

Spatiotemporal Annotation on Top of an Existing Treebank

Ineke Schuurman
K.U.Leuven
Centrum voor Computerlinguïstiek

Abstract

In this paper we discuss a spatiotemporal layer of annotation to be added to an existing (syntactic) treebank. Although our system, called MiniSTEx, was developed for Dutch, it will also work for other EU-languages. This may, however, ask for some adaptations to the database which is the centre of our system. Next to adaptations for other languages, we may need adaptations for specific situations, even when only one language is covered.

1 Introduction

Although it is not very common to have one layer of annotation handling at the same time both the spatial and temporal characteristics of a text, we decided to investigate the pros and cons of such an approach. It turned out that both layers can be handled in very much the same way, especially when geospatial annotation is involved, a subtype of spatial annotation.

The aim of our approach is to locate eventualities on a time-axis and to disambiguate (geo)spatial information such that such entities can be located on a map.

The system is largely database-driven, meaning that among others a lot of spatiotemporal world knowledge is contained in a spatiotemporal database. A first version of MiniSTEx was partially developed within D-Coi, one of the projects in the STEVIN programme (2005-2010).¹ In this programme a whole series of projects together are to constitute a large Corpus of Written Dutch (500M words). This whole corpus will be parsed with the Alpino-parser, cf. Van der Beek et al. (2005) and a subcorpus of 1M words will be manually corrected. That 1M will also be annotated for named entity recognition, coreference resolution, semantic roles, and spatiotemporal semantics. The latter will be done using MiniSTEx.

In the remainder of this paper we will first discuss some characteristic properties of MiniSTEx in section 2, amongst others addressing the possibilities of using this approach for other languages. In section (3) we will address the way MiniSTEx relates to other approaches.

¹Funded by the Dutch and Flemish Governments (<http://taaluniversum.org/taal/technologie/stevin/>).

2 Characteristics of MiniSTEx

MiniSTEx relies on information provided by a parser (like Alpino): which verbal forms do belong together (in Dutch they can be separated by many other elements), which are the clauses, NPs, and PPs, and what are their roles (subject, temporal modifier). This information largely simplifies its implementation and improves its accuracy. It is expected that Named Entity Classification will also prove to be helpful in order to detect whether an expression like *Brussel* functions as the name of a region, a town, the capital of Europe, a sporting club,...

In MiniSTEx we handle spatial and temporal expressions in combination because there are striking similarities: the temporal and spatial NPs and PPs functioning as anchors, the relationships à la Allen (1984) that can be formulated for spatial expressions as well although they are not just two-dimensional.

The MiniSTEx database contains lots of spatial and temporal expressions (from single tokens to full expressions). Among them many temporal and (geo)spatial homonyms, and their disambiguation depends largely on the factors `background` and `intended audience`, cf section 2.3.

For both temporal and (geo)spatial annotation it seems that not the language is the most relevant issue when using the system in another situation, for example when using another language, but rather this background and intended audience. Whether a Dutch text originates from the Netherlands or Flanders² does influence the interpretation of this text when relevant homonyms do occur.

A last reason pro is that for quite a number of expressions both temporal and geospatial characteristics are relevant, cf section 2.6 on `geotemporal expressions`.

In MiniSTEx, as it is used in the Corpus of Written Dutch, only the geospatial component will be used, not the full spatial one because of the new characteristic interpretation (describing the *point of view*) that comes into play.

- (1) Jan staat achter de auto
Jan stands behind the car
Jan is standing behind the car

Is Jan standing at the rear end of the car, or in a position in which the car is in-between the speaker and Jan? In the first case we have an *intrinsic* interpretation (i.e. taking into account the coordinate system of the car), in the latter an *extrinsic* interpretation (i.e. taking into account the coordinate system of the speaker), cf. Helbig (2006). Note that in the latter case (extrinsic interpretation), the point of view can also be that of a third person (the addressee).

Because of this complication, we decided to concentrate on geospatial annotation for the moment.³ We will try to tackle the full spatial component a later version.

²In both Dutch is the official language.

