

# Syntaktická analýza

# Syntaktická analýza

- další rovina anotace korpusu
- předpoklad – **morfologicky** značkováný (a desambiguovaný) korpus
- několik rovin značkování
- výsledek – **závislostní / složkové stromy**
- **treebank** (závislostní korpus, stromová banka)
- automatická, poloautomatická, ruční anotace

# Základní pojmy

- věta (sentence S, clause)
- nominální fráze (NP)
- verbální fráze (VP)
- závislostní stromy (tree)
- závislostní korpus (treebank)

# K čemu to potřebujeme?

- další rovina popisu jazyka
- treebank – referenční data pro automatické nástroje
- synchronní a diachronní studie
- frekvenční studie
- strojový překlad
- navazující aplikace

# Syntaktická analýza – Praha

- historické pozadí
- lingvistický strukturalismus, Pražská škola
- Pražský lingvistický kroužek (1926, Mathesius, Jakobson, Trnka)
- funkčně generativní popis (FGD, Sgall, 60. léta)

# Syntaktická analýza – Praha

- <http://ufal.mff.cuni.cz>
- **PDT 2.0** (*Prague Dependency Treebank, Pražský závislostní korpus*)
- 2 mil. slovních jednotek
- rovina anotace:
  - slovní
  - morfologická
  - syntaktická (analytická)
  - sémantická (tektogramatická)
  - aktuální členění věty, koreferenční vztahy
- Vallex – valenční slovník

# Syntaktická analýza – Brno

- syntaktický analyzátor klara, synt, set, zuzana
- morfologicky značkový korpus
- formální popis gramatiky
- anotace na úrovni částí věty
  - sentence (věta)
  - clause (věty v souvětí)
  - nominální fráze (NP)
  - verbální fráze (VP)
  - koordinace (COORD)

# Odkazy

- synt
- <http://nlp.fi.muni.cz/projekty/wwwsynt/>
- anotace frází
- <http://nlp.fi.muni.cz/trac/nlpunix/wiki/NP>
- zuzana
- <http://nlp.fi.muni.cz/projekty/zuzana/>
- The Penn Treebank
- <http://www.cis.upenn.edu/~treebank/>