Annotation 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

Universität Konstanz

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**
- Crosslinguistically 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

t1: Morphosyntactic abstraction – maps morphosyntax onto semantic features, e.g.
TENSE past $\rightarrow$ TEMP-REF past : $R \prec S$

t2: Semantic construction – composition of features, e.g.
TENSE past $\wedge$ PERF + $\rightarrow$
TEMP-REF past : $E \prec R \prec S$

t3: Pragmatic reasoning – composition with context, e.g.
TEMP-REF future : $R \prec E \wedge \text{ctx}(R_{past}) \circ$
TEMP-REF future : $R \prec E \prec S$

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

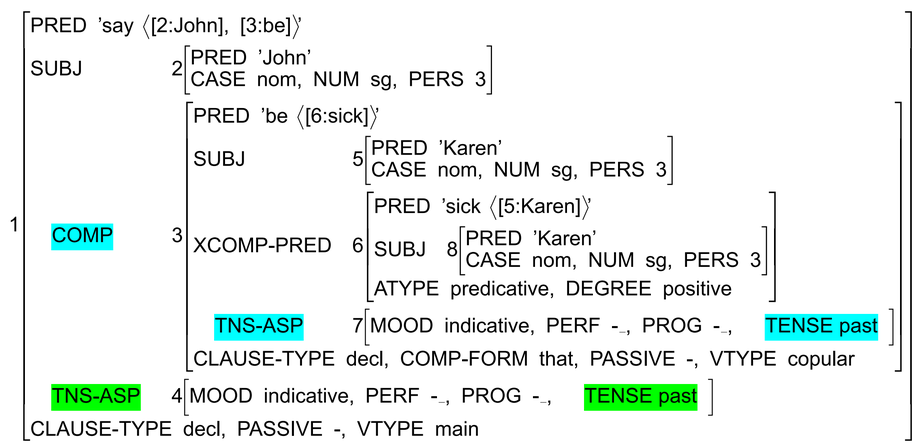
1. #g TNS-ASP #h TENSE past &
   #g COMP #i TNS-ASP #j TENSE past **LFG**
2. #g Tense past &
   #g CCOMP #i COP #j Tense past **DEP**

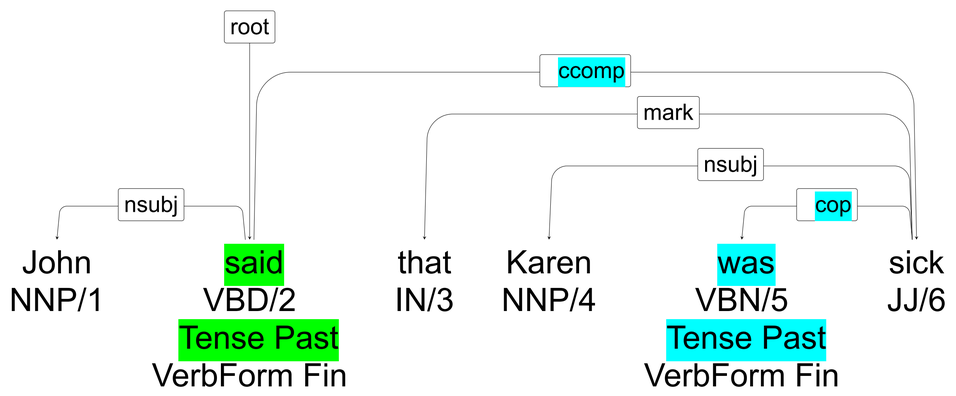
$\rightarrow$ Allows for computer-assisted annotation of t1 and t2 (but not t3)

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

t1 - $\text{TENSE}_{\text{matrix}}$ past $\wedge$ MOOD indicative $\rightarrow$ $\text{TEMP-REF}_{\text{matrix}}$ 'past' : $t \prec t_0$
   - $\text{TENSE}_{\text{comp}}$ past $\wedge$ MOOD indicative $\rightarrow$ $\text{TEMP-REF}_{\text{comp}}$ 'past' : $t' \prec t_0$

t2 - TEMP-REF $_{\text{matrix}}$ 'past' $\wedge$ TEMP-REF $_{\text{comp}}$ 'past' $\wedge$ COMP(E_{matrix},E_{comp})
   $\rightarrow$ $\text{TEMP-REF}_{\text{comp}}$ 'non-future' : $(\!|t'' \prec t', t'' \otimes t'|\!)$

Formal semantic computation

- $[\![\text{PAST}]\!]^{w,g} = \lambda P.\lambda t.t \prec t_0 \wedge P(t)$
  $[\![\text{Tom said that Q}]\!]^{w,g}$ $\lambda t.t \prec t_0 \wedge say(t, tom, Q)$
- $[\![\text{NON-FUTURE}]\!]^{w,g}$ =
  $(\!|\lambda P.\lambda t'.\lambda t.t' \prec t \wedge P(t), \lambda P.\lambda t'.\lambda t.t' \circ t \wedge P(t)|\!)^{w,g}$
- $[\![\text{Karen was sick}]\!]^{w,g}$= $[\![\text{Q}]\!]^{w,g} = \lambda t.t' \prec t \wedge be-sick(t', karen)$
  $[\![\text{Q'}]\!]^{w,g} = \lambda t.t' \circ t \wedge be-sick(t', karen)$
- $[\![\text{Tom said that Karen was sick}]\!]^{w,g}$ =
  $(\!|\lambda t.t \prec t_0 \wedge say(t, tom, \exists t'[t' \prec t \wedge be-sick(t', karen)]),$
  $\lambda t.t \prec t_0 \wedge say(t, tom, \exists t'[t' \circ t \wedge be-sick(t', karen)])|\!)$

Computational representation

```
<doc time="now"/>
<timeref xml:id="t0" target="doc time"/>      S

<event xml:id="e1" target="#token2">          said
<!-- conceptual description -->
<concept xml:id="c0">
<tempref="past" semantics="r1"/> ...
</concept></event>

<event xml:id="e2" target="#token6">          sick
<!-- conceptual description -->
<concept xml:id="c1">
<tempref="non-future" semantics="r2"/> ...
</concept></event>

<!-- temporal realization -->
<trel xml:id="r1" relation="e1≺t0"/>
                                   R_said ≺ S
<trel xml:id="r2" relation="e2<e1 || e2=e1"/>
                   R_sick ≺ R_said ∨ R_sick ⊗ R_said
```

Embedding in the temporal context

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

- $R_{said} \subset [\![$ *yesterday* $]\!] \rightarrow \text{yesterday}(R_{said})$
- $R_{sick} \subset [\![$ *last week* $]\!] \rightarrow \text{last\_week}(R_{sick})$

$\rightarrow$ yesterday(R_{said}) $\wedge$ $\text{TEMP-REF}_{\text{comp}}$ 'past' : $t' \prec t_0 \circ$
$\text{TEMP-REF}_{\text{comp}}$ yesterday(R_{said})
yesterday(R_{sick}) $\wedge$ $\text{TEMP-REF}_{\text{comp}}$ 'non-future' : $(\!|t'' \prec t', t'' \otimes t'|\!) \circ$
$\text{TEMP-REF}_{\text{comp}}$ past : last_week(R_{sick}) **past-under-past**

CASE 2: *John said (yesterday$_t$) that Karen was sick (then$_t$).*

- $R_{said} \subset [\![$ *yesterday* $]\!] \rightarrow \text{yesterday}(R_{said})$
- $R_{sick} \subset [\![$ *then* $]\!] \rightarrow \text{yesterday}(R_{sick})$

$\rightarrow$ yesterday(R_{said}) $\wedge$ $\text{TEMP-REF}_{\text{comp}}$ 'past' : $t' \prec t_0 \circ$
$\text{TEMP-REF}_{\text{comp}}$ yesterday(R_{said})
yesterday(R_{sick}) $\wedge$ $\text{TEMP-REF}_{\text{comp}}$ 'non-future' : $(\!|t'' \prec t', t'' \otimes t'|\!) \circ$
$\text{TEMP-REF}_{\text{comp}}$ non-future : yesterday(R_{sick}) **non-future**

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

	Total	German	Italian	Polish	Urdu	Indonesian
Sentences	196	56	50	45	47	48
Implication rules	9	3	3	1	3	2
Compatibilities	191	45	39	34	36	37

NUANCE FOUNDATION

Selected references:
Dick Crouch et al (2017): XLE Documentation. Palo Alto Research Center. • Victoria Rosén, Koenraad De Smedt, Paul Meurer, and Helge Dyvik (2012): An open infrastructure for advanced treebanking. In: Jan Hajič, Koenraad De Smedt, Marko Tadić, and António Branco (eds.) *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, pages 22–29, Istanbul, Turkey, May 2012. • Sebastian Sulger et al (2013): The ParGram Parallel Treebank. In: *ACL*, pages 550–560.