Building a Temporally Annotated Corpus for Information Extraction

Andrea Setzer and Robert Gaizauskas

Department of Computer Science
University of Sheffield
Regent Court
211 Portobello Street
Sheffield S1 4DP, UK
{A.Setzer, R.Gaizauskas}@dcs.shef.ac.uk

Abstract

Although there is now considerable availability of what might be termed 'low-level' (e.g. part-of-speech) annotated corpora, such resources are only of limited use for Information Extraction (IE); however, there is only a very limited availability of 'high-level' annotated corpora, which are of direct use. Over time it has become evident that 'intermediate-level' annotations for, e.g., named entities and coreference chains are also needed in order to build better IE systems. We share this view and believe that other intermediate level annotations are necessary. One such intermediate level is that which marks temporal reference and relations in text. To this end we have devised an annotation scheme for annotating those features and relations in texts which enable us to determine the relative order and, if possible, the absolute time, of the events reported in them. The scheme is being used to construct a temporally annotated corpus. This paper describes the annotation scheme, how we are building this corpus and the corpus that has so far resulted. We have also conducted a small scale experiment to ascertain to what extend superficial IE techniques are likely to be able to determine temporal relations in a text, which illustrates the insights for IE that can accrue from constructing annotated corpora.

1. Introduction

While there is now considerable availability of what might be termed 'low-level' annotated corpora – for example, part-of-speech and constituent structure tagged corpora such as the British National Corpus (Burnard, 1995) and Penn TreeBank (Marcus et al., 1993) – these resources are of indirect and limited use for Information Extraction (IE). At the same time there is only very limited availability of 'high-level' annotated corpora, such as those created for the Message Understanding Conferences (MUC), which are of direct use for IE. The MUC resources are best known for associating target filled templates with texts. But over time the MUC community came to realise that 'intermediate-level' annotations for, e.g., named entities and coreference chains are also needed in order to build better IE systems. We share this view and believe that other intermediate level annotations are necessary. One such intermediate level is that which marks temporal reference and relations in text.

It is widely acknowledged that annotating a corpus enhances the usefulness of the corpus for research and development in general (see, e.g., **?**)) and for IE in particular (see, e.g., **?**)). Although temporal relations are very important for IE tasks, only minimal work has been done in either the IE community or by corpus linguists concerning the annotation and extraction of temporal information. This is evidenced, for example, by tasks set in recent Message Understanding Conferences (MUC). The MUC-6 named entity subtask required the identification of absolute time expressions in text (MUC, 1995), and the MUC-7 named entity subtask extended this requirement to include relative time expressions (MUC, 1998), but none of these tasks required placing events in time, or temporally relating events to each other. The MUC-5 and -7 scenario tasks did require participants to assign a calendrical time to certain specified event types (joint venture announcements and rocket launchings, respectively). However, this task is quite limited and the scores were low, indicating its difficulty.

The importance of extracting temporal information from texts, together with the difficulty of the task, suggest that a concerted effort be made to analyse how temporal information is actually conveyed in real texts. To this end we have devised an annotation scheme for annotating those features and relations in texts which enable us to determine the relative order and, if possible, the absolute time, of the events reported in them. The scheme is being used to construct an annotated corpus which will lead to: a better understanding of the phenomena of concern (both their variety and distribution); a resource for training adaptive algorithms to automatically identify features and relations of interest; and, a resource for evaluating algorithms which purport to identify the features and relations of interest. Our effort to date has been concentrated on newswire texts. These are a popular source for IE systems because of the wide range of potential applications. Such texts exhibit a rich variety of temporal phenomena, but we are aware of the limitations of working with a single text genre. However, we are optimistic that a temporal annotation scheme developed for this genre will not only prove useful for IE applications but will also be extensible to other genres.

In Setzer and Gaizauskas (2000) we describe the annotation scheme in some detail. Here we briefly recap this scheme (sections 2. and 3.) and then go on to discuss how we have begun to annotate a corpus using it (section 4.), the corpus that has so far resulted (section 5.), and a small scale experiment to ascertain to what extent superficial IE techniques are likely to be able to determine temporal relations in a text (section 6.). This experiment illustrates the insights for IE that can accrue from constructing annotated corpora.

