Announcement of the Syntax/Semantics Interface as Bridge between Deep Linguistic Parsing and TimeML

Mark-Matthias Zymla
University of Konstanz
Mark-Matthias.Zymla@uni-konstanz.de

Introduction – Semantic Annotation and Treebanks

Semantic annotation of treebanks
ParTMA semantic annotation of tense/aspect based on formal semantic insights:
- Annotation of syntax/semantics interface based on morphosyntactic input as provided by treebanks
- Cross-linguistically viable due to abstraction over morphosyntax based on semantic and pragmatic reasoning
- Inspired by glue semantics (first developed in LFG) and the XLE transfer system (Crouch, 20017)

Three tiers of annotation

1. Morphosyntactic abstraction – maps morphosyntax onto semantic features, e.g. TENSE past \rightarrow TEMP-REF past: \( R < S \)
2. Semantic construction – composition of features, e.g. TENSE past \& PERS \rightarrow TEMP-REF past: \( E < R < S \)
3. Pragmatic reasoning – composition with context, e.g. TEMP-REF future: \( R < E \wedge \text{ctx}(R_{past}) \rightarrow \text{TEMP-REF future}: \( R < E < S \)

Treesbank technology for annotation

Annotation of the syntax/semantics interface inspired by INESS treesbank infrastructure (Rosen et al., 2016). Applicable to LFG structures and dependency structures.

1. TNS \& DEP \rightarrow LFG
2. TEMPOFF \& DEP \rightarrow LFG

From Morphosyntax to Context-Sensitive Temporal Annotation: An Example

Semantic annotation is multi-layered in order to model differences found in complex semantic phenomena, e.g. sequence of tense.

(1) John said that Karen was sick.
   a. John said: “Karen is sick.”
   b. John said: “Karen was sick.”

Rule-based annotation

1. TENSE tenses past \& MODD indicative \rightarrow TEMP-REF matrix ‘past’ \( t < t_0 \)
2. TEMP-REF matrix ‘past’ \& TEMP-REF comp ‘past’ \& COMP matrix, \text{Rcomp} \rightarrow TEMP-REF comp ‘non-future’ \( (t^t < t', t'' \circ t'') \)

Formal semantic computation

- \([\text{PAST}]^n_s = \lambda P. \lambda t. t < t_0 \wedge P(t)
- \text{[Tom said that Q]}^n_s = \lambda P. \lambda t. t < t_0 \wedge \text{say}(t, \text{tom}, Q)
- \text{[Tom said that Q]}^n_s = \lambda P. \lambda t. t < t_0 \wedge P(t)\text{[tom]}
- \text{[Tom said that Q]}^n_s = \lambda P. \lambda t. t < t_0 \wedge \text{be} \rightarrow \text{sick}(t', \text{karen})
- \text{[Tom said that Q]}^n_s = \lambda P. \lambda t. t < t_0 \wedge \text{say}(t, \text{tom}, \text{3rd}[t' < t \wedge \text{be} \rightarrow \text{sick}(t', \text{karen})], \text{last} < t_0 \wedge \text{say}(t, \text{tom}, \text{3rd}[t' < t \wedge \text{be} \rightarrow \text{sick}(t', \text{karen})])

Computational representation

<doc time="now"/>
<timeref id="10" target="doc\_time"/>
<event id="e1" target="#token2"> said
</event>
<concept id="c0"> conceptual description -->
<tempref id="past" semantics="r1"/>
</concept>
</event>
<event id="e2" target="#token6"> sick
</event>
<concept id="c0"> conceptual description -->
<tempref id="non\_future" semantics="r2"/>
</concept>
</event>
</doc>

Embedding in the temporal context

CASE 1: John said (yesterday) that Karen was sick (last week).

- R_c < \{ yesterday \} \rightarrow \text{yesterday}(R_c)
- R_c \in \{ \text{last week} \} \rightarrow \text{last\_week}(R_c)
- \text{yesterday}(R_c) \wedge \text{TEMP-REF comp} ‘past’ \( t' < t_0 \circ\)
- \text{TEMP-REF comp yesterday}(R_c)

CASE 2: John said (yesterday,) that Karen was sick (then).

- R_c < \{ yesterday \} \rightarrow \text{yesterday}(R_c)
- R_c \in \{ \text{then} \} \rightarrow \text{yesterday}(R_c)

ParTMA Annotation and Cross-Linguistic Variation

- Annotation of a diverse set of languages from different language families, e.g. Russian, Norwegian, Japanese
- The three different tiers provide insights into cross-linguistic variation on different levels

Selected references: