

A Pilot Study on Acquiring Metric Temporal Constraints for Events

Inderjeet Mani and Ben Wellner

The MITRE Corporation

202 Burlington Road, Bedford, MA 01730, USA

and

Department of Computer Science, Brandeis University

415 South St., Waltham, MA 02254, USA

{**imani, wellner**}@mitre.org

Abstract

Previous research on temporal anchoring and ordering has focused on the annotation and learning of temporal relations between events. These qualitative relations can be usefully supplemented with information about metric constraints, specifically as to how long events last. This paper describes the first steps in acquiring metric temporal constraints for events. The work is carried out in the context of the TimeML framework for marking up events and their temporal relations. This pilot study examines the feasibility of acquisition of metric temporal constraints from corpora.

1 Introduction

The growing interest in practical NLP applications such as question-answering and text summarization places increasing demands on the processing of temporal information. In multi-document summarization of news articles, it can be useful to know the relative order of events so as to merge and present information from multiple news sources correctly. In question-answering, one would like to be able to ask when an event occurs, or what events occurred prior to a particular event. A wealth of prior research by (Passoneau 1988), (Webber 1988), (Hwang and Schubert 1992), (Kamp and Reyle 1993), (Lascarides and Asher 1993), (Hitzeman et al. 1995), (Kehler 2000) and others, has explored the different knowledge sources used in inferring the temporal ordering of events, including temporal adverbials, tense, aspect, rhetorical relations, pragmatic conventions, and background knowledge. For example, the narrative convention of events being described in the order in which they

occur is followed in (1), but overridden by means of a discourse relation, Explanation in (2).

- (1) Max stood up. John greeted him.
- (2) Max fell. John pushed him.

While there has been a spurt of recent research addressing the event ordering problem, e.g., (Mani and Wilson 2000) (Filatova and Hovy 2001) (Schilder and Habel 2001) (Li et al. 2001) (Mani et al. 2003) (Li et al. 2004) (Lapata and Lascarides 2004) (Boguraev and Ando 2005) (Mani et al. 2006), that research relies on qualitative temporal relations. Qualitative relations (e.g., event A BEFORE event B, or event A DURING time T) are certainly of interest in developing timelines of events in news and other genres. However, metric constraints can also be potentially useful in this ordering problem. For example, in (3), it can be crucial to know whether the bomb landed a few minutes to hours or several years BEFORE the hospitalization. While humans have strong intuitions about this from commonsense knowledge, machines don't.

- (3) An elderly Catholic man was *hospitalized* from cuts after a Protestant gasoline bomb *landed* in his back yard.

Fortunately, there are numerous instances such as (4), where metric constraints are specified explicitly:

- (4) The company *announced* Tuesday that third quarter sales *had fallen*.

In (4), the falling of sales occurred over the three-month period of time inferable from the speech time. However, while the announcement is anchored to a day inferable from the speech

time, the length of the announcement is not specified.

These examples suggest that it may be possible to mine information from a corpus to fill in extents for the time intervals of and between events, when these are either unspecified or partially specified. Metric constraints can also potentially lead to better qualitative links, e.g., events with long durations are more likely to overlap with other events.

This paper describes some preliminary experiments to acquire metric constraints. The approach extends the TimeML representation (Pustejovsky et al. 2005) to include such constraints. We first translate a TimeML representation with qualitative relations into one where metric constraints are added. This representation may tell us how long certain events last, and the length of the gaps between them, given the information in the text. However, the information in the text may be incomplete; some extents may be unknown. We therefore need an external source of knowledge regarding the typical extents of events, which we can use when the text doesn't provide it. We accordingly describe an initial attempt to bootstrap event durations from raw corpora as well as corpora annotated with qualitative relations.

2 Annotation Scheme and Corpora

TimeML (Pustejovsky et al. 2005) (www.timeml.org) is an annotation scheme for markup of events, times, and their qualitative temporal relations in news articles. The TimeML scheme flags tensed verbs, adjectives, and nominals with EVENT tags with various attributes, including the class of event, tense, grammatical aspect, polarity (negative or positive), any modal operators which govern the event being tagged, and cardinality of the event if it's mentioned more than once. Likewise, time expressions are flagged and their values normalized, based on an extension of the ACE (2004) (tern.mitre.org) TIMEX2 annotation scheme (called TIMEX3).

For temporal relations, TimeML defines a TLINK tag that links tagged events to other events and/or times. For example, given sentence (4), a TLINK tag will **anchor** the event instance of announcing to the time expression *Tuesday* (whose normalized value will be inferred from context), with the relation IS_INCLUDED. This is shown in (5).

(5) The company <EVENT event-tID=e1>announced</EVENT> <TIMEX3

```
tid=t1 value=1998-01-08>Tuesday
</TIMEX3> that <TIMEX3 tid=t2
value=P1Q3 beginPoint=t3 end-
Point=t4>third-quarter</TIMEX3>
sales <EVENT eventID=e2> had
fallen</EVENT>.
<TLINK eventID=e1 relatedToEventID=e2 relType=AFTER/>
<TLINK eventID=e1 relatedToTimeID=t1
relType=IS_INCLUDED/>
<TIMEX3 tid=t3 value=1997-07/>
<TIMEX3 tid=t4 value=1997-09/>
```

The representation of time expressions in TimeML uses TIMEX2, which is an extension of the TIMEX2 scheme (Ferro et al. 2005). It represents three different kinds of time values: points in time (answering the question “when?”), durations (answering “how long?”), and frequencies (answering “how often?”)¹.

TimeML uses 14 temporal relations in the TLINK relTypes. Among these, the 6 inverse relations are redundant. In order to have a non-hierarchical classification, SIMULTANEOUS and IDENTITY are collapsed, since IDENTITY is a subtype of SIMULTANEOUS. (An event or time is SIMULTANEOUS with another event or time if they occupy the same time interval. X and Y are IDENTICAL if they are simultaneous *and* coreferential). DURING and IS_INCLUDED are collapsed since DURING is a subtype of IS_INCLUDED that anchors events to times that are durations. (An event or time INCLUDES another event or time if the latter occupies a proper subinterval of the former.) IBEFORE (immediately before) corresponds to MEETS in Allen's interval calculus (Allen 1984). Allen's OVERLAPS relation is not represented in TimeML.

The above considerations allow us to collapse the TLINK relations to a disjunctive classification of 6 temporal relations $TReIs = \{\text{SIMULTANEOUS, IBEFORE, BEFORE, BEGINS, ENDS, INCLUDES}\}$. These 6 relations and their inverses map one-to-one to 12 of Allen's 13 basic relations (Allen 1984).

Formally, each TLINK is a constraint of the general form $x R y$, where x and y are intervals, and R is a disjunct $\bigvee_{i=1,\dots,6}(r_i)$, where r_i is a relation in $TReIs$. In annotating a document for Ti-

¹ Our representation (using t3 and t4) grounds the fuzzy primitive P1Q3 (i.e., a period of one 3rd-quarter) to specific months, though this is an application-specific step. In analyzing our data, we normalize P1Q3 as P3M (i.e., a period of 3 months). For conciseness, we omit TimeML EVENT and TIMEX3 attributes that aren't relevant to the discussion.

meML, the annotator adds a TLINK iff she can commit to the TLINK relType being unambiguous, i.e., having exactly one relType r.

Two human-annotated corpora have been released based on TimeML²: TimeBank 1.2 (Pustejovsky et al. 2003) with 186 documents and 64,077 words of text, and the Opinion Corpus (www.timeml.org), with 73 documents and 38,709 words. TimeBank 1.2 (we use 1.2.a) was created in the early stages of TimeML development, and was partitioned across five annotators with different levels of expertise. The Opinion Corpus was developed recently, and was partitioned across just two highly trained annotators, and could therefore be expected to be less noisy. In our experiments, we merged the two datasets to produce a single corpus, called OTC.

