

GAK – CJBB84

čt. 12.30-14.00

3.11. 2011

Vyhledávání jevů na hranici slovník/gramatika

- tvarová homonymie – noční můra automatizace v NLP
- homonymie způsobená homonymií gramatických a slovotvorných formantů – konkrétně prefix *po-*
- otevřené seznamy – korpus jako zdroj dat pro doplnění okrajových jevů gramatiky do slovníků automatických morfologických analyzátorů

Futurum v češtině a korpusy

- dokonavá slovesa
- opisné futurum nedokonavých sloves
- syntetické futurum

Značkování gramatické kategorie času v korpusech ČNK

Pozice 9 - Čas (TENSE)

- - - neurčuje se
- F - futurum (budoucí čas)
- H - minulost nebo přítomnost (P/R)
- P - prézens (přítomný čas)
- R - minulý čas
- X - libovolný čas (F/R/P) - - neurčuje se

Implicitní atribut tag
dotazovací řádek VB.....F.*
Nfiltr lemma být

Frekvenční distribuce

Výsledek frekvenční distribuce

Limit: 0 Řádků k zobrazení: 22

lemma: 0<KWIC zobrazit součet schovat

lemma:	##
být	5302509
jít	181280
trvat	85600
jet	82167
růst	14078
nést	12481
běžet	10456
letět	5800
téci	3710
vést	2555
vést	2278
hrnout	969
cestovat	961
plout	758
lézt	755
stěhovat	406
hnát	382
řítit	22
plavat	21
plazit	19
krást	9
kvést	4

Zavřít Uložit

Korpus SYN (22)

● být	5302509	jit	181280
● trvat	85600	jet	82167
● rüst	14078	nést	12481
● běžet	10456	letět	5800
● téci	3710	vézt	2555
● vést	2278	hrnout	969
● cestovat	961	plout	758
● lézt	755	stěhovat	406
● hnát	382	řítit	22
● plavat	21	plazit	19
● krást	9	kvést	4

Korpusy ?

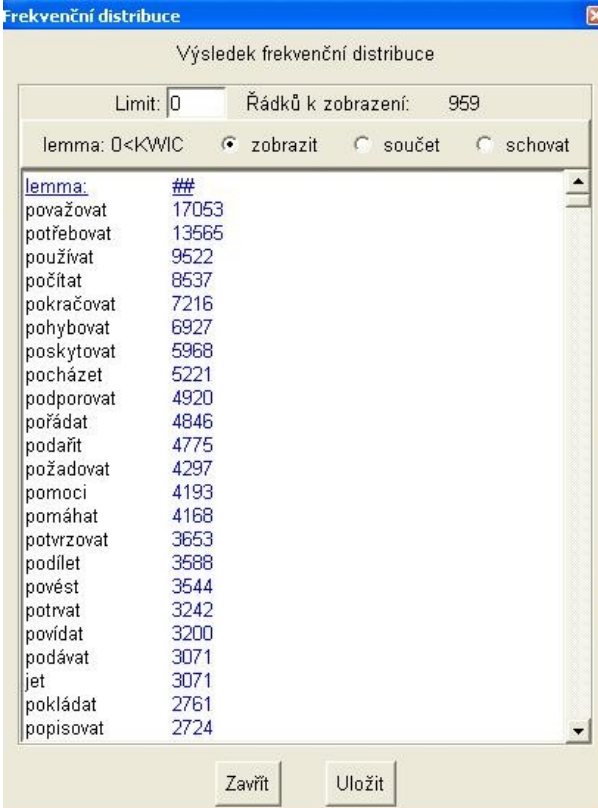
- Pokuste si vzpomenout na další možnosti.
- Ověřte, zda se vyskytují v korpusech.

Implicitní atribut **word**
dotazovací řádek **po.***

Implicitní atribut **tag**

Pfiltr **VB.***

Konkordance – statistiky – **lemma**



Frekvenční distribuce

Výsledek frekvenční distribuce

Limit: 0 Řádků k zobrazení: 959

lemma: 0<KWIC zobrazit součet schovat

lemma:	##
považovat	17053
potřebovat	13565
používat	9522
počítat	8537
pokračovat	7216
pohybovat	6927
poskytovat	5968
pocházet	5221
podporovat	4920
pořádat	4846
podařit	4775
požadovat	4297
pomoci	4193
pomáhat	4168
potvrzovat	3653
podílet	3588
pověst	3544
potrvat	3242
povídat	3200
podávat	3071
jet	3071
pokládat	2761
popisovat	2724

Zavřít Uložit

poválí

? povalit ? valit

The screenshot shows the Bonito software interface. At the top, there is a menu bar with options: Manažer, Korpus, Dotaz, Konkordance, Zobrazení, Výběr, and Nápověda. Below the menu bar, there is a search bar with the text "Nový dotaz" and a dropdown menu for "jméno:" containing the value "syn2005". The main area displays a concordance search for the word "poválí". The results are shown in a table with columns for the word, its grammatical form, and the context. The word "poválí" is highlighted in red in the original image. The context is a paragraph of text from a document. At the bottom of the window, there is a status bar that reads "Zobrazeno: vše/15 Řádek: 12 Vybráno: 5".

