Annotation of Temporal Relations with Tango

Marc Verhagen*, Robert Knippen*, Inderjeet Mani†, James Pustejovsky*

*Computer Science Department, Brandeis University
South Street, Waltham, USA
{marc,knippen,jamesp}@cs.brandeis.edu
†The MITRE Corporation
202 Burlington Rd. Bedford, USA
imani@mitre.org

Abstract

Temporal annotation is a complex task characterized by low markup speed and low inter-annotator agreements scores. Tango is a graphical annotation tool for temporal relations. It is developed for the TimeML annotation language and allows annotators to build a graph that resembles a timeline. Temporal relations are added by selecting events and drawing labeled arrows between them. Tango is integrated with a temporal closure component and includes features like SmartLink, user prompting and automatic linking of time expressions. Tango has been used to create two corpora with temporal annotation, TimeBank and the AQUAINT Opinion corpus.

1. Introduction

Temporal annotation poses specific challenges for annotation tools. The high density of temporal relations makes the temporal annotation task a complex one, characterized by low markup speed, hard-to-avoid inconsistencies, and low inter-annotator agreement. The high density is due to the fact that the set of possible temporal relations is essentially quadratic to the number of events and time expressions in a document. This large space of possible relations has two consequences: (i) only a fraction of all possible relations will be marked up, and (ii) inter-annotator agreement scores are low because two annotators will often choose to not add relations between the same events.[1] Annotation of temporal relations requires more reflection than, for example, annotation of part-of-speech tags, and is therefore slower. Syntactic tags and many semantic tags, such as entity tags and event tags, can be added in a strictly linear fashion. Temporal relations are different because they require us to specify attributes of pairs of objects, and the objects involved may not be close to each other in the text. Also, the value of one temporal link imposes restrictions on other temporal links. Keeping an annotated document consistent can be difficult and, indeed, experiments with consistency checking tools showed that it is hard to annotate a one-page document without introducing inconsistencies.

Tango is an annotation tool that was designed to cope with the challenges of temporal annotation. We will introduce its functionality and give some preliminary evaluation results in section 3, but first some words on temporal annotation.

[1]The event pairs and event-timex pairs that are selected by two annotators for adding temporal links are in many cases not the same. The exact inter-annotator scores in cases like this are partially determined by the annotation guidelines, depending on how forcefully they prescribe what events and times should be linked. The latest agreement score for annotation of temporal links is 55% average precision and recall. That is, two annotators choose to specify a temporal relation between the same events and/or times in 55% of the cases. Of those 55%, the annotators agree in 77% of the cases as to what temporal relation to specify.

2. TimeML

TimeML is a recent approach to temporal annotation originally developed in the context of two ARDA-AQUAINT workshops. In TimeML, temporally relevant entities are marked up with EVENT and TIMEX3 tags. The EVENT tag is used to annotate those elements in a text that mark the semantic events described by it. Syntactically, events are typically verbs, although event nominals, such as "crash" in "...killed by the crash" are events as well. The TIMEX3 tag is used to mark up explicit temporal expressions, such as times, dates, and durations. Temporally significant relations between events and times can be specified with ALINK, SLINK and TLINK tags. Link tags do not span any text extent, they are empty and simply state relations between text entities.

Consider the sentences in (1) for some examples. An aspectual link (ALINK), exemplified in (1a), indicates an aspectual relation like initiates, terminates or continues. A subordinating link (SLINK) indicates a temporally loaded relation in a sub-ordinating context. Sentence (1b) introduces a modal context between *seeking* and *mediate*, and sentence (1c) contains a reporting context between *said* and *retaliate*. Temporal links proper (TLINKs) denote relations like before, includes and simultaneous. In (1d), *rise* occurs after *invasion*.[2]