3 Translation

3.1 Introduction

The first step is to translate a TimeML representation with qualitative relations into one where metric constraints are added. This translation needs to produce a consistent metric representation. The temporal extents of events, and between events, can be read off, when there are no unknowns, from the metric representation. The problem, however is that the representation may have unknowns, and the extents may not be minimal.

3.2 Mapping to Metric Representation

Let each event or time interval x be represented as a pair of start and end time points $\langle x_1, x_2 \rangle$. For example, given sentence (4), and the TimeML representation shown in (5), let x be *fall* and y be *announce*. Then, we have $x_1 = 19970701T00$, $x_2 = 19970930T23:59$, $y_1 = 19980108Tn_1$, $y_2 = 19980108Tn_2$ (here T represents time of day in hours).

To add metric constraints, given a pair of events or times x and y , where $x = \langle x_1, x_2 \rangle$ and $y = \langle y_1, y_2 \rangle$, we need to add, based on the qualitative relation between x and y , constraints of the general form $(x_i - y_j) \leq n$, for $1 \leq i, j \leq 2$. We follow precisely the method ‘Allen-to-metric’ of (Kautz and Ladkin 1991) which defines metric constraints for each relation R in *TREls*. For example, here is a qualitative relation and its metric constraints:

$$(6) \quad x \text{ is BEFORE } y \text{ iff } (x_2 - y_1) < 0.$$

²More details can be found at timeml.org.

In our example, where x is *fall* and y is the *announce*, we are given the qualitative relationship that x is BEFORE y , so the metric constraint $(x_2 - y_1) < 0$ can be asserted.

Consider another qualitative relation and its metric constraint:

$$(7) \quad z \text{ INCLUDES } y \text{ iff } (z_1 - y_1) < 0 \text{ and } (y_2 - z_2) < 0.$$

Let y be *announce* in (4), as before, and let $z = \langle z_1, z_2 \rangle$ be the time of *Tuesday*, where $z_1 = 19980108T00$, and $z_2 = 19980108T23:59$. Since we are given the qualitative relation y IS_INCLUDED z , the metric constraints $(z_1 - y_1) < 0$ and $(y_2 - z_2) < 0$ can be asserted.

3.3 Consistency Checking

We now turn to the general problem of checking consistency. The set of TLINKs for a document constitutes a graph, where the nodes are events or times, and the edges are TLINKs. Given such a TimeML-derived graph for a document, a temporal closure algorithm (Verhagen 2005) carries out a transitive closure of the graph. The transitive closure algorithm was inspired by (Setzer and Gaizauskas 2000) and is based on Allen’s interval algebra, taking into account the limitations on that algebra that were pointed out by (Vilain et al. 1990). It is basically a constraint propagation algorithm that uses a transitivity table to model the compositional behavior of all pairs of relations in a document. The algorithm’s transitivity table is represented by 745 axioms. An example axiom is shown in (8):

$$(8) \quad \text{If relation}(A, B) = \text{BEFORE} \ \&\& \ \text{relation}(B, C) = \text{INCLUDES} \\ \text{then infer relation}(A, C) = \text{BEFORE}.$$

In propagating constraints, links added by closure can have a disjunction of one or more relations in *TREls*. When the algorithm terminates, any TLINK with more than one disjunct is discarded. Thus, a closed graph is consistent and has a single relType r in *TREls* for each TLINK edge. The algorithm runs in $O(n^3)$ time, where n is the number of intervals.

The closed graph is augmented so that whenever input edges $a \ r_1 \ b$ and $b \ r_2 \ c$ are composed to yield the output edge $a \ r_3 \ c$, where r_1 , r_2 , and r_3 are in *TREls*, the metric constraints for r_3 are added to the output edge. To continue our example, since the *fall* x is BEFORE the *Tuesday* z

and z INCLUDES y (*announce*), we can infer, using rule 8, that x is BEFORE y , i.e., that *fall* precedes *announce*. Using rule (6), we can again assert that $(x_2 - y_1) < 0$.