2. Conceptualising Time

Before an annotation scheme for a temporally annotated corpus can be defined, it is necessary to make clear what kind of temporal entities and relations we suppose exist. Although we have profited from the discussions about which temporal ontology and which temporal relations are appropriate for analysing temporal phenomena in natural language (Allen, 1984; Galton, 1990; Steedman, 1997), our goal is not one of philosophical truth or description of linguistic completeness but rather the pragmatic one of providing a framework which will enable us to classify expressions in real texts in a way that enables us to gain useful insights into how temporal information is conveyed in written language. Our aim is to develop an annotation scheme with which we can build a useful temporally annotated corpus of newswire articles.

With this in mind, we will now summarise our descriptive framework. We presume the world contains the following primitives: events, states, times, temporal relations and subevent relations. Each primitive is described briefly below.

Events Intuitively an event is something that happens, something that one can imagine putting on a time map. Events can be ongoing or conceptually instantaneous, we do not distinguish between these. What defines an event is very much dependent on the application and domain, but generally events have to be anchorable in time and they are usually conveyed by finite verbs or by nominalisations. Examples of events are:

> *A small single-engine plane crashed into the Atlantic Ocean.*
>
> *The 1996 crash of the TWA 747 remains unexplained.*

Event Classes During our analysis of newswire articles it became clear that events can be classified into groups which will be beneficial when interpreting the temporal information contained in a text building a time map. Event classifications are not a new idea, (see Halliday (1985)), but existing ones were not tailored towards our goal of analysing temporal information in newswire articles.

We will give a short overview of our classification; for more information see Setzer and Gaizauskas (2000).

The most common events are what we call *occurrence events* – these are the events we want to place on a time map. Examples are:

> *A small single-engine plane crashed into the Atlantic Ocean about eight miles off New Jersey on Wednesday.*

Another class of events frequently found in newswire articles are *reporting events*, whose main function is to associate the source of information with an (occurrence) event. Examples are:

> ***The Coast Guard reported*** *finding aircraft debris and fuel slick.*

The following example illustrates another type of event:

> *The plane* ***was seen*** *hitting the water shortly after 11 a.m.* ***by a fisherman.***

We call these events *perception events* and although they are relatively rare, the benefits of annotating them justify their being included in the scheme.

Attitude events are similar to *reporting* and *perception* events in that they all take another event as an argument. Examples are *John hopes to go to New York on Friday* or *Mary believes that the plane hit the the Atlantic Ocean*.

The final class of events we distinguish is *aspectual events*, such as

> *The first interruption* ***began*** *1 minute and 39 seconds after the sound and lasted 1 minute and 17 seconds; the second was just before the end of the tape.*

which involve *aspectual verbs* like *start, stop, finish etc*. Their temporal consequence is that the aspectual event indicates the start or ending of the related event.

Times Like events, times can be viewed as having extent (intervals) or as being punctual (points). Rather than trying to reduce one perspective to the other, as has happened in much philosophical discussion on time, we shall simply treat both as *time objects*. A time object must, however, be capable of being placed on a time line (fictional or real).

Following general convention, and the approach taken in MUC, we distinguish between two classes of time objects, DATES and TIMES, times which are larger or smaller than a day, respectively.

States A state is a relation between entities or the holding of an attribute of an entity which, while capable of change, is ongoing over a time span, usually longer than the time span covered by the article. Examples are:

> *The plane,* ***which can carry four people****, ...*
>
> *The water is about 125 feet deep in that area.*

Typically, a change of state constitutes an event. At this point we are less interested in states, and we have not taken them into account in our annotation scheme.

Temporal Relations Events stand in certain temporal relations to other events and to times. Times may be temporally related to other times as well, although this does not happen very often in the articles we analysed so far.

> *The plane crashed* ***after*** *the pilot and his crew ejected.*
>
> *A small single engine plane crashed into the Atlantic Ocean* ***on*** *Wednesday.*

The full set of temporal relations we suppose at present is:

included *The plane crashed on Wednesday*

includes *By midafternoon, several vessels were combing the area*

after *The plane crashed after the pilot and his crew ejected.*

before *...before the craft fell, its three rotor blades shot off*

simultaneous *All 75 people on board the Aeroflot Airbus died when it ploughed into a Siberian mountain in March 1994*

This is a minimal set we defined after analysing a number of newswire articles and it can easily be expanded should it prove necessary or beneficial.

Subevent Relations Subevent and event identity (which can be viewed as a special case of subevent) are quite commonly used in newswire articles where it is common practise to introduce events and them come back to them later in the same text and provide more information. This is exhibited in the following example:

> *A small single-engine plane* ***crashed*** *into the Atlantic Ocean about eight miles off New Jersey on Wednesday.*
>
> *[. . .]*
>
> *The plane, which can carry four people, was seen* ***hitting*** *the water shortly after 11 a.m. by a fisherman, ...*

The first sentence introduces the event and the second event refers to a subevent (*hitting the water*), with the subevent providing more precise information, i.e. that the plane crashed happened around 11 a.m. on the Wednesday in question.

3. Annotating Temporal Information in Text

Now that we have a conceptual framework, we can propose an annotation scheme which will enable us to mark up events, times, and temporal and subevent relations. The annotation scheme is defined in SGML. First, we briefly describe the annotation scheme (for more detailed information see Setzer and Gaizauskas (2000)). We then discuss specific annotation issues regarding the annotation of implicit temporal information, which will lead us to a discussion of comparing the results of different annotators, for which we will introduce the notion of a minimal temporal model of a text. Finally, we discuss the actual process of annotation.

3.1. The Annotation Scheme

3.1.1. Annotating Events

Events are marked by annotating a representative in the clause conveying the event. The first choice for a representative is the head of the finite verb group. If a nominalisation conveys the event, then the head of the nominalisation serves as the representative and in the rare case of an event being conveyed by a non-finite clause, the non-finite verb is marked as the representative.

An events carries attributes for some or all of the following properties: unique event ID, event class (see section 2.), verb tense, verb aspect, which event it is related to and by which temporal relation, which time object it is related to and by which temporal relation, the word(s) by which the temporal relation is signalled, and the ID of events it might have as an argument. For example:

A small single-engine plane
<event eid=16 class=OCCURRENCE tense=past>
crashed
</event>
into the Atlantic Ocean about eight miles off New Jersey

3.1.2. Annotating Times

Time expressions are uniquely identified by an ID. As mentioned in section 2., we distinguish between the types TIME and DATE, as the following example shows.

<timex tid=5 type=DATE> Tuesday </timex>
<timex tid=5 type=TIME> 11 a.m. </timex>

3.1.3. Annotating Temporal Relations

Events and times can be related to other events or times. If two events are related then one of the events carries the OD of the other as well as the temporal or subevent relation in which they stand to each other. If an event is related to a time then the event carries the ID of the time object and the temporal relation. In either case, if the relation is signalled explicitly in the text, then the ID of this signal is an attribute as well, as the following two examples illustrate [1].

All 75 people on board the Aeroflot Airbus
<event eid=4 class=OCCURRENCE tense=past
relatedToEvent=5 eventRelType=simultaneous
signal=7>
died </event>
<tr_signal sid=7> when </tr_signal>
it
<event eid=5 class=OCCURRENCE tense=past >
ploughed </event>
into a Siberian mountain.

A small single-engine plane
<event eid=9 class=OCCURRENCE tense=past
relatedToTime=5 timeRelType=included
signal=9>
crashed </event>
into the Atlantic Ocean about eight miles off
New Jersey
<tr_signal sid=9> on </tr_signal>
<timex tid=5> Wednesday </timex>.

