Annotating Event Anaphora: A Case Study

Tommaso Caselli, Irina Prodanof

ILC - CNR
Via G. Moruzzi, 1 56124 Pisa
firstName.secondName@ilc.cnr.it

Abstract

In recent years we have resgitered a renewed interest in event detection and temporal processing of text/discourse. TimeML (Pustejovsky et al., 2003a) has shed new lights on the notion of event and developed a new methodology for its annotation. On a parallel, works on anaphora resolution have developed a reliable methodology for the annotation and pointed out the core role of this phenomenon for the improvement of NLP systems. This paper tries to put together these two lines of research by describing a case study for the creation of an annotation scheme on event anaphora. We claim that this work could have consequences for the annotation of eventualities as proposed in TimeML and on the use of the `<event>` tag and on the study of anaphora and its annotation. The annotation scheme and its guidelines have been developed on the basis of a coarse grained bottom up approach. In order to do this, we have performed a small sampling annotation which has highlighted shortcomings and open issues which need to be resolved.

1. Introduction

This paper aims at describing a proposal for an XML annotation scheme for marking up anaphoric relations between eventualities. Eventualities represent the building blocks of the informative content of a document and, as textual entities, they give rise to relations which create a rich informative network. For instance consider these examples from the Italian Treebank (ISST (Montemagni et al., 2003)):

1. Gli Usa continuano a premere su tedeschi e giapponesi perche' **espandano**$_j$ le loro economie e questi recalcitrano alla sola *idea*$_j$.(ISST els019)
   The U.S. keeps to put pressure on Germany and Japan [to expand]$_j$ their economies and they resist the very [idea]$_j$.

2. Secondo il governo di Pechino, **le accuse in base alle quali due diplomatici cinesi sono stati espulsi la settimana scorsa dagli Stati Uniti, sono una montatura**$_j$. *Lo*$_j$ ha detto ieri un portavoce del ministero degli Esteri. (ISST els075)
   According to Beijing Government, [the charges on the basis of which two Chinese diplomats have been banned last week from the United States, are a frame]$_j$. [It]$_j$ was reported yesterday by a spokesman of the Foreign Ministry.

As it can be noticed, in example 1. the NP "*idea*" [idea] assumes an eventive reading due to the fact that it is anaphorically linked to the event "*espandono*" [expand]. Establishing the link between these two elements is necessary in order to retrieve and extract information with respect to the resistance of the German and Japanes governments to the pressure of the U.S. In 2. the pronoun "*lo*" [it] works as topic placeholder and corresponds to the content of the statement of the spokeman of the Foreign Ministry.

In recent years, different annotation schemes for the markup of eventualities have been developed (Setzer, 2001; Katz and Arosio, 2001). So far, the most complete and with the highest representative power is TimeML (Pustejovsky et al., 2003a). TimeML is now becoming an ISO

standard, namely ISO-TimeML, under the ISO TC37/SC 4 initiative. Though the standardization effort has overcome some shortcomings of the annotation scheme, so far the issue of how anaphoric relations between eventualities should be treated has not been tackled yet. The study of such phenomenon is part of the more general problem of the study of text/discourse cohesion and coherence. Nevertheless, on the one hand, it could have consequences for the annotation of eventualities as proposed in TimeML and on the use of the `<event>` tag, and, on the other hand, it sheds new lights on the study of anaphora both by widening the set of possible markables with respect to classical studies on anaphoras, which have mainly concentrated on NPs, and by posing issues on the identification and extent of the anchoring elements. A core aspect of this work is represented by the use of a coarse grained bottom-up approach for the development of the annotation scheme, the use of non-expert annotators, and the linguistic aspects of the event anaphora phenomenon.

The remaining of the paper is structured as followed: in section 2. we will describe the methodology used to annotated event instances in a corpus of Italian, the Italian TimeBank and how the TimeML specifications are limited when coming to the annotation of anaphoric relations. In section 3. we will describe the annotation scheme, its representation format and how the task has been conceived. In section 4. we will present a preliminary evaluation performed on a restricted portion of the corpus and by means of non-expert or *naive* annotators. Finally, in section 5. we will illustrate the modifications to the scheme which have been adopted as a consequence of the analysis and describe some open issues.

