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

Motivation

Research presented in the paper

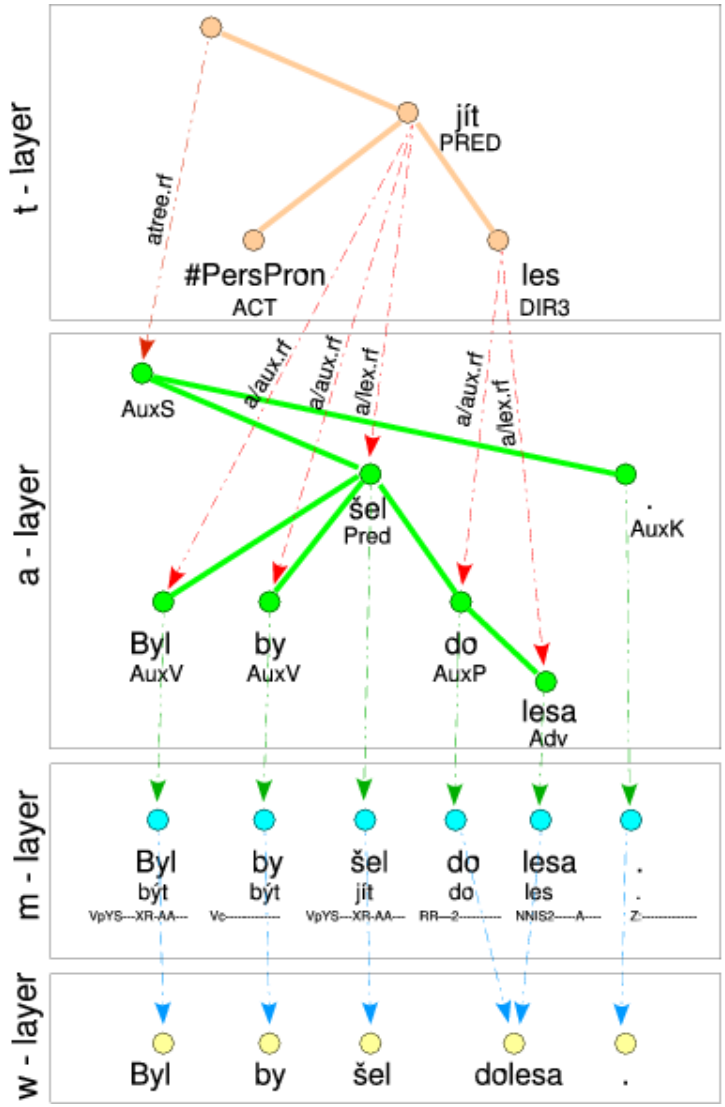
map PDT annotation $\mapsto$ RMRS structures

- 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

Prague Dependency Treebank 2.0

- 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."

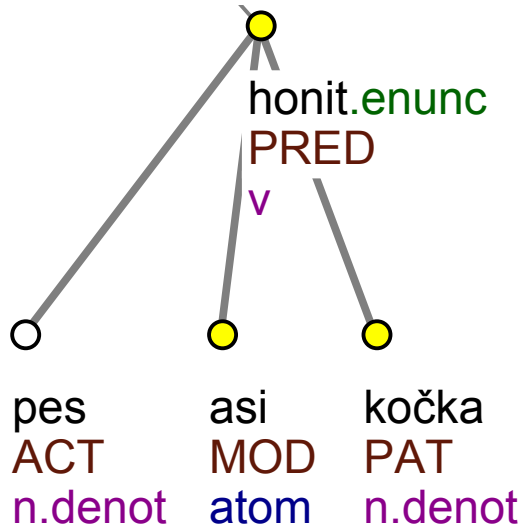


Input of the mapping (1)

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."

