expression with respect to the discourse model.

[2] In fact, the use of the term 'coreference annotation' would not be completely misguided. van Deemter and Kibble (2000) assume the definition of 'reference' typically found in formal semantics, but in functional linguistics, the term 'referring expression' is used to indicate expressions that introduce new discourse entities in a discourse model or that denote an old one (see, e.g., (Gundel et al., 1993)).

[3] It was assumed that the tags for 'coreference' annotation would be part of a special namespace, COREF–i.e., that the actual name of these tags are ⟨coref:de⟩, ⟨coref:link⟩, etc. We omit the namespace indication in this paper.

In these cases, the multiple anchors mechanisms allows each of the possibilities to be marked by means of a separate ⟨anchor⟩ element. In (2a), for example, the pronun *it* in 15.16 could refer equally well to engine E3 or the tanker car. With the MATE mechanism, both antecedents can be annotated, as shown in (2b).

(2) a.
```
15.12    : we're gonna take the engine E3
15.13    : and shove it over to Corning
15.14    : hook it up to the tanker car
15.15    : _and_
15.16    : and send it back to Elmira
```

b. coref.xml:

```
15.12    : we're gonna take
             <de ID="de_15">the engine E3</de>
15.13    : and shove <de ID="de_16"> it </de>
             over to Corning
15.14    : hook <de ID="de_17">it</de> up to
             <de ID="de_18">the tanker car</de>
15.15    : _and_
15.16    : and send <de ID="de_19">it</de>
             back to Elmira

<link href="coref.xml#id(de_16)" type="ident">
  <anchor href="coref.xml#id(de_15)"/>
</link>
<link href="coref.xml#id(de_17)" type="ident">
  <anchor href="coref.xml#id(de_16)"/>
</link>
<link href="coref.xml#id(de_19)" type="ident">
  <anchor href="coref.xml#id(de_17)"/>
  <anchor href="coref.xml#id(de_18)"/>
</link>
```

2.3 Instantiations of the Meta-Scheme

As said above, the markup elements just discussed were meant to be general enough to support different types of annotation. Three such examples were considered.

The Core Scheme In the most basic type of coreference scheme, only anaphoric relations between NPs are considered, and only identity relations. Schemes of this type can be implemented by having just one anaphoric relation, IDENT. The remaining differences between the schemes have then mostly to do with the instructions to annotators–for example, which types of anaphoric relations to be considered as cases of 'identity' (see (van Deemter and Kibble, 2000) for some problems with the choices made in MUCCS). In the comments for the designers of a scheme, it was suggested that some of the cases marked as coreference in MUCCS, such as the relation between *the temperature* and *90 degrees* in *the temperature rose to 90 degrees before dropping to 70 degrees*, would be best marked as function-value relations (viewing *the temperature* as a function from objects and time points into values, rather than an individual-denoting term).

Extended Relations In DRAMA, a number of associative relations are considered, such as SUBSET or PART, together with instructions how to annotate them. This types of anaphora can be annotated in the MATE markup scheme using additional relations, as in (3), where the discourse entity realized by *LES FUSEES QUI ONT BIEN VOLE*‘ denotes a subset of the set denoted by discourse entity DE_88, *LES MODELES DE FUSEES*.

(3) a.
```
F: Alors donc / vous avez / ici /
   LES MODELES DE FUSEES /
M: Oui
F: Et vous allez essayer de vous
   mettre d'accord sur un classement
   /hein classer
   LES FUSEES QUI ONT BIEN VOLE‘ ou
   QUI ONT MOINS BIEN VOLE‘
```

b.
```
F: Alors donc / vous avez / ici /
   <de ID="de_88"> les mode'les de fuse'es </de>
M: Oui
F: Et vous allez essayer de vous mettre d'accord
   sur un classement /hein classer
   <de ID="de_89"> les fuse'es qui ont
   bien vole' </de>
   ou <de ID="de_90"> qui ont
   moins bien vole' </de>

<link href="coref.xml#id(de_89)">
  <anchor href="coref.xml#id(de_88)"
          type="subset " />
</link>
<link href="coref.xml#id(de_90)"
      type="subset " >
  <anchor href="coref.xml#id(de_88)"/>
</link>
```

It was pointed out, however, that the results of Poesio and Vieira (1998) indicated that this type of annotation could be highly unreliable.