2. Event and co-reference annotation in TimeML

The proposal of annotating anaphoric relations between eventualities emerged as part of the annotation process that we have undertaken for the creation of the Italian TimeBank. The corpus is composed by 149 newspaper articles for a total of 63397 tokens extracted both from the ISST

and from the corpus PAROLE. The corpus has been annotated with an adapted version of the TimeML guidelines to Italian by six annotators. The annotation is a multiple-step process. In the detection phase, the annotators with different backgrounds in linguistics have separately annotated the markables, namely the tag EVENT, TIMEX3 and SIGNAL, in different moments. After this phase, two judges have checked the consistency of the annotations and resolved some contradictions. In order to obtain reliable results, in addition to the annotation guidelines the annotators have been provided with a detailed set of decision trees which give them a formal way for annotating events realized by more than one token such as light verbs constructions (3.), complex NPs of the form "NP + di + NP" (4.), multiwords and idiomatic expressions (5.) and (6.):

3. fare la spesa [*to do shopping*].

   ```
   <EVENT>fare</EVENT> la
   <EVENT>spesa</EVENT>
   ```

4. il fallimento del progetto di acquisizione [*the failure of the acquisition project*].

   ```
   il <EVENT>fallimento</EVENT> del
   <EVENT>progetto</EVENT> di
   <EVENT>acquisizione</EVENT>
   ```

5. (dare) il via libera [*(to give) the green light*].

   ```
   il <EVENT mw=''1''>via</EVENT >
    libera
   ```

6. stare a cuore [*to care about*].

   ```
   <EVENT mw=''1''>stare</EVENT >
   a cuore
   ```

In Table 1 we report the agreement scores in terms of precision, recall and K-statistics. The annotation has been performed by means of BAT, the Brandeis Annotation Tool. The next steps towards the creation and release to the Italian TimeBank are the completion of the annotation of the attributes of the markables (i.e. the EVENT and TIMEX3 tag) and that of the three links (temporal, subordinating and aspectual).

Markables	P&R	K-value
EVENT	0.84	0.83
TIMEX3 - detection	0.97	0.97
TIMEX3 - normalization	0.96	0.96
SIGNAL	0.89	0.89

Table 1: Italian TimeBank evaluation for markable annotation.

2.1. Linking events: the use of the TLINK

TimeML presents three links to put in relations the markables. We concentrate on the use of the TLINK between events. Apart from the annotation of temporal relations, the TLINK tag has a special value which is not at all temporal, namely the `identity` value. This special value is used both to link the same event instance when it is realized by more than one token, like in example 3., reported here as 7., but also to mark up cases of set/subset anaphoric relations between two events as illustrated in 8., where the event "*adempimenti*" [fulfillments], in bold, is referred back by the event "*approvazione*" [approval] in italics, as signaled by the pendix "*j*"; e.g.:

7. fare la spesa [*to do shopping*].

   ```
   <EVENT id="e1">fare</EVENT> la
   <EVENT id="e2">spesa</EVENT>
   <TLINK lid="l1" eventInstanceID="e1"
   relatedToEventInstance="e2"
   relType="IDENTITY"/>
   ```

8. La sessione privata servira' a tre **adempimenti**$_j$. Innanzitutto, all' *approvazione*$_j$ della proposta di Abete (ISST sole006).
   The private session will be used for three [fulfillments]$_j$. First, the [approval]$_j$ of the proposal of Abete.

   ```
   La <EVENT id="e1">sessione</EVENT>
   privata
   <EVENT id="e2">servira'</EVENT> a tre
   <EVENT id="e3">adempimenti</EVENT>.
   <SIGNAL id="s1">Innanzitutto</SIGNAL>,
   all'<EVENT id="e4">approvazione</EVENT>
   della
   <EVENT id="e5">proposta</EVENT>di Abete.
   <TLINK lid="l1" eventInstanceID="e3"
   relatedToEventInstance="e4"
   relType="IDENTITY"/>
   ```

However, this use of the value "`identity`" is not completely satisfying since it is not homogeneous. Moreover, during the annotation effort we noticed that the cases in which this value could be applied is wider than simple set/subset relations, including also synonyms, hypernyms and pure coreference (i.e. repetition of the same lemma) relations.