3.4 Reading off Temporal Extents

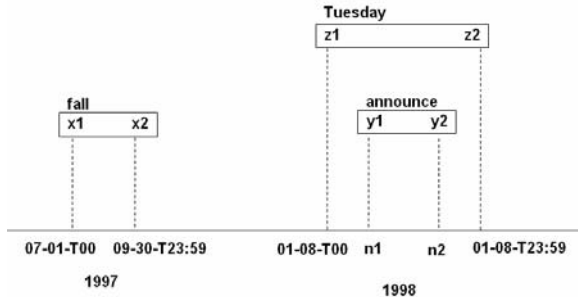


Figure 1. Metric Constraints

We now have the metric constraints added to the graph in a consistent manner. It remains to compute, given each event or time $x = \langle x_1, x_2 \rangle$, the values for x_1 and x_2 . In our example, we have *fall* $x = \langle 19970701T00, 19970930T23:59 \rangle$, *announce* $y = \langle 19980108Tn1, 19980108Tn2 \rangle$, and *Tuesday* $z = \langle 19980108T00, 19980108T23:59 \rangle$, and the added metric constraints that $(x_2 - y_1)$, $(z_1 - y_1)$, and $(y_2 - z_2)$ are all negative. Graphically, this can be pictured as in Figure 1.

As can be seen in Figure 1, there are still unknowns (n_1 and n_2): we aren't told exactly how long *announce* lasted -- it could be anywhere up to a day. We therefore need to acquire information about how long events last when the example text doesn't tell us. We now turn to this problem.

4 Acquisition

We started with the 4593 event-time TLINKs we found in the unclosed human-annotated OTC. From these, we restricted ourselves to those where the times involved were of type TIMEX3 DURATION. We augmented the TimeBank data with information from the raw (un-annotated) British National Corpus. We tried a variety of search patterns to try and elicit durations, finally converging on the single pattern "lasted". There were 1325 hits for this query in the BNC. (The public web interface to the BNC only shows 50 random results at a time, so we had to iterate.) The retrieved hits (sentences and fragments of sentences) were then processed with components from the TARSQI toolkit (Verhagen et al. 2005) to provide automatic TimeML annotations. The TLINKs between events and times that were

TIMEX3 DURATIONS were then extracted. These links were then corrected and validated by hand and then added to the OTC data to form an integrated corpus. An example from the BNC is shown in (9).

(9) The `<EVENT>storm</EVENT>`
`<EVENT>lasted</EVENT>` `<TIMEX3`
`VAL="P5D">five days</TIMEX3>`.

Next, the resulting data was subject to morphological normalization in a semi-automated fashion to generate more counts for each event. Hyphens were removed, plurals were converted to singular forms, finite verbs to infinitival forms, and gerundive nominals to verbs. Derivational ending on nominals were stripped and the corresponding infinitival verb form generated. These normalizations are rather aggressive and can lead to loss of important distinctions. For example, sets of events (e.g., storms or bombings) as a whole can have much longer durations compared to individual events. In addition, no word-sense disambiguation was carried out, so different senses of a given verb or event nominal may be confounded together.

5 Results

Number of durations	Number of Events	Normalized Form of Event
26	1	<i>lose</i>
16	1	<i>earn</i>
10	1	<i>fall</i>
9	1	<i>rise</i>
8	1	<i>drop</i>
7	2	<i>decline, increase</i>
6	4	<i>end, grow, say, sell</i>
5	2	<i>income, stop</i>
4	9	...
3	17	...
2	40	...
1	176	...

Table 1. Frequencies of event durations

The resulting dataset had 255 distinct events with the number of durations for the events as shown in the frequency distribution in Table 1. The granularities found in news corpora such as OTC and mixed corpora such as BNC are domi-

nated by quarterly reports, which reflect the influence of specific information pinpointing the durations of financial events. This explains the fact that 12 of the top 13 events in Table 1 are financial ones, with the reporting verb *say* being the only non-financial event in the top 13.

The durations for the most frequent event, represented by the verb *to lose*, is shown in Table 2. Most losses are during a quarter, or a year, because financial news tends to quantize losses for those periods.

