05151 Summary—Annotating, Extracting and Reasoning about Time and Events

G. Katz (Univ. Osnabrück, DE)
J. Pustejovsky (Brandeis Univ. Waltham, US)
F. Schilder (Thomson Legal & Regulatory, Eagan MN, US)

1 Introduction

Newspaper articles and other natural-language texts describe actions, events, and states of affairs. A crucial first step toward the automatic extraction of information from these texts—for use in such applications as automatic question answering or summarization—is the capacity to identify what events are being described and to make explicit when these events occurred and which temporal relations hold among them. There has recently been a renewed interest in making use of this kind of temporal and event-based information, with a wide variety of proposals and applications having been presented at recent conferences and workshops. [4, 3, 5, 6]. The central goal of the seminar was to consolidate the insights that have been made in recent years and to identify and address issues concerning annotation, temporal reasoning and event identification that remain unresolved.

Much of the temporal information conveyed in a natural language text is left implicit. Significant recent work has focused on developing schema for making this information explicit, typically via annotation. An important result of contemporary research has been the adoption of a *de facto* standard for time and event annotation: TimeML [5, 6, 7]. This XML-based markup language is specifically designed for annotating texts with tags that make explicit the temporal and event-based information conveyed by the text and has been adopted by a number of researchers in this domain. Much of our seminar was concerned with issues specific to this annotation scheme.

There are three basic types of tags used by the TimeML language: TIMEX tags are used to annotate temporal expressions and provide them with a normalized value (e.g. ⟨TIMEX tid="t1" val="2005-04-21"⟩ *April 21st, 2005* ⟨TIMEX⟩); EVENT tags are used to annotate event expressions, providing "hooks" to relate them to other events and times introduced in the text (e.g. ⟨EVENT eid="ei"⟩*opened*⟨/EVENT⟩); So-called TLINK tags indicate the temporal relations that hold between times and events (e.g. *the stock market opened on April 21st, 2005 at 10:00pm* ⟨TLINK event="e1" relatedTime="t1" relation=INCLUDED-BY⟩). Other tags are used to capture more subtle se-

Dagstuhl Seminar Proceedings 05151
Annotating, Extracting and Reasoning about Time and Events
http://drops.dagstuhl.de/opus/volltexte/2005/354

mantic relations. SLINK tags, for example, are used to indicate various kinds of subordination relations, such as the negation in *The stock market did not open on April 21st, 2005 at 10:00pm* or the only potential event in *Investors hoped that the stock market would open on April 21st, 2005 at 10:00pm*. A small corpus of TimeML annotated documents (*TimeBank*) has been generated, and can be browsed at timeml.org.

The main focus of the seminar was on TimeML-based temporal annotation and reasoning. We were concerned with three main points: determining how effectively one can use the TimeML language for consistent annotation, determining how useful such annotation is for further processing, and determining what modifications should be applied to the standard to improve its usefulness in applications such as question-answering and information retrieval.

2 Summary of talks

After a brief introduction to the TimeML markup-language by James Pustejovsky we had one block on approaches to automatically extracting temporal expressions and linking them to event expressions. Bran Boguraev (IBM research) reported on work he did with Rie Kubota Ando. Using the method described in [2] they showed interesting results for automatically generating TLINKs from TimeBank using machine learning. The second talk on automatic markup was given by David Ahn (University of Amsterdam). He presented work on tagging temporal expressions by comparing a statistical machine-learning approach (conditional random fields) with a rule-based approach (JAPE). His experiments showed that the machine-learning approach beats the rule-based approach and can also improve the rule-based normalization task for temporal expressions (i.e. the derivation of ISO time stamps for temporal expressions).

