

ZADÁNÍ SEMINÁRNÍHO ÚKOLU

CJBB85 POČÍTAČOVÉ NÁSTROJE PRO ČEŠTINU

podzim 2014

Hlavním úkolem je připsat valenční rámce ke slovesům obsaženým v sémantických třídách.

1. V učebních materiálech v ISu je vložen soubor Rozdeleni_semantickych_trid, ve kterém jsou každému studentovi zadány sémantické třídy či podtřídy ke zpracování (použijte verzi v Excelu – xls, automaticky vygenerovaný soubor pdf nezobrazuje obsah správně).
2. Seznam sémantických tříd najdete na adrese:
<http://nlp.fi.muni.cz/~ydana/Semtridy.html>
3. Zkopírujte sémantické třídy/podtřídy, které vám byly zadány, do textového editoru (Word, LibreOffice). Zkontrolujte, zda jde opravdu o správné třídy/podtřídy.
4. Upravte definici významu třídy/podtřídy co nejuvýstižněji. Pokud definice zcela chybí, doplňte ji (výstižnou a stručnou).
5. Ke slovesům obsaženým v zadané třídě/podtřídě zapište valenční rámec na morfosyntaktické úrovni ve tvaru:

zájmeno s číslem pádu VERB (předložka) zájmeno s číslem pádu

např. **kdo1 VERB co4, kdo1 VERB co4 pro koho4**

U všech doplnění rozlišujte životnost a neživotnost.

(např. sloveso *rozstříknout* nemůže mít valenční rámec kdo1 VERB koho4)

Pod každý rámec doplňte seznam sloves ze zadané třídy/podtřídy, pro která rámec platí. Slovesa pište za sebe a odděluje je čárkami. (Uvažujte sloveso vždy pouze ve významu dané sémantické třídy, nezapívejte se jeho dalšími či přenesenými významy.)

Zapisujete tedy:

- levostranné valenční doplnění, typicky substantivum, zde zastoupené zájmenem s číslem pádu (nejčastěji **kdo1**)
- pravostranné valenční doplnění, typicky substantiva v nepředložkových či předložkových pádech (**co4, na koho/co4, za kým7, komu3, bez koho/čeho2** atd.)
- pravostranným valenčním doplněním může být také – adjektivum (**adj**), infinitiv (**inf**), vedlejší věta (**spojka**)

- zapisujte maximálně **2** pravostranná doplnění
 - nezapisujte doplnění **adverbiem** (příslovečná určení času, místa, způsobu), pokud jej sloveso nutně nevyžaduje
 - pokud je sloveso bezvalenční, zapíšete pouze **kdo1 VERB**
6. Pokud v sémantické třídě/podtřídě naleznete slovesa, která:
- svým významem do třídy nespádají
 - jsou zastaralá, zřídka užívaná, nářeční, hovorová
 - mají v korpusu czTenTen12 nulovou nebo velmi nízkou frekvenci,

uvedte je na závěr každé sémantické třídy/podtřídy za **Nepatří sem:**.

7. Na rozdíl od instrukcí uvedených na seminářích **NEMUSÍTE** ověřovat, zda jsou slovesa obsažena také ve VerbaLexu a přidávat další slovesa.
8. Použijte nástroje:
- **DEBDict** – ověření významu slovesa, slovníkové definice, stylových příznaků (doplněk Firefoxu, instaluje se z adresy <http://deb.fi.muni.cz>)
 - **Sketch Engine** – korpus **CzTenTen12** – ověření frekvence výskytu sloves, ověření významu slovesa a jeho valence (sledování kontextu). Najdete na adrese <http://ske.fi.muni.cz>.

Hotové práce vkládejte do **Odevzdávárny** v ISu do **30. 1. 2015**.