³There still may be a few occasions where *point of view* will become relevant in the STEVIN

2.1 Time and place of location, eventuality and perspective

In our approach we distinguish between *time of location* (t_l), *time of perspective* (t_p), and *time of eventuality* (t_i) on the one hand,⁴ and *place of location* (p_l), *place of perspective* (p_p) and *place of eventuality* (p_i) on the other hand. Most of the information used to determine especially t_l , p_l and p_p ⁵ in a text is contained in a database. The approach was designed for Dutch, keeping in mind that it should be 'Euroversal', i.e. useful for all the official languages of the European Union.⁶

It turns out that especially the tense and aspect characteristics coming with verbs really differ per language (the properties of the 'onvoltooid tegenwoordige tijd' in Dutch and corresponding 'simple present' in English are for example not identical). Things even get more complicated by a 'shift of perspective' that may occur.

2.2 Shift of perspective

Suppose in the following the Document Creation Time (DCT) to be 2003-03-23:

- (2) Jan zei eergisteren dat hij zijn moeder morgen zou bezoeken.
Jan said the day before yesterday that he his mother tomorrow would visit.

The day before yesterday Jan said he would visit his mother tomorrow.

- (3) Jan zei eergisteren: "Morgen zal ik mijn moeder bezoeken".
Jan said the day before yesterday: "Tomorrow will I my mother visit".

The day before yesterday Jan said: "Tomorrow I'll visit my mother".

From (2) it can be deduced that Jan intended to visit his mother the 24th of March, but from (3) that the visit would be the 22nd.

In (2) the interpretation (visit the 24nd) would be the same when *eergisteren* (the day before yesterday) is replaced by *gisteren* (yesterday). Doing so in (3) does change the intended date into the 23rd.

This so-called shift of perspective is in this case triggered by the combination of a colon and opening marks, the new t_p being *eergisteren*. The closing marks

treebank. In that case the tag `interp="unknown"` is used, instead of `interp="intrinsic"` or `interp="extrinsic"`.

⁴Inspired by DRT, and especially Van Eynde (1999). Tense and aspect properties, for example, of eventualities are expressed as relations between t_i , t_l and t_p . Note that we simplified things (not using for example *time of utterance* but only *time of perspective*).

⁵ t_p (often DCT) is likely to be mentioned in the metadata coming with a text (like the date of publication of an item in a newspaper).

⁶In fact, as shown in section 2.3, 'language' is not really the crucial notion when dealing with texts in several languages, whereas 'intended audience' is. Which date, for example, is to be associated with *Thanksgiving*?

at the end of the sentence trigger a new shift of perspective, back to the original one (i.e. DCT). Also with respect to (geo)spatial phenomena a shift of perspective may occur when deictic expressions are used. Suppose the Document Creation Location (DCL) to be EU::BE::Vlaanderen.

- (4) Jan zei gisteren in Amsterdam dat hij hier graag woonde.
Jan said yesterday in Amsterdam that he here willingly lived.
Yesterday in Amsterdam Jan said that he liked living here.
- (5) Jan zei gisteren in Amsterdam: “Ik woon hier graag”.
Jan said yesterday in Amsterdam: “I live here willingly”.
Yesterday in Amsterdam Jan said: “I like living here”.

In sentence (4) Jan likes living in Flanders, whereas in (5) he likes living in Amsterdam.

Note that the verbal strings in the odd and even sentences are not the same. In (2, 4) a past tense has been used, and in (3, 5) the corresponding present tense. In order to get the correct temporal interpretation, in both (2) and (4) the string is to be converted.

2.3 Background and intended audience

The vital property of a text seems having an intended audience: a medical text written for British GPs is not likely to be fully understandable for either aerospace engineers, teachers or linguists. Nor for Norwegian GPs. And in a Belgian newspaper a reference to *Haren* will be to the *Haren* in the Brussels Capital Region, not to the ones in the Netherlands (or Germany). Otherwise it should have been mentioned explicitly. This is the case because every speaker (author) will apply conversational maxims as formulated by Grice (1975), often paraphrased as “Don’t say too much and don’t say too little.” without as much as thinking.

At the moment the database which is one of the core components of MiniSTEx is centered around The Netherlands and Flanders when it comes to

- Culture specific characteristics (e.g. religious holidays)
- Location specific characteristics (e.g. northern vs southern hemisphere)
- History specific characteristics (e.g. beginning vs end of WWII)

Therefore an adaption of the database is necessary when the approach is used for another language/country. This sounds worse as it is, because a lot of information is already available, cf. the *background* in table 1, which is to be related with the tables 2 and 3: *Christmas* in a country with a christian tradition (cf. tabel 3) will be on the 25th of December (dbid 1310 in table 1) no matter whether the text refers to Xmas, Christmas, Kerst, or Weihnachten.