Another block was concerned with possible future applications of time and event annotation. Frank Schilder discussed what kind of legal documents (legal narratives, transactional documents, statutes) may benefit from temporal information extraction and pointed out where the current version of TimeML has to be expanded in order to handle, for example, temporal information in normative law. Andrea Setzer (University of Sheffield) presented a project from the medical domain. In this project temporal information from patient notes dictated by doctors is to be extracted and mapped on to a database containing records of interventions (e.g. surgery) and investigations (e.g. X-RAY) performed on the patient.

Four talks (Robert Gaizauskas, University of Sheffield; Marc Verhagen, Brandeis University, Waltham; Tom Bittner, Univ. des Saarlandes; Inderjeet Mani, Georgetown University, Washington) were concerned with the issue of which temporal relations among events should be annotated. They all pointed to the necessity of providing for the annotation of underspecified temporal relations, instead of using the full set of Allen's (1983) interval relations [1].

Issues concerning what should be annotated were also raised by Jerry Hobbs (USC/ISI - Marina del Rey), who presented the OWL-Time ontology that was

developed in conjunction with the DARPA Agent Markup Language (DAML) program. In addition, he discussed results of an annotation study where annotators had to estimate and annotate metric information about event and inter-event durations. This study showed very low agreement among naive annotators. A detailed annotation guide helped to improve inter-annotator agreement significantly, however.

Logical frameworks for temporal reasoning were presented by Mark Steedman (University of Edinburgh) as well as Ian Pratt-Hartmann (Manchester University). Steedman presented a calculus for planning and reasoning about action called the "calculus of affordance". Pratt-Hartmann defined in his talk a grammar fragment of English including the temporal constructions such as *during every X* or *for the first time*. He then showed how to map the sentences recognized by this grammar to formulas of a temporal interval logic. Finally, he analyzed the expressive power of the temporal logic. Both of these talks addressed the problem of annotating and extracting temporal expressions based on the tractability of the underlying logical framework.

Practical tools for reasoning with temporal information were presented by Benjamin Han (CMU - Pittsburgh) and Hans-Jürgen Ohlbach (Universität München). They presented implementations that do reasoning with temporal information, such as computing the current date plus 2 months. Their implementations are written in Python (Han) or C++ (Ohlbach).

Questions regarding the veridicality of events were addressed by Annie Zaenen and Lauri Karttunen (both PARC - Palo Alto) in their talk. Discussing a variety of data, they showed that the veridicality of assertions of the author and conversational implications depend on not only the matrix verb, but also some other aspect of the syntax (e.g. *Bush forgot to read the report* vs. *Bush forgot that he read the report*). A detailed analysis of the scope of temporal locating adverbials in discourse was presented by Laure Vieu (LOA - Trento). Her data showed that temporal adverbials that occur as sentential adjuncts serve as a frame for the following discourse introducing a new discourse topic. This observation has important implications both for the annotations task as well as the automatic extraction and linking of temporal expressions. Moreover, she suggested that it could be helpful to annotate discourse relations and discourse topics as well.

A future perspective for TimeML with respect to the annotation of arguments was provided by James Pustejovsky. He discussed whether the explicit annotation of participants in an event (the agent or the theme of an action, for example) would be useful for future versions of TimeML. So far only the main verb or noun that indicates an event type (e.g. *buy* or *purchase*) are annotated with the EVENT label. His conclusion was that this extra annotation should be left to others, since it would make it very difficult to reach useful inter-annotator-agreement scores. Graham Katz presented a denotational semantics for a TimeML fragment, and showed that, since embedding is not allowed in TimeML, it is difficult to represent scoping appropriately. A variety of solutions for this problem were discussed.

3 Highlights

One of the highlights of the seminar was an annotation exercise which was carried out by all participants in groups. This served both as a touchstone for discussing issues that came up in the course of the seminar and as a source of examples of difficulties to be addressed. As the "target text" we choose a newspaper article from the *Seattle Times* describing the wedding of Prince Charles and Camilla Parker Bowles, an event that had just occurred.