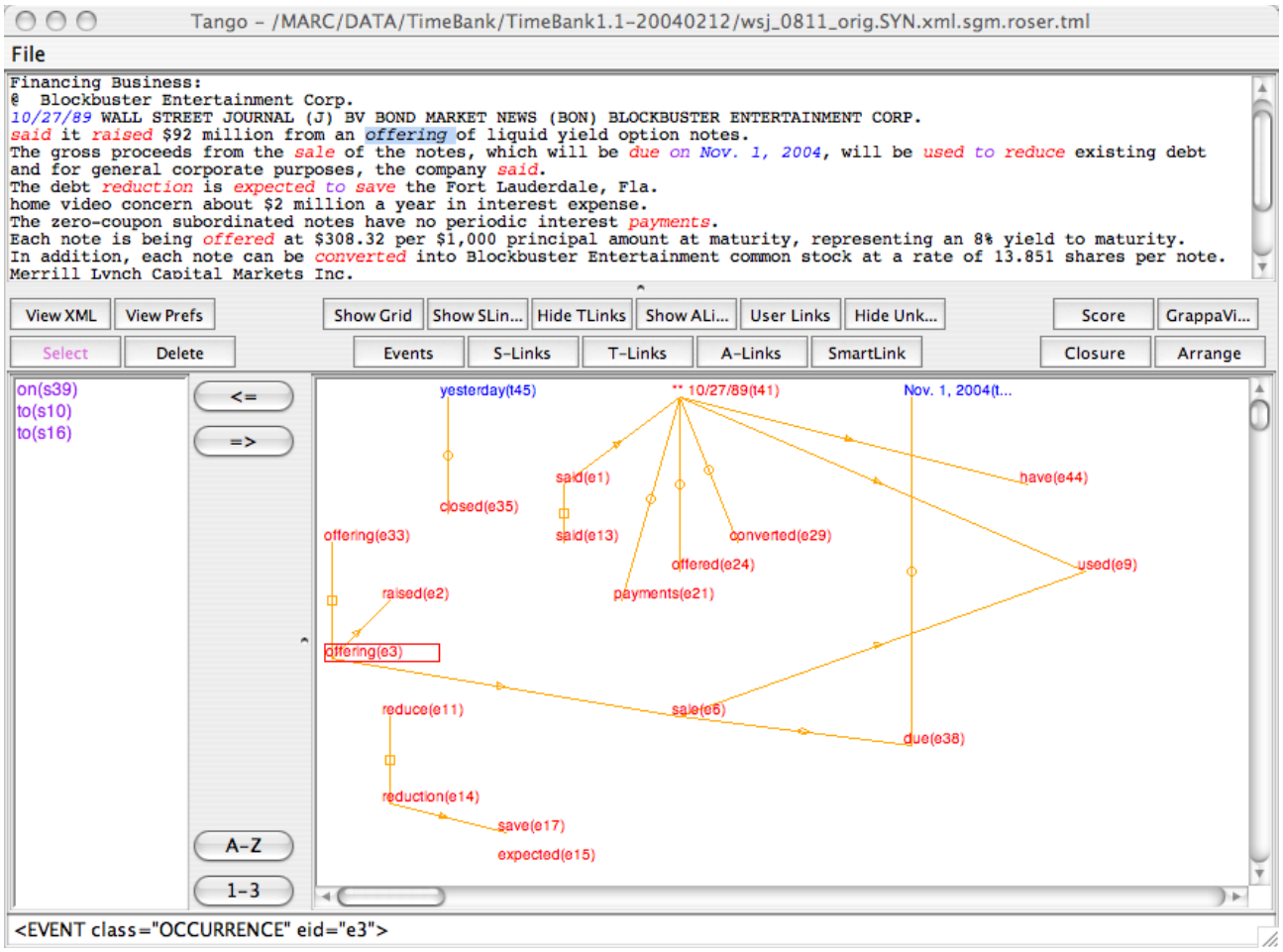
(1) a. The U.S. military *buildup* in Saudi Arabia *continued* at fever pace

b. King Hussein of Jordan arrived in Washington *seeking* to *mediate* the Persian Gulf crisis

c. The Iraqi ambassador to Venezuela *said* on Tuesday that Iraq would *retaliate* against Venezuela

d. A steep *rise* in world oil prices followed the Kuwait *invasion*

[2]All examples are taken from Associated Press document AP900815-0044. This document, and its TimeML annotation, can be viewed with the TimeBank browser, available at http://timeml.org/site/timebank/browser.html



The TimeBank corpus was created as a resource that functioned as an illustration and proof of concept for the TimeML annotation language. TimeBank, soon to be released by the Linguistic Data Consortium, is a relatively small corpus with about 7900 events, 1400 time expressions, 2900 subordinating links and 6400 temporal links. See (Pustejovsky et al., 2003a; Saurí et al., 2006; Pustejovsky et al., 2003b) and http://timeml.org/ for more details on TimeML and the TimeBank corpus.

