

The TIMEBANK Corpus

James Pustejovsky
Patrick Hanks
Roser Saurí
Andrew See
Dept. of Computer Science
Brandeis University
jamesp@brandeis.edu
patrick.hanks@rcn.com
roser@cs.brandeis.edu
aragorn@cs.brandeis.edu

Robert Gaizauskas
Andrea Setzer
Dept. of Computer Science
University of Sheffield
R.Gaizauskas@dcs.shef.ac.uk
A.Setzer@dcs.shef.ac.uk
Fax: +44 (0) 114 22 21810

Dragomir Radev
University of Michigan
radev@umich.edu

Beth Sundheim
SPAWAR Systems Center
beth.sundheim@navy.mil

David Day
Lisa Ferro
Marcia Lazo
The MITRE Corporation
day@mitre.org
lferro@mitre.org
mlazo@mitre.org

Abstract

While the mechanisms for conveying temporal information in language have been extensively studied by linguists, very little of this work has been done in the tradition of corpus linguistics. In this paper we discuss the outcomes of a research effort to build a corpus, called TIMEBANK, which is richly annotated to indicate *events*, *times*, and *temporal relations*. We describe the annotation scheme, the corpus sources and tools used in the annotation process, and then report some preliminary figures about the distribution of various phenomena across the corpus. TIMEBANK represents the most fine-grained and extensively temporally annotated corpus to date, and will be a valuable resource both for corpus linguists interested in time and language, and for language engineers interested in applications such as question answering and information extraction for which accurate knowledge of the position and ordering of events in time is of key importance.

1. Introduction

During 2002 an extended workshop, called TERQAS¹, was held to address the problem of how to answer temporally based questions about the events and entities in text, especially news articles. It was motivated by the fact that questions like the ones below are not supported in a unified way by existing question answering (QA) systems:

1. Is Gates currently CEO of Microsoft?
2. When did Iraq finally pull out of Kuwait during the war in the 1990s?
3. Did the Enron merger with Dynegy take place?

Such questions cannot be answered without paying attention to the temporal properties, modalities and ordering of events.

There has recently been a renewed interest in temporal and event-based reasoning in language, in particular when applied to information extraction and reasoning tasks (cf. *ACL Workshop on Temporal and Spatial Reasoning*, 2001; LREC workshop *Annotation Standards for Temporal Information in Natural Language*, 2002). One of the major differences between this work and earlier work in AI and computational linguistics on temporal reasoning and temporal semantics has been a focus on the identification of the real linguistic phenomena that are used to convey temporal information in text. This has led to a growing movement to build corpora annotated with temporal information – not simply times and dates, but events and temporal relations between events. Clearly an understanding of these phenomena, as they occur in real texts, is necessary to be able to design systems to answer questions such as those above.

¹ The TERQAS (Time and Event Recognition for Question Answering Systems) workshop was held at MITRE Bedford and Brandeis University. It was funded through the US National Regional Research Centre (NRRRC) and fully sponsored by the US Advanced Research and Development Agency (ARDA).

To advance this understanding, the TERQAS workshop had the following major objectives:

1. To define and design a common meta-standard for the mark-up of events, their temporal anchoring, and how they are related to each other in news articles (to be called TimeML, for Time Markup Language).
2. To create a gold-standard human-annotated corpus marked up for temporal expressions, events, and temporal relations, based on the TimeML specification (TIMEBANK).

In this paper, we give a brief overview of the TimeML annotation scheme and its development (section 2), but we will focus mainly on the description of the TIMEBANK corpus (section 3). TIMEBANK contains 300 newswire articles with careful, detailed annotations of terms denoting events, temporal expressions, and temporal signals, and, most importantly, of links between them denoting temporal relations. This collection, the largest temporal-event annotated corpus to date, provides a solid empirical basis for future research into the way texts actually express and connect series of events. It will support research into areas as diverse as the semantics of tense and aspect, the explicit versus implicit communication of temporal relational information, and the variation in typical event structure across narrative domains (e.g. striking differences were observed by annotators between the typical event structure of business reports and the typical event structure of political, social, and military events). Finally, it will support research efforts to construct practical QA and information extraction systems which are sensitive to temporal information.

