Mapping between Dependency Structures and Compositional Semantic Representations

LREC 2010

Max Jakob, Markéta Lopatková, Valia Kordoni

ÚFAL at Charles University in Prague, Czech Republic
German Research Centre for Artificial Intelligence (DFKI GmbH)
Dept. of Computational Linguistics, Saarland University, Germany
Research presented in the paper
map PDT annotation $\mapsto$ RMRS structures

- manually annotated corpora are valuable
- barrier: difference in formal descriptions
  $\rightarrow$ usage of a resource remains limited
- $\Rightarrow$ precisely relate formalisms to overcome these limitations

Benefits: flexibility, availability, RMRSs for Czech
Czech newspaper and magazine articles

theoretical backgr.: Functional Generative Description theory (Sgall et al., 1986)

3 annotation layers

“Byl by šel do lesa.”
“He would have gone into the woods.”
Tectogrammatical trees:
(Hajič et al., 2006)

- highest level of abstraction
- sub-layers:
  - structure and dependencies
  - morphological categories
  - coreferences
  - topic-focus

“Pes asi honí kočku.”

“The dog probably chases a cat.”
PDT Valency Dictionary:

(Hajič et al., 2003)

- separate data source from PDT
- comprises obligatory and optional valency modifications
  - does not contain free modifications

minout  \textit{(to pass)}

- ACT, PAT  \textit{(the bullet passed/missed the victim)}
- ACT  \textit{(the holidays have passed)}
(Robust) Minimal Recursion Semantics:
(Copestake et al., 2005; Copestake, 2007)

- flat, underspecified representation
- no semantic theory

\[
< [l0, e2], \\
\{ l1: \_every_q(x1, h1, h2), l2: \_white_adj(x1), \\
    l2: \_cat_n(x1), l3: \_probably_adv(e1, h3), \\
    l4: \_eat_v(e2[tense:past], x1, x2), l5: \_a_q(x2, h4, h5), \\
    l6: \_mouse_n_1(x2) \}, \\
\{ h1 = q l2, h3 = q l4, h4 = q l6 \} >
\]

“Every white cat probably ate a mouse.”
- adapt theoretical background of PDT
- rule-based approach

\[
< [\mathit{0}, a_1, e_1],
\{ l_1: a_1: \_honit\_v\_1(e_1),
    l_2: a_2: \_pes\_n\_denot(x_1),
    l_4: a_4: \_kočka\_n\_denot(x_2),
    l_3: a_3: \_asi\_atom(e_2),
    l_1: a_5: \_MOD(e_1) \},
\{ a_1: \_ACT(x_1),
    a_1: \_PAT(x_2),
    a_3: \_ARG1(h_1),
    a_5: \_ARG1(e_2) \},
\{ h_1 =_q l_1 \} >
\]
node-RMRS

represents a subtree as RMRS

honit
PRED
< [l3, a1, e1],
{l1: a1: _honit_v_1(e1), l2: a2: _pes_n.denot(x1), l3: a3: _kočka_n.denot(x2),
l4: a4: _asi_atom(e2), l1: a5: MOD(e1)},
{a1: ACT(x1), a1: PAT(x2), a4: ARG1(h1), a5: ARG1(e2)},
{h1 = q /1}>

pes
ACT
< [l2, a2, x1],
{l2: a2: _pes_n.denot(x1)},
{ }>,
{ }>

asi
MOD
< [l4, a4, e2],
{l4: a4: _asi_atom(e2)},
{a4: ARG1(h1)},
{ }>,
{ }>

kočka
PAT
< [l3, a3, x2],
{l3: a3: _kočka_n.denot(x2)},
{ }>,
{ }>
node-RMRS Initialization

\[
< \ [l3, a3, x2],
\{ \ l3:a3:_kočka_n.denot(x2[number:sg,gender:fem]),
\ l4:a4:udef_q(x2) \},
\{ \ a4:RSTR(h1), a4:BODY(h1) \},
\{ \ h1 =_q l3 \} >
\]

- relation name: _lemma_POS_index
- variable features: morphological category
- set hook values
- introduce a quantifier for nominal objects
node-RMRS Composition

- **valency modification**
  - add argument to governing lexical EP

- **free modification**
  - add EP that relates lexical EPs

- **coordination**
  - add a coordination EP

- add constraints, update hook
- build union of all involved sets
**Algorithm Sketch**

**Input:** tectogrammatical tree (, valency dictionary)
**Output:** RMRS structure

---

**get\_node-RMRS(node)**

1. **initialize** node-RMRS
2. **for all** relevant dependent nodes
3. dep\_node-RMRS ← get\_node-RMRS(dep.)
4. treat dep\_node-RMRS as a member of a *coordination* or as a *valency modification* or as a *free modification*
5. merge dep\_node-RMRS with node-RMRS
6. **return** node-RMRS
no manual checking
← corpus size & lack of sufficient Czech skills

**Structurally valid MRSs**

1. must be a **net** (Flickinger *et al.*, 2005).
2. must have **at least one configuration**.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>40120/44725</th>
<th>89.70 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>40120/49431</td>
<td>81.16 %</td>
<td></td>
</tr>
</tbody>
</table>

(skipped 4706 trees: 9.52 %)
mapping of PDT dependency trees onto flat RMRS structures is feasible

mapped: structure and dependencies, morphological categories, some grammatical coreferences

future work: word order, quantifier representation, rest of grammatical coreference, textual coreference, topic-focus articulation

Benefits:
- treebank data available in (R)MRS
- towards formalism independence
- compositional semantics structures for Czech