3. Tango

The first version of TimeBank was annotated with the Alembic workbench (Day et al., 1997). It became quickly apparent that Alembic was well-suited to annotation of events and time expressions, but that annotation of links was cumbersome and error-prone. Alembic uses a table metaphor to display temporal relations. From the table, it is almost impossible to get a view of the temporal structure of the text as a whole and annotation reduces to a case-by-case inspection of event pairs.[3] The complexity and high density of temporal relation requires a more intuitive annotation tool. The ARDA-funded Tango project (Pustejovsky et al., 2003c) addressed this need.

3.1. Tango Functionality

Tango provides an editable and fully graphical display of events and temporal relations, a screenshot is provided in figure 1. A TimeML annotation is presented as a graph where events and times are the nodes and temporal links are the arcs. Events that have not been linked yet are listed in a pending list on the left. Annotation proceeds by drawing arrows between events and labeling them with relation types. Events can be moved to any desired position on the canvas. The display is made to resemble a timeline by placing the time expressions at the top and ordering them according to their ISO values. Additional features of Tango are:

- SMARTLINK. Any event or time expression can be selected to be the focus element, and dropping one or more objects to the left or right of the focus element automatically creates before or after links.
- AUTO LAYOUT. Tango also provides an auto-arrange feature that distributes events left-to-right based on

[3] Alembic has since been replaced by Callisto (see http://callisto.mitre.org), which, although arguably a better annotation tool, still uses the same table metaphor that does not serve temporal annotation very well. However, Callisto is set up in a much more modular fashion and may some day allow integration of Tango. For TimeBank annotation, Callisto and Tango have often been used in tandem, annotating events and times with Callisto and adding links with Tango.

their links, using topological sort. This feature is very useful for TimeML documents that have never been displayed in Tango before: in these documents, all events are crowded to the far left of the canvas.

- TIMEX LINKING. TLINKs can be created automatically between those time expressions that have been annotated with specific ISO values.

- TEMPORAL CLOSURE. Tango includes a temporal closure component that derives new implied links from existing links. It is based on the interval algebra of (Allen, 1983) and the point algebra of (Vilain et al., 1990). Closure ensures consistency and greatly increases the number of TLINKs in an annotation. For example, applying closure to version 1.1 of the TimeBank corpus more than quadruples the number of TLINKs. See (Verhagen, Forthcoming) for more details.

- USER PROMPTING. This feature aids the annotator in achieving a complete annotation. If switched on, user-prompting will ask annotators to provide temporal relations for specific, not yet annotated, event pairs. Each time a new TLINK is added, temporal closure will run to derive all implied relations and the cycle will continue with a new user prompt and closure application until all TLINKs are specified.

The user can choose to hide or display certain links. This is most useful for hiding links that were derived using temporal closure, in many case the large amount of inferred TLINKs can crowd the display.

3.2. Evaluation

The first experiences with Tango are satisfactory. A bit surprisingly, Tango did not make the annotation process faster than it was with Alembic. This seems to be explained by the willingness of annotators to sit down and really dive into the temporal structure of a document. Also, with Alembic, annotators quickly zeroed in on a very specific strategy: they go through the text linearly and add table entries that typically link an event to a nearby event or time. This, although requiring an elaborate sequence of mouse clicks, made for a relatively quick workflow. However, it does not facilitate creation of non-local links, that is, the vast majority of links in Alembic-annotated documents are within a sentence or cross one sentence boundary only. This is not the case with Tango.

Annotation with Tango did reduce the number of mistakes made by the annotators, a mini-experiment conducted during the TANGO project indicated that the percentage of TLINKs that received the correct relation type went up from 64% to 76% when Tango was used as the annotation tool rather than Alembic.

