

Sémantická analýza



Proseminář z počítačové lingvistiky
Podzim 2019

Markéta Masopustová

Termíny

- Sémantika;
- Hyponyma;
- Hyperonyma;
- Kohyponyma;
- Meronyma;
- Synset.



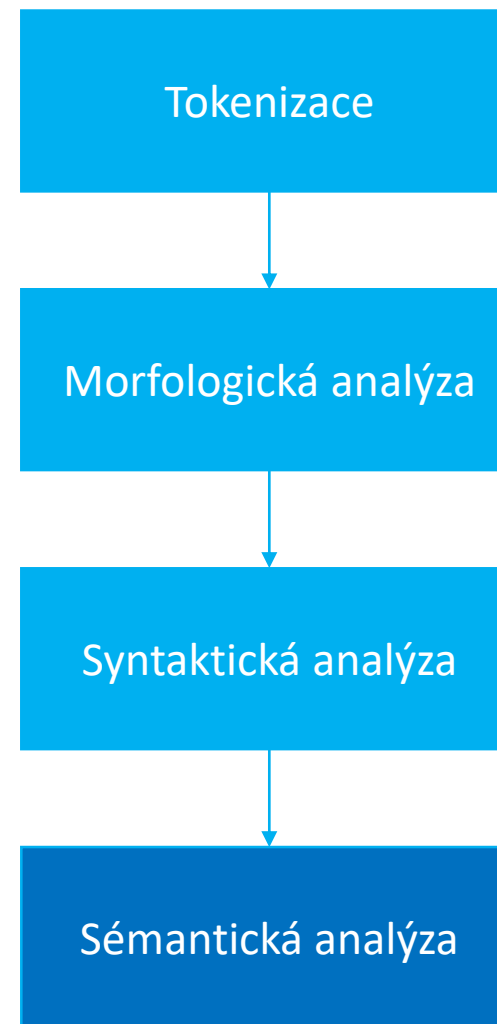
Termíny

- Sémantika – nauka o významu;
- Hyponyma – podřazená slova;
- Hyperonyma – nadřazená slova;
- Kohyponyma – významově shodná slova;
- Meronyma – označuje část celku;
- Synset – synonymická řada.



Sémantická analýza

- Snaží se o formální popis významu.
- Měla by být jazykově nezávislá.
- Počátky v ontologiích.