The entire seminar was split up into groups of four or five researchers and each group carried out the annotation in two parts. In the first part, we attempted to identify, making use of the TimeML guidelines, the events and times which were described by the article and to identify the relations that hold among them. We found there to be very clear agreement about what events there were. Issues of event identity (is the *waving* the same as the *greeting*?) were the foremost problems. Also the temporal relations were fairly well agreed upon. Here again there was very little in the way of disagreement, with the major problems being those surrounding the differentiation among simultaneity, overlap and immediate precedence. What was striking, however, was that there were far more events described (and for which TimeML guidelines require annotation) than participants judged would be likely to be useful for any application.

In a second part of the annotation exercise the same groups attempted to do metrical annotation, of the type described by Hobbes. Here we tried to specify how long each of the events was and how long the intervals between events were. In contrast, here there was wide variation in some cases (how long does the state of the couple being *newly married* hold?), but in other cases fairly close agreement. The highlight of this exercise came when we compared our consensus interpretation of the text to the BBC video of the event described. The very low correlation between our estimated durations for events (the waving, the walking to the car) and their actual durations as shown on the video raised questions, less for the value of annotation, but for the veracity of newspaper texts.

4 Schedule

MONDAY April 11, 2005

9:15 Introduction and TimeML Tutorial
Graham Katz, Frank Schilder, and James Pustejovsky

10:15 BREAK

10:30 TimeML Annotation Exercise (in Groups)
At long last, Charles weds Camilla, 4/10/05 News Article

12:15 LUNCH

2:00 Discussion of Annotation, Problems

2:45 Branimir Boguraev
TimeBank-driven TimeML Analysis

3:30 BREAK

4:00 Frank Schilder
Temporal Information Extraction from Legal Documents

4:45 Andrea Setzer
TimeML in a Medical Application

6:00 DINNER

TUESDAY April 12, 2005

9:15 Jerry Hobbs
A Temporal Ontology for the Semantic Web

10:15 BREAK

10:45 Lauri Karttunen and Annie Zaenen
Veridicity and Commitment?

11:30 Laure Vieu
Scope of Temporal Adverbials in Discourse

12:15 LUNCH

2:00 Annotation of Event Durations

3:15 David Ahn
Towards Task-based Temporal Extraction and Recognition

4:00 BREAK

4:30 Benjamin Han
Understanding Times: An Constraint-based Approach

5:15 Tom Bittner
Approximate Qualitative Temporal Reasoning

6:00 DINNER

WEDNESDAY April 13, 2005

9:15 Mark Steedman
The Calculus of Affordance

10:15 BREAK

10:45 Marc Verhagen
Drawing TimeML Relations

11:30 Hans-Jürgen Ohlbach
Computational Treatment of Temporal Notions the CTTN System

12:15 LUNCH

2:00 TRIP WITH DINNER

THURSDAY April 14, 2005

9:15 Rob Gaizauskas
Getting Closure: Vagueness and Disjunction in TimeML

10:00 BREAK

10:30 Graham Katz
The Semantics of TimeML

11:15 James Pustejovsky
Event Arguments in TimeML

12:15 LUNCH

2:00 Ian Pratt-Hartmann
Temporal Prepositions and their Logic

2:45 Inderjeet Mani
Chronoscopes: A theory of Underspecified Temporal Relations

3:30 BREAK

4:00 Working Group Discussions (Defining Core TimeML)

6:00 DINNER

8:00 Discussion of WG (Prepare 2 Slides)

FRIDAY April 15, 2005

9:15 Presentation of WG proposals on Core TimeML

10:00 BREAK

10:30 Road Map for Future Research

12:15 LUNCH and Goodbyes

5 Conclusion

At the closing session we discussed problems with the current version of TimeML and pointed to several new research directions. The discussion started off with a renewed call for a TimeML Lite version that should only encompass the core ideas of temporal annotation. In order to focus on the main temporal information most preprocessing of other informations has to be done separately from the actual TimeML annotation or should be supplied by other resources. PropBank was suggested as a good starting point for event annotation but also stemming, tense, aspect, word sense information as well as phrasal verbs, idioms and light verbs coult provide important information.

