Increasing Informativeness in Temporal Annotation

James Pustejovsky
Department of Computer Science
Brandeis University MS 018
Waltham, Massachusetts, 02454 USA
jamesp@cs.brandeis.edu

Amber Stubbs
Department of Computer Science
Brandeis University MS 018
Waltham, Massachusetts, 02454 USA
astubbs@cs.brandeis.edu

Abstract

In this paper, we discuss some of the challenges of adequately applying a specification language to an annotation task, as embodied in a specific guideline. In particular, we discuss some issues with TimeML motivated by error analysis on annotated TLINKs in TimeBank. We introduce a document level information structure we call a *narrative container (NC)*, designed to increase informativeness and accuracy of temporal relation identification. The narrative container is the default interval containing the events being discussed in the text, when no explicit temporal anchor is given. By exploiting this notion in the creation of a new temporal annotation over TimeBank, we were able to reduce inconsistencies and increase informativeness when compared to existing TLINKs in TimeBank.

1 Introduction

In linguistic annotation projects, there is often a gap between what the annotation schema is designed to capture and how the guidelines are interpreted by the annotators and adjudicators given a specific corpus and task (Ide and Bunt, 2010; Ide, 2007). The difficulty in resolving these two aspects of annotation is compounded when tasks are looked at in a potentially incomplete annotation task; namely, where the guideline is following a specification to a point, but in fact human annotation is not even suggested as complete because it would be infeasible. Creating temporal links to represent the timeline of events in a document is an example of this: human annotation of every possible temporal relationship between events and times in a narrative would be an overwhelming task.

In this paper, we discuss how temporal relation annotation must be sensitive to two aspects of the task that were not mentioned in the TimeBank guideline (Pustejovsky et al., 2005): (a) sensitivity to the genre and style of the text; and (b) the interaction with discourse relations that explicitly reference the flow of the narrative in the text. We believe that making reference to both these aspects in the text during the annotation process will increase overall informativeness and accuracy of the annotation. In the present paper, we focus primarily on the first of these points, and introduce a document level information structure we call a *narrative container (NC)*.

Because of the impossibility of humans capturing every relationship, it is vital that the annotation guidelines describe an approach that will result in maximally informative temporal links without relying on standards that are too difficult to apply. With this in mind, we have been examining the TimeBank corpus (Pustejovsky et al., 2003) and the annotation guideline that created it, and have come to these realizations:

(1) • The guideline does not specify certain types of annotations that should be performed;
• The guideline forces some annotations to be performed when they should not always be.

Additionally, we have discovered some inconsistencies in the TimeBank corpus related to temporal links. Furthermore, upon examination, we have become aware of the importance of the text style and

Proceedings of the Fifth Law Workshop (LAW V), pages 152–160,
Portland, Oregon, 23-24 June 2011. ©2011 Association for Computational Linguistics

genre, and how readers interpret temporally unahchored events.

This gave rise, in examining the genres that are most frequent in TimeBank (namely news and finance), to the possibility that readers of news articles and narratives have possible default assumptions about when unanchored events take place. It seems reasonable for a reader to assume in a sentence such as: *Oneida Ltd. declared a 10% stock dividend, payable Dec. 15 to stock of record Nov. 17*, that the "declared" event took place soon before the article's Document Creation Time (DCT).

Exactly how soon before may be related to some proximate interval of time associated with both the publication time and frequency. That is, it appears that just as importantly, if not more so, than the DCT, is a related and dependent notion of the salient interval surrounding the creation time, for interpreting the events that are being reported or written about. We will call this the *Narrative Container*. There seems to be a default value for this container affected by many variables. For example, a print newspaper seems to associate in the content and style a narrative container of approximately 24 hours, or one business day. A newswire article, on the other hand, has a narrative container of 2-10 hours. Conversely, weekly and monthly publications would likely have a narrative container of a much longer duration (a week or more).

Along with the narrative container, there are two related concepts that proved useful in framing this new approach to temporal annotation. The *Narrative Scope* describes the timespan described in the document, with the left marker defined by the earliest event mentioned in the document, and the right by the event furthest in the future. The other important concept is that of *Narrative Time*. A Narrative Time is essentially the current temporal anchor for events in a document, and can change as the reader moves through the narrative.

