An Annotation Scheme for Reichenbach's Verbal Tense Structure

Leon Derczynski and Robert Gaizauskas

University of Sheffield

11 January 2011

The University Of Sheffield.

Outline

The
University
Of
Sheffield.

Introduction to RTMML

We present RTMML - a markup schema for a model of tense in language - that helps us better automatically process temporal information in text.

Temporal Annotation

- Relating time is an essential part of discourse.
- To automatically process time information in documents, we can record data with a temporal annotation.
- Data annotated might include: times, events, the relations between them.
- Existing standards - TIMEX, TimeML
- Existing corpora - TimeBank, AQUAINT TimeML corpus, WikiWars
- We focus on two sub-tasks: relating events, and processing temporal expressions.

The
University
Of
Sheffield.

Relating events

- Temporal expressions and events may be seen as **intervals**.
- Defining relations between two intervals permits temporal ordering.
- Human readers can usually temporally relate events in a discourse using cues.
- Automatic labelling of temporal links is a difficult research problem.

The
University
Of
Sheffield.

Temporal expressions

- `Wednesday; December 4, 1997; for two weeks`
- A temporal expression is text that describes a time or period.
- The process of mapping temporal expressions to an absolute calendar is **normalisation**.
- `Wednesday` $\Rightarrow$ `2011/01/12 T 00:00:00 - 2011/01/12 T 23:59:59`

The
University
Of
Sheffield.

Outline

1 Introduction

2 Reichenbach's Model of Tense

3 RTMML

4 Using the markup

5 Conclusion

The
University
Of
Sheffield.

The Model

- When describing an event, we can call the time of the event E.
- The event is described at speech time S.
- *"I will walk the dog."* $\Rightarrow S < E$
- Reichenbach introduces a **reference point**, an abstract time from which events are viewed.
- *"I ran home"* $\Rightarrow E < S$
- *"I had run home"* $\Rightarrow E < R < S$

The
University
Of
Sheffield.

Special properties of the reference point

- **permanence**: when sentences or clauses are combined, grammatical rules constrain the set of available tenses. These rules work such that R has the same position in all cases.
- *"I had run home when I heard the noise"*
- **positional use**: when a time is found in the same clause as a verbal event, R is bound to the time.
- *"It was six o'clock and John had prepared"* – the preparation, E, occurs before six o'clock, R.

The
University
Of
Sheffield.

Motivation

- At least 10% of normalisation errors are due to incorrect R[1].

- There is a *"need to develop sophisticated methods for temporal focus tracking if we are to extend current time-stamping technologies"*[2].

- If can relate the S and R of multiple event verbs, we can reason about and label the relation between event times.

- Tracking S, E and R according to Reichenbach's model helps generate correct tense and aspect for NLG[3].

---

[1]Mani & Wilson, 2000
[2]Mazur & Dale, 2010
[3]Elson & McKeown, 2010

The University Of Sheffield.

Outline

The
University
Of
Sheffield.

What to annotate?

- RTMML is intended to describe Reichenbach's verbal event structure.

- First primitive: **verb groups** describing an event (*had snowed*, *is running*, *will have given*).

- Second primitive: text describing a **time** (*next two weeks*, *last April*).

- We are more concerned with the relations between times than the absolute position of the times.

The
University
Of
Sheffield.

Annotation Schema

- RTMML is an XML-based standoff annotation standard that may be standalone or integrated with TimeML.

- `<doc>` defines document creation/utterance time, which is also the default speech time.

- `<verb>` denotes verb groups, including tense, and relations between S, E and R. These points may be expressed in terms of other times in the annotation or new labels.

- `<timerefx>` denotes a time-referring expression, which can be optionally specified in TIMEX3 format.

The
University
Of
Sheffield.

Example markup

```
<rtmml>
Yesterday, John ate well.
<seg type="token" />
<doc time="now" />
<timerefx xml:id="t1" target="#token0" />
<verb xml:id="v1" target="#token3"
view="simple" tense="past"
sr=">" er="=" se=">"
r="t1" s="doc" />
</rtmml>
```

RTMML links

- Although we can relate primitives with the current syntax, three common types of link are catered for with `<rtmlink>`:

- POSITIONS, describing positional use of the reference point;

- SAME_TIMEFRAME, where events share a commonly situated reference point;

- REPORTS, for reported speech.

```
<rtmlink xml:id="l1" type="POSITIONS">
<link source="#t1" />
<link target="#v1" />
</rtmlink>
```

The
University
Of
Sheffield.

Outline

The
University
Of
Sheffield.

Example - Relations between events

Saddam **appeared** *to accept a border demarcation treaty he* **had rejected** *in peace talks.*

- Verb 1 - *appeared* - is simple past. $E_{v1} < S_{v1}, E_{v1} = R_{v1}$
- `<verb xml:id="v1" view="simple" tense="past" />`
- Verb 2 - *had rejected* - is anterior past. $E_{v2} < R_{v2}, R_{v2} < S_{v2}$
- `<verb xml:id="v2" view="anterior" tense="past" />`
- These occur in the same timeframe, sharing a reference point. $R_{v1} = R_{v2}$
- `<rtmlink type="SAME_TIMEFRAME">`
  `<link target="v1" />`
  `<link target="v2" />`
  `</rtmlink>`
- We can infer from this that $E_{v2} < E_{v1}$.

The
University
Of
Sheffield.

Context

- Speech and reference points tend to persist in discourse - until changed by a shift in *context*.

- *Emmanuel had said* "This will explode!", *but later changed his mind.*

- Positions of S and R persist throughout each context.

- Changing the $S - R$ relationship inside a context leads to awkward text;
  "By the time I ran, John will have arrived".

The
University
Of
Sheffield.

Outline

The
University
Of
Sheffield.

Conclusion

We have presented Reichenbach's tense model, a markup for representing its time points, and shown how it can be applied.

The
University
Of
Sheffield.

Improvements

- Guidelines for languages other than English.
- Does not cater for event intervals or progressive aspect.
- No mechanism for integrating nominal events, which may be used positionally as a `timerefx`.

RTMBank

- Current work in progress.
- Goal is to create a corpus of fifty to sixty RTMML-annotated documents.
- Documents chosen for annotation are already in TimeBank.

The
University
Of
Sheffield.

Questions

Thank you for your time. Are there any questions?