References to the Visual Situation A special ⟨universe⟩ element was suggested for MapTask-style annotations of references to visible objects. The ⟨universe⟩ element containing one ⟨ue⟩ element for each object in the visual scene; including such elements in an annotation makes it possible to use ⟨link⟩ elements to annotate references to such objects.[4] Cases in which the participants to a conversation have different visual situations, as in the MapTask dialogues, can be handled by having separate universes, one for each participant to the conversation. In addition, a WHO-BELIEVES attribute of ⟨link⟩ elements was proposed to represent situations in which only one participant believes that a particular anaphoric relation holds, as in example (7) (Appendix A), where it's only the follower to believe that *a gold mine* refers to the same object as *diamond mine*.

2.4 Instructions for Identifying Markables

Because the goal of the MATE annotation proposals was to provide a set of tools that could be used to implement a variety of options, rather than to identify a specific scheme appropriate for all applications, it didn't make sense to specify detailed instructions for annotation. However, a substantial effort was made to provide an exhaustive inventory of the options for identi-

[4]The ⟨universe⟩ mechanism is based on the notion of 'anchor' developed in Discourse Representation Theory (DRT), although simplified in a number of ways.

fying markables that were available to the designers of a scheme for anaphoric annotation. These suggestions were in part derived from MUCCS and from Passonneau's DRAMA scheme, but a number of additional problems were considered as well.

As in MUCCS, it was assumed that annotation of anaphora is best separated in two steps: first the *markables* (the text constituent that realize semantic objects that may enter in anaphoric relations) are agreed upon, then anaphoric relations between them are marked.

Concerning markable identification, the main suggestions were to concentrate on anaphoric expressions realized as NPs and their antecedents; and to rely on the output of a parser as much as possible. But because of the assumption that only NPs evoking discourse entities should be considered, it was suggested that not all NP should be treated as markables: for example, it was recommended that NPs in post-verbal position in predicative clauses (such as *a policeman* in *John is a policeman*) should be excluded. This recommendation was later reconsidered (see below).

One of the novel aspects of the MATE instructions was the concern for markable identification in languages other than English. One such issue was how to deal with incorporated clitics and empty subjects; the suggestion was to use a separate element, ⟨seg⟩, to turn verbs into non-nominal markables, as in the following example:

(4)
```
coref.xml:

A: Dov'e' <de ID="de_157">Gianni?</de>
   [Where is Gianni?]
B: <seg type="pred" ID="seg_158 >e'
   andato a mangiare </seg>
   [_ went to have lunch]

<link href="coref.xml#id(seg_158)"
      type="ident">
   <anchor href="coref.xml#id(de_157)"/>
</link>
```

It was also proposed that the ⟨seg⟩ element could be used in more ambitious schemes as general mechanism for specifying non-nominal markables –e.g., in ellipsis, to indicate the antecedents of discourse deixis, etc.[5]

3 Work based on the MATE proposals

Ideas from the MATE 'scheme' have been adopted and tested both in annotation projects and by the developers of annotation tools. In this section we review some of these activities and summarize the conclusions concerning advantages and disadvantages of the MATE scheme that can be drawn from them.

[5] A second range of issues considered in the MATE scheme had to do with dialogue phenomena, such as non-contiguous elements; we will not consider these issues here.

3.1 Annotation work related to the GNOME project

The most direct application of the ideas discussed above was found in the annotation work undertaken as part of the GNOME project. GNOME was concerned with the empirical investigation of the aspects of discourse that appear to affect generation, especially salience (Pearson et al., 2000; Poesio et al., 2000; Poesio and Di Eugenio, 2001; Poesio and Nissim, 2001; Poesio et al., 2004c). Particular attention was paid to the factors affecting the generation of pronouns (Pearson et al., 2000; Henschel et al., 2000), demonstratives (Poesio and Nygren-Modjeska, To appear) possessives (Poesio and Nissim, 2001) and definites in general (Poesio, 2004). These results, and the annotated corpus, were applied to the development of both symbolic and statistical natural language generation algorithms with the application of these empirical results to natural language generation, from sentence planning (Poesio, 2000a; Henschel et al., 2000; Cheng et al., 2001), to aggregation (Cheng, 2001) and text planning (Kibble and Power, 2000; Karamanis, 2003). The empirical side of the project involved both psychological experiments and corpus annotation, based on a scheme based on the MATE proposals, as well as on a detailed annotation manual (Poesio, 2000b), the reliability of whose instructions was tested by extensive experiments (Poesio, 2000a). More recently, the corpus has also been used to develop and evaluate anaphora resolution systems, with a special focus on the resolution of bridging references (Poesio, 2003; Poesio and Alexandrov-Kabadjov, 2004; Poesio et al., 2004b).

