

CJBB75

14.10.2010

8.20 - 9.50

L11

Je možné hledat v korpusu substantiva (i jiné slovní druhy) podle vzoru ?

- Korpus →
- Implicitní atribut →
- word
- lemma
- tag
- POS
- ? vzor ?

Co je to vzor ?

- Vzor není GK (vyjádření gramatického významu závazným a stabilním systémem prostředků / formálních)
- Vzor je pouze systém formálních vlastností, z nich lze ovšem vyjít při formulaci dotazu (dotaz – formální pravidlo)

Vyberte jeden vzor subst. flexe

- např. žena
- Jak formulujeme dotaz ?
- Můžeme najít všechna substantiva rodu ženského
- Můžeme z nich vybrat všechna, jejichž základní tvar končí na –a
- Platí: V češtině neexistuje substantivum rodu ženského zakončené v n.sg. na –a, které by se neskloňovalo podle vzoru žena.

Korpus → Implicitní atribut → tag NNF.*

The screenshot shows the Bonito software interface. The window title is "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The search bar contains "Nový dotaz" and "jméno:". The search results are displayed in a table with three columns: the original text, a list of highlighted NNF tags, and the text with those tags replaced by their corresponding attribute values.

Jiří Mlčoušek ÚVODNÍ SLOVO Příběhy v této . Kolem nich se točí celý děj . Jsou to míst jsou ale mírně pozměněny . Vyprávěné svatého Huberta . Berme to tak , že jsou to , že jsou to postavy zosobňující určitou na svůj stav a cech . Hajný Vítězslav byl Vítězslav byl postavy neveliké . Oč byl v byl postavy neveliké . Oč byl v délce a postavy neveliké . Oč byl v délce a hmotnosti přírodou ošizen , o to více jej tato všemocná dárkyně života obdařila smyslem pro humor a dobře , hajný Vítězslav byl pořádný kopec Čankajšek . Snad proto , že na nose nosil proto , že na nose nosil brýle úctyhodné	knize postavy události postavy kategorii postavy délce hmotnosti přírodou dárkyně recesi srandiček brýle sily	mají dva protagonisty . Hajného Vítězslava skutečné , názvy popisovaných míst jsou se opravdu staly , i když je někdy něco zosobňující určitou kategorii starých lesáků starých lesáků , ortodoxních myslivců , neveliké . Oč byl v délce a hmotnosti přírodou a hmotnosti přírodou ošizen , o to více přírodou ošizen , o to více jej tato všemocná ošizen , o to více jej tato všemocná dárkyně života obdařila smyslem pro humor a recesi . Zkrátka a dobře , hajný Vítězslav byl . Lidé z okolí mu říkali Čankajšek . Snad úctyhodné síly a že očička za nimi vypadala a že očička za nimi vypadala malá a trochu
---	--	---

Počet výskytů: 12359807
> Query : "NNF.*"

Zobrazeno: 1+50/12359807 (0%) Řádek: 1

Korpus → Implicitní atribut → lemma .*a

- P-filtr

The screenshot shows the Bonito software interface. The window title is "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The main interface has a search bar with "P filtr" and "jméno:" fields, and a dropdown menu set to "syn2000". Below the search bar are controls for "První", "Poslední", "Od", "Pozic", "od", "< KWIC", "Do", "Pozic", "od", and "KWIC >". The main text area displays a concordance entry for the lemma ".*a". The text is split into three columns: the left column contains the original text, the middle column contains the lemma in red, and the right column contains the text with the lemma highlighted. The lemma is "knize postavy postavy postavy délce přírodou srandiček síly pusy startkou hospodě manželku přílbu ženu". The text in the left column is: "iří Mlčoušek ÚVODNÍ SLOVO Příběhy v této . Kolem nich se točí celý děj . Jsou to svatého Huberta . Berme to tak , že jsou to svůj stav a cech . Hajný Vítězslav byl Vítězslav byl postavy neveliké . Oč byl v postavy neveliké . Oč byl v délce a hmotnosti dobře , hajný Vítězslav byl pořádný kopec proto , že na nose nosil brýle úctyhodné za nimi vypadala malá a trochu šikmá . Z stále trčela dřevěná cigáršpička s hořící ze sebe nesundal ani v kanceláři , ani v , že v tomto dresu obšťastňoval i svoji manželku . Pouze prý si sundal motoristickou výstroji . Hajný Vítězslav žil se svou". The text in the right column is: "mají dva protagonisty . Hajného Vítězslava skutečné , názvy popisovaných míst jsou zosobňující určitou kategorii starých lesáků neveliké . Oč byl v délce a hmotnosti přírodou a hmotnosti přírodou ošizen , o to více ošizen , o to více jej tato všemocná dárkyně . Lidé z okolí mu říkali Čankajšek . Snad a že očička za nimi vypadala malá a trochu mu stále trčela dřevěná cigáršpička s hořící . Hubertus a ruksak ze sebe nesundal ani , ale ani v lese , i když bylo třeba třicetistupňové . Pouze prý si sundal motoristickou přílbu , která rovněž patřila k jeho celodenní a jedinou dcerou na hájence uprostřed lesů". At the bottom, the status bar shows "Zobrazeno: 1+50/6931105 (0%) Řádek: 1".

Počet výskytů: 6931105
> Query : "NNF.*"
> P-filtr: ".*a"

Zobrazeno: 1+50/6931105 (0%) Řádek: 1

Konkordance → Statistiky → Frekvenční ..

Frekvenční distribuce

Výsledek frekvenční distribuce

Limit: 0 Řádků k zobrazení: 13780

lemma: 0<KWIC zobrazit součet schovat

lemma:	##
strana	130835
doba	111053
Praha	99695
firma	83608
vláda	81887
cena	81879
koruna	80002
hodina	77122
žena	59524
cesta	57620
skupina	55898
republika	53915
ruka	53446
řada	52063
škola	49609
většina	49299
otázka	48635
banka	45879
voda	43166
služba	42783
Evropa	41121
změna	40715
hlava	40189

Zavřít Uložit

Pokuste se podobným způsobem najít substantiva dalších vzorů

- Který vzor/ které vzory substantiv by se hledaly obtížněji než uvedený vzor žena ?
- Vyberte obtížnější případ a hledejte způsob, jak jej řešit.

Analogicky lze postupovat při hledání jednotek, které jsme schopni formálně definovat

- Nemáme-li označovaný korpus, lze pro hledání tvarů využít jejich formální vlastnosti
- Slovtvorná analýza vycházející z formálních pravidel derivace
- Analýza založená na grafických vlastnostech (hledání proprií, substandardních tvarů, cizích slov atd.)