In MiniSTEx, we assume that a text always provides the (intended) reader with all information necessary to understand this text. If not, i.e. when human readers

Table 1: Concepts (simplified)

concept	dbid	background	tag	rank	parts
Haren	142	BE::BR	<geo type="place" val="EU::BE::BR:::Haren />	2	
Haren	143	NL::GR	<geo type="place" val="EU::NL:::GR::Haren />	1	
Haren	144	NL::NB	<geo type="place" val="EU::NL:::NB::Haren />	3	
vaderdag (father's day)	1500	EU::(NL UK FR)	<temp type="cal" form="XXX-06-D07,15..21"/>		
vaderdag	1501	EU::BE	<temp type="cal" form="XXX-06-D07,08..14"/>		
vaderdag	1502	BE::AN	<temp type="cal" form="XXX-03-19"/>		
Thanksgiving	210074	NA::VS	<temp type="cal" form="XXX-11-D04,22..28"/>	1	
Thanksgiving	210075	NA::CA	<temp type="cal" form="XXX-10-D01,08..14"/>	2	
avond (evening)	1302	DU	<temp type="clock" form="T18/24"/>		
nacht (night)	1303	DU	<temp type="clock" form="T22/06"/>		
middag (afternoon)	1291	EU::NL	<temp type="clock" val="T12/18"		
namiddag (afternoon)	1292	EU::NL	<temp type="clock" val="T16/18"		
namiddag	1293	EU::BE	<temp type="clock" val="T12/18"		
Kerst (Christmas)	1310	chr	<temp type="cal" form="XXX-12-25"/>		
Kerst	1311	orth	<temp type="cal" form="XXX-01-07"/>		
winter (winter)	100562	north	<temp type="cal" form="XXX-12/02"/>		
Rio de Janeiro	101	BR::RJ	<geo type="place" val="SA::BR::RJ:::Rio de Janeiro"/>	1	
Rio de Janeiro	141	SA::BR	<geo type="region" val="SA::BR::Rio de Janeiro"/>	2	
UNCED	500010	UN conf	<stex> <temp type="cal" val="1992-06-3/14"/> </stex>		101

belonging to the intended audience fail to understand a text, a system can neither be blamed for failing. MiniSTEx handles texts by using the background and world knowledge the intended audience is supposed to have, cf. Schuurman (2007b).

2.4 Determination of intended audience and spatiotemporal background knowledge

As far as the intended audience is concerned, note that our approach is not designed to primarily deal with web pages, but rather with digital archives (broadcasting companies, news agencies), corpora and the like. Of the latter kind of resources the background is usually known. This is very important as it helps us a lot in determining both the intended audience and the spatiotemporal background knowledge this audience may be supposed to possess.

In case the background is unknown, a first clue is provided by the language used: a text written in Dutch is usually meant for Dutch and/or Flemish readers. For texts in English, the intended audience is more difficult to determine as these are either meant for a British (or an American, Australian, Canadian,...) audience, i.e. the text has a national scope, or for "the rest of the world" (global scope).

But, especially for the smaller languages, data with respect to the intended audience can be derived even when details about the source of the text are unknown. However, for known resources many more details are available, making use of the spatiotemporal data associated with the title (like *De Morgen*, *Daily Telegraph*, *Boston Globe*, *www.vlaanderen.be* etc.), cf. table 2.⁷

Table 2: Background-doc

concept	dbid	status	geo	trad	cal	lang	scope
De Morgen	220000	newspaper	Brussel			Dutch	national
De Telegraaf	220003	newspaper	Amsterdam				national
Ref. Dagblad	220009	newspaper	Apeldoorn	orth-ref			
Vloverheid	230000	web	Brussel			Dutch	regional
Vloverheid	230000	web	Brussel			English	global

Other information relevant for determining the intended audience are `tradition` (Christian, Islamic, Jewish, Eastern Orthodox, ...), and `calendar`: (Gregorian, Hebrew, Hindu, ...).

Note that sometimes a ranking is added. When the background of the intended audience is not decisive (for example when in *De Morgen Rio de Janeiro* is mentioned without further specifications) the reference with the lowest rank will be taken (in this case the town, not the region). It is especially the ranking that should be adapted when the database is used for another intended audience.