Input of the mapping (2)

PDT Valency Dictionary:

(Hajič *et al.*, 2003)

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

minout *(to pass)*

- ACT, PAT *(the bullet passed/missed the victim)*
- ACT *(the holidays have passed)*

Output of the mapping

(Robust) Minimal Recursion Semantics:

(Copestake *et al.*, 2005; Copestake, 2007)

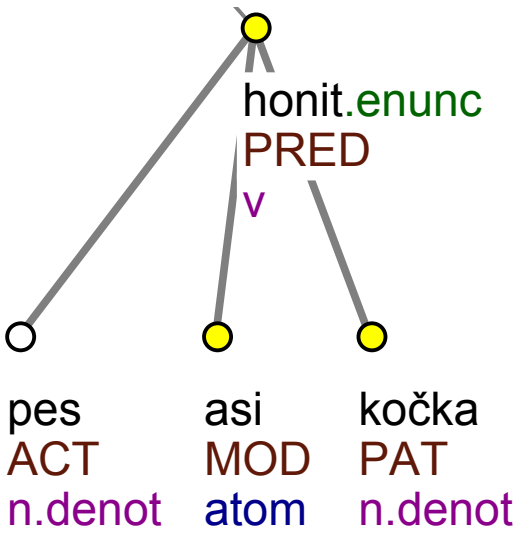
- flat, underspecified *representation*
- no semantic theory

$$
\begin{array}{l}
< [l0, e2], \\
\quad \{\ l1\colon \text{\_every\_q}(x1, h1, h2),\ l2\colon \text{\_white\_adj}(x1), \\
\quad\ \ l2\colon \text{\_cat\_n}(x1),\ l3\colon \text{\_probably\_adv}(e1, h3), \\
\quad\ \ l4\colon \text{\_eat\_v}(e2_{[tense:past]}, x1, x2),\ l5\colon \text{\_a\_q}(x2, h4, h5), \\
\quad\ \ l6\colon \text{\_mouse\_n\_1}(x2)\ \}, \\
\quad \{\ h1 =_q l2,\ h3 =_q l4,\ h4 =_q l6\ \} >
\end{array}
$$

"Every white cat probably ate a mouse."