With these as initial assumptions we did some cursory inspection of the TimeBank data to determine if there was a correlation between Narrative Container length and genre, and found it to be a compelling assumption. With that in mind, we determined that TLINK creation should be focused on relationships to the narrative container, rather than to the DCT.

Our goal is, to the extent possible, to see how we can use a container metaphor, albeit somewhat underspecified, to left-delineate the container within which unanchored events might be in relation to.

2 Identifying Temporal Relations

While low-level temporal annotation tasks such as identifying events and time expressions are relatively straightforward and can be marked up with high consistency, high-level tasks such as arranging events in a document in a temporal order have proved to be much more challenging. The temporal ordering of events in a document, for example, is accomplished by identifying all distinct event-event pairings. For a document that has n events, this requires the annotation of $\binom{n}{2}$ events pairs. Obviously, for general-purpose annotation, where all possible events are considered, the number of event pairs grows essentially quadratically to the number of events, and the task quickly becomes unmanageable.

There are, however, strategies that we can adopt to make this labeling task more tractable. First we need to distinguish the domains over which ordering relations are performed. Temporal ordering relations in text are of three kinds:

(2) a. A relation between two events;
b. A relation between two times;
c. A relation between a time and an event.

TimeML, as a formal specification of the temporal information conveyed in language, makes no distinction between these ordering types. But a human reader of a text does make a distinction, based on the discourse relations established by the author of the narrative (Miltsakaki et al., 2004; Poesio, 2004). Temporal expressions denoting the local *Narrative Container* in the text act as embedding intervals within which events occur. Within TimeML, these are event-time anchoring relations (TLINKs). Discourse relations establish how events relate to one another in the narrative, and hence should constrain temporal relations between two events. Thus, one of the most significant constraints we can impose is to take advantage of the discourse structure in the document before event-event ordering relations are identified.

Although, in principle, during an annotation a temporal relation can be specified between any two events in the text, it is worth asking what *informativeness* a given temporal relation introduces to the annotation. The informativeness of an annotation will be characterized as a function of the information contained in the individual links and their closure. We can distinguish, somewhat informally for now, two sources of informativeness in how events are temporally ordered relative to each other in a text: (a) externally and (b) internally. Consider first *external informativeness*. This is information derived from relations outside the temporal relation constraint set, e.g., as coming from explicit discourse relations between events (and hence is associated with the relations in (2a) above). For example, we will assume that, for two events, e_1 and e_2, in a text, the temporal relation between them is more informative if they are also linked through a discourse relation, e.g., a PDTB relation (Prasad et al., 2008). Making such an assumption will allow us to focus in on the temporal relations that are most valuable without having to exhaustively annotate all event pairs.

Now consider *internal informativeness*. This is information derived from the nature of the relation itself, as defined largely by the algebra of relations (Allen, 1984; Vilain et al., 1986). First, we assume that, for two events, e_1 and e_2, a temporal relation R_1 is more informative than R_2 if R_1 entails R_2. More significantly, however, as noted above, is to capitalize on the relations that inhere between events and the times that anchor them (i.e., (2c) above). Hence, we will say that, given an event, e_1 and a time t_1, a temporal relation R is more informative the more it anchors e_1 to t_1. That is, a containment relation is more informative than an ordering relation, and the smaller the container, the more informative the relation.[1]

The Document Creation Time (DCT) as designed in TimeML is introduced as a reference time, against which the mentioned events and time expressions in the document can be ordered. Consider the text fragment below.

[1] We defer discussion of the formal definition of informativeness for the present paper, as we are focusing on initial results over re-annotated data in TimeBank.

4-10-2011
Local officials **reported** *yesterday that a car* **exploded** *in downtown Basra.*

The TimeML annotation guideline (AG) suggests identifying relations between the DCT and textual events. Hence standard markup as in TimeBank results in the following sort of annotation:

(3) a. DCT= t_1, val=10-04-2011
b. t_2 = yesterday, val=09-04-2011
b. e_1 = report
c. e_2 = explode
d. TLINK$_1$ = before(e_1, t_1)
e. TLINK$_2$ = before(e_2, t_1)
f. TLINK$_3$ = includes(t_2, e_1)

This is a prototypical annotation fragment. Notice that by focusing on the link between events and the DCT, the annotator is forced to engage in a kind of periodic "back-and-forth" evaluation of the events in the text, relative to the DCT. While there is a container TIMEX3 that bounds e_1, there is no information given grounding the actual time of the event of interest, namely, the explosion, e_2. By following the AG literally and through no fault of their own, the annotators have missed an opportunity to provide a more informative markup; namely, the identification of the TLINK below:

(4) TLINK$_4$ = includes(t_2, e_2)

That is, the explosion occurred on the date valued for *yesterday*, i.e., "09-04-2011".

The point of this paper is to discuss the difference encountered when applying a specification given a particular guideline for annotating a body of text. The example we want to discuss is the manner in which events are linked (related) to the Document Creation Time (DCT) in TimeML. These considerations have arisen in the context of new annotation problems in different genre and domains, hoping to apply the principles of TimeML.

3 Narrative Scope

As previously mentioned, the Narrative Scope of a document is the temporal span over which the events in a document occur, as defined by the timexes in a

document. While not every event in a document will necessarily occur inside the Narrative Scope (some may still occur before or after any dates that are specifically mentioned), the Narrative Scope provides a useful container for describing when events discussed most likely occurred. The narrative scope was not considered as part of the annotation task, but it did help to ground the concepts of Narrative Containers and Narrative Times.

4 Narrative Time

As a reader moves through a document, the introduction of a new TIMEX will often shift the temporal focus of the events to be anchored to this new time point (Smith, 2003). These temporal anchors are what we refer to as Narrative Times, and function in much the same way as newly introduced locations in spatial annotation.

However, consider how we can use Narrative Times to increase accuracy of the TLINKS over a document in TimeML. As mentioned above, we distinguish three types of temporal orderings in a text: time-time, event-time, and event-event. The first identifies orderings between two TIMEX3 expressions and is performed automatically. The second identifies what the local Narrative Time for an event is, i.e., how an EVENT is anchored to a TIMEX3. Event-event pairings, for the purposes of this paper, will not be discussed, though they are a vital and complex component of temporal annotation, largely involving discourse relations.

To illustrate our proposed strategy, consider the news article text shown below.

> *April 25, 2010 7:04 p.m. EDT* **-t0**
>
> **S1:** President Obama *paid*-**e1** tribute *Sunday* **-t1** to 29 workers *killed*-**e2** in an *explosion* **-e3** at a West Virginia coal mine *earlier this month*- **t2**, *saying*-**e4** they *died*-**e5** "in pursuit of the American dream."
>
> **S2:** The *blast*-**e6** at the Upper Big Branch Mine was the worst U.S. mine disaster in nearly 40 years.

There are three temporal expressions in the above text: the Document Creation time, **t0**; and two TIMEXes, **t1** and **t2**. Each of these TIMEXes functions as a Narrative Time, as they are clearly providing temporal anchors to nearby events. In this case, all the events are located within the Narrative Time appropriate to them. Hence, the number of orderings is linearly determined by the number of events in the document, since each is identified with a single Narrative Time. Knowing the narrative time associated with each event will allow us to perform limited temporal ordering between events that are associated with different narrative times, which, as mentioned above, is significantly more informative than if events were only given partial orderings to the DCT or to each other.

5 Narrative Containers

So far we have examined sentences that contain specific temporal anchors for the events discussed. Consider, however, the following sentences from article wsj_1031.tml in TimeBank:

> 10-26-1989
> 1 Philip Morris Cos., New York, *adopted* a defense measure *designed* to *make* a hostile *takeover* prohibitively expensive.
>
> 2 The giant foods, tobacco and brewing company *said* it will *issue* common-share purchase rights to shareholders of *record* Nov. 8.

Aside from the DCT, the only TIMEX in these two sentences is Nov. 8, which is only anchoring *issue* and *record*. The other events in the sentences can only be connected to the DCT, and presumably only in a 'before' or 'after' TLINK—in the absence of other information, any reader would assume from the past tenses of *adopted* and *said* that these events occurred before the article was published, and that any events associated with the future (*make, takeover*) are intended to happen after the DCT.

However most readers, knowing that the Wall Street Journal is published daily, will likely assume that any event mentioned which is not specifically associated with a date, occurred within a certain time frame—it would be extremely unusual for a newspaper to use the construction presented above if the events actually occurred, for example, a year or even a week prior to the publication date. We call this assumed window the Narrative Container, as it provides left and right boundaries for when unan-

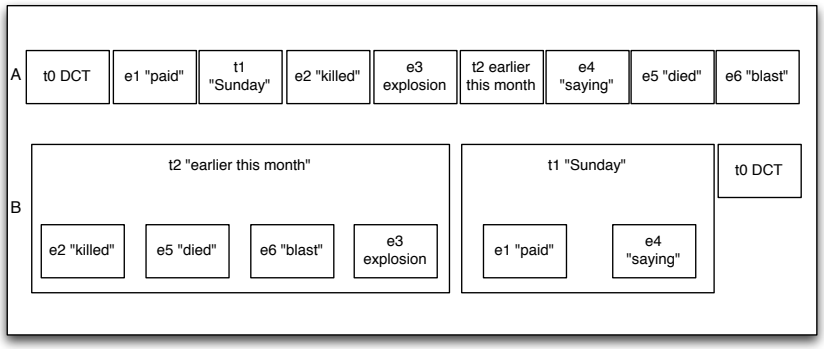

Figure 1: A: Times and events as appearing in the text; B: events grouped into their appropriate Narrative Times.

chored events most likely occurred, where in previous TimeML annotations these events would usually be given one-sided relationships to the DCT. In most cases, the right boundary of the Narrative Container is the DCT. The left boundary, however, requires other factors about the article to be taken into account before it can be given a value. The primary factor is how frequently the source of the document is published, but other aspects of the article may also determine the Narrative Container size.

5.1 Style, Genre, Channel, and Anchors

In order to determine what factors might influence the interpretation of the size of a Narrative Container, we asked an undergraduate researcher to categorize each of the articles in TimeBank according to the following characteristics (Lee, 2001; Biber, 2009).

(5) • Channel: is the document written or spoken?
• Production circumstances: how was the document distributed? broadcast, newswire, daily publication;
• Style: what format was used to present the information?
• Presence of a temporal anchor: Whether an article contained a Narrative Time in the first sentence of the document.

In general, we felt that the production circumstances would be the most relevant in determining the duration of the Narrative Container. The distributions of the different categories in TimeBank are shown in Table 1. There is a 100% overlap between the "broadcast" and "spoken" subcategories—all of those articles are word-for-word transcripts of television news reports. The "style" category proved the most difficult to define—the 'quiz' article is a broadcast transcript of a geography question asked during the evening news, while the 'biography' articles are overviews of people's lives. The editorials include a letter to the editor of the Wall Street Journal and an editorial column from the New York Times.

Category	number	percent
Production Circ.		
broadcast	25	13.7%
daily paper	140	76.5%
newswire	18	9.8%
Channel		
spoken	25	13.7%
written	158	86.3%
Style		
biography	2	1.1%
editorial	2	1.1%
finance	135	73.8%
news	43	23.5%
quiz	1	0.5%
Temporal Anchor		
no	138	75.4%
yes	45	24.6%

Table 1: Distributions of categories in TimeBank

6 Preliminary Studies

In order to assess the validity of our theories on Narrative Containers, Time, and Scope, we asked three undergraduate researchers to re-annotate TimeBank using the Narrative Container theory as a guide.

Each annotator evaluated all of the events in TimeBank by identifying the temporal constraint that anchored the event. If the annotators felt that the event was not specifically anchored, they could

place it within the Narrative Container for the document, or they could give the event a simple "before" or "after" value related to the Narrative Container or Document Creation Time. We also asked them to assign start and end times to the Narrative Container for each document.

The annotation here was not intended to be as complete as the TimeBank annotation task, or even the TempEval tasks—rather, the goal was to determine if the Narrative Container theory could be applied in a way that resulted in an increase in informativeness, and whether the annotators could work with the idea of a Narrative Container. Because these annotations are not comprehensive in their scope, the analysis provided here is somewhat preliminary, but we believe it is clear that the use of a Narrative Container in temporal annotations is both informative and intuitive.

6.1 Narrative container agreement

Each annotator was asked to assign a value to the narrative container of each document. They were given limited directions as to what the size of an NC might be: only some suggestions regarding possible correlations between type and frequency of publication and size of the narrative container. For example, it was suggested that a news broadcast might have a narrative container of only a few hours, a daily newspaper would have one of a day (or one that extended to the previous business day), and a newswire article would have a narrative container that extended back 24 hours from the time of publication.

All the annotators agreed that an NC would not extend forward beyond the document creation time (DCT), and that in most cases the NC would end at the DCT. Because the annotators gave their data on the size of the NC in free text (for example, an annotator would say "1 day" to indicate that the NC for an article began the day before the article was published) the comparison of the narrative containers was performed manually by one of the authors to determine if the annotators agreed on the size of the NC.

Agreement was determined using a fairly strict matching criterion—if the narrative containers given were clearly referring to the same interval they were interpreted to be the same. If, however, there was ambiguity about the date or one annotator indicated a smaller time period than another, then they were judged to be different. A common example of ambiguity was related to newspaper articles that were written on Mondays—annotators could not always determine if the events described occurred the day before, or on the previous business day For evaluation purposes, the ambiguous cases were given "maybe" values, but were not included in analysis that relied on the NCs being the same.

Overall, using the strict agreement metric all the annotators agreed on the size of the narrative container in 95 out of 183 articles—slightly over 50% of the time. However, the annotators only completely disagreed on 6 of the 183 articles—in all other cases there was some level of agreement between pairs of annotators.

6.2 NCs and Document Classifications

We compared Narrative Container agreements against the categories outlined above: style, channel, production circumstances, and temporal anchorings in order to determine if any of those attributes lent themselves to agreement about the size of the Narrative Container. We disregarded the biography, quiz, and editorial classifications as those categories were too small to provide useful data.

For the most part, no one category stood out as lending itself to accuracy—newswire had the highest levels of agreement at 72%, while daily papers came in at 58%. Written channels had 60% agreement, and the finance style had 59%. Articles with temporal anchors in the beginning of the document were actually slightly less likely to have agreement on the Narrative Container than those that didn't—48% and 53%, respectively.

While the higher disagreement levels over Narrative Container size in the presence of a temporal anchor seems counter-intuitive, it stems from a simple cause: if the temporal anchor overlapped with the expected narrative container but was not exactly the same size, sometimes one annotator would use that anchor as the Narrative Container, while the others would not. This sometimes also happened with a Narrative Time that was not at the start of the document or sometimes even the Narrative Scope would be used as the Narrative Container. While in some articles it is the case that a Narrative Time anchors more events than the Narrative Container,

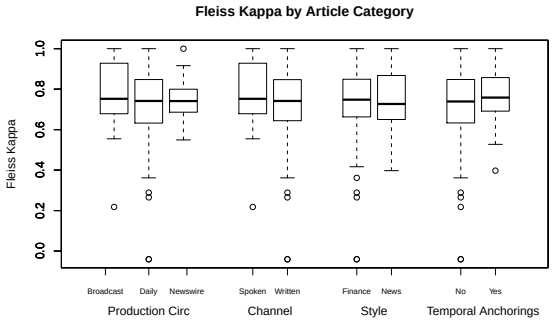

Figure 2: Distributions of Fleiss Kappa scores over TimeBank categories

that does not make that Narrative Time the Narrative Container for the document—the Narrative Container is always the interval during which an unanchored event would be assumed to have taken place. This point of confusion can easily be clarified in the guidelines.