With Tango, it seems to be easier to achieve a complete annotation since it leads to more densely connected graphs than Alembic does, and more connected graphs make complete annotations easier to achieve. The numbers to back this up are preliminary. An earlier version of TimeBank was annotated in large part with Alembic, running closure over that version resulted in a fourfold increase in the number of TLINKs. The AQUAINT Opinion corpus was annotated using Tango, here the number of TLINKs after closure increased by a factor ten. One should be careful to not read too much into this however, for example, differences in genre may account for part of these differences.

There are no solid data to support claims that annotation with Tango increases inter-annotator agreement. In footnote 1, we quoted 55% as the current inter-annotator agreement on selection of event pairs that are temporally linked by annotators. The only figure available for TLINK annotation with Alembic is 20%. However, this number is from an experiment with novice annotators at a time that the TimeML specifications were still under constant revision.

TANGO has been used effectively to annotate 186 TimeBank and 73 Opinion corpus documents (the latter from the AQUAINT Program). It has also served well in visualizing and correcting output from automatic TimeML annotations from the TARSQI system (Verhagen et al., 2005). Tango is available for download at http://timeml.org/tango/.

3.3. Current Work

One of the main motivations behind Tango was to create an annotation tool that displays the temporal structure of a document as clearly as possible. The connected graph is certainly easier to read than a table, and it also very easy to see how complete an annotation is. Yet the display still has a couple of drawbacks: (i) the display can get crowded, especially in those cases where closure has been applied, (ii) placement of events relative to each other is not predictable and the auto layout feature does not completely solve this, and (iii) it is hard to distinguish between the relation types. The main problem is that the dimensions and directions of the graph are not unambiguously mapped to a predictable temporal interpretation and that potential clutter is not controlled. The example fragment in figure 2 illustrates these problems, even though the annotator took pains to provide correct positioning of events in the canvas.

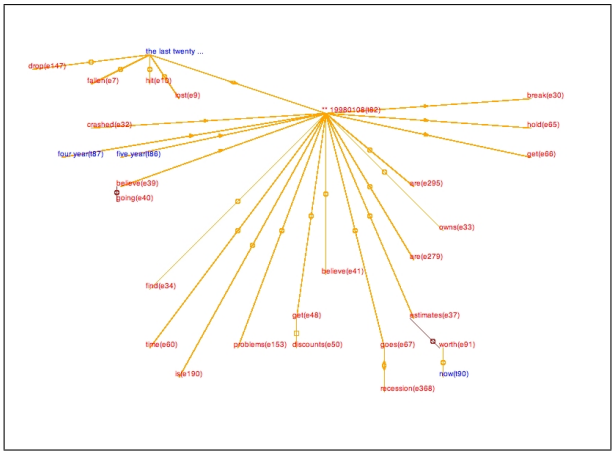


Figure 2: An annotation as a connected graph.

T-BOX is an alternative display where the annotation is not presented as a graph but as boxes where box inclusion indicates temporal inclusion, arrows indicate before relations

and stacked boxes indicate simultaneity. The same annotation from figure 2 is presented in figure 3 for comparison. In this figure, a duration, ”last twenty four hours”, includes

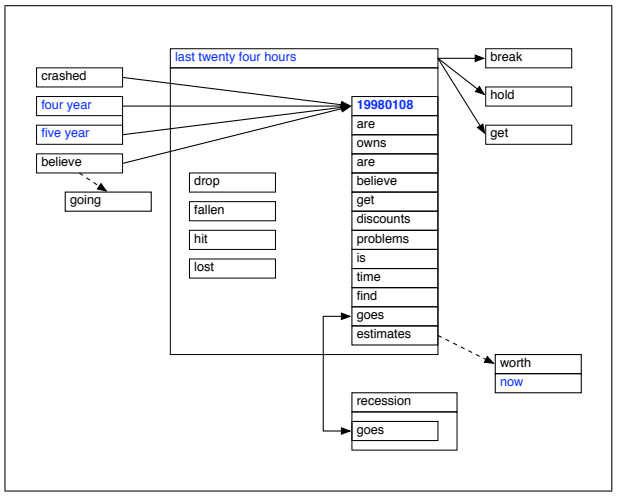


Figure 3: An annotation in T-BOX style.

