

# Tectogrammatical-to-AMR conversion: current status

#### CLAMR team – JHU Workshop 2014

Ondřej Bojar

with

Jan Hajič, Silvie Cinková, Roman Sudarikov Zdeňka Urešová, Martin Popel, Ondřej Dušek Institute of Formal and Applied Linguistics Computer Science School Faculty of Mathematics and Physics Charles University in Prague Malostranské náměstí 25 11800 Prague 1 http://ufal.mff.cuni.cz



## Motivation



CLAMR Goal: Machine Translation using AMRs

- Well, maybe not quite reachable, at least some steps to get there:
- Graph transformation algorithms, ML
- Generation/Parsing from/into AMRs in Czech/English, ...

All you need is ... Data

- Parallel Czech-English AMRs: 100 sentences
- Monolingual (being annotated for English)
- Parallel Czech-English (t-layer only)
  - **PCEDT 2.0**, 49k sents, manual t-layer for en+cs
  - CzEng 1.0, 15M sents (200MW per lang.), auto t-layer





## TR to AMR Example: What We Want



3

# **Our Approach and Process**

- 1. Get all AMR examples from the AMR manual
  - 259 in total; have been ID-ed (three digit number, 001 259)
- 2. Get all phrases (segments) for these (n:1 mapping)
  - 401 of them; related to the AMR subtrees by ID (001a, 001b, ...)
- 3. Imagine the t-(sub)tree for those phrases (Silvie Cinková)
  - Formulate it as a query to a t-tree (there will be 401 of them, now 183 done)
- 4. Run the queries on a treebank (CzEng, En side) (Roman Sudarikov)
  - Results in nodes being marked by a AMR rule ID
- 5. Rule disambiguation (to solve conflicting rule matches)
  - manual on a subset, later ML-ed (if refinement of rules impossible)
- 6. Clone the t-tree, replace subtrees marked by AMR trees (RS)
  - Restructure the result get a full AMR tree (first implementation ready)







# 1&2: AMR Example to Tree Query UFAL

#### A linguist considers AMR tree + all expression forms:

- 138 >AMR (a / adjust-01
- 138 >AMR :ARG0 (g / girl)
- 138 >AMR :ARG1 (m / machine))
- 138 >
- 138 > The girl adjusted the machine.
- 138 > The girl made an adjustment to the machine.





# 3: AMR Example to Tree Query



#### ...and formulates what the corresponding t-tree would like:

```
#138a-The_girl_adjusted_the_machine
t-node $a_adjust_01 := [gram/sempos="v",
    t-node $g_girl := [gram/sempos~"^n..*$", functor="ACT"],
    t-node $m_machine := [gram/sempos~"^n..*$", functor="PAT"]]
```

```
#138b-The_girl_made_an_adjustment_to_the_machine
t-node $make_DEL := [gram/sempos="v",
    t-node $g_girl := [gram/sempos~"^n..*$", functor="ACT"],
    t-node $adjustment_DEL:= [gram/sempos~"^n..*$",
        t_lemma in {"N_V"}, functor="CPHR"]]
```

















July 7, 2014

# 4. Search for 'Guideline Patterns'

PML-TQ "rules" mark nodes where matched:



Israel has launched a new effort to prove...



July 7, 2014

# 5. Which Rules to Apply?



#### ...work in progress...

- In practice, many guidelines rules apply at the same position.
- Proposed approach:
  - Manually disambiguate a small set.
  - Apply ML on the full data.
- Status:
  - Tentative UI for annotation ready.



#### Israel has launched a new effort to prove...





# Relation of the Matched Rule and Target AMR



But State Department officials accuse Israel of leaking questionable claims to embarrass the U.S.



July 7, 2014

# 6. The Conversion



- Procedure:
  - With rules selected (disjoint or causing no conflict), restructure the subtree to mimic the "Guideline AMR shape".
  - Apply default mapping elsewhere.
- Status:
  - First implementation ready (Perl, hand-coded)
- Room for generalization:
  - Some toolkit for tree rewriting? (Subtree replacement for dependency trees.)





# Summary



- Data and tools available:
  - Manual 100 parallel AMRs
  - Manual 49k parallel t-trees
  - XML-based 'treex' format, Treex toolkit, TrEd editor
- T-to-AMR conversion status:
  - Sort-of finished pipeline (Roman Sudarikov).
  - Rule disambiguation much needed.
  - Nothing evaluated yet.