The corpus The GNOME corpus currently includes texts from three domains, about 3000 NPs were annotated in each domain. The museum subcorpus consists of descriptions of museum objects, generally with an associated picture, and brief texts about the artists that produced them. The pharmaceutical subcorpus is a selection of leaflets providing the patients with legally mandatory information about their medicine.

Several layers of information were annotated, including layout in the case of text and rhetorical structure in the case of tutorial dialogues, sentences and potential utterances, noun phrases, a variety of attributes of the objects denoted by noun phrases,[6] and anaphoric relation. We concentrate here on anaphoric information, and refer the reader to the manual for the other types of annotation.

Markup scheme The markup scheme for markables and anaphoric relations adopted in GNOME follows very

[6] E.g., whether an NP denoted generically or not; whether it denoted an animate or inanimate entity, as well as other ontological properties; and whether it denoted a discourse entity, a quantifier, or a predicate. In the case of a discourse entity, we also annotated whether it denoted an atom, a set, or a mass term; and whether it denoted uniquely or not.

closely that proposed in MATE, except that the ⟨de⟩ element was renamed ⟨ne⟩ (since all NPs were marked), and the ⟨link⟩ element was renamed ⟨ante⟩. More substantial differences are the decision not to use standoff, and the introduction of new elements necessary for the study of salience, such as elements that could be used to investigate the notion of UTTERANCE used in Centering (Poesio et al., 2004c).

Although standoff is a clear improvement over including all annotation levels in a single file, our own experiences during the creation of the GNOME corpus being further proof of this, it's only really possible when tools are available both to create the annotation and–crucially–later to 'knit back' the separate levels when needed. As neither the MATE workbench nor any other tools based on standoff were available by the time the GNOME annotation started,[7] in GNOME we didn't use standoff, but integrated all levels of annotation in one file; an Emacs mode was developed for the annotation. This decision made it very easy to use the annotated corpus for a number of studies, but did resulted in a number of problems, the main among which were that the annotators had to be very careful not to damage other annotations; that annotators working on one level were occasionally confused by annotations for other levels; and that the annotation work had to be organized in a careful sequential way even for levels that could have been annotated independently.

The main new aspect of the markup scheme, especially as far as our studies of salience were concerned, are the elements used to annotate potential utterances in the sense of Centering (Grosz et al., 1995). In order not to prejudge the answer to the question of which text constituents are best viewed as utterances, we used a 'generic' element called ⟨unit⟩ to mark up finite and non-finite clauses, but also parentheticals and appositions, elements of bulleted lists, etc.

The following example illustrates both the use of ⟨unit⟩ elements and of the elements ⟨ne⟩ and ⟨ante⟩ replacing ⟨de⟩ and ⟨link⟩:

(5)
```
<unit finite='finite-yes' id='u227'>
 <ne id='ne546' gf='subj'>  The drawing of
   <ne id='ne547' gf='np-compl'>the corner
     cupboard </ne>
 </ne>
 <unit finite='no-finite' id='u228'>,
   or more probably
   <ne id='ne548' gf='no-gf'> an engraving of
      <ne id='ne549' gf='np-compl'>  it </ne>
   </ne>
 </unit>,
...
</unit>
<ante current="ne549" rel="ident">
<anchor ID="ne547">
</ante>
```

[7] In the end lack of time prevented the inclusion of a tool for anaphoric annotation in the released MATE workbench.

Bridging References Apart from the basic anaphoric relations of identity, in GNOME we were concerned with bridging references, hence our annotation scheme incorporated aspects of the 'Extended Relations' and the 'MapTask' instantiations of the MATE meta-scheme.

One of our aims was to continue the work on bridging references annotation and interpretation in (Poesio and Vieira, 1998), which showed that marking up bridging references is quite hard. In addition, work such as (Sidner, 1979; Strube and Hahn, 1999) suggested that indirect realization can play a crucial role in maintaining the CB. After testing a few types of associative reference (Hawkins, 1978), we decided to annotate only three non-identity relations, as well as identity. These relations are a subset of those proposed in the 'extended relations' version of the MATE scheme: set membership (ELEMENT), subset (SUBSET), and 'generalized possession' (POSS), which includes both part-of relations and ownership relations.