Text	Forma	Text
na mávání opilýho hašíšáka , ale	poválí/povalit/VB-S---3P-AA---P	ho a poválí to dál . Příkladali m
opilýho hašíšáka , ale poválí ho a	poválí/povalit/VB-S---3P-AA---P	to dál . Příkladali mu , aby si se
víc než pět minut , až se přes ně	poválí/povalit/VB-S---3P-AA---P	povodeň . Na to vám přísahán
e jenom poddej proudu , voda tě	poválí/povalit/VB-S---3P-AA---P	zároveň s kamením , ale tišně
dy , jež , nemajíce doma co jíst ,	poválí/povalit/VB-S---3P-AA---P	se na Západ , vylupující kraj .
řejít ulice bez obavy , že tě něco	poválí/povalit/VB-P---3P-AA---P	nebo zahltí spálenou naftou . 7
o zase začne . Od zítřka se zase	poválí/povalit/VB-P---3P-AA---P	transporty , tak zase do toho .
ol , domnívající se , že se potom	poválí/povalit/VB-S---3P-AA---P	vlastní důležitostí sama . Obra
ní týden , pak tě Indie přemůže ,	poválí/povalit/VB-S---3P-AA---P	tě svou vahou . Čím déle cest
ní době je stále častěji vytrhají a	poválí/povalit/VB-P---3P-AA---P	na zem . Řidiči , obzvlášt' poka
ávažnější kolize . " Kolikrát nám	poválí/povalit/VB-S---3P-AA---P	plot nebo i polámou stromy a p
ologové věděli , že se na Čechy	poválí/povalit/VB-S---3P-AA---P	shůry mimořádné množství vo

povede

? *povést* ? *vést*

„Zůstaň u nás, budeš nám vařit,
a když všechno budeš dělat,
povede se ti dobře !“

(Jan Vodňanský: Analýza pohádky o Sněhurce a sedmi
trpaslících)

Mluvnice & korpusy

- Seznamy uváděné v příručkách (viz bibliografie) jsou rozsáhlejší než uvedený seznam
- Předpoklad: tvoří patrně otevřené množiny
- Závěr : korpusy mohou přispět k jejich rozšíření
- Korpusy mohou zpřesnit stávající popisy

Implicitní atribut **word** dotazovací řádek **povede** Zobrazení – atributy – **lemma+tag**

The screenshot shows the Bonito software interface. At the top, there is a menu bar with options: Manažer, Korpus, Dotaz, Konkordance, Zobrazení, Výběr, and Nápověda. Below the menu bar, there is a search bar with the text "Nový dotaz" and a dropdown menu. To the right of the search bar, there is a field labeled "jméno:" with a dropdown menu and a field containing "syn2000".

The main area of the interface displays a list of search results. Each result consists of a snippet of text followed by the word "povede" and its grammatical attributes: /pověst/VB-S---3P-AA---. The text snippets are:

- Kolikrát s ní musíte někdy jít, než se to ale už nejsou. Něco nebo někdo přijde a
- ím v žádné výpovědi, ale třeba se nám to ležet a čekat, jestli se transplantace rekonstruovat. Nyní věřím, že se nám to
- na thumič. Nevadí, příště se to určitě jenom vymetám smetl a pochybuji, že to řici, že bez ohledu na to, jakým směrem
- letvař se tak pochybovačně. Třeba se to řejší duchovní a mravní impotenci národa
- totální nasazení, které, jak věděli, kdo uší, slyš. Jestliže kdo do vězení
- nejakou volbu? Má - li kdo uší, slyš. Jestliže kdo do vězení
- půjde; zabije - li kdo mečem, musí mečem zabit být

At the bottom of the interface, there is a status bar that reads: "Zobrazeno: 1+50/2862 (1%) Řádek: 12 Vybráno: 5".

Řadu sloves, která uvádějí tištěné zdroje v korpusech nejsou

- *capat/pocapá*
- *drát se/podere*
- *jechat/pojechá*
- *kulhat/pokulhá*
- *kvačit/pokvačí*
- *lít/pole(i)jeje*
- *loudat se/poloudá*
- *pršet/poprší*
- *rázovat/porázuje*
- *štrádovat/poštráduje*
- *vinout se/povine*
- *vrávorat/povrávorá*
- *cupat/pocupá*
- *chvátat/pochvátá*
- *klíčit/poklíčí*
- *kulit se/pokulí*
- *kvapit/pokvapí*
- *linout se/poline*
- *pálit/popálí*
- *rajzovat/porajzuje*
- *šířit se/pošíří*
- *vát/pověje*
- *viset/povisí*

korpusy dokládají navíc (tag=„VB.*“)