Spoken/broadcast articles had the lowest agreement on Narrative Container size, with none of those articles having complete agreement between annotators. This was largely caused by our annotators not agreeing on how much time those categories would encompass by default–two felt that the narrative containers for broadcast news would extend to only a few hours before going on air, and the other felt that, like a daily paper, the entire previous day would be included when dealing with unanchored times.

As for the question of how large a Narrative Container should be for broadcast articles, the size of all Narrative Containers will need to be studied more in depth in order to determine how widely they can be applied— it is possible that in general, the actual size is less important than the simple concept of the Narrative Container.

6.3 Agreement over event anchors

The annotators were asked to read each article in TimeBank and "create links from each event to the nearest timex or to the DNC." They were asked specifically to not link an event to another event, only to find the time that would be used to anchor each event in a timeline. The annotators were also asked to use only three relationship types: before, after, and is_included (which also stood in for "overlap"). This was done in order to keep the annotation as simple as possible: we wanted to see if the narrative container was a useful tool in temporal annotation, not produce a full gold standard corpus.

This differs from the TimeML annotation guidelines, which suggested only that "A TLINK has to be created each time a temporal relationship holding between events or an event and a time needs to be annotated." (Saurí et al., 2006) Examples given were for sentences such as "John drove to Boston on Monday"—cases where an event was specifically related to a time or another event. However, because such examples were relatively rare, and temporal relationships are not always so clearly expressed, this annotation method resulted in a corpus that was not optimally informative. TimeML also uses a fuller set of temporal relations.

The NC annotations, on the other hand, are much richer in terms of informativeness. Annotators most often linked to the NC, often with an "is_included" relationship (as in: e1 is_included NC). In fact, roughly 50% of the events were linked to the narrative container and had "is_included" as the relationship type. In previous TimeML annotations, most of those events would have been annotated as simply occurring before or overlapping with the document creation time, which is a significantly less informative association. Clearly the narrative container was an intuitive concept for the annotators, and one that was relevant to their annotations.

6.3.1 Inter-annotator agreement

We used Fleiss' kappa (Fleiss, 1971) to obtain values for agreement between the three annotators: first, we compared the number of times they agreed what the temporal anchor for an event should be, then we compared whether those links that matched had the same relation type. Data analysis was done in R with the irr package (R Team, 2009; Gamer et al., 2010).

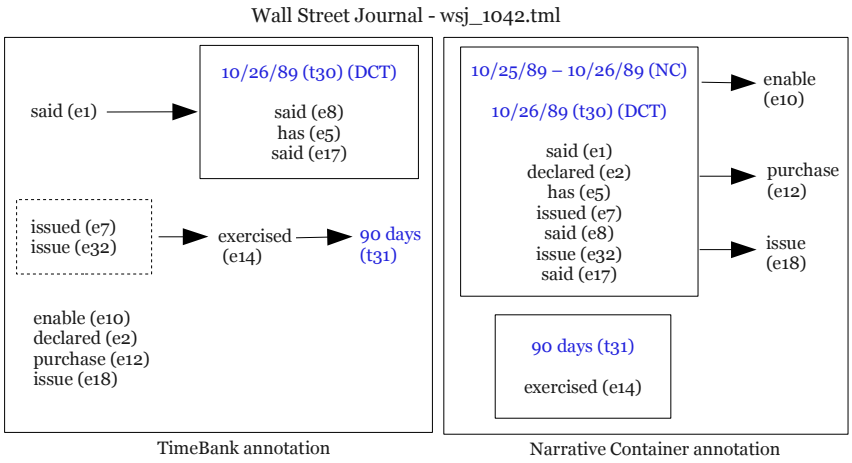

Figure 3: Visual depictions of the TLINK annotations in TimeBank and with the Narrative Container annotations. Solid lines indicate events and times in the box have IS_INCLUDED relationships with the timex at the top, and dotted lines indicate events that were given IDENTITY relationships

When looking at the kappa scores for the temporal anchor, it should be noted that these scores do not always accurately reflect the level of agreement between annotators. Because of the lack of variability, Fleiss' Kappa will interpret any article where an annotator only linked events to the NC received negative agreement scores. These values have been left in the tables as data points, but it should be noted that these annotations are entirely valid—some articles in TimeBank contain no temporal information other than the document creation time (and by extension, the narrative container), making it only natural for the annotators to annotate events only in relation to the narrative container. The average Fleiss' Kappa scores for the temporal anchors was .74, with a maximum of 1 and a minimum of -.04.

