Coreference in the PDT 2.0

Zdeněk Žabokrtský
Institute of Formal and Applied Linguistics
Charles University in Prague
What is coreference?

- multiple expressions in a sentence or document can refer to the same thing
Coreference in PDT

- links between tectogrammatical nodes
- technically: pointer from an anaphor t-node to its antecedent t-node
- links can form chains
Two types of coreference

according to Functional Generative Description, two types of coreference distinguished:

- **grammatical** coreference
  - (partially) determined by grammar rules

- **textual** coreference
  - determined only by text meaning
Grammatical coreference (1)

- relative pronouns
- "The man, who…", "The man, whose ..."
- typical local configuration:

... noun modified by the relative clause

main verb of the relative clause

relative pronoun

...
Grammatical coreference (2)

- reflexive pronouns
- in Czech, pronouns referring to clause subject have reflexive form
- typical local configuration:

![Diagram showing reflexive pronoun in a clause structure]
Grammatical coreference (3)

- reconstructed (surface-unexpressed) actor of infinitive verbs
- "He started to sing." "They asked him to come."
- typical local configuration:

```
#Cor.ACT - reconstructed coreferential actor
```

```
control verb
```

```
infinitive verb
```

```
... (reconstructed coreferential actor)
```
Textual coreference

- **anaphors:**
  - personal pronouns
  - possessive pronouns
  - reconstructed pronouns (pro-drop)
Special cases

- multiple antecedent:
  - two or more parallel links from a plural anaphor (*Peter and Paul ... they...*)

- cataphora
  - left-to-right links

- segm - vague reference to the previous context

- exoph - exophora
Amount of data

- manually annotated coreference in 50,000 sentences
- around 45,000 coreference links
Summary

- coreference in PDT 2.0
  - one of the largest coreference resources
  - two types of coreference links
    - grammatical coreference
    - textual coreference
- anaphors:
  - pronouns (personal, possessive, relative, reflexive)
  - reconstructed nodes (pro-drops, actants of infinitive verbs, ...)

PDT 2.0