Coder manual Perhaps the most important aspects of the annotation work in GNOME are the development of detailed instructions for annotators and the reliability experiments testing several aspects of the scheme, particularly the annotation of bridging references.

The identification of sentences, units and markables was done entirely by hand, without encountering particular problems. (The Emacs mode, an extension of SGML-mode, provides some support for introducing new elements, marking regions, and attribute editing, as well as anaphoric annotation.) Unlike in MATE, all NPs were tagged as ⟨ne⟩. The instructions for ⟨unit⟩s were based on Marcu's proposals for discourse units annotation (Marcu, 1999). All attributes of sentences, ⟨unit⟩s and ⟨ne⟩s in the final version of the scheme, including DEIX, can be annotated reliably.

In order to achieve reliability on anaphoric annotation, the range of anaphoric phenomena considered was restricted in many ways. Apart from marking a limited number of associative relations, the annotators only marked relations between objects realized by noun phrases and not, for example, anaphoric references to actions, events or propositions implicitly introduced by clauses or sentences. We also gave strict instructions to our annotators concerning how much to mark. They were told to mark all identity relations, but to mark associative relations only if either (i) no IDENT relation could be marked for the anaphoric expression, or (ii) an IDENT relation with an entity not mentioned in the previous ⟨unit⟩. Furthermore, preferences were specified, e.g., for appositions: for example, in *Francois, the Dauphin*, the embedding NP would be chosen as an antecedent of subsequent anaphoric references, rather than the NP in appositive position.

We found a reasonable, although by no means perfect, agreement on identity relations. In a typical analysis (two annotators looking at the anaphoric relations between 200 NPs) we observed no real disagreements; 79.4% of these relations were marked up by both annotators; 12.8% by only one of them; and in 7.7% of the cases, one of the annotators marked up a closer antecedent than the other. Limiting the relations did limit the disagreements among annotators on associative relations (only 4.8% of the relations are actually marked differently) but only 22% of bridging references were marked in the same way by both annotators; 73.17% of relations are marked by only one or the other annotator. Reaching agreement on this information involved several discussions between annotators and more than one pass over the corpus (Poesio, 2000a).

3.2 Annotation tools

Although no annotation tool implementing the MATE or GNOME schemes as described exists, in the years after the development of the MATE guidelines tools supporting XML standoff annotation for coreference have appeared, including MMAX from EML (Müller and Strube, 2003) and the Annotator from ILSP. Although the format used for storing anaphoric information by these tools is not entirely satisfactory, the files they produce can be easily converted into MATE format.

MMAX, for example, is based on a simplified standoff format, in which three main files are maintained for each annotated file in the corpus: a base file containing the words, a file identifying sentences, and a file identifying markables. Anaphoric information is stored as attributes of the markables. Two special attributes are used for this purpose, and recognized by MMAX: the `MEMBER` attribute, used to indicate membership in a coreference chain (a coreference equivalence class), and the `POINTER` attribute, used to mark up to one associative anaphoric relation for each anaphoric expression. We discuss the use of MMAX in the VENEX project below.

3.3 The VENEX Corpus

The VENEX corpus is an anaphorically annotated corpus of Italian being created in a joint project between the Universitá di Venezia and the University of Essex. The corpus includes both texts (newspaper articles) and dialogues (an Italian version of the MapTask corpus). This project widened our experiences of annotation with the MATE scheme in a number of respects. First of all, a number of proposals contained in the MATE guidelines but not relevant for GNOME, including the suggestions for dealing with misunderstandings and for incorporated anaphoric expressions such as clitics, were tested. Secondly, in this project we are attempting to identify markables automatically as far a possible, and data are stored in a standoff format, using a modern annotation tool (MMAX) for the annotation.

Markup Scheme As MMAX doesn't support ⟨`link`⟩ elements, and anaphoric information is stored with markables, it is necessary to use markable attributes to represent information that would have been encoded as part of the links. We used a separate attribute to specify the type of associative relation used by `POINTER` attribute, and a `SPACE` attribute to encode the information stored in the `WHO-BELIEVES` attribute of links (see below). In addition, only one `MEMBER` and `POINTER` attributes can be specified for each markable.