2. Annotating the Corpus

Unlike most previous attempts at event and temporal specification, TimeML separates the representation of event and temporal expressions from the anchoring or ordering dependencies that may exist in a given text. There are four major data structures that are specified in TimeML (Ingria and Pustejovsky, 2002, Pustejovsky et al., 2003): EVENT, TIMEX3, SIGNAL, and LINK. These are described in some detail below. The features distinguishing TimeML from most previous attempts at event and time annotation are summarized below:

1. Builds on the TIMEX2 guidelines for temporal expressions (Ferro et al., 2001) but introduces, for example, temporal functions to allow intensionally specified expressions: *three years ago*, *last month* (TIMEX2 annotates these as well, but requires the annotator to supply an ISO value).
2. Identifies signals determining interpretation of temporal expressions:
 - a. Temporal Prepositions: *for*, *during*, *on*, *at*;
 - b. Temporal Connectives: *before*, *after*, *while*.
3. Identifies all classes of event expressions:
 - a. Tensed verbs: *has left*, *was captured*, *will resign*;
 - b. Stative adjectives: *sunken*, *stalled*, *on board*;
 - c. Event nominals: *merger*, *Military Operation*, *Gulf War*.
4. Creates dependencies between events and times:
 - a. Anchoring: *John left on Monday*.
 - b. Orderings: *The party happened after midnight*.
 - c. Embedding: *John said Mary left*.

2.1. The Conceptual and Linguistic Basis

Before the annotation scheme is described in detail, it is necessary to make clear what kind of temporal entities and relations we suppose exist. The main entities are similar to the ones in Setzer (2001), on which the development of TimeML was based. They are **events**, **times**, and **relations** holding between events and events and times.

Events are taken to be situations that *occur* or *happen*. They can be punctual (*John reached the summit*) or last for a period of time (*John walked up a mountain*). Events are generally expressed by the means of tensed or untensed verbs, nominalisations, adjectives, predicative clauses, or prepositional phrases.

Predicates describing states or circumstances in which something obtains or holds true are also considered as events. However, only those stative predicates are marked up that participate in an opposition structure in a given text. For example, one can imagine that the state described in the expression *New York is on the east coast* holds true for a time span that is considerably longer than the time span containing the events of focal interest in a typical newswire text. This expression would not be annotated. An expression like *the people on board the airbus*, appearing in a newswire text about a crash of that airbus, can be seen as describing a state that would change within the time span covered by the text – and would thus be annotated.

Times may be either points, intervals, or durations, and may be referred to by fully specified temporal expressions like *June 11, 1989*, underspecified temporal expressions such as *Monday*, intensionally specified expressions like *last week*, and duration expressions like *three months* or *two years*.

Relations can hold between events and events and times. In order to correctly identify the temporal order of events, it is necessary to identify more than temporal relations in a text. Subordinate relations, for example, can indicate that an event is hypothetical and thus should receive a different temporal interpretation. There are also aspectual relations (*John started to run*), or relations representing polarity (*John did not run*) and temporal quantification (*John ran three times*).

There are three types of relations included in TimeML: temporal relations, subordinate relations, and aspectual relations. The full set is detailed in section 2.3.

2.2. The Annotation Scheme

With the conceptual framework in place, we can now proceed to present the XML-based TimeML annotation scheme. This is a brief overview of the annotation scheme; for more detailed information please see Pustejovsky et. al. (2003).

Annotating Events

Events are marked up by annotating a representative of the event expressions, usually the head of the verb phrase. Note that generics like *Use of corporate jets for business travel is legal* are not tagged as events. The attributes of events are a unique identifier, the event class, tense, and aspect. The following fully annotated example for an event expression illustrates the even annotation.

```
All 75 passengers <EVENT eid="1" class="OCCURRENCE" tense="past" aspect="NONE"> died
</EVENT>.
```

Events are subcategorised into the following classes: occurrence, perception, reporting, aspectual, state, intensional state, intensional action, and modal.

Annotating Times

The annotation of times was designed to be as compatible with the TIMEX2 time expression annotation guidelines (Ferro et al., 2001) as possible. The XML tag for time expressions within TimeML is called TIMEX3 to distinguish it from both the tags in Setzer (2001) and TIMEX2. TIMEX3 has the following attributes:

- A unique identifier: **tid**,
- The attribute **functionInDocument** indicates the function of the TIMEX3 in providing a temporal anchor for other temporal expressions. For example, the date of a newspaper article is often used to anchor underspecified temporal referring expressions.
- A TIMEX3 object is of a certain **type**, either a date, a time, or a duration.

- Time expressions like *last week* are analysed by using temporal function and the attribute **temporalFunction** indicates that the TIMEX3 is of this kind. Which temporal function is to be used, is registered in the attribute **TemporalFunctionID**.
- The actual value of a temporal expression, for example *1998*, is stored in the **value** attribute.
- Temporal expressions often have to be analysed with respect to another time or an event. The ID of these anchors can be found in **anchorTimeID** and **anchorEventID**.

Straightforward time expressions like July 1966 are annotated as follows:

```
<TIMEX3 tid="1" type="DATE" value="1966-07" temporalFunction="false"> July 1966
</TIMEX3>
```

This kind of annotation is well known from earlier annotation schemes adopted in TIDES (Ferro et. al., 2001) and the MUC competitions (Message Understanding Conferences, see MUC6 (1995) and MUC7 (1998)). One of the major contributions of TimeML is the treatment of temporal functions, which allows time-value dependent algorithms to delay the computation of the actual (ISO) value of the expression. The following informal paraphrase of some examples illustrates this point, where DCT is the Document Creation Time of the article.

- *last week* = (predecessor (week DCT))
That is, we start with a temporal anchor, in this case, the DCT, coerce it to a week, than find the week preceding it.
- *the week before last* = (predecessor (predecessor (week DCT)))
Similar to the previous expression, except that here we go back two weeks.

Annotating Signals

The SIGNAL tag is used to annotate sections of text, typically function words, that indicate how temporal objects are to be related to each other. The functionality of the SIGNAL tag was introduced by Setzer (2001). In TimeML it is also used to mark polarity indicators such as *not*, *no*, *none*, etc., as well as indicators of temporal quantification such as *twice*, *three times*, and so forth. Signals have only one attribute, a unique identifier. The following annotated example illustrates the annotation of a signal:

```
Two days <SIGNAL sid="1" before </SIGNAL> the attack ...
```

Annotating Temporal and other Relations

To annotate the different types of relations that can hold between events and events and times, the LINK tag has been introduced. This marks a significant departure from Setzer (2001). There are three types of link tags.

1. A TLINK or Temporal Link represents the temporal relationship holding between events or between an event and a time, and establishes a link between the involved entities making explicit if they are: *before*, *after*, *includes*, *is included*, *holds*, *simultaneous*, *immediately after*, *immediately before*, *identity*, *begins*, *ends*, *begun by*, *ended by*.
2. An SLINK or Subordination Link is used for contexts introducing relations between two events, or an event and a signal. SLINKs are of one of the following sorts: **Modal**, **Factive**, **Counter-factive**, **Evidential**, **Negative evidential**, **Negative**.
3. An ALINK or Aspectual Link represents the relationship between an aspectual event and its argument event. Examples of the aspectual relations to be encoded are: *initiation*, *culmination*, *termination*, *continuation*.

A TLINK records the relation between two events, or an event and a time. Thus, the attributes include the IDs of the source and the target entity, the ID of the signal representing the type of relation, and the relation type. The attributes for an SLINK include the IDs of the subordinating and the subordinated events, the ID

the signal indicating the type of subordination, and the type itself. The attributes of an ALINK include the ID of the aspectual event, the ID of the argument event, and the ID of a signal indicating the link.

Examples of LINK annotations are:

- The expression *John taught on Monday* states that there is an event *taught*, a time *Monday* and that the temporal relation holding between these two entities is that *taught* is temporally included in *Monday* – as signalled by the signal *on*. This is captured by the annotation `<TLINK eventInstanceID="2" relatedToTime="4" signalID="4" relType="IS_INCLUDED"/>`².
- The expression *John said he taught* includes a *said* event, and a *taught* event, and an evidential subordinate relation holds between them. This is annotated as follows: `<SLINK eventInstanceID="3" subordinatedEvent="4" relType="EVIDENTIAL"/>`.
- In the expression *John started to read*, the *started* event is of an aspectual nature and indicates that the *read* event has started (but not necessarily finished). This can be annotated like this: `<ALINK eventInstanceID="5" relatedToEvent="6" relType="INITIATES"/>`.

A full worked example can be seen in figure 1.

```

There was <SIGNAL sid="8"> no </SIGNAL> <EVENT EID="57" CLASS="STATE"
TENSE="PAST" ASPECT="NONE"> hint of trouble </EVENT> <SIGNAL id="11">
in </SIGNAL> the last <EVENT class="OCCURRENCE" aspect="NONE"
eid="10" tense="NONE"> conversation </EVENT> between controllers and
TWA pilot Steven Snyder. But <TIMEX3 TID="58" val="PT1M30S"
type="DURATION" temporalFunction="false"> a minute and a half
</TIMEX3> <SIGNAL SID="59"> later </SIGNAL>, a pilot from a nearby
flight <EVENT aspect="NONE" eid="18" tense="PRESENT"
CLASS="REPORTING"> calls </EVENT> in.

<SLINK eventInstanceID="eiid57" signalID="8" relType="NEGATIVE"/>
<TLINK eventInstanceID="eiid57" relatedToEvent="eiid10"
signalID="s11" relType="IS_INCLUDED"/>
<TLINK eventInstanceID="eiid18" relatedToEvent="eiid10"
signalID="s59" relType="AFTER" magnitude="t58"/>

```

Figure 1: A full worked example of TimeML annotation

3. The Corpus

3.1. The Text Sources

The texts in the TIMEBANK corpus were chosen to cover a wide variety of media sources from the news domain and come from a variety of sources:

- DUC (TIPSTER) texts from the Document Understanding Conference corpus cover areas like biography, single and multiple events (for example dealing with news about earthquakes and Iraq). This covers 12% of the corpus, or 35 texts;
- Texts from the Automatic Content Extraction (ACE) program come from transcribed broadcast news (ABC, CNN, PRI, VOA) and newswire (AP, NYT). These are 17% (50) texts and 16% (49) texts of the corpus, respectively. One of the reasons for choosing the ACE data is that they were already tagged according to the TIDES time expression guidelines, which was helpful in developing the TIMEX3 tag.

² Note that the event instance ID is automatically created by the annotation tool used.

- Propbank (Treebank2) texts are Wall Street Journal newswire texts, covering 55% or 166 texts of the corpus.

3.2. The Annotation Effort

The annotation of TIMEBANK has been carried out in two separate stages: in the first stage, 70% of the corpus was annotated (210 documents), and in the second stage 30% of the documents were annotated (90 documents). The initial stage was carried out by 5 annotators of remarkably different profiles with regards to their linguistic background. All of them however had participated in the development of the TimeML annotation scheme. The group of annotators for the second stage comprised 45 computer science undergraduate and graduate students, from a course on Commonsense and Temporal Reasoning at Brandeis University.

The annotation in the initial stage was carried out during several annotation-intensive weeks, which were preceded by some sessions of plenary discussion in order to attain a maximum level of agreement among annotators. As for the second stage, the annotation was developed by students who had no prior familiarization with TimeML, and thus some training by the previous annotators was required. In addition, each of the documents annotated at this stage was reviewed to guarantee the quality of the annotation.

The annotation of each document involved a pre-processing step in which some of the events and temporal, modal and negative signals were tagged. When possible, the information concerning event class, tense and aspect of events was also introduced at that point.

The output resulting from the pre-processing was checked during the following human annotation step, and completed with the introduction of other signals and events, time expressions, and the appropriate links among them. The average time to annotate a document of 500 words by a trained annotator is 1 hour.

3.3 The Annotation Tool

The TimeML data annotated within the TERQAS Workshop were annotated using a variety of annotation and pre-processing tools³. There were two major pre-processing steps. One step involved performing temporal expression tagging that annotated simple temporal expressions according to the TimeML guidelines.

This was performed to reduce the amount of manual labour required of the annotator and to help call their attention to relevant temporal expressions. A second pre-processing step involved running a modified version of the Alembic NLP system (Vilain and Day 1996) to generate likely event anchors, including verb phrases and their associated TimeML modalities (based on tense and aspect), as well as a modest number of nominalised event references. This data was also pre-processed running software developed to look inside the verb chunks and label them for tense, aspect and voice.

At this point the data could be loaded into a modified version of the Alembic Workbench annotation tool. The "relation" tagging table provided in the Workbench was used to edit and create new event tags, create SLINK tags where appropriate, and apply a first pass of temporal relation tags to the textual data. When a text is loaded into the tool, the text itself is shown in one window, which shows the results of the pre-processing via coloured tags. These tags can be edited or deleted, and new tags can be introduced. Links, which are shown in a second window, can be created by selecting tags in the text window and inserting these into the link window. Figures 2 and 3 show screenshots of the tool.

³ The authors would like to acknowledge Jose Castaño, Marc Verhagen, and Luc Bélanger for their work on developing the tools described in this section.

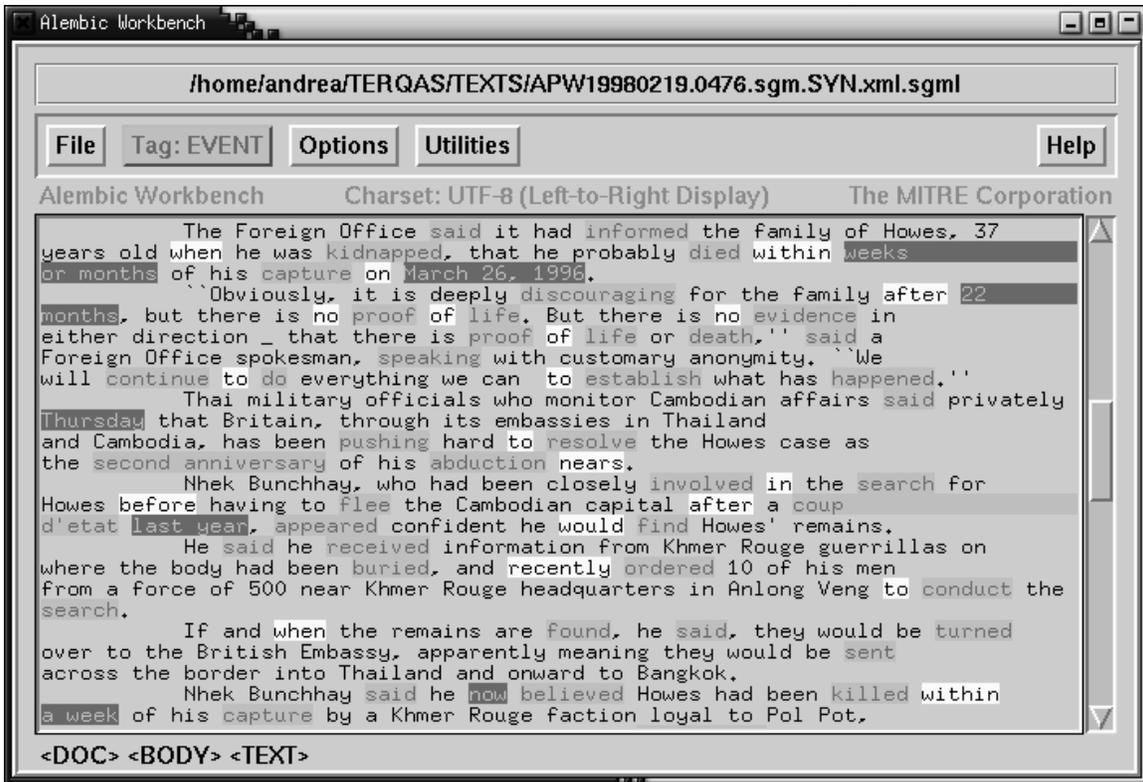


Figure 2: Screenshot 1 of the annotation tool

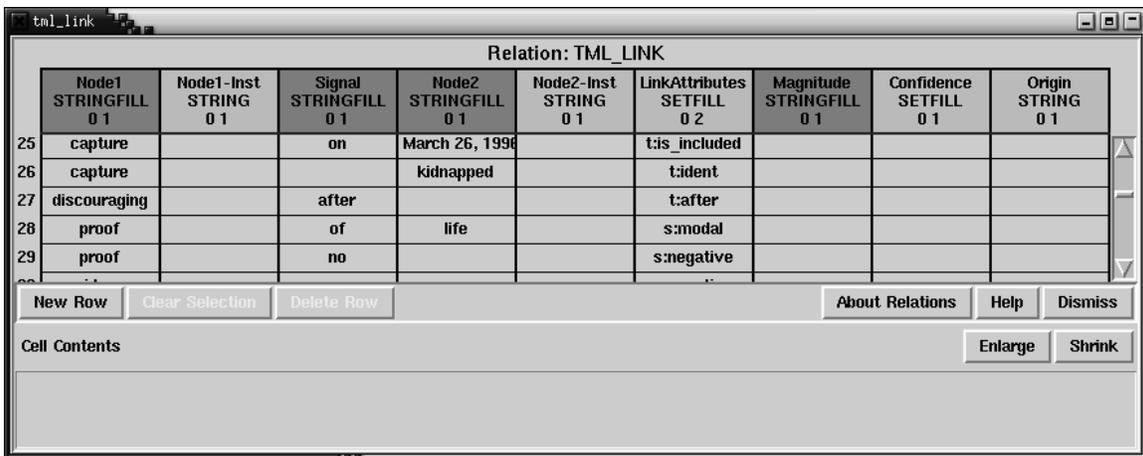


Figure 3: Screenshot 2 of the annotation tool

One aim of the annotation exercise is to allow the temporal relation between *every* pair of events or of events and times to be determined, insofar as this information is in the text. Because of the large number of possible temporal relation pairs that can be derived from even relatively short texts, the data entered in the steps described above was then passed on to a customized annotation tool, the semi-graphical temporal constraint tagger. This tool was arranged so that for any given anchoring event, all the other events are arrayed in lists to the left, right, above, or below, depending on the tools ability to establish their temporal relations. As the annotator assigns distinct temporal relations to particular event pairs, the tool applies a temporal closure algorithm to compute all the other temporal relations entailed from this new constraint,

thereby reducing the actual number of pairwise constraints that must be directly specified by the annotator (see Setzer (2001) for more details, on which this tool is based).

