

Gramatika a korpus II plin032

Středa: 7.30-9.00 G13

24. 2. Úvodní hodina: Homonymie a chybovost morfologického značkování

(opakování termínů: tokenizace, lemmatizace, disambiguace, typy značkování poziční/atributový systém, pražský a brněnský slovník a analyzátoři, korpusy SYN a webové korpusy, cvičení: hledání v korpusu podle vzoru/formální vlastnosti tvarotvorných paradigmat).

Cvičení

Jak vyhledat v korpusu substantiva podle vzoru: *feminina*.

Jednoduše vyhledatelná substantiva podle vzorů *žena* a *růže*.

žena

[lemma=".*a" & tag="N.F.*"]

růže

[lemma=".*[eě]" & tag="N.F.*"]

pozor na substantiva typu *dítě* i další substantiva (napětí mezi přirozeným a gramatickým rodem a problémy technického řešení na úrovni morfologického značkování)

? *píseň/kost*

[lemma!=".*[aeěiyouáéíóúůý]" & tag="N.F.*"]

2. 3. Dů: Jak vyhledat v korpusu substantiva podle vzoru *píseň* a *kost*