Duration	Frequency
1 day (P1D)	1
2 months (P2M)	1
unspecified weeks (PXW)	1
unspecified months (PXM)	1
3 months (P3M)	9
9 months (P9M)	3
1 year (P1Y)	6
5 years (P5Y)	1
1 decade (P1Y)	1
TOTAL	26

Table 2. Distribution of durations for event *to lose*

Ideally, we would be able to generalize over the duration values, grouping them into classes. Table 3 shows some hand-aggregated duration classes for the data. These classes are ranges of durations. It can be seen that the temporal span of events across the data is dominated by granularities of weeks and months, extending into small numbers of years.

Duration Class	Count
<1 min	1
5-15 min	12
1-<24 hr	20
1 day	14
2-14 days	49
1-3 months	120
7-9 months	48
1-6 years	97
1 decade - < 1 century	30
1-2 centuries	2
vague (unspecified mins/days/months, continuous present, indefinite future, etc.)	69

Table 3. Distribution of aggregated durations

Interestingly, 67 events in the data correspond to ‘achievement’ verbs, whose main characteristic is that they can have a near-instantaneous duration (though of course they can be iterated or extended to have other durations). We obtained a list of achievement verbs from the LCS lexicon of (Dorr and Olsen 1997)³. Achievements can be marked as having durations of PTXS, i.e., an unspecified number of seconds. Such values don’t reinforce any of the observed values, instead extending the set of durations to include much smaller durations. As a result, these hidden values are not shown in our data

6 Estimating Duration Probabilities

Given a distribution of durations for events observed in corpora, one of the challenges is to arrive at an appropriate value for a given event (or class of events). Based on data such as Table 2, we could estimate the probability $P(\textit{lose}, P3M) \cong 0.346$, while $P(\textit{lose}, P1D) \cong 0.038$, which is nearly ten times less likely. Table 2 reveals peaking at 3 months, 6 months, and 9 months, with uniform probabilities for all others. Further, we can estimate the probability that losses will be during periods of 2 months, 3 months, or 9 months as $\cong 0.46$. Of course, we would prefer a much large sample to get more reliable estimates.

One could also infer a max-min time range, but the maximum or minimum may not always be likely, as in the case of *lose*, which has relatively low probability of extending for “P1D” or “P1E”. Turning to earnings, we find that $P(\textit{earn}, P9M) \cong 4/16 = 0.25$, $P(\textit{earn}, P1Y) \cong 0.31$, but $P(\textit{earn}, P3M) \cong 0.43$, since most earnings are reported for a quarter.

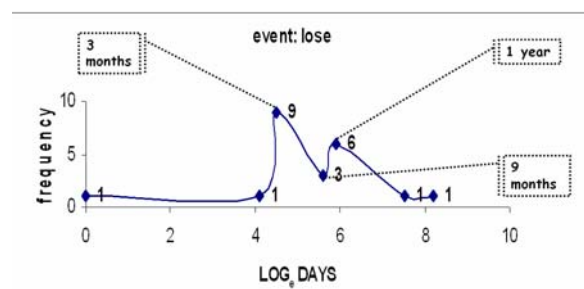


Figure 2. Distribution of durations of event *to lose*

So far, we have considered durations to be discrete, falling into a fixed number of categories. These categories could be atomic TimeML DU-

³See www.umiacs.umd.edu/~bonnie/LCS_Data-base_Documentation.html.

RATION values, as in the examples of durations in Table 2, or they could be aggregated in some fashion, as in Table 3. In the discrete view, unless we have a category of 4 months, the probability of a loss extending over 4 months is undefined. Viewed this way, the problem is one of classification, namely providing the probability that an event has a particular duration category.

The second view takes duration to be continuous, so the duration of an event can have any subinterval as a value. The problem here is one of regression. We can re-plot the data in Table 2 as Figure 2, where we have plotted durations in days on the x-axis in a natural log scale, and frequency on the y-axis. Since we have plotted the durations as a curve, we can interpolate and extrapolate durations, so that we can obtain the probability of a loss for 4 months. Of course, we would like to fit the best curve possible, and, as always, the more data points we have, the better.

7 Possible Enhancements

One of the basic problems with this approach is data sparseness, with few examples for each event. This makes it difficult to generalize about durations. In this section, we discuss enhancements that can address this problem.

7.1 Converting points to durations

More durations can be inferred from the OTC by coercing TIME3 DATE and TIME expressions to DURATIONS; for example, if someone announced something in 1997, the maximum duration would be one year. Whether this leads to reliable heuristics or not remains to be seen.

7.2 Event class aggregation

A more useful approach might be to aggregate events into classes, as we have done implicitly with financial events. Reporting verbs are already identified as a TimeML subclass, as are aspectual verbs such as *begin*, *continue* and *finish*. Arriving at an appropriate set of classes, based on distributional data or resource-derived classes (e.g., TimeML, VerbNet, WordNet, etc.) remains to be explored.

7.3 Expanding the corpus sample

Last but not least, we could expand substantially the search patterns and size of the corpus searched against. In particular, we could emulate the approach used in VerbOcean (Chklovski and Pantel 2004). This resource consists of lexical relations mined from Google searches. The min-

ing uses a set of lexical and syntactic patterns to test for pairs of verbs strongly associated on the Web in a particular semantic relation. For example, the system discovers that *marriage* happens-before *divorce*, and that *tie* happens-before *untie*. Such results are based on estimating the probability of the joint occurrence of the two verbs and the pattern. One can imagine a similar approach being used for durations. Bootstrapping of patterns may also be possible.

8 Conclusion

This paper describes the first steps in acquiring metric temporal constraints for events. The work is carried out in the context of the TimeML framework for marking up events and their temporal relations. We have identified a method for enhancing TimeML annotations with metric constraints. Although the temporal reasoning required to carry that out has been described in the prior literature, e.g., (Kautz and Ladkin 1991), this is a first attempt at lexical acquisition of metrical constraints. As a pilot study, it does suggest the feasibility of acquisition of metric temporal constraints from corpora. In follow-on research, we will explore the enhancements described in Section 7.

However, this work is limited by the lack of evaluation, in terms of assessing how valid the durations inferred by our method are compared with human annotations. In ongoing work, Jerry Hobbs and his colleagues (Pan et al. 2006) have developed an annotation scheme for humans to mark up event durations in documents. Once such enhancements are carried out, it will certainly be fruitful to compare the duration probabilities obtained with the ranges of durations provided in that corpus.

In future, we will explore both regression and classification models for duration learning. In the latter case, we will investigate the use of constructive induction e.g., (Bloedorn and Michalski 1998). In particular, we will avail of operators to implement attribute abstraction that will cluster durations into coarse-grained classes, based on distributions of atomic durations observed in the data. We will also investigate the extent to which learned durations can be used to constrain TLINK ordering.

References

- James Allen. 1984. Towards a General Theory of Action and Time. *Artificial Intelligence*, 23, 2, 123-154.

- Eric Bloedorn and Ryszard S. Michalski. 1998. Data-Driven Constructive Induction. *IEEE Intelligent Systems*, 13, 2.
- Branimir Boguraev and Rie Kubota Ando. 2005. TimeML-Compliant Text Analysis for Temporal Reasoning. *Proceedings of IJCAI-05*, 997-1003.
- Timothy Chklovski and Patrick Pantel. 2004. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. *Proceedings of EMNLP-04*. <http://semantics.isi.edu/ocean>
- B. Dorr and M. B. Olsen. Deriving Verbal and Compositional Lexical Aspect for NLP Applications. *ACL'1997*, 151-158.
- Janet Hitzeman, Marc Moens and Clare Grover. 1995. Algorithms for Analyzing the Temporal Structure of Discourse. *Proceedings of EACL'95*, Dublin, Ireland, 253-260.
- Feng Pan, Rutu Mulkar, and Jerry Hobbs. Learning Event Durations from Event Descriptions. *Proceedings of Workshop on Annotation and Reasoning about Time and Events (ARTE'2006)*, *ACL'2006*.
- C.H. Hwang and L. K. Schubert. 1992. Tense Trees as the fine structure of discourse. *Proceedings of ACL'1992*, 232-240.
- Hans Kamp and Uwe Ryle. 1993. *From Discourse to Logic (Part 2)*. Dordrecht: Kluwer.
- Andrew Kehler. 2000. Resolving Temporal Relations using Tense Meaning and Discourse Interpretation, in M. Faller, S. Kaufmann, and M. Pauly, (eds.), *Formalizing the Dynamics of Information*, CSLI Publications, Stanford.
- Henry A. Kautz and Peter B. Ladkin. 1991. Integrating Metric and Qualitative Temporal Reasoning. *AAAI'91*.
- Mirella Lapata and Alex Lascarides. 2004. Inferring Sentence-internal Temporal Relations. In *Proceedings of the North American Chapter of the Association of Computational Linguistics*, 153-160.
- Alex Lascarides and Nicholas Asher. 1993. Temporal Relations, Discourse Structure, and Commonsense Entailment. *Linguistics and Philosophy* 16, 437-494.
- Wenjie Li, Kam-Fai Wong, Guihong Cao and Chunfa Yuan. 2004. Applying Machine Learning to Chinese Temporal Relation Resolution. *Proceedings of ACL'2004*, 582-588.
- Inderjeet Mani and George Wilson. 2000. Robust Temporal Processing of News. *Proceedings of ACL'2000*.
- Inderjeet Mani, Barry Schiffman, and Jianping Zhang. 2003. Inferring Temporal Ordering of Events in News. Short Paper. *Proceedings of HLT-NAACL'03*, 55-57.
- Inderjeet Mani, Marc Verhagen, Ben Wellner, Chong Min Lee, and James Pustejovsky. 2006. Machine Learning of Temporal Relations. *Proceedings of ACL'2006*.
- Rebecca J. Passonneau. A Computational Model of the Semantics of Tense and Aspect. *Computational Linguistics*, 14, 2, 1988, 44-60.
- James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, David Day, Lisa Ferro, Robert Gaizauskas, Marcia Lazo, Andrea Setzer, and Beth Sundheim. 2003. The TimeBank Corpus. *Corpus Linguistics*, 647-656.
- James Pustejovsky, Bob Ingria, Roser Sauri, Jose Castano, Jessica Littman, Rob Gaizauskas, Andrea Setzer, G. Katz, and I. Mani. 2005. The Specification Language TimeML. In I. Mani, J. Pustejovsky, and R. Gaizauskas, (eds.), *The Language of Time: A Reader*. Oxford University Press.
- Roser Sauri, Robert Knippen, Marc Verhagen and James Pustejovsky. 2005. Evita: A Robust Event Recognizer for QA Systems. Short Paper. *Proceedings of HLT/EMNLP 2005*: 700-707.
- Frank Schilder and Christof Habel. 2005. From temporal expressions to temporal information: semantic tagging of news messages. In I. Mani, J. Pustejovsky, and R. Gaizauskas, (eds.), *The Language of Time: A Reader*. Oxford University Press.
- Andrea Setzer and Robert Gaizauskas. 2000. Annotating Events and Temporal Information in Newswire Texts. *Proceedings of LREC-2000*, 1287-1294.
- Marc Verhagen. 2004. *Times Between The Lines*. Ph.D. Dissertation, Department of Computer Science, Brandeis University.
- Marc Verhagen, Inderjeet Mani, Roser Sauri, Robert Knippen, Jess Littman and James Pustejovsky. 2005. Automating Temporal Annotation with TARSQI. Demo Session, *ACL 2005*.
- Marc Vilain, Henry Kautz, and Peter Van Beek. 1989. Constraint propagation algorithms for temporal reasoning: A revised report. In D. S. Weld and J. de Kleer (eds.), *Readings in Qualitative Reasoning about Physical Systems*, Morgan-Kaufman, 373-381.
- Bonnie Webber. 1988. Tense as Discourse Anaphor. *Computational Linguistics*, 14, 2, 1988, 61-73.