Mapping PDT $\mapsto$ RMRS

- adapt theoretical background of PDT
- rule-based approach

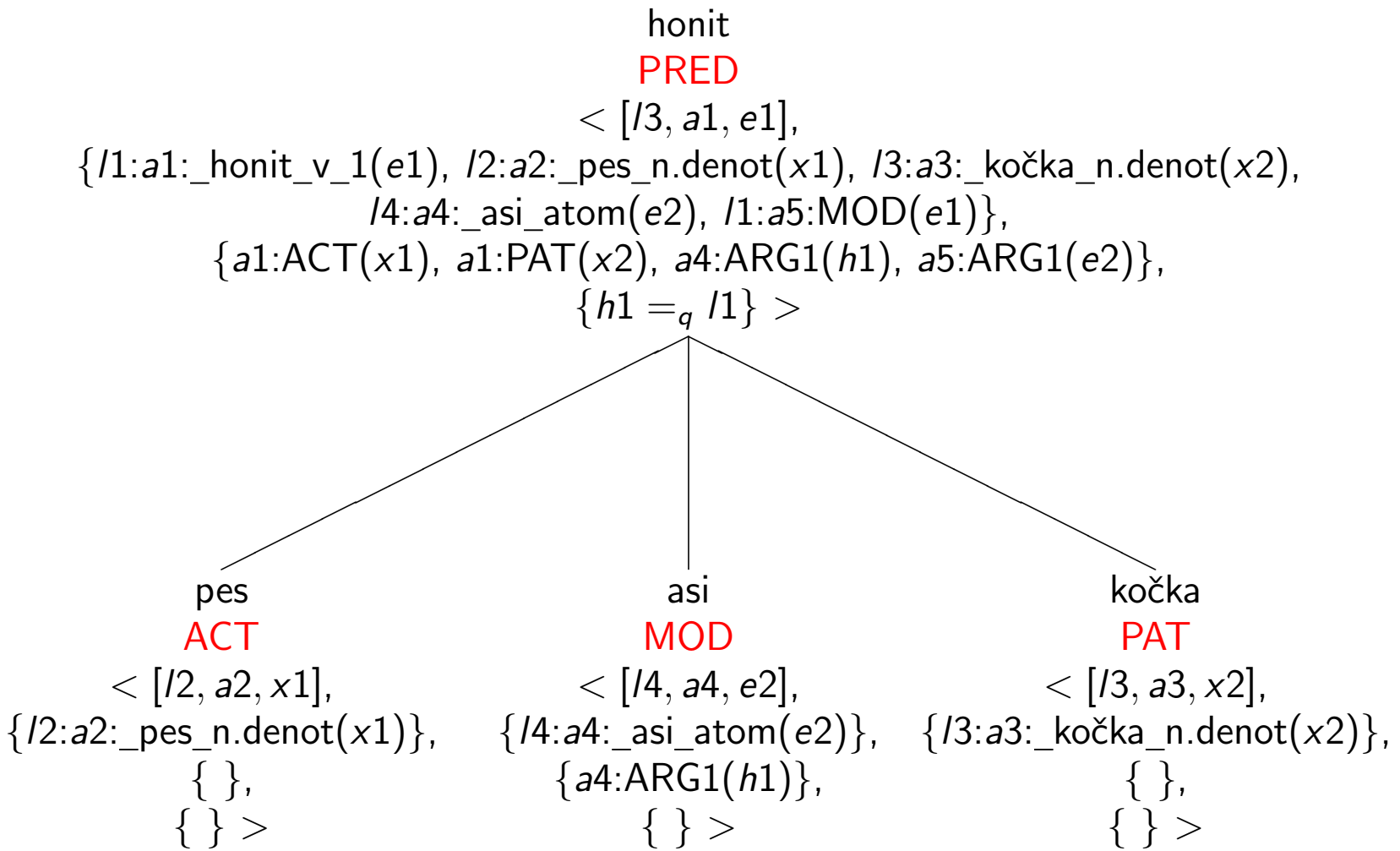


$\mapsto$

$$
\begin{aligned}
&< [l0, a1, e1], \\
&\quad \{\ l1{:}a1{:}\_honit\_v\_1(e1), \\
&\qquad l2{:}a2{:}\_pes\_n.denot(x1), \\
&\qquad l4{:}a4{:}\_kočka\_n.denot(x2), \\
&\qquad l3{:}a3{:}\_asi\_atom(e2), \\
&\qquad l1{:}a5{:}\mathrm{MOD}(e1)\ \}, \\
&\quad \{\ a1{:}\mathrm{ACT}(x1), \\
&\qquad a1{:}\mathrm{PAT}(x2), \\
&\qquad a3{:}\mathrm{ARG1}(h1), \\
&\qquad a5{:}\mathrm{ARG1}(e2)\ \}, \\
&\quad \{\ h1 =_q l1\ \} >
\end{aligned}
$$

node-RMRS

- represents a subtree as RMRS



node-RMRS Initialization

$$
\begin{aligned}
&< [l3, a3, x2],\\
&\quad \{\ l3{:}a3{:}\text{\_kočka\_n.denot}(x2_{[number:sg,gender:fem]}),\\
&\qquad l4{:}a4{:}\text{udef\_q}(x2)\ \},\\
&\quad \{\ a4{:}\text{RSTR}(h1),\ a4{:}\text{BODY}(h1)\ \},\\
&\quad \{\ h1 =_q l3\ \} >
\end{aligned}
$$

- relation name: _lemma_POS_index
- variable features: morphological categor.
- 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

Evaluation

- 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**.

Precision	40120/44725	89.70 %
Recall	40120/49431	81.16 %

(skipped 4706 trees: 9.52 %)

Conclusion

- 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

Paper Reference

Max Jakob, Markéta Lopatková, Valia Kordoni. Mapping between Dependency Structures and Compositional Semantic Representations. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta.