

The Annotation of Temporal Information in Natural Language Sentences

Graham Katz and Fabrizio Arosio

SFB 411
Univertität Tübingen
Nauklerstr. 35
D-72074 Tübingen, Germany
graham.katz@uni-tuebingen.de
fabrizio.arosio@uni-tuebingen.de

Abstract

The aim of this paper is to present a language-neutral, theory-neutral method for annotating sentence-internal temporal relations. The annotation method is simple and can be applied without special training. The annotations are provided with a well-defined model-theoretic interpretation for use in the content-based comparison of annotations. Temporally annotated corpora have a number of applications, in lexicon/induction, translation and linguistic investigation. A searchable multi-language database has already been created.

1 Introduction

In interpreting narratives the most essential information to be extracted is who did what where, when and why, the classic journalistic imperatives. The ‘who’ and ‘what’ information is usually expressed overtly, and this has made it possible to apply empirical techniques to problems in this domain (such as word-sense classification and argument structure mapping). The ‘when’ and ‘where’ information is, however, often left implicit, or, at least, only partially specified, making it difficult to apply such techniques to this domain.

Formal semantic theories of temporal interpretation (e.g. Kamp & Reyle 1993; Ogihara 1996; Abusch 1997) have been quite successful at specifying the contribution that such overt markers as tenses and temporal

adverbials make to the meaning of a sentence or discourse. Investigations into the interpretation of narrative discourse (Lascarides & Asher 1993, Reyle & Rossdeutscher 2000) have, however, shown that very specific lexical information plays an important role in determining temporal interpretation. As of yet it is not clear how this kind of lexical information could be automatically acquired. The most promising avenue for acquiring lexical information appears to be automatic induction from very large annotated corpora (Rooth, et. al. 1998). In order to apply these techniques to the temporal domain it is necessary that the temporal information be made explicit. Our task here is to provide a system of temporal annotation that fulfills this requirement.

The systems for temporal annotation we are familiar with have been concerned either with absolute temporal information (Wiebe, et. al. 1998, Androutsopoulos, Rithie & Thanisch 1997), or with the annotation of overt markers (Setzer & Gaizauskas 2000). Much temporal information, however, is not absolute but relative and not overtly marked but implicit. We are frequently only interested (and only have information about) the order events occurred in. And while there are sometimes overt markers for these temporal relations, the conjunctions *before*, *after* and *when* being the most obvious, usually this kind of relational information is implicit. The examples in (1) illustrate the phenomenon.

- (1) a. John kissed the girl he met at the party.
- b. Leaving the party, John walked home.
- c. He remembered talking to her and asking her for her name.

Although there are no obvious markers for the temporal ordering of the events described in these sentences, native speakers have clear intuitions about what happened when: we know that the kissing took place after the meeting and that the asking was part of the talking. But how do we know this? And – more importantly – how could this information be automatically extracted from these sentences? These are the questions that motivate the development of our annotation system.

We believe that the creation of a large scale treebank annotated with relational temporal information as well as standard morphological and syntactic information will make it possible to investigate these issues productively. The annotated treebank must be large scale for the obvious reason that the application of stochastic methods requires this. It should be syntactically as well as semantically annotated because we consider it quite likely that syntactic as well as lexical information plays a role in temporal interpretation. We don't know *a priori* whether in (1a) it is the lexical relationship between *kiss* and *meet* that is crucial to determining the temporal interpretation, or whether the fact that *meet* is in a subordinate clause – the syntactic relation – also plays a role. To answer these kinds of questions it is necessary to encode the temporal information conveyed by a sentence in a way which makes addressing such questions possible.

What we describe below is a practical system for encoding relational temporal information that is suited to large-scale hand annotation of texts. This system has a number of applications beyond this, both in the domain of cross-linguistic investigation and in empirical NLP.

2 Temporal annotation

The idea of developing a treebank enriched with semantic information is not new. In particular such semantically annotated corpora have been used in research on word sense disambiguation (*wordNet*, *Eagles*, *Simple*) and semantics role interpretation (*Eagles*). The public availability of large syntactically annotated treebanks (*Penn*, *VerbMobil*, *Negra*) makes such work attractive, particularly in light of the success that empirical methods have had (Kilgarriff & Rosenzweig 2000). Traditional semantic representational formalisms such as DRT (Kamp & Reyle 1993)

are ill suited to semantic annotation. Since these formalisms are developed in the service of theories of natural language interpretation, they are – rightly – both highly articulated and highly constrained. In short, they are often too complex and sometimes not expressive enough for the purposes at hand (as the experience of Poesio et. al. (1999) makes clear). Our proposal here is to adopt a radically simplified semantic formalism which, by virtue of its simplicity, is suited the temporal-annotation application.

The temporal interpretation of a sentence, for our purposes, can simply be taken to be the set of temporal relations that a speaker naturally takes to hold among the states and events described by the verbs of the sentence. To put it more formally, we associate with each verb a temporal interval, and concern ourselves with relations among these intervals. Of the interval relations discussed by Allen (1984), we will be concerned with only two: precedence and inclusion. Taking t_{talk} to be the time of talking, t_{ask} to be the time of asking and t_{remember} to be the time of remembering, the temporal interpretation (1c), for example, can be given by the following table:

	t_{talk}	t_{ask}	t_{remember}
t_{talk}			<
t_{ask}	\subseteq		<
t_{remember}			

Such a table, in effect, stores the native speaker's judgement about the most natural temporal interpretation of the sentence.

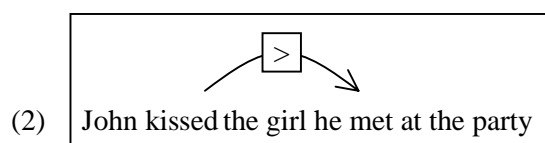
Since our goal was to annotate a large number of sentences with their temporal interpretations, with the goal of examining the interaction between the lexical and syntactic structure, it was imperative that the interpretation be closely tied to its syntactic context. We needed to keep track of both the semantic relations among times referred to by the words in a sentence and the syntactic relations among the words in the sentences that refer to these times, but not much more. By adopting existing technology for syntactic annotation, we were able to do this quite directly, by essentially building the information in this table into the syntax.

2.1 The annotation system

To carry out our temporal annotation, we made use of the *Annotate* tool for syntactic annotation

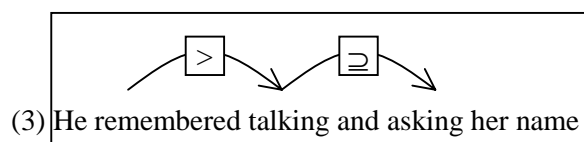
developed in Saarbrücken by Brants and Plaehn (2000). We exploited an aspect of the system originally designed for the annotation of anaphoric relations: the ability to link two arbitrary nodes in a syntactic structure by means of labeled “secondary edges.” This allowed us to add a layer of semantic annotation directly to that of syntactic annotation.

A sentence was temporally annotated by linking the verbs in the sentence via secondary edges labeled with the appropriate temporal relation. As we were initially only concerned with the relations of precedence and inclusion, we only had four labels: “<”, “⊆”, and their duals. Sentence (1a), then, is annotated as in (2).



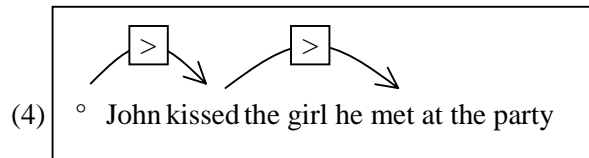
The natural ordering relation between the kissing and the meeting is indicated by the labeled edge. Note that the edge goes from the verb associated with the event that fills the first argument of the relation to the verb associated with the event that fills the second argument of the relation.

The annotation of (1c), which was somewhat more complex, indicates the two relations that hold among the events described by the sentence.

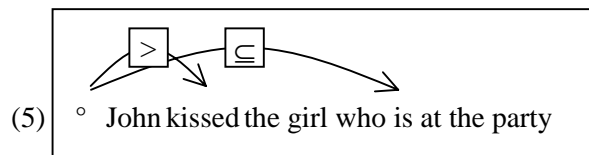


In addition to encoding the relations among the events described in a sentence, we anticipated that it would be useful to encode also the relationship between these events and the time at which the sentence is produced. This is, after all, what tenses usually convey. To encode this temporal indexical information, we introduce into the annotation an explicit representation of the speech time. This is indicated by the “°” symbol, which is automatically prefaced to all sentences prior to annotation.

The complete annotation for sentence (1a), then, is (4).