10 sloves + 1 BMK

- *vanout/povane*
- *svištět/posviští*
- *kráčet/pokráčí*
- *táhnout/potáhne*
- *mazat/pomaže*
- *valit (se)/poválí (se)*
- *pást (se)/popase (se)*
- *šinouť (se)/pošine (se)*
- *trénovat/potrénuje*
- *trvat/potrývá*
- *péct/popeče (s kým)*

korpusy dokládají (tag!=„V.*“)

9 sloves (doložena též na internetu)

- šupat/pošupe
- brodit se/pobrodí se
- crčēt/pocrčí
- šupajdit/pošupajdí
- harcovat/poharcuje
- koučovat/pokoučuje
- veslovat/povesluje (SYNPUB2006)
- šupačit/pošupačí (SYNPUB2006)
- prýštit/poprýští (SYNPUB2006)

Zastoupení lemmat v tištěných zdrojích, korpusech a na internetu

tištěné zdroje	60 lemmat	72,29 %
pouze korpusy	20 lemmat	24,1 %
korpusy i tištěné zdroje	37 lemmat	44,58 %
pouze tištěné zdroje	23 lemmat	27,71 %
pouze internet	3 lemmata	3,61 %
tištěné zdroje + korpusy	83 lemmat	100 %

Upřesnění definice významu sloves schopných tvořit tvary syntetického futura

- Pohyb v konkrétním slova smyslu (přemísťování z místa na místo jednosměrně/ v určeném směru) : (*půjde, pojedě, poběží, poplazí se, pokrácí, pošine se, povalí, pomaže, ...*)
- pohyb bez určení směru (různými směry) : (*pošíří se, poline se, ...*).
- Proces zahrnující pohyb
- a) přírodní (*povane, pověje, pol(i/e)je, poprší, ...*);
- b) s neživým nositelem (rostlina: *pokvete, poroste, poplazí se*, strom: *ponese (ovoce), ...*; hudba: *poline se, povalí ...*; vůně: *pošíří se, poline se, ...*; voda: *poteče, povalí, pocrcí, poprýští...*, zpráva: *pošíří se, ...*);
- c) mentální (*popálí mu to*)
- d) s živým nositelem (*poroste, popase se, ...* pták: *ponese (vejce), ...*);
- e) ve významu fungovat (v přeneseném významu *jít*) (správně) *poklape, pošlape, povalí ...*; vycházet s někým, (*poklape, pošlape, popeče, ...*).
- f) ve významu vést (*pokoučuje, potrénuje, povládne*)

Gramatika a korpus

- rozšíření seznamu sloves, která mohou v češtině vyjadřovat futurum synteticky
- upřesnění definic jejich významů
- stabilní jádro
- proměnlivá periferie

Bibliografie

- ČECHOVÁ A KOL. (1996): *Čeština, řeč a jazyk*. Praha : ISVN.
- ČERMÁK, F. (1990): *Syntagmatika a paradigmatica českého slova*. Praha : Karlova Univerzita.
- KOMÁREK, M. a kol. (1986): *Mluvnice češtiny II.*, Praha : Academia.
- OSOLSOBĚ, K. (2007): Syntetické futurum v češtině – gramatiky, slovníky, korpusy. *Přednášky a besedy z XL. běhu LŠSS*. Brno, s. 131-144.
- OSOLSOBĚ, K. (2007): Popis gramatických významů jednoduchých slovesných tvarů v anotacích českých (slovenských) korpusů. *SPFFMU A, 55, v tisku*.
- OSOLSOBĚ, K.: Značkování gramatických kategorií v korpusech ČNK a jejich zachycení v gramatice a ve slovníku (syntetické futurum, stupňování adjektiv, neurčité číslovky a příslovce míry). In Štícha, F. *Grammar & Corpora / Gramatika a korpus 2007*. Academia : Praha, 2008, s. 407-416.
- ŠMILAUER, V. (1972): *Nauka o českém jazyku*. Praha : SPN.
- Elektronická verze *Slovníku spisovného jazyka českého (SSJČ)* a *Slovníku spisovné češtiny pro školu a veřejnost (SSČ)* dostupná z <chrome://debdict/content/debdict.xul>
- *Český národní korpus - SYN2000, SYN2005, SYN2006PUB, BMK, PMK, ORAL2006*. Ústav Českého národního korpusu FF UK, Praha 2000. Dostupné z WWW: <http://ucnk.ff.cuni.cz>. *Morfologický analyzátor Ajka* dostupný z <http://nlp.fi.muni.cz/projekty/ajka/ajkacz.htm>.

Úkol na 10.10. 2011

- Vyhledejte v korpusu SYN tvary slovesa *mazat* (*pomažu, ...*)v syntetickém futuru a ve futuru opisném (*budu mazat, ...*).
- Porovnejte frekvenci výskytu synonymních forem.