Table 3: Background-geo

concept	dbid	status	trad	cal	hem	UTC ⁸	lang	partof	division
Spanje (Spain)	109	cntry	chr	Greg	north	+1	ES	EU	2=region, 3=province
Nederland (The Netherlands)	146	cntry	chr	Greg	north	+1	DU	EU	2=—, 3=province
België (Belgium)	137	cntry	chr	Greg	north	+1	DU, FR, GE	EU	2=region, 3=province
VS (US)	199	cntry	chr	Greg	north	-(5/10)	EN, ES	NA	2=state, 3=county
Vlaanderen (Flanders)	102	region					DU	BE	

The MiniSTEx database consists of more tables than presented in this paper, and all tables are linked: in table 2 the `geo`-column refers to geospatial entities. Via table 1 these entities can be linked with entities in table 3. This table defines the spatiotemporal backgroundknowledge associated with a geospatial entity, unless it is superseded by information in table 2 itself. These columns in table 2 are only filled out in case they contain information that overrules the general information. So, *Reformatorisches Dagblad* is said to belong to the orthodox-reformatoric tradition instead of the more general christian tradition. For *De Morgen* and *De*

⁷For convenience of the reader most tables as they are presented here contain the concepts. This is only for matter of presentation. In reality the only column all tables contain is the one with the `dbid`. The real tables also contain more columns, i.e. more types of data.

Table 4: Name-variants of concepts

concept	dbid	NL	EN	DE	FR
Den Haag	135	's Gravenhage hofstad	The Hague	Den Haag	La Haye
Apeldoorn	145		hofstad		
Rijn	510		Rhine	Rhein	Rhin
vaderdag (father's day)	1500		father's day	Vatertag	Fêtes des Pères
vaderdag	1501		father's day	Vatertag	Fêtes des Pères
vaderdag	1502		father's day	Vatertag	Fêtes des Pères

Telegraaf the values for `geo` and `trad` are those of *Brussel* and *Amsterdam* respectively. For *De Telegraaf* `lang` is also that of *Amsterdam*, whereas for *De Morgen* the values for *Brussel* are overruled by the statement that only *Dutch* is used.

2.5 Role of language

In table 4 other names for concepts are presented, sometimes within the same language. Note that *Rijn* refers to the same concept, whether it is called *Rijn*, *Rhein* or *Rhine*. This is clearly not the case for *vaderdag* (father's day). But the crux is not the language used, but the background related to the various concepts. When a Flemish woman says “*ik had op vaderdag een ongeluk*” (I had an accident on father's day) she had an accident on the second Sunday in June, not on the third (father's day in the UK).⁹ When translating this sentence into another language, like English, such 'details' often get lost.¹⁰ But also a Dutch person is inclined to think that this accident took place the third Sunday in June. So it is a matter of background,¹¹ not of language.

There is also a rather small table with language-sensitive concepts (table 5). Above we have explained that in general all and every of the background factors is of greater importance than the language.¹² There are just a few exceptions, in which a language only allows one value to be associated with a concept, while in other languages these concepts are associated with other values. An example that comes to mind is *avond* – *evening* vs *nacht* – *night*.¹³

⁹See table 1.

¹⁰Especially in machine translation.

¹¹And when this Flemish woman lived in Antwerp, a province of Flanders, the accident would have been in March!

¹²Although the language is important in determining the intended audience.

¹³In these cases the boolean feature `noise` would be used as the hours mentioned are just roughly correct, cf. Schuurman and Monachesi (2006)

Table 5: Language-sensitive concepts

concept	dbid	language	tag
avond (evening)	1302	Dutch	<temp type="cal" val="T18/24">
nacht (night)	1303	Dutch	<temp type="cal" val="T22/06">
evening	1308	English	<temp type="cal" val="T18/21">
night	1309	English	<temp type="cal" val="T21/06">

2.6 Geotemporal annotation

Quite often a town, date or incident is associated with both a geographical entity and a date. An example is *9/11*, which for the Flemish/Dutch audience is associated with the 11th of september 2001 (although in Dutch the order day-month-year is used (9/11 = 9th of november), and with the town of New York. Another one is the hurricane *Katrina*, associated with 2005-08-29, New Orleans. In the example in section (2.7) *de Rio-conferentie* (the Rio-conference) is related with its official name (UNCED), and the year 1992 (known as it was mentioned earlier in the test). Once more, the associations made are biased towards a Flemish/Dutch audience. People from Louisiana might disagree with associating Katrina primarily with New Orleans. In fact, also in our database Katrina will be associated with the Gulf Coast as well, albeit with a ranking that favours the interpretation mentioned above.