3. Developing an annotation scheme for event anaphora

Previous works in developing annotation schemes and guidelines for anaphoric relations and events and related resources are quite limited. Most works are for NPs coreference rather than event coreference. In recent years, a number of annotation schemes for marking up coreference and anaphoric relations between NPs have been proposed. Among them we mention the MUC annotation scheme

(Hirschman and Chinchor, 1997) used in MUC-7 coreference task, the MATE scheme (Davies et al., 1998), as a de-facto standard for anaphoric and coreference annotation, and the MATE/GNOME scheme (Poesio, 2004). On the other hand, the work in event annotation is quite limited. One of the main reason is related to the definition of the notion of "event" and to which realizations of events in a text/discourse to mark-up. Moreover, most work in event annotation (Setzer, 2001; Katz and Arosio, 2001; Pustejovsky et al., 2003a) are related to capturing the temporal aspects of events and temporal information.

On the specific subject of event coreference annotation we have identified two preminent previous works, namely (Hasler et al., 2006) and (Bejan and Harabagiu, 2008). In (Hasler et al., 2006) a pilot project for the annotation of NP and event coreference is investigated. In this work, the authors have developed a new methodology for the marking of NPs and event coreference on the basis of detailed guidelines and annotation schemes. The authors have used a bottom-up approach to develop the annotation scheme and resolve possible inconsitences in order to ensure that future annotation based on the proposed methodology would capture the phenomenon in a reliably and detailed way. As for the event annotation, the authors have used as starting point the ACE guidelines. In ACE eight event types or frames (e.g. LIFE, MOVEMENT, TRANSACTION, BUSINESS, CONFLICT, CONTACT, PERSONNEL, JUSTICE). Each event type is also annotated for sub-types, polarity markers, trigger word, genericity, tense, and also for their arguments. The corpus used is composed by domain specific newswire texts (related to terrorism/security) extracted from the Reuters corpus (Rose et al., 2002). As for NP annotation, only coreferential relations have been annotated by means of a reduced set of relations. Instances of indirect anaphora have not been annotated. As for the event annotation, not all events instances have been annotated but only those which correspond to the domain.

In (Bejan and Harabagiu, 2008) the annotation of event coreference was a side-effect of a study to discover event structure relations. By means of a clustering algorithm an event structure is assigned to each event instance. In this case event coreference is considered when two events are expressed by the same predicate, or they are synonyms or hyponyms, or when both predicates express the same arguments. The event annotation has been performed on the basis of the TimeML specifications.

In our work we have adopted a different perspective with respect to both (Hasler et al., 2006) and (Bejan and Harabagiu, 2008). Firstly, we have not taken into account event frames or templates but all event instances (nominal, verbal, adjectival and also prepositional) which are marked up in TimeML and then we have adopted an open domain text/discourse analysis with a particular attention on the compatibility of the TimeML event annotation with issues related to anaphora resolutions and how this could facilitate or improve the annotation of temporal relations.

3.1. The annotation scheme and guidelines

The annotation scheme we have developed in a preliminary version is just a starting point, since we have adopted a bottom-up approach to study such phenomenon. In Table 2 we illustrate the tags used and their attributes.

Markables	Attributes
MARKABLE	ID, DEFINITENESS, POS, CLASS
EMPTY	ID
TOPIC	ID
LINK	ID, ANAPHORTYPE, SRC

Table 2: Tags for the preliminary annotation of event anaphora.

The tags `<markable>`, `<empty>` and `<topic>` represent the markable elements. The tag `<markable>` is a cover tag which is used to annotate all event instances and which roughly corresponds to the `<event>` tag in TimeML. In this phase of the research we have preferred not to use the same tag as in TimeML in order to keep distinct this level of annotation in the perspective of developing an autonomous layer of annotation for event anaphors. Though very similar, the tag `<markable>` has some differences with respect to the TimeML `<event>` tag, in particular as far as the set of possible parts-of-speech which can be marked is concerned. In particular, this tag is used to annotate all possible linguistic realizations of events and linguistic elements which may assume an eventive reading, including pronouns and adverbs, which are currently out of the scope of the TimeML annotation. The attribute "class" is inherited from the TimeML specifications and it is responsible for the assignment of a TimeML class to the annotated elements. The attributes "definiteness" and "pos" are quite straightforward. The first is used to make explicit the definiteness value of the annotated element, and its value are "yes", "no" and "not applicable", for those elements like verbs, adjectives and adverbs for which the category of definiteness is not applicable. The attribute "pos" is used to assign the correct part of speech to the annotated element, i.e. verb, noun, adjective, pronoun, adverb and preposition. The tag `<empty>` is used to annotated zero anaphora or cases of ellipsis, which are quite widespread in Italian documents due to the fact that Italian is a pro-drop language. Finally, the tag `<topic>` is used to annotated entire portions of texts. This tag has been introduced in order to provide an anchor for those linguistic elements which have the capabilities of referring back to multiple event instances or entire posrtions of the text/discourse. The tag `<topic>` can have other nested tags in its scope, such as `<markable>` and `<empty>` tags. To clarify its use consider the example reported in 9., where the adverb "così" [so] refers back to the entire content of the interview, annotated by means of the tag `<topic>`, which contains also each single event instance, here annotated with the tag `<markable>`:

9. "**Stiamo ancora parlando, come certamente deve essere, e continueremo a consultarci**"$_j$. James Baker, segretario al Tesoro americano, ha commentato *così*$_j$ i risultati dell'assemblea. (ISST els019)
"[We are still speaking, as it should be, and we will keep consulting]"$_j$. James Baker, the American Trea-

sure secretary, commented [so]ⱼ the results of the assembly.

```
''<TOPIC id="t1">Stiamo ancora
<MARKABLE id="m1">parlando</MARKABLE>,
come certamente
<MARKABLE id="m2">deve</MARKABLE>
<MARKABLE id="m3">essere</MARKABLE>, e
<MARKABLE id="m4">
continueremo
</MARKABLE>
a <MARKABLE id="m5">
consultarci
</MARKABLE>
''</TOPIC>.
 James Baker,
segretario al Tesoro americano, ha
<MARKABLE id="m6">
commentato<
/MARKABLE>
<MARKABLE id="m7">
<LINK id="l1" anaphorType="topic"
src="t1"/>
cosi'
</MARKABLE> i
<MARKABLE id="m8">
risultati
</MARKABLE>
dell'
<MARKABLE id="m9">
assemblea
</MARKABLE>.
```

The annotation of the anaphoric relations between events is performed by means of the `<link>` tag. This tag is devoted to the annotation of anaphoric relations between events, a level of annotation which could be included as an autonomous level of annotation in TimeML, thus eliminating the non-temporal uses of the TLINK tag. As illustrated in example 9., the `<link>` tag is a nested tag inside the anaphoric element. The attribute "`src`" is use to signal the anchor element of the anaphor. Finally, the attribute "`anaphorType`" is responsible for the identification of the particular relation between the anaphoric event and its anchor(s). The preliminary set of values of this attributes are based on previous annotation schemes for anaphora annotation such as MATE (Davies et al., 1998) and MATE/GNOME (Poesio, 2004) and are listed below:

- *coref*: two events stand in a coreferential relation if they share the same participants and if the anaphoric event is exactly the same as the anchor;

- *synonym*: two events stand in a synonym relation if they share the same participants and the anaphoric event is a synonym or a near-synomyn of the anchor;

- *part_of*: two events stand in a part_of relation if the anaphoric event is a part of or a sub-event the anchor;

- *hypernymy*: two events stand in a hypernymy relation if the anaphoric event is a subclass of the anchor;

- *topic*: this value indicates that the anaphoric element or event refers back to a portion of texts or to a set of events previously mentioned;

- *unclassified*: two events stand in an unclassified relation when the none of the previous relations applies. Notice that the under this value lots of cases of bridging anaphors could be reported. Such a strategy has been developed, on the one hand, to avoid the loss of information and, on the other hand, to develop a set of data-based heuristics which exploit processing requirements for the analysis of this class of anaphors.

In order to facilitate the task, we have imported the annotated events from BAT into another tool, PALinkA (Orasan, 2003). In this way the annotators[1] have not to decide what was an event, which is *per sé* a relative difficult task. The annotators, three students of a M.A. in Linguistics, have been provided with a reduced set of guidelines, which provided a description of the task (i.e. annotation of anaphoric relations between events) and of the tags. The only rule was related to the choice of the anchoring element which stated that whenever possible the nearest anchor was to be preferred with a lookback of maximum of five sentences before. However, this was not an inviolable rule. The annotators were allowed to overcome it. Although this possibility is not a good practice in anaphora annotation, since it may reduce the overall agreement and, thus, making unreliable the resulting annotated resource, we were interested in discovering if eventive elements, regardless their POS, have stronger referential properties than entity denoting elements. As for the arguments of an event, contrary to previous experiments, they were not annotated unless they were events themselves.