As we see in (5), this coding scheme enables us to represent the different interpretations that past tensed and present tensed clauses have.



Notice that we do not annotate the tenses themselves directly.

Note that in the case of reported speech, the time associated with the embedding verb plays, for the embedded sentence, much the same role that the speech time plays for the main clause. Formally, in fact, the relational analysis implicit in our notation makes it possible to avoid many of the problems associated with the treatment of these constructions (such as those discussed at length by von Stechow (1995)). We set these issues aside here.

It should be clear that we are not concerned with giving a semantics for temporal markers, but rather with providing a language within which we can describe the temporal information conveyed by natural language sentences. With the addition of temporal indexical annotation, our annotation system gains enough expressive power to account for most of the relational information conveyed by natural language sentences. Left out at this point is temporal-metrical information such as that conveyed by the adverbial “two hours later.”

2.2 Annotation procedure

The annotation procedure itself is quite straightforward. We begin with a syntactically annotated treebank and add the speech time marker to each of the sentences. The annotator then simply marks the temporal relations among verbs and the speech time for each sentence in the corpus. This is accomplished in accordance with the following conventions:

- (i) temporal relations are encoded with directed “secondary edges”;

- (ii) the edge goes from the element that fills the first argument of the relation to the element that fills the second;
- (iii) edge labels indicate the temporal relation that holds;
- (iv) edge labels can be “>”, “<”, “ \subseteq ” and “ \supseteq ”

Annotators are instructed to annotate the sentences as they naturally understand them. When the treebank is made up of a sequence of connected text, the annotators are encouraged to make use of contextual information.

The annotation scheme is simple, explicit and theory neutral. The annotator needs only to exercise his native competence in his language and he doesn't need any special training in temporal semantics or in any specific formal language; in pilot studies we have assembled small temporal annotated databases in few hours. Our current database consists of 300 sentences from six languages.

2.3 Comparing annotations

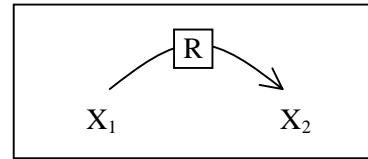
It is well known that hand-annotated corpora are prone to inconsistency (Marcus, Santorini & Marcinkiewicz, 1993) and to that end it is desirable that the corpus be multiply annotated by different annotators and that these annotations be compared. The kind of semantic annotation we are proposing here introduces an additional complexity to inter-annotation comparison, in that the consistency of an annotation is best defined not in formal terms but in semantic terms. Two annotations should be taken to be equivalent, for example, if they express the same meanings, even if they use different sets of labeled edges.

To make explicit what semantic identity is, we provide our annotations with a model theoretic interpretation. The annotations are interpreted with respect to a structure $D, <, \subseteq$, where D is the domain (here the set of verbs tokens in the corpus) and $<$ and \subseteq are binary relations on D . Models for this structure are assignments of pairs of entities in D to $<$ and \subseteq satisfying the following axioms:

- $\forall x, y, z. x < y \ \& \ y < z \rightarrow x < z$
- $\forall x, y, z. x \subseteq y \ \& \ y \subseteq z \rightarrow x \subseteq z$
- $\forall w, x, y, z. x < y \ \& \ z \subseteq x \ \& \ w \subseteq y \rightarrow z < w$
- $\forall w, x, y, z. x < y \ \& \ y < z \ \& \ x \subseteq w \ \& \ z \subseteq w \rightarrow y \subseteq w$

Thus $<$ and \subseteq have the properties one would expect for the precedence and inclusion relation. We are assuming that in the cases of interest verbs refer to simply convex events. Intuitively, the set of verb tokens in the corpus corresponds the set of times at which an event or state of the type indicated by the verb takes place or holds. In our corpus the number of sentences that involved quantified or generic event reference was quite low.

An annotated relation of the following form



is satisfied in a model iff the model assigns $\langle X_1, X_2 \rangle$ to R if R is $<$ or \subseteq , or $\langle X_2, X_1 \rangle$ to R if R is $>$ or \supseteq . A sentence is satisfied by a model iff all relations associated with the sentence are satisfied by the model. Intuitively an annotated is satisfied by a model if the model assigns the appropriate relation to the verbs occurring in the sentence.