2.7 A simplified example

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<alpino_de version="1.1" status="simplified">
<stex version="1.0" status="light">
  <node cat="top" rel="top">
    <node cat="smain" rel="-">
      <node cat="np" rel="su">
        <stex id="1" conc="UNCED">
          <temp id="2" role="tl" type="cal" val="1992"/>
          <geo id="3" type="place" val="BR>Rio de Janeiro"/>
        </stex>
        <node pos="det" rel="det" word="De"/>
        <node pos="noun" rel="hd" word="Rio-conferentie"/>
      </node>
      <node pos="verb" rel="hd" word="legde"/>
      <node pos="det" rel="obj1" word="dit"/>
      <node pos="part" rel="svp" word="vast"/>
    </node>
    <temp id="4" form="at(ti,tl)" role="ti" ta="ovt" type="eventuality" val="1992"/>
    <node pos="punct" rel="-" word=":"/>
  </node>
</stex>
</alpino_de>
<sentence>De Rio-conferentie legde dit vast.</sentence>
```

</stex>
</alpino_ds>

Example: De Rio-conferentie legde dit vast (The Rio-conference established this)

3 Other approaches

At the moment MiniSTEx handles geospatial and temporal expressions, as well as geotemporal expressions, i.e. expressions associated with a combination of geospatial and temporal properties. How does it relate to other approaches?

3.1 Relation to TimeML

MiniSTEx covers more or less the same (temporal) phenomena as TimeML (cf (Sauri et al., 2006), (Pustejovsky et al., 2007)). There are, however, a few issues when adopting (and adapting) a scheme like TimeML for the Flemish/Dutch STEVIN programme:

1. we want to make use of information available through other layers like Syntactic Analysis (SA) and Part of Speech tagging (PoS) when annotating the sentences (semi-)automatically,
2. the semantic foundation should provide a basis for a more extended treatment of Tense & Aspect,
3. the proposed analysis in TimeML wrt temporal expressions is sometimes not precise enough in order to allow for spatiotemporal reasoning.

With respect to point 1, TimeML, like most annotation schemes around, seems to start from scratch, not really taking into account other layers of annotation. For MiniSTEx, we wanted to make use of all information available (such as part of speech, syntactic analysis, named entity recognition, . . .).

Regarding point 2, in TimeML, *states* are considered particular types of *events*, which is incorrect from a more semantic point of view: they are at the same level, and they both are subtypes of *eventualities*.¹⁴ We therefore will not use this part of TimeML, although we do see the merits of a characterization of the verbs involved in an eventuality in order to rate the relevance of this eventuality (cf. our common tag, (Schuurman, 2007a)).

An elaborated tense and aspect component is often not considered necessary for applications, especially when the annotated corpus consists of news items (Setzer, 2001). We nevertheless want to make use of a more elaborate theory of tense and

¹⁴The problem seems to be the various readings of the notion *event*. Within the temporal domain however, event and state have a fixed meaning, cf. Mani, Pustejovsky, and Gaizauskas (2005), p. 491, suggesting that in TimeML *eventuality* might have been used instead of *event*.

aspect than the one used in TimeML as we consider this of importance for reasoning in texts (like the properties of `events` vs `states` when dealing with texts instead of sentences in isolation), especially as we are to annotate all kinds of texts, e.g. news items, fiction, wikipedia.¹⁵ We therefore want to merge TimeML with theories like Discourse Representation Theory (DRT) (Kamp and Reyle, 1993). Although within the current version of MiniSTEx only *events* and *states* are used, nothing in its design prohibits an extension with *processes*.

The last point refers to the fact that no months (or the like) are associated with expressions like *summer*, *autumn*. This way, the system is not able to locate such periods on a time-axis. In MiniSTEx, expressions like these are associated with specific months, taking into account the location (like northern vs southern hemisphere), cf. section 2.3. Sometimes the interpretation is also too strict: “*two weeks ago*” is interpreted as “*exactly two weeks ago*”, i.e. the sloppy way people express themselves (not only in informal discussions, but also in editorial items in newspapers and the like) is not at all taken into account. We introduced the features `noise` and `soft` to deal with this to some extent (Schuurman and Monachesi, 2006).

3.2 Relation to more linguistically oriented approaches

The more linguistically oriented part of our approach is inspired by the way tense & aspect is handled in (Allen, 1984), DRT (Kamp and Reyle, 1993), HPSG (especially Van Eynde (1999)).