6.4 Informativeness in NC Annotation

As we previously described, Narrative Containers are theoretically more informative than Document Creation Times when trying to place unanchored events on a timeline. In practice, they are as informative as we anticipated: compare the visualizations of TLINK annotations between TimeBank and the NC links in Figure 3. These were created from the file wsj_1042.tml, one that had complete agreement between annotators about both the size of the NC (one day before the DCT through the DCT) and all the temporal anchors and temporal relations.

Clearly, the NC task has resulted in a more informative annotation–all the events have at least one constraint, and most have both left and right constraints.

7 Conclusions and Future Work

Narrative Containers, Narrative Times, and Narrative Scopes are important tools for temporal annotation tasks. The analysis provided here clearly shows that annotating with an NC increases informativeness, and that the concept is sufficiently intuitive for it to not add confusion to the already complicated task of temporal annotation. However, the work in this area is far from complete. In the future we intend to study where the left boundary of the NC should be placed for different genres and publication frequencies. Another annotation task must be performed, requiring a more comprehensive TLINK creation guideline, using both event-time and event-event links. Finally, the use of all three concepts for automated annotation tasks should be examined, as they may prove as useful to machines as they are to humans.

Acknowledgements

This work has been supported by NSF grant #0753069 to Co-PI James Pustejovsky. Many thanks to Chiara Graf, Zac Pustejovsky, and Virginia Partridge for their help creating the annotations, and to BJ Harshfield for his R expertise. We would also like to acknowledge Aravind Joshi, Nianwen Xue, and Marc Verhagen for useful input.

References

James Allen. 1984. Towards a general theory of action and time. *Arificial Intelligence*, 23:123–154.

Douglas Biber. 2009. *Register, Genre, and Style*.

J. L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Matthias Gamer, Jim Lemon, and Ian Fellows Puspendra Singh ¡puspendra.pusp22@gmail.com¿, 2010. *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.83.

Nancy Ide and Harry Bunt. 2010. Anatomy of annotation schemes: Mappings to graf. In *In Proceedings 4th Linguistic Annotation Workshop (LAW IV)*.

Nancy Ide. 2007. Annotation science: From theory to practice and use: Data structures for linguistics resources and applications. In *In Proceedings of the Bienniel GLDV Conference*.

David Lee. 2001. Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the bnc jungle. *Language Learning & Technology*, 5(3.3):37–72.

Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2004. The penn discourse treebank. In *In Proceedings of LREC 2004*.

Massimo Poesio. 2004. Discourse annotation and semantic annotation in the gnome corpus. In *In Proceedings of the ACL Workshop on Discourse Annotation*.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The penn discourse treebank 2.0. In *In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.

James Pustejovsky, Patrick Hanks, Roser Saurì, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. 2003. The timebank corpus. In Dawn Archer, Paul Rayson, Andrew Wilson, and Tony McEnery, editors, *Proceedings of the Corpus Linguistics 2003 conference*, pages 647–656, Lancaster University. UCREL.

James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Saurí. 2005. Temporal and event information in natural language text. *Language Resources and Evaluation*, 39:123–164, May.

R Team, 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Roser Saurí, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky, 2006. *TimeML Annotation Guidelines*, version 1.2.1 edition, January.

Carlota Smith. 2003. *Modes of Discourse*. Cambridge University Press, Cambridge, UK.

Marc Vilain, Henry Kautz, and Peter Beek. 1986. Constraint propagation algorithms for temporal reasoning. In *Readings in Qualitative Reasoning about Physical Systems*, pages 377–382. Morgan Kaufmann.