many events: some (”drop”, ”fallen”, ”hit”, and ”lost”) are properly temporally included in the duration (that is, their begin points occur after the beginning of the duration and their end points occur before the end of the duration); others are lined up hugging the right border of the box, the duration is ended by these events. It is important to note that vertical placement by itself means nothing, for example, the three events in the top right corner of figure 3 do not stand in any specific relation to each other. Had there been such a relation, then arrows, box stacking, or box inclusion would have been added. One of the advantages of this display is that annotation errors become obvious. Figure 3 is a bit odd with the massive amount of stacked events inside the duration, and it is rather suspicious that the date ”19980108” is said to end the duration ”last twenty four hours”. These problems are not immediately obvious in figure 2.

There is a mechanical procedure to go from an annotation graph to a T-BOX representation. This procedure includes temporal closure, link normalization, graph reduction, and a bottom-up mapping from graph elements to T-BOXes. Refer to (Verhagen, 2005) for more details on T-BOX in general and the above-mentioned procedure in particular. It should be noted that some of the mechanisms used for T-BOX can also help to prevent graph clutter in the standard Tango display. The issue remains though that the standard display does not have any semantics associated to horizontal and vertical placement.

T-BOX has been recently added to Tango as an alternative to the standard display. At the time of this writing, editing capabilities for the T-BOX canvas have not yet been implemented.

4. Conclusion

We have presented Tango, a graphical annotation tool for temporal relations. Tango displays a temporal annotation as a graph where the events and times are the nodes and where relations between events and times are represented by labeled arcs. An alternative view exploiting the T-BOX representation has been added recently. Tango has been successfully used to mark up texts with TimeML tags, most notably for the TimeBank and AQUAINT Opinion corpora.

5. Acknowledgements

We would like to thank Linda van Guilder, Andrew See, and Alex Baron for their work on the Tango tool. The work presented in this paper has been partially supported by the ARDA AQUAINT program, grant number NBCHC040027-MOD-0003.

6. References

James Allen. 1983. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11):832–843.

David Day, John Aberdeen, Lynette Hirschman, Robyn Kozierok, Patricia Robinson, and Marc Vilain. 1997. Mixed-Initiative Development of Language Processing Systems. In *Fifth Conference on Applied Natural Language Processing Systems*, pages 88–95, Washington D.C., U.S.A.

James Pustejovsky, José Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003a. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *IWCS-5 Fifth International Workshop on Computational Semantics*.

James Pustejovsky, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. 2003b. The TIMEBANK Corpus. *Proceedings of Corpus Linguistics*.

James Pustejovsky, Inderjeet Mani, Luc Belanger, Linda van Guilder, Robert Knippen, Andrew See, Jon Schwarz, and Marc Verhagen. 2003c. TANGO Final Report. Technical report, The MITRE Corporation, Bedford, Massachusetts.

Roser Saurí, Marc Verhagen, and James Pustejovsky. 2006. Annotating and Recognizing Event Modality in Text. In *The 19th International FLAIRS Conference*, Melbourne, Florida, USA.

Marc Verhagen, Inderjeet Mani, Roser Saurí, Robert Knippen, Jessica Littman, and James Pustejovsky. 2005. Automating Temporal Annotation with TARSQI. In *Proceedings of the ACL 2005*, Ann Arbor, USA.

Marc Verhagen. 2005. Drawing TimeML Relations with T-BOX. In James Pustejovsky, Graeme Katz, and Frank Schilder, editors, *International Workshop on Annotating, Extracting and Reasoning with Time*, Dagstuhl, Germany.

Marc Verhagen. Forthcoming. Temporal Closure in an Annotation Environment. In James Pustejovsky and Robert Gaizauskas, editors, *Language Resources and Evaluation*. Springer.

Marc Vilain, Henry Kautz, and Peter van Beek. 1990. Constraint propagation algorithms: A revised report. In D. S. Weld and J. de Kleer, editors, *Qualitative Reasoning about Physical Systems*, pages 373–381. Morgan Kaufman, San Mateo, California.