We are using notions like *time of eventuality*, *time of perspective* and *time of location* (cf. Reichenbach (1947), DRT), the relationships between temporal intervals (cf. Allen), as well as the spatial counterparts of all these. It will be clear that we do not follow the originals to the dot.

3.3 Relation to topological approaches

Geospatial annotation as such is far less widespread and standardized than temporal annotation.¹⁶ The subtask of disambiguation, however, is also a subject in geographic information extraction, cf. Leidner (2006) and Volz, Kleb, and Mueller (2007).¹⁷ The approaches described in their papers influenced our work (cf. especially Schuurman (2007b)).

In most of the approaches cited above a town like Bergen (Norway) would be represented as `Bergen>Hordaland>Norway`, ‘`x>y`’ meaning ‘`x` part-of `y`’. We are using another notation, partly because we want the same order (from larger entities to smaller ones) as in the temporal part of the system, and we also want to

¹⁵We need to be able to deal with a phenomenon like *shift of perspective* which often occurs in fiction, but also is to be expected in newspapers (a.o. interviews).

¹⁶We are not aware of work functioning as a standard for such annotation, cf TimeML for temporal annotation.

¹⁷Note that we annotate more phenomena than covered in these papers, cf. section 3.3.

express the continent and such entities as both regions and provinces (as is relevant in for example Belgium). We also prefer to have a fixed number of fields, some of which may remain empty.

- (6) place="continent::country::region::province::town"
- (7) place="EU::NL::-::GR::Groningen"
- (8) place="EU::BE::FL::VB::Tienen"
- (9) place="EU::BE::BR::-::Brussel"

Whether a particular field represents a county or a province, or a state or a region is reflected in the `Background-geo` component of our database. The fields are separated by '::', 'x::y' meaning 'y part-of x'.

4 Conclusion and plans for the future

MiniSTEx is an annotation scheme offering new possibilities to anchor both temporal and/or geospatial expressions that can in principle be used for all EU-languages. It also allows for extensions like handling of processes, next to events and states. In the future a full spatial component will be added. A previous version has been used to annotate a small corpus by hand, (semi-)automatic annotation of more material is planned for the near future.

References

- Allen, F. 1984. Towards a General Theory of Action and Time. *Artificial Intelligence*, 23:123–154.
- Grice, H.P. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Speech Acts*, volume 3 of *Syntax and Semantics*. Academic Press, New York, pages 43–58.
- Helbig, Hermann. 2006. *Knowledge Representation and the Semantics of Natural Language*. Springer.
- Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, volume 42 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers, Dordrecht, Boston, London.
- Leidner, Jochen. 2006. Toponym Resolution: A First Large-Scale Comparative Evaluation. Technical report, School of Informatics, University of Edinburgh, July.
- Mani, Inderjeet, James Pustejovsky, and Rob Gaizauskas, editors. 2005. *The Language of Time. A Reader*. Oxford University Press.

- Pustejovsky, J., R. Knippen, J. Littman, and R. Saurí. 2007. Temporal and Event Information in Natural Language Text. In H. Bunt and R. Muskens, editors, *Computing Meaning*, Language Resources and Evaluation 39. Springer, Dordrecht, pages 123–164.
- Reichenbach, H. 1947. *Elements of Symbolic Logic*. The Free Press, New York. reprint (1966).
- Sauri, Roser, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky, 2006. *TimeML Annotation Guidelines, version 1.2.1*.
- Schuurman, Ineke. 2007a. MiniSTEx Protocol, version 0.2. KULeuven 2007, March.
- Schuurman, Ineke. 2007b. Which New York, which Monday? The role of background knowledge and intended audience in automatic disambiguation of spatiotemporal expressions. In *Proceedings of CLIN 17*.
- Schuurman, Ineke and Paola Monachesi. 2006. The contours of a semantic annotation scheme for Dutch. In *Proceedings of CLIN 2005*.
- Setzer, Andrea. 2001. *Temporal Information in Newswire Articles: An Annotation Scheme and Corpus Study*. Ph.D. thesis, University of Sheffield.
- Van der Beek, L., G. Bouma, J. Daciuk, T. Gaustad, R. Malouf, M.-J. Nederhof, G. Van Noord, R. Prins, and B. Villada. 2005. Algorithms for Linguistic Processing. Technical report, RijksUniversiteit Groningen.
- Van Eynde, Frank. 1999. *Extracting temporal information from texts*. Unpublished.
- Volz, R., J. Kleb, and W. Mueller. 2007. Towards ontology-based disambiguation of geographical identifiers. In *WWW2007*, Banff, Canada, May 8-12.