There are four semantic relations that can hold among between annotations. These can be defined in model-theoretic terms:

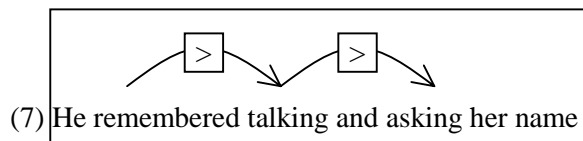
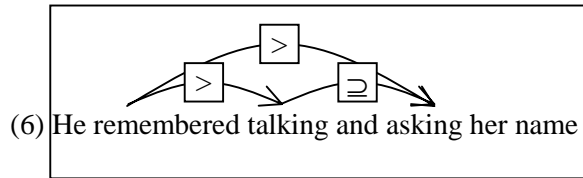
- Annotation A and B are **equivalent** if all models satisfying A satisfy B and all models satisfying B satisfy A .
- Annotation A **subsumes** annotation B iff all models satisfying B satisfy A .
- Annotations A and B are **consistent** iff there are models satisfying both A and B .
- Annotations A and B are **inconsistent** if there are no models satisfying both A and B .

We can also define the minimal model satisfying an annotation in the usual way. We can then compute a distance measure between two annotations by comparing set of models satisfying the annotations. Let M_A be the models satisfying A and M_B be those satisfying B and M_{AB} be those satisfying both (simply shorthand for the intersection of M_A and M_B). Then the distance between A and B can be defined as:

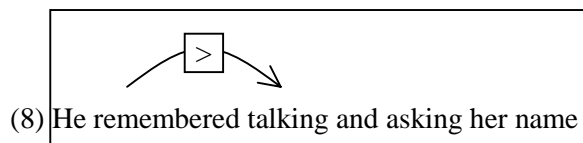
$$d(A, B) = (|M_A - M_{AB}| + |M_B - M_{AB}|) / |M_{AB}|$$

In other words, the distance is the number of relation pairs that are not shared by the annotations normalized by the number that they

do share. We can use this metric to quantify the “goodness” of both annotations and annotators. Consider again (1c). We gave one annotation for this in (3). In (6) and (7) there are two alternative annotations.



As we can compute on the basis of the semantics for the annotations (6) is equivalent with (3) – they are no distance apart, while (7) is inconsistent with (3) – they are infinitely far apart. The annotation (8) is compatible (7) and is a distance of 1 away from it.



As in the case of structural annotation, there are a number of ways of resolving inter-annotator variation. We can chose the most informative annotation as the correct one, or the most general. Or we can combine annotations. The intersection of two compatible annotations gives an equally compatible annotation which contains more information than either of the two alone. We do not, as of yet, have enough data to determine which of these strategies is most effective.

In preliminary work, we had two annotators annotate 50 complex sentences extracted randomly from the BNC. The results were quite encouraging. Although the annotations were identical in only 70% of the cases, the annotations were semantically consistent in 85% of the cases.

3 Applications of temporal annotation

There are any number of applications for a temporally annotated corpus such as that we

have been outlining. Lexicon induction is the most interesting, but, as we indicated at the outset, this is a long-term project, as it requires a significant investment in hand annotation. We hope to get around this problem. But even still, there are a number of other applications which require less extensive corpora, but which are of significant interest. One of these has formed the initial focus of our research, and this is the development of a searchable multilingual database.

3.1 Multilingual database

Our annotation method has been applied to sentences from a variety of languages, creating a searchable multi-language treebank. This database allows us to search for sentences that express a given temporal relation in a language. We have already developed a pilot multilingual database with sentences from the *Verbmobil* database (see an example in fig. 1) and we have developed a query procedure in order to extract relevant information.