The choice of structuring this annotation experiment in a quite open way is strictly related to the idea of developing an annotation scheme by applying a coarse grained bottom-up approach in order to discover, on the one hand, short-comings in the annotation scheme and how to develop reliable annotation procedures, and, on the other hand, to study the phenomenon of event anaphora. Under this perspective, an evaluation of the annotation scheme in terms of classical statistical measure is out of the purpose of this work.

3.2. The tool: PALinkA

The annotation of the anaphoric relations was carried out by the annotation tool PALinkA which has been previously used in other coreference annotation tasks. Due to reason of time we did not had access to the accommodated version used in (Hasler et al., 2006). However, this did not result in real issues for the annotation (e.g. impossibility to annotate relations or similar), although the newer version of the tool would have been more user-friendly.

4. Analysis, evaluation and feedback

We have performed a preliminary analysis of the phenomenon of event anaphora and of the annotation scheme on a set of three texts (ISST cs031, cs037 and els003) for a

---

total of 1,792 words.

The first result is related to the anaphora detection. This type of analysis has no specific classes of analysis using the kappa measure or other agreement measure, such as precision and recall. In Table 3 we report the anaphora detection agreement. The acronyms used in the table mean: A1: anaphors detected by annotator 1; A2: anaphors detected by annotator 2; A3: anaphors detected by annotator 3; CA: detected anaphors by all 3 annotators.

	cs031	cs037	els003
A1	33	15	21
A2	38	17	15
A3	30	13	13
CA	14	5	9

Table 3: Anaphora detection agreement.

The results illustrated in Table 3 are not satisfying and partly in-line with our expectations. In particular, the relative low number of common anaphoras is a clear evidence that the anaphoric relations which may arise between events are multiple and seem to require more information available to the annotators in order to improve their identification. However, it is interesting to point out that detected anaphoric elements on which there is no agreement are not mistakes. This means that, on the one hand, event anaphora is a relatively widespread phenomenon in text/discourse and, on the other hand, that guidelines and specifications must be reinforced since in their current version have prevented a uniform application of the criteria to identify anaphoric elements. Moreover, the identification of the participants of the events is an essential feature in order to recognize when two event stand in an anaphoric relation. Cases of ellipsis or zero anaphora which refer to the participants of events have certainly influenced the identification of anaphoric instances. A further element which may have influenced the identification of the anaphoric elements is the commonsense knowlege. The way an event is perceived is very important. A possible solution to avoid the influence of commonsense knowledge could be represented by the use of linguistic theories of event structure, like Frame Semantics . This does not mean that the annotation scheme must mirror a linguistic theory, but in order to obtain good results, both in the task definition and in the developement of guidelines and specifications detailed instructions inspired by a linguistic theory *à la* FrameNet would be very useful. As already stated, this study was an explorative and speculative study on the phenomenon of event anaphora also as a possible task to be done by *naive* annotators in platforms like Amazon Mechanical Turk. The very poor results obtained in the identification of anaphoric elements do not prevent the idea of developing it as a collaborative internet-based annotation task, but due to its difficulty, the use of experts and well instructed annotators would be more appropriate.

We have also analyzed the agreement on the anchor identification, which corresponds to the attribute "`src`" of the `<link>` tag. The results are illustrated in Table 4.

Anchor	cs031	cs037	els003
Complete Agreement	9	3	9
Partial Agreement	3	0	0
No Agreement	2	2	0

Table 4: Anchor detection agreement.

As it can be observed, the annotators obtained a relative good level of agreement in the identication of the anchor. This suggests that the five sentence window is an appropriate text span for the identification of event anaphora. However, going into the details of the disagreement and of partial agreement instances, we have observed that some annotators have not respected the text span, signaling antecedents which are at a larger distance (one of them has identified a case of anaphoric relation with an anchor whose distance is of 31 sentences before).

As for the anaphoric relations, the results are very poor. With the exception of the "`coref`" value, on which annotators always agree, the agreement on all other anaphoric relation types is very poor.

5. Conclusion and future work

In this work we have described an experiment on the annotation of event anaphora. With respect to previous works, we have adopted a bottom-up strategy to find out shortcomings and strategies to improve the feasability of the task. The results obtained are very poor in terms of standard evaluation techniques for annotation schemes and language resources. Nevertheless, they have been extremely informative on how to improve the task and to develop annotation specifications which can path the way to a reliable annotation. In any case, the task is not a trivial one. Our annotators can be considered as almost experts and have found lots of difficulties to accomplish this task.