3.2. Annotating Implicit Temporal Relations

Temporal relations are not always made explicit in a text. For example, simple juxtaposition can be used to relate an event to a time object:

> *A senior investigator who looked at the cockpit wreckage Tuesday ...,*

Further, instead of expressing subevent relations directly (which is rarely done), the reader is generally presumed to have the script-like knowledge of stereotypical scenarios needed to infer the temporal relations. Although there are

[1] We are aware of the problem of relating an event to more than one other event or time. Our annotation scheme can easily be adapted in numerous ways to accommodate this.

many ways of temporally relating events in an implicit way, e.g. narrative sequence, but they are either not widely used in newspaper articles or too difficult to define in an annotation scheme. To date, we are concentrating only on the two mechanisms mentioned above.

Implicit temporal relations are easily annotated within the annotation scheme proposed in section 3.1.. However, equivalent implicit relations can be marked up in a variety of ways. For instance, if A and B are *simultaneous* and C is implicitly *after* A and B, then C might be annotated as *after* A or *after* B or both. Hence it is a very difficult task to produce an annotation guideline which would result in the same annotation independent of the annotator. Since, as the preceding example shows, much of the temporal information that could be marked is redundant, we would expect high inter-annotator disagreement.

This problem of annotating implicit temporal information is similar to the problem of coreference annotation in MUC-6 and -7 (1995; 1998). There too the relation being annotated (coreference) could be marked in superficially different, but semantically identical ways (equivalent coreference chains can be specified by annotating different links). The solution adopted in MUC was to define a model-theoretic scoring scheme which relied upon comparing the equivalence classes of the set of linked entities produced by the annotator 1 with that produced by annotator 2 (**?**). Our approach is similar, but the problem more complex because there are more relations and not all of the relations are equivalence relations.

3.3. Comparing Temporal Annotations

The events and times annotated in a text, or rather their respective IDs, form two sets, E and T, respectively. Since all of our temporal annotations are binary relations relating events or times to other events or times, the denotation of each relation as specified in the text can be viewed as a subset of $(E \cup T) \times (E \cup T)$. For each temporal relation certain formal properties pertain. For example *simultaneous* is an equivalence relation, while *before*, *includes* and *subevent* are transitive, but asymmetric and irreflexive. Therefore, given a partially specified model of the temporal relations in a text (specified as a set of pairs comprising a part of the denotation of the relation) the deductive closure of each relation can be computed to arrive at a total model. If the deductive closures of two partially specified models are identical, then these two models are equivalent, though not themselves identical. Further if any (partial) model $\mathcal{M}$ is such that no proper subset of $\mathcal{M}$ has an equivalent deductive closure to $\mathcal{M}$, then $\mathcal{M}$ is a *minimal* model of the temporal relations in the text.

Let us denote sets of pairs from $(E \cup T) \times (E \cup T)$ which constitute the denotations of *simultaneous, before, includes* and *subevent* by S, B, I, and E respectively. The set of inference rules we need to compute the deductive closure is not yet complete, but will at least contain rules to exploit the transitivity of all the relations and that *simultaneous* is an equivalence relation, as well as rules like the following: $\forall x, y, z \in (E \cup T) \times (E \cup T)$:

- $(x, y) \in B \wedge (y, z) \in S \Rightarrow (x, z) \in B$
- $(x, y) \in I \wedge (y, z) \in S \Rightarrow (x, z) \in I$
- $(x, y) \in S \wedge (y, z) \in I \Rightarrow (x, z) \in S$
- $(x, y) \in S \wedge (x, z) \in B \Rightarrow (y, z) \in B$
- $(x, y) \in I \wedge (y, z) \in B \Rightarrow (x, z) \in B$
- $(x, y) \in E \Rightarrow (y, x) \in I$

We then can denote the deductive closure of S, B, I, and E by $S^{\models}$, $B^{\models}$, $I^{\models}$, and $E^{\models}$ respectively. Given these definitions we are now in a position to specify what precision and recall mean in this framework. Letting S_k and S_r denote the annotated *simultaneous* relations in the answer key and system response respectively and $S_k^{\models}$ and $S_r^{\models}$ their deductive closures, respectively (and similarly for B, I and E). The recall and precision for the *simultaneous* relation is given by:

$$R = \frac{|S_k^{\models} \cap S_r^{\models}|}{|S_k^{\models}|}$$

$$P = \frac{|S_k^{\models} \cap S_r^{\models}|}{|S_r^{\models}|}$$

Recall and precision measures can be defined in a parallel fashion for the other relations. An overall recall and precision measure for all temporal relation could be defined as follows (note that subeventness is factored out in this equation because, as the final inference rule above implies, the members of $E^{\models}$ are also included in $I^{\models}$ and they must not be counted twice):

$$R = \frac{|S_k^{\models} \cap S_r^{\models}| + |B_k^{\models} \cap B_r^{\models}| + |I_k^{\models} \cap I_r^{\models}|}{|S_k^{\models}| + |B_k^{\models} + |I_k^{\models}|}$$

$$P = \frac{|S_k^{\models} \cap S_r^{\models}| + |B_k^{\models} \cap B_r^{\models}| + |I_k^{\models} \cap I_r^{\models}|}{|S_r^{\models}| + |B_r^{\models} + |I_r^{\models}|}$$

4. The Process of Annotation

Using this scheme we have annotated a small trial corpus (see section 5. below). The annotation took place in several stages. During the first stage, all events and times were annotated, without paying attention to whether they were related or not. In a second phase we annotated all explicit temporal relation signals (e.g. *after, when, on*) and recorded the entities (events or times) that were related by those. This results in having all explicit relations included in the annotated text. The tedious task of annotating events and times could be supported by a system which automatically pre-annotates these expressions and only asks the annotator to confirm or edit these.

The third stage involves marking those implicit temporal relations we described in section 3.2. and is currently carried out manually. However, one could envisage supporting it with a program to prompt the annotator about missing relations.

In order to build a complete model of the temporal information contained in a text (as described in section 3.3.),

	sentences	words	avg. word/sentence
text1	16	332	20.75
text2	13	268	20.61
text3	13	212	16.30
text4	11	263	23.90
text5	20	489	24.45
text6	18	457	25.38
text7	10	210	21.00
text8	20	501	25.05
sum	121	2732	177.44
avg.	15.12	341.5	22.18

Table 1: Size of newswire texts.

	all	distinct	explicit	implicit	times
text1	31	27	4	11	5
text2	19	14	5	7	3
text3	30	28	2	7	1
text4	24	23	8	5	3
text5	42	39	7	4	2
text6	25	23	7	7	11
text7	16	16	2	3	4
text8	13	13	10	2	5
sum	200	183	45	46	34
average	25	22.87	5.62	5.75	4.25

Table 2: Number of events in the corpus.

we first analyse the annotated information, draw all possible inferences and assign as many pairs as possible from $(E \cup T) \times (E \cup T)$ to one of the relations in S, B, I and E. Because of the formal properties of the relations (e.g. antisymmetry of *before*) certain pairs are also *excluded* from certain relations (e.g. if event A is before event B then B cannot be before A). The annotator is then prompted for any unknown pairs. To minimise the number of questions that need to be asked, we will try and develop a scheme which asks those questions first which allow as many inferences as possible. This model then will allow us to evaluate inter-annotator agreement, as well as computing precision and recall (as described earlier).

To enable simple and user friendly annotation, we have developed a graphical Annotation Tool in Perl/Tk which allows the annotator to select a portion of the text with the mouse and click on the appropriate type. At this point the annotator is prompted to enter all necessary attributes, some of which he or she can choose from a selection box; Figure **??** shows a screen shot of the implemented system.

5. The Corpus

Our trial corpus currently comprises 8 New York Times newswire articles, manually annotated with our annotation scheme. Table 1 shows the size of the texts, in number of sentences and number of words. Table 2 shows the number of events in each text (column *all*) and the number of these which were distinct events, i.e. not identical to another event (column *distinct*). The third column (*explicit*) contains how many events were explicitly related to either another event or time and column *implicit* shows how many events were implicitly related by the means we discussed in section 3.2.. The last column (*times*) shows the number of time objects contained in the text. Of all events in the texts, 22% were annotated as *explicitly* and 23% as *implicitly* standing in a temporal or subevent relation to another event or time. This leaves 109 events being annotated as not being temporally related to another event or time. The interesting result here is, that the implicitly related events are as great in number as the explicitly related ones.