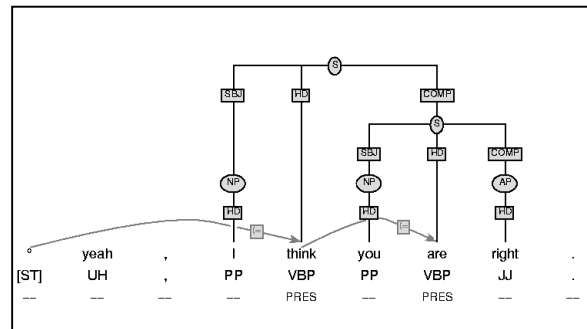
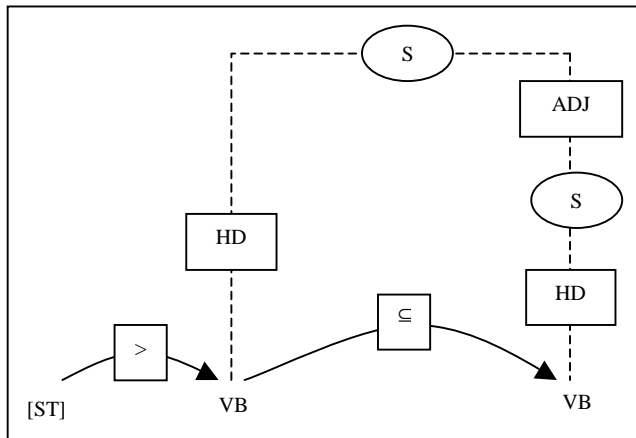


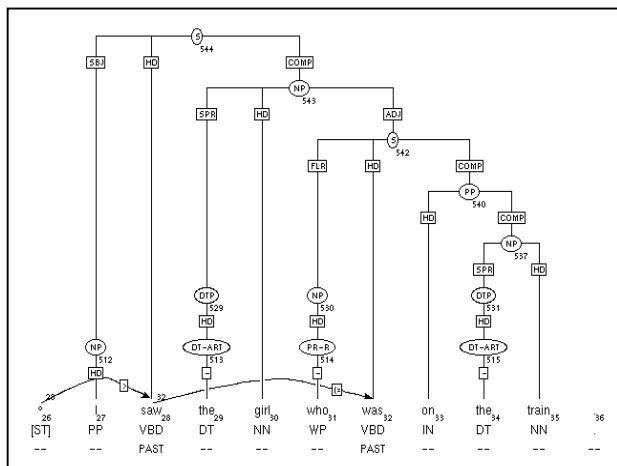
Fig.1 A temporally annotated sentence from the *Verbmobil* English treebank as displayed by @nnotate.

As can be seen, the temporal annotation is entirely independent of the syntactic annotation. In the context of the *Annotate* environment a number of tools have been developed (and are under development) for the querying of structural relations. Since each sentence is stored in the relational database with both syntactic and temporal semantic annotations, it is possible to make use of these querying tools to query on structures, on meanings, and on structures and meanings together. For example a query such as: “Find the sentences containing a relative clause which is interpreted as temporally overlapping the main clause” can be processed. This query is

encoded as a partially specified tree, as indicated below:



In this structure, both the syntactic configuration of the relative clause and the temporal relations between the matrix verb and the speech time and between the matrix verb and the verb occurring in the relative clause are represented. Querying our temporally annotated treebank with this request yields the following result:



The application to cross-linguistic research should be clear. It is now possible to use the annotated tree-bank as an informant by storing the linguistically relevant aspects of the temporal system of a language in a compact searchable database.

3.2 Aid for translation technology

Another potential application of the annotation system is as an aid to automatic translation systems. That the behaviour of tenses differ from language to language makes the translation of tenses difficult. In particular, the application of example-based techniques faces serious difficulties (Arnold, et. al. 1994). Adding the

intended temporal relation to the database of source sentences makes it possible to moderate this problem.

For example in Japanese (9a) is properly translated as (10a) on one reading, where the embedded past tense is translated as a present tense, but as (10b) on the other, where the verb is translated as a past tense.

(9) a. Bernard said that Junko was sick

(10)a. Bernard-wa Junko ga byookida to it-ta
lit: Bernard said Junko is sick

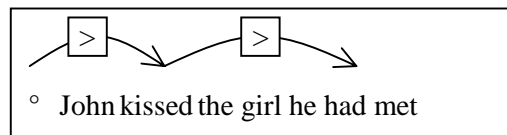
b. Bernard-wa Junko-ga byookidata to it-ta.
lit: Bernard said Junko was sick.

Only the intended reading can distinguish these two translations. If this is encoded as part of the input, we can hope to achieve much more reasonable output.

3.3 Extracting cues for temporal interpretation

While we see this sort of cross-linguistic investigation as of intrinsic interest, our real goal is the investigation of the lexical and grammatical cues for temporal interpretation. As already mentioned, the biggest problem is one of scale. Generating a temporally annotated treebank of the size needed is a serious undertaking.

