Základy využití korpusu v praxi **cjbb75**

**Středa: 9.10-10.50 G13**

**20. 4. 2016**

**Je možné získat z korpusů informace o slovotvorbě?**

1. Hledání slov dle počátečního/koncového řetězce

Mějme např. za úkol vyhledat slova tvořená příponou *-tel*. Má-li jít o slova tvořená touto příponou, pak je můžeme hledat dle kombinace lemmatu s koncovým řetězcem *tel.*

[lemma=".\*tel"]

[https://kontext.korpus.cz/first?shuffle=1&reload=&corpname=omezeni%2Fsyn2010&usesubcorp=&queryselector=cqlrow&iquery=&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=[lemma%3D%22.\*tel%22]&default\_attr=word&fc\_lemword\_window\_type=both&fc\_lemword\_wsize=5&fc\_lemword=&fc\_lemword\_type=all&fc\_pos\_window\_type=both&fc\_pos\_wsize=5&fc\_pos\_type=all&sca\_opus.nazev](https://kontext.korpus.cz/first?shuffle=1&reload=&corpname=omezeni%2Fsyn2010&usesubcorp=&queryselector=cqlrow&iquery=&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=%5blemma%3D%22.*tel%22%5d&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&sca_opus.nazev)=



Vidíme, že nechceme např. lemma *postel* (slovo je utvořeno ze slovesa *stlát*) a řetězec *tel* není sufix, ale sč. kořene.

Dotaz přegenerovává.

Co lze dále využít?
Jaké vlastnosti mají jména tvořená příponou *–tel*?

Podíváme-li se do konkordančního seznamu, vidíme, že jsou to názvy osob jako (*věřitel, objevitel, učitel*).

Lze tohoto pozorování (empirický výzkum) využít?

Ano, můžeme omezit dotaz a vyhledat pouze maskulina životná (využít tagu).

P/filtr

[lemma=".\*tel" & tag="NNM.\*"]

[https://kontext.korpus.cz/filter?within=0&q=~YVQLpvoE&ctxattrs=word&pagesize=40&refs=doc%2C%3Dopus.nazev&viewmode=kwic&attrs=word%2Clemma%2Ctag&corpname=omezeni%2Fsyn2010&structs=s&attr\_allpos=kw&pnfilter=p&filfl=f&filfpos=0&filtpos=0&inclkwic=True&queryselector=cqlrow&iquery=&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=[lemma%3D%22.\*tel%22+%26+tag%3D%22NNM.\*%22]&default\_attr=word&sca\_opus.nazev](https://kontext.korpus.cz/filter?within=0&q=~YVQLpvoE&ctxattrs=word&pagesize=40&refs=doc%2C%3Dopus.nazev&viewmode=kwic&attrs=word%2Clemma%2Ctag&corpname=omezeni%2Fsyn2010&structs=s&attr_allpos=kw&pnfilter=p&filfl=f&filfpos=0&filtpos=0&inclkwic=True&queryselector=cqlrow&iquery=&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=%5blemma%3D%22.*tel%22+%26+tag%3D%22NNM.*%22%5d&default_attr=word&sca_opus.nazev)=



Jak jsou tato podstatná jména utvořena?

Co je základové slovo?

Jaký mají význam?

Je-li základové slovo sloveso, které označuje činnost, jíž se zaobírá osoba pojmenovaná odvozeným podstatným jménem, pak se tažme:

Patří sem slovo *přítel/nepřítel*?

Od kterého slovesa je odvozeno?

Existuje nějaké pravidlo, jak se od sloves jména tvoří?

Co je součástí přípony/ od kterého slovesného kmene se tvoří takováto podstatná jména?

Co provází odvozování takových substantiv?

Mohou na –*tel* končit i slova, která nesplňují podmínky, které jsme formulovali výše a která přesahují vlastnosti, jež jsme na základě pozorování dat učinili?

Mohou na *tel* končit substantiva náhodně?

Jsou sufixem *–tel* tvořeny jenom jména osob?

[lc=".\*tel"]



Existují nějaké formální (odhlédnuto od významu) vlastnosti jmen osob tvořených ze sloves sufixem *-tel*, jimiž jsme na základě pozorování korpusových dat schopni zobecnit, proč např. *kostel/postel* nepatří mezi příslušná substantiva?

Mohou na *tel* končit tvary slov, která nejsou substantiva?

[lc=".\*tel"]



Je to, že přestože *–l* je jednou výraznou tvarovou koncovkou sloves v češtině, tak na *tel* žádné české l-ové příčestí nekončí, náhoda?



**Otázka pro chytré hlavičky:**

**Jak by musel vypadat infinitiv takového slovesa?**

2. Polysémní sufixy a náhodné shody řetězců na počátku/konci slov - homonymie částí sufixů (*vykrajovátko/koťátko, břicháč/zelenáč/pekáč, nadejít/nadefinovat, sedadlo/sídlo/sádlo*)

Vyhledej

[lemma=".\*átko"]

[lemma=".\*áč"]

[lemma=".\*dlo"]

[lemma="nade.\*"]

Sleduj, kdy jde o sufix/jakýa kdy jde jen o koncový řetězec, kdy jde o prefix/počáteční řetězec.

Snažte se formulovat vlastní závěry.

3. Představení nástroje *Morfio*

Zadání dotazu

Jednoduché čtení výsledků





Která dvojice do seznamu nepatří?

Co je to *kostelka*?



**Dotaz pro chytré hlavičky:**

**Existuje ještě nějaká podobná dvojice?**

**ÚKOL na příště (26. 4. 2017)**

Ve většině českých mluvnic i dalších příruček se pracuje s dvojicí sufixů *–ský/-cký*. Vyhledejte v korpusu SYN2010 adjektiva zakončená na *cký* a na *ský* a pozorujte je. Pokuste se argumentovat proč je varianta *cký* poněkud sporná. Všimněte si, od jakých základů (slov) se adjektiva zakončená na *cký* tvoří a definujte v rámci těchto slov výrazné skupiny, které mají nějaké formální i sémantické společné vlastnosti.