At the end of this process all of the TimeML information is represented in XML tags embedded directly in the original source document. A separate utility was able to render this XML into a web-based graphical representation. This graphical rendering allows the user to make subsequent graphical modifications to the visualization – but not to edit the underlying temporal or event information. A new fully graphical annotation tool that will support all of these distinct annotation tasks/steps within a single interface is under construction.

3.4 Corpus Statistics

The following figures give an idea of the density of the annotation and the distribution of the different tags throughout the corpus. The data presented below was obtained from 60% of the corpus comprising only texts annotated by the 5 initial annotators. We would expect this sample to be representative for the whole corpus. Table 1 shows the total number of Event, Timex3, and Signal tags, together and individually, and the proportion of words that fall into these classes.

	<i>Count</i>	<i>Proportion of words</i>
<i>Words</i>	68555	100%
<i>Tags (all 3 kinds)</i>	11206	16.3%
<i>Events</i>	7571	11.0%
<i>Timex3</i>	1423	2.1%
<i>Signal</i>	2212	3.2%

Table 1: Proportion of tags

The distribution of the different kinds of events and time expressions (Timex3 tags) can be observed in Tables 2 and 3, respectively. There are 1966 signals in the part of TIMEBANK which was analysed, but they are not shown here in a separate table because TimeML does not explicitly distinguish signals into different classes. They are, however, implicitly subcategorised by the links in which they occur. For example, signals being referred to by TLINKS (or temporal links) can be classed as temporal signals. This subcategorisation of signals has not been done yet.

<i>Event Class</i>	<i>Count</i>	<i>Frequency</i>
Occurrence	3933	51.90%
State	1151	15.20%
Reporting	954	12.60%
i-action	636	8.50%
i-state	562	7.40%
Aspectual	256	3.40%
Perception	40	0.50%
missing_class	39	0.50%
total	7571	100.00%

Table 2: Distribution of event classes

<i>Timex3 Type</i>	<i>Count</i>	<i>Frequency</i>
Date	975	68.50%
Duration	314	22.10%
Time	80	5.60%
missing_type	54	3.80%
total	1423	100.00%

Table 3: Distribution of TIMEX3 types

Finally, Table 4 gives the total count and relative frequency for each kind of link:

<i>Link type</i>	<i>Count</i>	<i>Relative Frequency</i>
<i>Link</i>	8242	100.00%
<i>Tlink</i>	5132	62.2%
Tlink:Before	1183	14.3%
Tlink:IsIncluded	973	11.8%
Tlink:After	701	8.5%
Tlink:Simultaneous	633	7.7%
Tlink:Identical	624	7.6%
Tlink:Included	568	6.9%
Tlink:Holds	282	3.4%
Tlink:Ends	67	0.8%
Tlink:Begins	41	0.5%
Tlink:I-After	32	0.4%
Tlink:I-Before	28	0.3%
<i>Slink</i>	2857	34.7%
Slink:Modal	1437	17.5%
Slink:Evidential	1072	13.0%
Slink:Negative	271	3.3%
Slink:Counter-Fact	50	0.6%
Slink:Neg-Evidential	27	0.3%
<i>ALink</i>	253	3.1%

Table 4: Distribution for LINK types

3.5 The Question Corpus

One of the research areas to which TimeML aims to contribute is Question Answering (QA) – temporal question answering in particular. Temporal questions can be broadly classified into two categories:

1. Questions that ask for a temporal expression as an answer, like *When was Clinton president of the United States? When was Lord of the Rings: The Two Towers released? At what time does the summit start? How often did Elvis perform live on stage?* We call this type explicit.
2. Questions that either use temporal expression to ask for a non-temporal answer or that ask about the relations holding between events. Examples are *Who was president of the United States in 1990? Did world steel output increase during the 1990s? Did sugar prices rise during WWI? Who was CEO of IBM when it bid to take over Apple?* We call this type of temporal question implicit.

To evaluate the usefulness of TimeML for (temporal) QA, a question corpus of 50 questions has been created, which was annotated according to a specially developed annotation scheme. This scheme allows features such as the type of the expected answer and the volatility of the answer (i.e. how often it changes) to be annotated. The questions contained in the corpus cover both types mentioned above. Example questions are:

- When did the war between Iran and Iraq end?
- When did John Sununu travel to a fundraiser for John Ashcroft?
- How many Tutsis were killed by Hutus in Rwanda in 1994?
- Who was Secretary of Defense during the Gulf War?
- What was the largest U.S. military operation since Vietnam?

4. Conclusion and Future Work

In the foregoing we have briefly described the TimeML annotation scheme for annotating times, events and temporal relations in text and we have presented the TIMEBANK corpus, a 300 text corpus annotated according to the TimeML scheme. A first version of TIMEBANK will be released later in 2003 for general use via the Linguistic Data Consortium (LDC).

Since the corpus creation effort is just coming to a close, there has as yet been no time to analyse the corpus; the statistics given in section 3.4 do not represent such an analysis, but only a very preliminary scoping of the corpus. However, we anticipate that the corpus will allow a whole new range of questions, both theoretical and practical, to be pursued. For example, it will become possible to study to what extent temporal ordering of events is conveyed explicitly through signals, such as temporal subordinating conjunctions, and to what extent implicitly through the lexical semantics of the verbs or nominalizations expressing the events. From a practical computational perspective, it will become possible to consider training and evaluating algorithms which determine event ordering and time-stamping, and to explore their utility in question answering.

We are under no illusions that TimeML is as yet fully comprehensive and without anomalies or that TIMEBANK has been completely and correctly annotated. Further revision of the annotation scheme and of the corpus is inevitable. However, we believe that together TimeML and TIMEBANK form a significant contribution to the study of how temporal information is expressed in natural language, and a significant step for corpus linguistics in the direction of the semantic study language.

5. References

Ferro L, Mani I, Sundheim B, and Wilson G 2001 *TIDES Temporal Annotation Guidelines, Version 1.0.2* MITRE Technical Report, MTR 01W0000041.

Ingria R and Pustejovsky J 2002 *TimeML Specification 1.0* available at <http://www.time2002.org>.

MUC6 1995 Proceedings of the Sixth Message Understanding Conference (MUC-6), Morgan Kaufman.

MUC7 1998 Proceedings of the Seventh Message Understanding Conference (MUC-7), available at http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_toc.html.

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

Setzer, A. 2001 *Temporal Information in Newswire Articles: an Annotation Scheme and Corpus Study*, Unpublished PhD thesis, University of Sheffield.

Vilain M and Day D 1996 Finite-State Parsing by Rule Sequences International Conference on Computational Linguistics (COLING-96). Copenhagen, Denmark.