Table 3 gives an overview of the distribution of events over the type of clauses by which they are conveyed. The majority of events, 79%, are conveyed by tensed clauses and a much smaller number of events are conveyed by nominalisations (9.5%) and non-finite clauses (11.5%).

	tensed	nominal.	non-finite
text1	24	5	2
text2	24	4	2
text3	17	2	0
text4	17	2	5
text5	32	0	10
text6	19	4	2
text7	13	1	2
text8	12	1	0
sum	158	19	23
average	19.75	2.37	2.87

Table 3: Types of Clauses

6. Working with the Annotation Scheme

An interesting question is to what extent events in text can be ordered using only the information that is captured with our annotation scheme and to this end we have conducted an experiment in which the annotation scheme was applied to three New York Times newswire texts. The annotation was then separated from the text and analysed independently to determine how well events could be located in time without actual textual content, i.e. solely based on information like <event eid=16 class=OCCURRENCE tense=past>. We then manually created the deductive closure, as described in section 3.3., and compared the results to the complete model of the temporal information contained in the text. This complete model was also manually built, using all available knowledge (annotation, textual content and general world knowledge).

This method resulted in 37% Recall and, since no false temporal relations were annotated, 100% Precision; but more articles need to be examined to confirm these results. We recently revised the evaluation scheme and thus only three texts could be evaluated, but the experiment shows how the annotation scheme may provide insights into the mechanisms by which temporal information is conveyed in text.

7. Conclusions

The annotation scheme we propose is a general purpose scheme. It is flexible enough to be straightforwardly adapted to other applications by, e.g., adding different types of classes or temporal relations or expanding the ontology.

As illustrated in sections 5. and 6., the annotated corpus gives information about the variety and distribution of the temporal phenomena in newswire texts. With these figures we know the relative importance of the different mechanisms of conveying temporal information and we can exploit the most important ones. This will lead to better IE systems. In particular, these figures suggest that superficial IE techniques that use only surface signals to determine temporal relations are unlikely to be sufficient to locate the bulk of the events in time.

A better understanding could also lead to more user friendly annotation tools. Explicit temporal information, for example, could easily be automatically pre-annotated and would only need confirmation from the annotator (this idea is also suggested in **?**)).

We have begun to build a corpus, but it needs to be expanded in order to get more significant results. This, and also interannotator agreement, which needs to be determined, may result in a revision of the annotation scheme. We also aim at developing more tools to aid analysis of the corpus as well as supporting the actual process of annotation.

In this paper, we have concentrated on the genre of newswire articles and it would be interesting be to see how the annotation scheme would transfer to other genres.

Acknowledgements We would like to thank Mark Hepple, Nick Webb and Mark Stevenson for their help.

8. References

J.F. Allen. 1984. Towards a General Theory of Action and Time. *Artificial Intelligence*, 23:123–154.

L. Burnard, 1995. *Users Reference Guide for the British National Corpus*, May.

A. Galton. 1990. A Critical Examination of Allen's Theory of Action and Time. *Artificial Intelligence*, 42.

M.A.K. Halliday. 1985. *An Introduction to Functional Grammar*. Edward Arnold.

M.P. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1993. Building a large annotated corpus of english: The penn tree bank. *Computational Linguistics*, 19(2).

1995. *Proceedings of the Sixth Message Understanding Conference (MUC-6)*. Morgan Kaufman.

1998. *Proceedings of the Seventh Message Understanding Conference (MUC-7)*. Morgan Kaufman. Available at http://www.saic.com.

A. Setzer and R. Gaizauskas. 2000. Building a Temporally Annotated Corpus for Information Extraction. In *Proceedings of the Workshop Information Extraction Meets Corpus Linguistics held in conjunction with the Second International Conference on Language Ressources and Evaluation (LREC2000)*.

M. Steedman. 1997. Temporality. In J. van Bentham and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 895–938. Elsevier.