Before the discussion continued to focus on a roadmap discussion of different tags, a short presentation of the Web-interface of TimeBank was given by Marc Verhagen. Such a resource will probably be very valuable for future research efforts.

Different issues were raised in the subsequent session:

- Pilot study on parallel texts was suggested.
- SLinks:
  - How many RelTypes should be allowed.
  - Adding attribute to Slinks to indicate that (a) the event happened, (b) the event did not happen, or (c) we don't know.
- TLINKs:
  - Pilot study using a compacted set of temporal relations was discussed.
  - It was suggested to look at (dis-)agreement among annotators in TimeBank regarding temporal relations to extract equivalence classes.
  - Additional temporal relation: overlap.

  There was a lively discussion of the representation of temporal relations. Among other things, it was mentioned that an explicit disjunction between relations was needed. Moreover, some suggested that the beginning and end points of the event should be seen as an interval not not as points. It was also proposed that the actual source of TLINK derivation should be annotated.
- States
  - Should states be dropped from the annotation?
  - Should only fully predicative tensed statives be analyzed?
- What is a minimally compliant TimeML relation annotation of a text?

Various organizational efforts were identified:

1. Try to move TimeBank into the LDC. Facilitate interaction with the Penn Treebank.

2. TimeML working group. Possible subgroups are:

   (a) TLINK set

   (b) SLINK strategies

   (c) Temporal Functions, temporal aggregates

Finally, evaluation efforts of the annotation were discussed. The following three requirements were postulated:

1. Inter-annotator agreement values need to be defined.

2. The price of joining a Working Group is to annotate a collection of articles.

3. Annotation environment must be stable enough, and adopted by all sites doing annotation. Easy to use.

We concluded with a discussion on new directions and possible applications for TimeML were. One application we would like to focus on is the automatic generation of biographies extracted from newspaper articles. The idea is to identify life-changing events (e.g. job change events) and place them on a time line. It became very clear to all concerned, however, that the usefulness of having a single standard for time and event annotation for which data could be collected exceeded the problems with adopting a standard which might be less than optimal.

References

[1] Allen, J.F. (1983). Maintaining knowledge about temporal intervals. Communications of the ACM 26 (1983), 832–843.

[2] Ando, R. (2004). Semantic Lexicon Construction: Learning from Unlabeled Data via Spectral Analysis. In Proceedings of CoNLL-2004. Hwee Tou Ng and Ellen Riloff (eds.) pages 9-16.

[3] Harper, L., Mani, I. and Sundheim, B. (2001) (eds.). Proceedings on the Workshop on Temporal and Spatial Information Processing. ACL 2001 Conference, Toulouse, 7 July 2001.

[4] Mani, I. and Wilson, G. (2000). Robust temporal processing of news. In Proceedings of the 38th Annual Meeting of the ACL (Hong Kong, Oct. 2000).

[5] Pustejovsky, J. (2002). TERQAS: Time and Event Recognition for Question Answering Systems. ARDA Workshop, MITRE, Boston. `http://www.timeml.org/terqas/`

[6] Pustejovsky, J., Mani, I., Belanger, L., Boguraev, B., Knippen, B., Littman, J., Rumshisky, A., See, A., Symonen, S., Van Guilder, J., Van Guilder, L. and Verhagen, M. (2003). ARDA Summer Workshop on Graphical Annotation Toolkit for TimeML, MITRE, Boston. `http://nrrc.mitre.org/NRRC/TangoFinalReport.pdf`

[7] Saurí, R., Littman, J., Gaizauskas, R., Setzer, A., and Pustejovsky, J. (2004). TimeML annotation guidelines version 1.2. Technical report, TimeML document. available at `http://www.timeml.org`.