It would, of course, be of great help to be able to partially automate this task. To that end we are currently engaged in research attempting to use overt cues such as perfect marking and temporal conjunctions such as *before* and *after* to bootstrap our way towards a temporally annotated corpus. Briefly, the idea is to use these overt markers to tag a corpus directly and to use this to generate a table of lexical preferences. So, for example, the sentence (11) can be tagged automatically, because of the presence of the perfect marking.



This automatic tagging will allow us to assemble an initial data set of lexical preferences, such as that that would appear to hold between *kiss* and *meet*. If this initial data is confirmed by comparison with hand-tagged data, we can use

this information to automatically annotate a much larger corpus based on these lexical preferences. It may then be possible to begin to carry out the investigation of cues to temporal interpretation before we have constructed a large hand-coded temporally annotated treebank.

4 Conclusions

We have described a simple and general technique for the annotation of relational temporal information in naturally occurring sentences. This annotation system is intended for use in the large-scale annotation of corpora. The annotations are provided with a model theoretic semantics. This semantics provides the basis for a content-based system for comparing and evaluating inter-annotator agreement.

Temporally annotated corpora such as those we are developing have a number of applications in both corpus-based theoretical work, and in practical applications. Of particular interest to us is the promise such annotated databases bring to the problem of extracting automatically lexical information about stereotypical ordering relations among events that are used by native speakers to interpret sentences temporally. We see this application as the main future research to be carried out.

References

Abusch, Dorit. 1997. Sequence of Tense and Temporal De Re. *Linguistics and Philosophy*, 20: 1-50.

Allen, James F. 1984. A General Model of Action and Time. *Artificial Intelligence* 23, 2.

Androutsopoulos, Ion, Graeme D. Ritchie and Peter Thanisch. 1998. Time, Tense and Aspect in Natural Language Database Interfaces. *Natural Language Engineering*, vol. 4, part 3: 229-276, Cambridge University Press.

Arnold, Doug, Lorna Balkan, R. Lee Humphreys, Siety Meijer and Louisa Sadler. 1994. *Machine Translation: an introductory guide*. Blackwells/NCC, London.

Brants, Thorsten and Oliver Plaehn. 2000. Interactive Corpus Annotation. In *Second International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece.

Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht, Holland.

Kilgarriff, Adam and Joseph Rosenzweig. 2000. Framework and Results for English SENSEVAL. *Special Issue on SENSEVAL: Computers and the Humanities*, 34 (1-2): 15-48.

Lascarides, Alex and Nicholas Asher. 1993. A Semantics and Pragmatics for the Pluperfect. In *Proceedings of the Sixth European Chapter of the Association of Computational Linguistics*, Utrecht.

Marcus, Mitchell P., Beatrice Santorini and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational linguistics*, 19: 313-330.

Ogihara, Toshiyuki. 1996. *Tense, Scope and Attitude Ascription*. Kluwer, Dordrecht, Holland.

Poesio, Massimo, Renate Henschel, Janet Hitzeman, Rodger Kibble, Shane Montague, and Kees van Deemter. 1999. "Towards An Annotation Scheme For Noun Phrase Generation", *Proc. of the EACL Workshop on Linguistically Interpreted Corpora*. Bergen.

Reyle, Uwe and Antje Rossdeutscher. 2000. Understanding very short stories, ms. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Rooth, Mats, Stefan Riezler, Detlef Prescher, Sabine Schulte im Walde, Glenn Carroll and Franz Beil. 1998. *Inducing Lexicons with the EM Algorithm*. *AIMS Report 4(3)*. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

- Setzer, Andrea and Robert Gaizauskas. 2000. Annotating Events and Temporal Information in Newswire Texts. In *Second International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece.
- von Stechow, Arnim. 1995. The Proper Treatment of Tense. In *Proceedings of Semantics and Linguistic Theory V*. 362-386. CLC Publications. Cornell University. Ithaca, NY.
- Vazov, Nikolay and Guy Lapalme. 2000. Identification of Temporal Structure in French. *Proceeding of the Workshop of the 7th International Conference on Principles of Knowledge Representation and Reasoning, Breckenridge, Colorado*.
- Wiebe, Janyce, Tom O'Hara, Kenneth McKeever, and Thorsten Öhrström-Sandgren. 1998. An Empirical Approach to Temporal Reference Resolution. *Journal of Artificial Intelligence Research*, 9: 247-293.