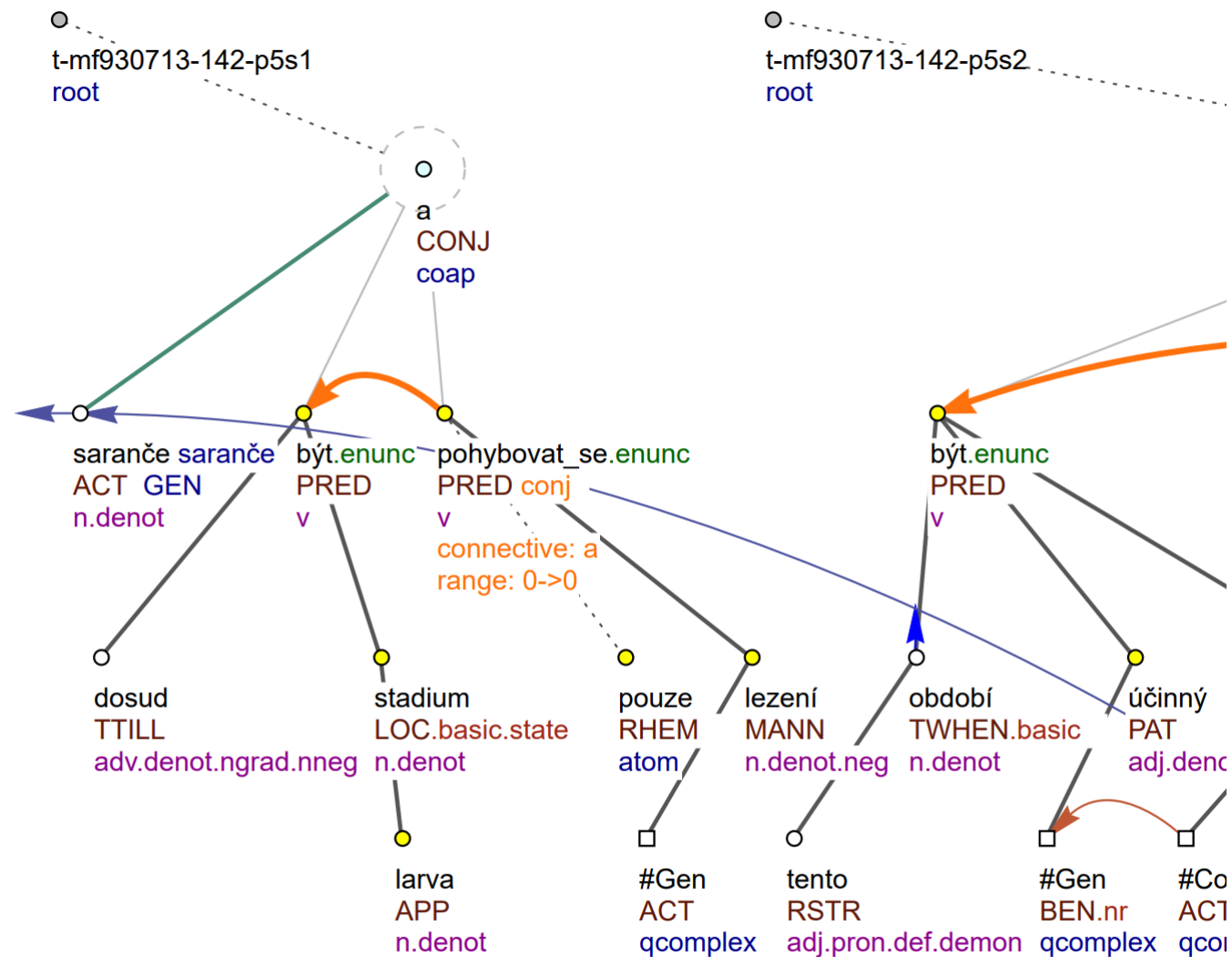
NER

- Rozpoznání pojmenovaných entit (Named Entity Recognition).
- Cílem je najít předem definované kategorie v nestrukturovaném textu.
- <https://nlp.fi.muni.cz/projekty/ner/v2/>
- <http://ufal.mff.cuni.cz/cnec/cnec2.0>

Pan Ing. Jan Novák z obce Chýně navrhl, aby se od 15. září 2014 městská rada (dále jen MR) scházela 4 km od památné lípy v centru obce. Svůj návrh podpořil citací § 6 zákona 121/2001 Sb. a svérázným tanečkem.

V rámci ÚFALu

- PDT:
 - Prague Dependency Treebank;
 - Nejedná se o význam, ale o významovou roli ve větě.
- Semantic Pattern Recognition;
- SEANCe – značkování sentimentu.



Sémantické sítě

- FrameNet

- <https://framenet.icsi.berkeley.edu/fndrupal/>

- VerbNet

- <https://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

- ConceptNet

- <http://conceptnet.io/>

Location of cat

en my lap →
en a bed →
en the windowsill →
en a chair →
en a table →
en a vet →
en the barn →
en the floor →
en your way →
en the backyard →
en bag →
en someone's home →
en the rug →
en an alley →
en a back yard →
en a cat box →
en a closet →
en a house →
en the roof →

cat is capable of...

en hunt mice →
en catch a mouse →
en drink water →
en climb up a tree →
en corner a mouse →
en look at a king →
en kill birds →
en mother her kittens →
en catch a bird →
en cleaning itself →
en drink milk →
en scratch →
en scratch furniture →
en sleep →
en wash its paws →
en eat cat food →
en eye a mouse →
en hide under the bed →
en meow →

WordNet

- G. A. Miller (*Princeton University*) – psycholog a psycholinguista.
- Základním je Princeton WordNet (1985), postupně vytvářeny národní Wordnety.
- <http://wordnet.princeton.edu>
- <http://globalwordnet.org/resources/wordnets-in-the-world/>