This latter limitation wasn't much of a problem, given that the annotation instructions used in VENEX are derived from those developed for GNOME and also attempt to limit annotators to mark at most one identity and one bridging relation for each anaphoric expression. The separation of attributes of links proved, however, a problem, as annotators often forget to annotate one or the other.

An additional problem is that the version of MMAX we used (0.92) only allows for one type of markable, meaning that ⟨`unit`⟩ elements could not be annotated, and instead of using separate ⟨`ne`⟩ and ⟨`seg`⟩ elements for nominal and non-nominal markables, a single markable had to be used (see below).[8]

Misunderstandings The MapTask part of the VENEX corpus contains numerous examples like (7), where the differences between Giver and Follower map lead to one participant believing that two objects are anaphorically related, while the other participant either is not aware of this or doesn't believe this to be the case. We found that after a few iterations of training, our annotators were able to handle these cases properly (a more formal evaluation is underway; we hope to report the results at the meeting). Again, the only problems were caused by the fact that these attributes had to be added to markables, which sometimes led to annotators forgetting to set them. (This was only required in case the default, that an anaphoric relation was in the common ground of both participants, didn't hold.)

4 Discussion

4.1 Aspects of the MATE proposals that have been proven useful

Our experience with multi-level annotation in GNOME suggests that standoff is clearly the way to go, allowing multiple annotators to work on the same files, and separating logically independent tasks, but appropriate tools are required. The annotation tools we have discussed, such as MMAX, are therefore useful even though they do

[8] The version of MMAX currently being developed will allow for multiple markables.

not implement all aspects of the MATE. Knitting back is also possible with the Discourse API.

Our experience with VENEX suggests that two of the most beneficial aspects of having a separate ⟨link⟩ element are ones that we had not originally considered: that they can be used to mark general semantic relations, not just anaphoric relations (for more complex types of semantic annotation); and that they make it harder for annotators to forget to fill in aspects of the annotation. Unfortunately, at the moment there is no tool that can be used to create this type of annotation directly.

4.2 Aspects already reconsidered

Predicative NPs During the GNOME and VENEX annotations we realized that the recommendation not to mark predicative NPs makes it impossible to do markable identification automatically. In addition, it's often difficult to decide whether an NP is used predicatively or referentially, especially in languages like Italian where subjects in such clauses are often used predicatively (as in *La soluzione e' questa*). In GNOME, a new attribute LF_TYPE was introduced to specify the type of semantic object denoted by an NP: term, quant and pred. The annotators were instructed to concentrate on term-denoting NPs. The instructions for classifying NPs according to their semantic type were based mostly on syntactic information, but the annotation was reliable. In the instructions for the VENEX annotation, the instructions for recognizing term-denoting NP are further developed.

Restricting the range of associative relations The range of associative relations tested in GNOME is much narrower than those considered in DRAMA, but they can be annotated reliably, at least in the sense that very few disagreements are observed. Extending the range of relations to include, for example, attributes (e.g, *I am not going to buy that. The price is too high.* or situational associations (*John entered a restaurant. The waiter approached him immediately*) has proven difficult.

Units and Utterances The study of Centering carried in GNOME indicated quite clearly that annotation of ⟨unit⟩ elements is essential for the study of anaphora.

4.3 Further Revisions

One aspect of the markup scheme that needs revision is the placement of the semantic relation. One problem we observed in GNOME is that often the ambiguity is not simply between two possible antecedents each of which stands in the same relation to the anaphoric expression, but between two antecedents which stand in different relations. In the pharmaceutical texts, for example, it is often unclear whether a particular mention of the medicine under consideration refers to the generic product, or to the particular instance that the user has in their hands. In this case, we would want annotators to mark the anaphoric expression as IDENT with one object, and ELEMENT of the other (ELEMENT is also used in GNOME for relations between instances and types), as follows, but this is not possible in either the original MATE scheme or in the GNOME markup scheme:

(6)
```
<ante current="ne1">
<anchor ID="ne2" rel="ident">
<anchor ID="ne3" rel="element">
</ante>
```

4.4 Open Issues

Ambiguity Offering annotators the opportunity to annotate anaphoric ambiguity is essential, especially for annotations used to study linguistic phenomena, but raises serious theoretical and practical problems. A coreference chain containing such links becomes a coreference (directed) *graph*, in which each of the paths across the graph is a potential interpretation. While having multiple paths is not a problem as far as evaluating the results of an anaphoric resolver (any path in the graph counts as a valid solution), it is a serious problems both for scripts attempting to ensure consistency (e.g., that all references to the same object are marked as either generic or non-generic– this is of course impossible when one of the possible antecedents is generic while the other isn't) as well for annotation tools (the problem is of course worsened when the tool only uses a single attribute to indicate membership in a coreference chain).

Revision A second difficult problem is caused by cases, common in the MapTask dialogues, in which after a while a participant realizes that their previous belief that an object was identical to another object is mistaken. In these cases, the participant is arguably revising their previous beliefs; it is not clear then what should be done with the annotation of the original anaphoric information.

References

F. Bruneseaux and L. Romary. 1998. Documents préparatoires pour le codage de dialogues multimodaux suivant les directives de la TEI. Available at http://www.loria.fr/~romary/Documents/index.html.

L. Burnard and C. M. Sperberg-McQueen. 2002. TEI lite: An introduction to text encoding for interchange. http://www.tei-c.org/Lite.

H. Cheng, M. Poesio, R. Henschel, and C. Mellish. 2001. Corpus-based NP modifier generation. In *Proc. of the Second NAACL*, Pittsburgh.

Hua Cheng. 2001. *Modelling Aggregation Motivated Interactions in Descriptive Text Generation*. Ph.D. thesis, University of Edinburgh.

B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):202–225.

J. K. Gundel, N. Hedberg, and R. Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

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

I. Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts at Amherst.

R. Henschel, H. Cheng, and M. Poesio. 2000. Pronominalization revisited. In *Proc. of 18th COLING*, Saarbruecken, August.

L. Hirschman. 1998. MUC-7 coreference task definition, version 3.0. In N. Chinchor, editor, *In Proc. of the 7th Message Understanding Conference*.

H. Kamp and U. Reyle. 1993. *From Discourse to Logic*. D. Reidel, Dordrecht.

N. Karamanis. 2003. *Entity coherence for descriptive text structuring*. Ph.D. thesis, University of Edinburgh.

R. Kibble and R. Power. 2000. An integrated framework for text planning and pronominalization. In *Proc. of the International Conference on Natural Language Generation (INLG)*, Israel, June.

D. Marcu. 1999. Instructions for manually annotating the discourse structures of texts. Unpublished manuscript, USC/ISI, May.

D. McKelvie, A. Isard, A. Mengel, M. B. Moeller, M. Grosse, and M. Klein. 2001. The MATE workbench - an annotation tool for XML corpora. *Speech Communication*, 33(1-2):97–112.

C. Müller and M. Strube. 2003. Multi-level annotation in MMAX. In *Proc. of the 4th SIGDIAL*, pages 198–207.

R. Passonneau. 1997. Instructions for applying discourse reference annotation for multiple applications (DRAMA). Unpublished manuscript., December.

J. Pearson, R. Stevenson, and M. Poesio. 2000. Pronoun resolution in complex sentences. In *Proc. of AMLAP*, Leiden.

M. Poesio and M. Alexandrov-Kabadjov. 2004. A general-purpose, off the shelf anaphoric resolver. In *Proc. of LREC*, Lisbon, May.

M. Poesio and B. Di Eugenio. 2001. Discourse structure and anaphoric accessibility. In Ivana Kruijff-Korbayová and Mark Steedman, editors, *Proc. of the ESSLLI 2001 Workshop on Information Structure, Discourse Structure and Discourse Semantics*.

M. Poesio and M. Nissim. 2001. Salience and possessive NPs: the effect of animacy and pronominalization. In *Proc. of AMLAP (Poster Session)*.

M. Poesio and N. Nygren-Modjeska. To appear. Focus, activation, and this-noun phrases: An empirical investigation. In A. Branco, R. McEnery, and R. Mitkov, editors, *Anaphora Processing*. John Benjamins.

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

M. Poesio, F. Bruneseaux, and L. Romary. 1999. The MATE meta-scheme for coreference in dialogues in multiple languages. In M. Walker, editor, *Proc. of the ACL Workshop on Standards and Tools for Discourse Tagging*, pages 65–74.

M. Poesio, H. Cheng, R. Henschel, J. M. Hitzeman, R. Kibble, and R. Stevenson. 2000. Specifying the parameters of Centering Theory: a corpus-based evaluation using text from application-oriented domains. In *Proc. of the 38th ACL*, Hong Kong, October.

M. Poesio, R. Delmonte, A. Bristot, L. Chiran, and S. Tonelli. 2004a. The VENEX corpus of anaphoric information in spoken and written Italian. Submitted.

M. Poesio, R. Mehta, A. Maroudas, and J. Hitzeman. 2004b. Learning to solve bridging references. Submitted.

M. Poesio, R. Stevenson, B. Di Eugenio, and J. M. Hitzeman. 2004c. Centering: A parametric theory and its instantiations. *Computational Linguistics*. To appear.

M. Poesio. 2000a. Annotating a corpus to develop and evaluate discourse entity realization algorithms: issues and preliminary results. In *Proc. of the 2nd LREC*, pages 211–218, Athens, May.

M. Poesio, 2000b. *The GNOME Annotation Scheme Manual*. University of Edinburgh, HCRC and Informatics, Scotland, fourth version edition, July. Available from `http://www.hcrc.ed.ac.uk/~gnome`.

M. Poesio. 2003. Associative descriptions and salience. In *Proc. of the EACL Workshop on Computational Treatments of Anaphora*, Budapest.

M. Poesio. 2004. An empirical investigation of definiteness. In S. Kepser, editor, *Proc. of the International Conference on Linguistic Evidence*, Tübingen, January. University of Tübingen, SFB 441.

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

M. Strube and U. Hahn. 1999. Functional centering–grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

K. van Deemter and R. Kibble. 2000. On coreferring: Coreference in MUC and related annotation schemes. *Computational Linguistics*, 26(4):629–637. Squib.

B. L. Webber. 1979. *A Formal Approach to Discourse Anaphora*. Garland, New York.

A An example with separate universes and single-space anaphoric links

(7) a.

```
GIVER: Do_you have diamond_mine.
FOLLOWER: Yes I've got a gold_mine.
GIVER: Ah. S--.
FOLLOWER: ....
GIVER: You don't have diamond_mine though.
FOLLOWER: No. It's a gold_mine according to
       this one.
       Presumably that's the same.
GIVER: Well I've got a gold_mine as well
       you see. (MT)
```

b. `coref.xml`:

```
<universe ID="common">
<ue ID="ue2"> gold mine </ue>
....
</universe>
<universe ID="GIVER_universe"
          modifies="common">
<ue ID="ue1"> diamond mine </ue>
...
</universe>
<universe ID="FOLLOWER_universe"
          modifies="common">
....
</universe>

GIVER: Do_you have
       <de ID="de_20"> diamond_mine. </de>
FOLLOWER: Yes I've got
       <de ID="de_21"> a gold_mine. </de>
GIVER: Ah. S--.
FOLLOWER: ....
GIVER: You don't have
       <de ID="de_22"> diamond_mine </de>
       though.
FOLLOWER: No.
       It's <de ID="de_23"> a gold_mine</de>
       according to this one.
       Presumably <de ID="de_24"> that's </de>
       the same.
GIVER: Well I've got
       <de ID="de_25"> a gold_mine </de>
       as well you see.

<link href="coref.xml#id(de_20)"  type="ident"
      who-believes="G">
   <anchor href="coref.xml#id(ue1)"/>
</link>
<link href="coref.xml#id(de_21)"  type="ident"
      who-believes="F" >
   <anchor href="coref.xml#id(ue2)"/>
</link>
<link href="coref.xml#id(de_21)"  type="ident"
      who-believes="F" >
   <anchor href="coref.xml#id(de_20)"/>
</link>
<link href="coref.xml#id(de_22)"  type="ident"
      who-believes="G">
   <anchor href="coref.xml#id(ue1)"/>
</link>
<link href="coref.xml#id(de_22)"  type="ident">
   <anchor href="coref.xml#id(de_20)"/>
</link>
<link href="coref.xml#id(de_23)"  type="ident"
      who-believes="F" >
   <anchor href="coref.xml#id(ue2)"/>
</link>
<link href="coref.xml#id(de_23)"  type="ident"
      who-believes="F" >
   <anchor href="coref.xml#id(de_22)"/>
</link>
<link href="coref.xml#id(de_24)"  type="ident"
      who-believes="F" >
   <anchor href="coref.xml#id(de_22)" />
</link>
```