So far some suggestions on how to modify the annotation scheme have emerged. In particular, the `<event>` tag and annotation methodology from TimeML can be adopted and integrated in our scheme, the `<markable>` tag is to be reserved only to pronouns and adverbial items. As for the use of this latter tag, standard anaphora resolution annotation procedure must be adopted, i.e. annotate all pronouns and adverbs, regardless of their status, and then resolve the possible anaphoric cases to events. Similarly, the tag `<empty>` should be maintained for zero anaphors and ellipsis in order not to loose information. The tag `<topic>` does not present particular issues, although its annotation is subordinated to the identification of an anaphoric relations. This may be a possible cause of inconsistencies.

As for the annotation of the anaphoric relation types, provided the not satisfying results, we propose to use coarse grained values, namley "`direct`" and "`indirect`". The first value has to be used for cases of coreference (repetition of the same token or synonyms), while the second applies to all cases of indirect anaphoric relations, such as part_of relations, hypernyms, set/subset and others.

Open issues are related to the annotation of the participants and classification of the events. The experiment has highlighted how the presence of this kind of additional informa-

tion could be useful but at the moment no practical solution has been identified and further experiments are under development.

An interesting result is represented by the compatibility of this annotation scheme with TimeML. In particular, the TimeML annotation methodology for events can be applied without modifications and the tags proposed in this scheme add a layer of information in TimeML. Moreover, the temporal information explicited in TimeML can be used to facilitate the event anaphora resolution task.

6. References

N. Asher and A. Lascarides. 1998. Bridging. *Journal of Semantics*, 15(1):83–113.

C. A. Bejan and S. Harabagiu. 2008. A linguistic resource for discovering event structures and resolving event coreference. In *Proceedings of the Sixth International conference on Language Resources and Evaluation (LREC-08)*.

E. Boldrini, M. Puchol-Basco, B. Navarro, P. Martínez-Barco, and C. Vargas-Sierra. 2009. Aqa: a multilingual anaphora annotation scheme for question answering. *Procesamiento del Linguaje Natural*, 42:97–104.

S. Davies, M. Poesio, F. Bruneseaux, and L. Romary. 1998. Annotating coreference in dialogues: proposal for a scheme for mate. first draft.

L. Hasler, C. Orasan, and K. Naumann. 2006. Nps for events: Experiments in coreference annotation. In *Proceedings of the 5th edition of the International Conference on Language Resources and Evaluation (LREC2006)*.

L. Hirschman and N. Chinchor. 1997. Muc-7 coreference task definition.

G. Katz and F. Arosio. 2001. The annotation of temporal information in natural language senteces. In *Proceedings of the ACL-EACL 2001 Workshop for Temporal and Spatial Information Processing*.

S. Montemagni, F. Barsotti, M. Battista, N. Calzolari, O. Corazzari, A. Lenci, V. Pirelli, A. Zampolli, F. Fanciulli, M. Massetani, R. Raffaelli, R. Basili, M. T. Pazienza, D. Saracino, F. Zanzotto, N. Mana, F. Pianesi, and R. Delmonte. 2003. The syntactic-semantic treebank of italian. an overview. *Linguistica Computazionale, Computational Linguistics in Pisa, special Issue*, XVIII-XIX:461–93.

C. Orasan. 2003. Palinka: A highly customisable tool for discourse annotation. In *Proceedings of the 4th SIGDial Workshop on Discourse and Dialogue*, pages 39–43.

M. Poesio. 2004. The mate/gnome scheme for anaphoric annotation, revisited. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, pages 154–162.

J. Pustejovsky, J. Castao, R. Ingria, R. Saur, R. Gaizauskas, A. Setzer, and G. Katz. 2003a. TimeML: Robust specification of event and temporal expressions in text. In *Fifth International Workshop on Computational Semantics (IWCS-5)*.

J. Pustejovsky, P. Hanks, R. Saur, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. 2003b. The TIMEBANK corpus. In *Corpus Linguistics 2003*.

T.G. Rose, M. Stevenson, and M Whitehead. 2002. The reuters corpus volume 1 - from yesterday's news to tomorrow's lnaguage resources. In *Proceedings of the Third International conference on Language Resources and Evaluation (LREC-02)*.

A. Setzer. 2001. *Temporal information in newswire articles: an annotation scheme and a corpus study*. Ph.D. thesis, University of Sheffield.