▲ Základní

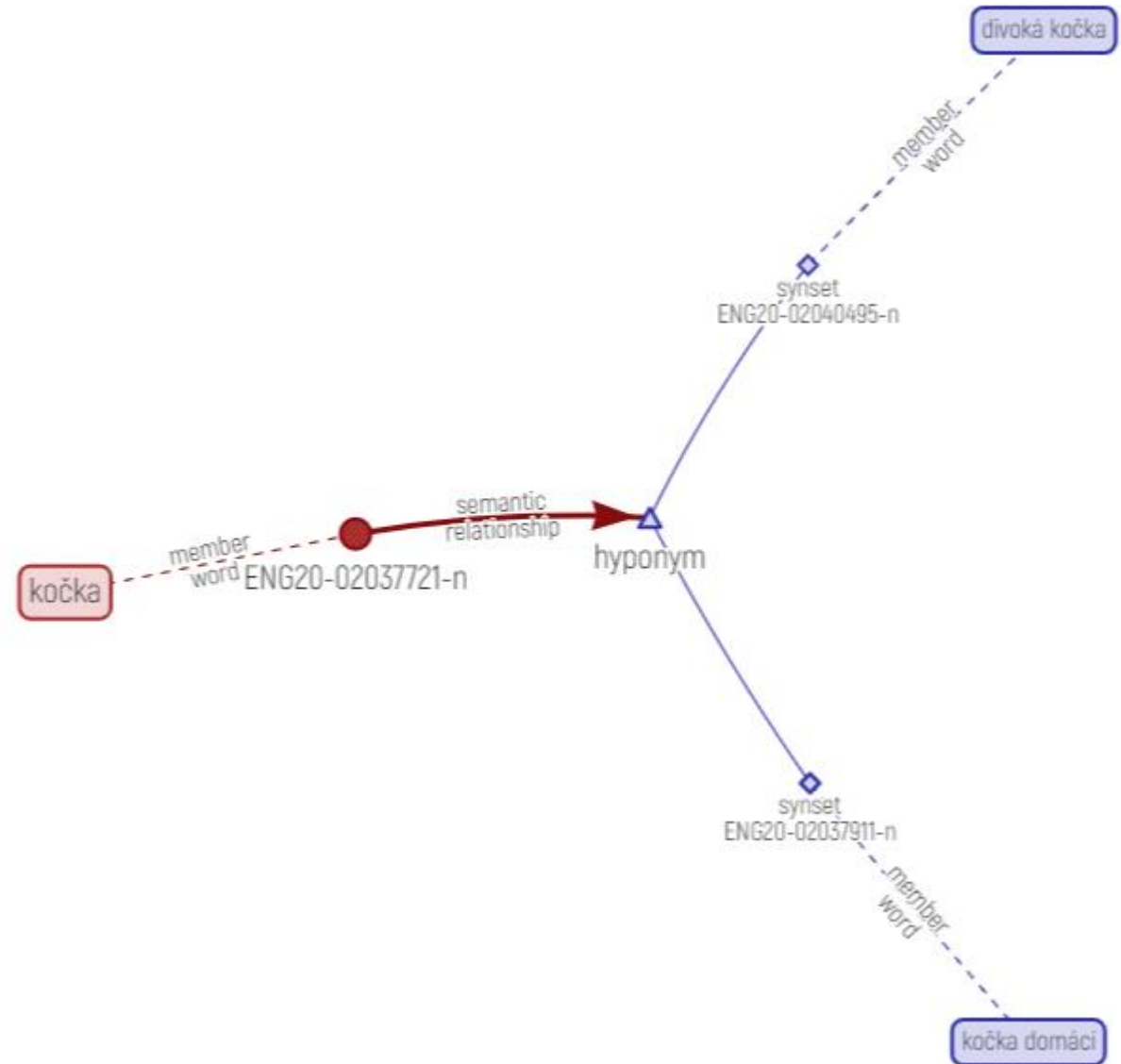
Doslova „Síť slov“. Lexikální databáze, ve které jsou slova a slovní spojení (tzv. *literály*) seskupena do synonymických řad neboli *synsetů* (z angl. *synonymical set*) a jednotlivé synsety jsou propojeny sémantickými vztahy. Každý sémantický vztah je spojnicí mezi dvěma synsety, díky čemuž tvoří W. síť (graf). Nejčastěji používanými sémantickými vztahy ve W. jsou *hyponymně-hyperonymní vztah*, *meronymně-holonymní vztah* a *opozitnost*. Původní ambicí projektu bylo vytvořit databázi modelující lidskou lexikální paměť, v průběhu času se však ukázalo, že může sloužit i v oblasti počítačového zpracování přirozeného jaz. jako *tezaurus* a svého druhu ontologie.

W. angličtiny je vytvářen od r. 1985 na Princetonské univerzitě. V současnosti (r. 2013) obsahuje 117 tis. synsetů. Na jeho základě vznikly W. pro jiné jaz. Od r. 1996 byly vytvořeny v projektu EuroWordNet „národní“ W., kde figuruje mj. i český W. Vývoj „národních“ W. (v současnosti více než 70 jaz. světa) mapuje a podporuje Global WordNet Association.

W. obsahuje *subst.*, *adj.*, *verb.*, *adverb.* Např. {kabriolet, sportřák} tvoří synset, jehož hyperonymem je synset {auto, vůz}, který je spojen vztahem hyperonymie s (jednoduchým) synsetem {motorové vozidlo}. Synset {motorové vozidlo} je spojen vztahem holonymie (tj. celek obsahuje část) se synsetem {spalovací motor}. V *počítačovém zpracování přirozeného jazyka* lze takovou síť využít k reprezentaci a odvozování znalostí. Např. ze vztahů „součástí motorového vozidla je spalovací motor“ a „kabriolet je druhem motorového vozidla“, které jsou ve W. obsaženy, lze odvodit novou znalost, tj. „kabriolet má spalovací motor“ (👉 Touretzky, 1986).

WordNet prakticky

- Přístup: demo/demo; read/read.
- DebVisDic 2:
 - https://deb.fi.muni.cz/proj_debvisdic-cs.php
- RAW viewer:
 - <https://deb.fi.muni.cz/raw-viewer/rawviewer.html>



Děkuji za pozornost.

