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

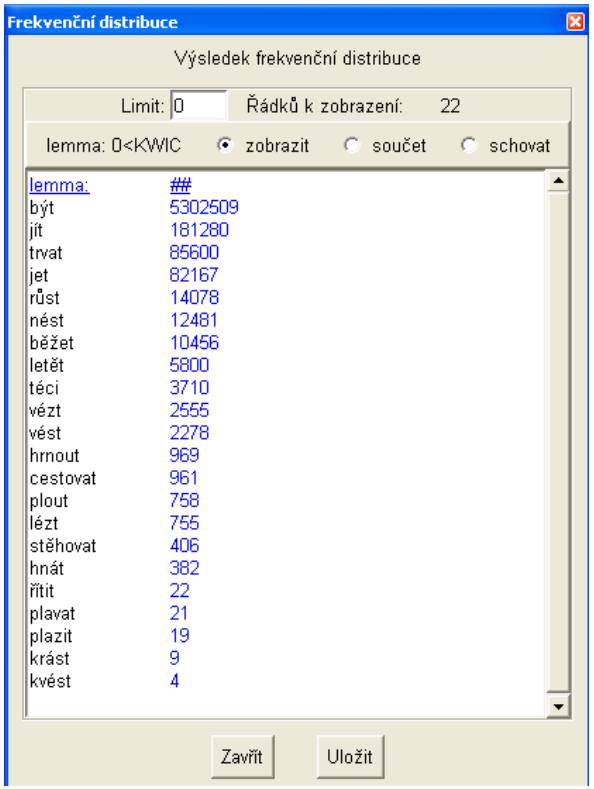
Futurum v češtině a korpusy

- 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



Korpus SYN (22)

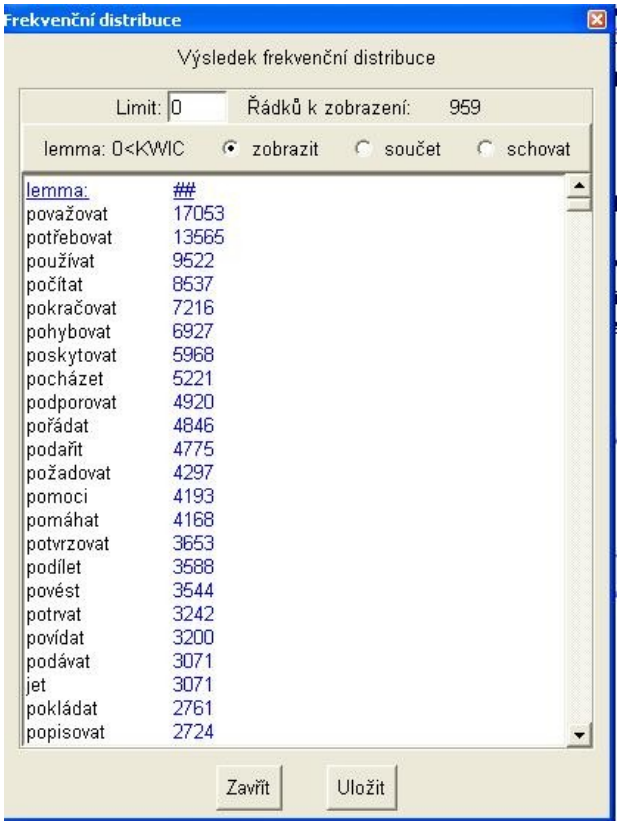
být	5302509	jít	181280
trvat	85600	jet	82167
růst	14078	nést	12481
běžet	10456	letět	5800
téci	3710	vézt	2555
vést	2278	hrnout	969
cestovat	961	plout	758
lézt	755	stěhovat	406
hnát	382	řítit	22
plavat	21	plazit	19
krást	9	kvést	4

CJBB84

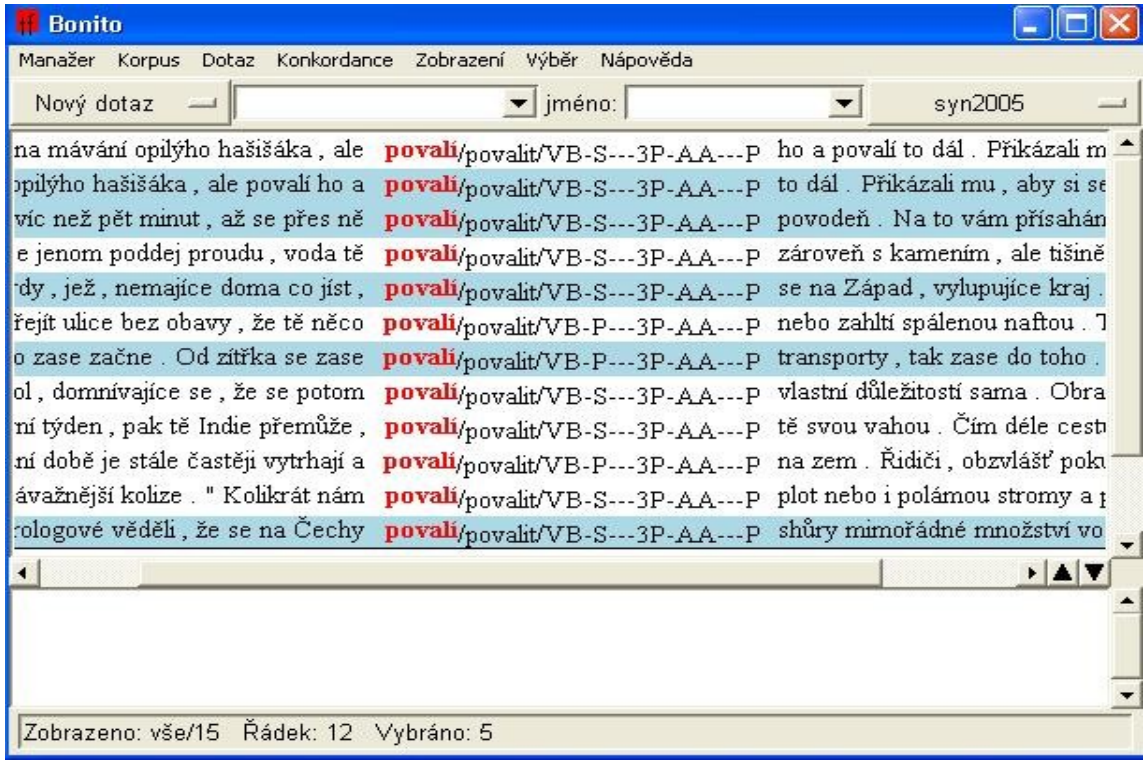
Korpusy ?

- 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



povalí
? povalit ? valit



povede
? povést ? vést

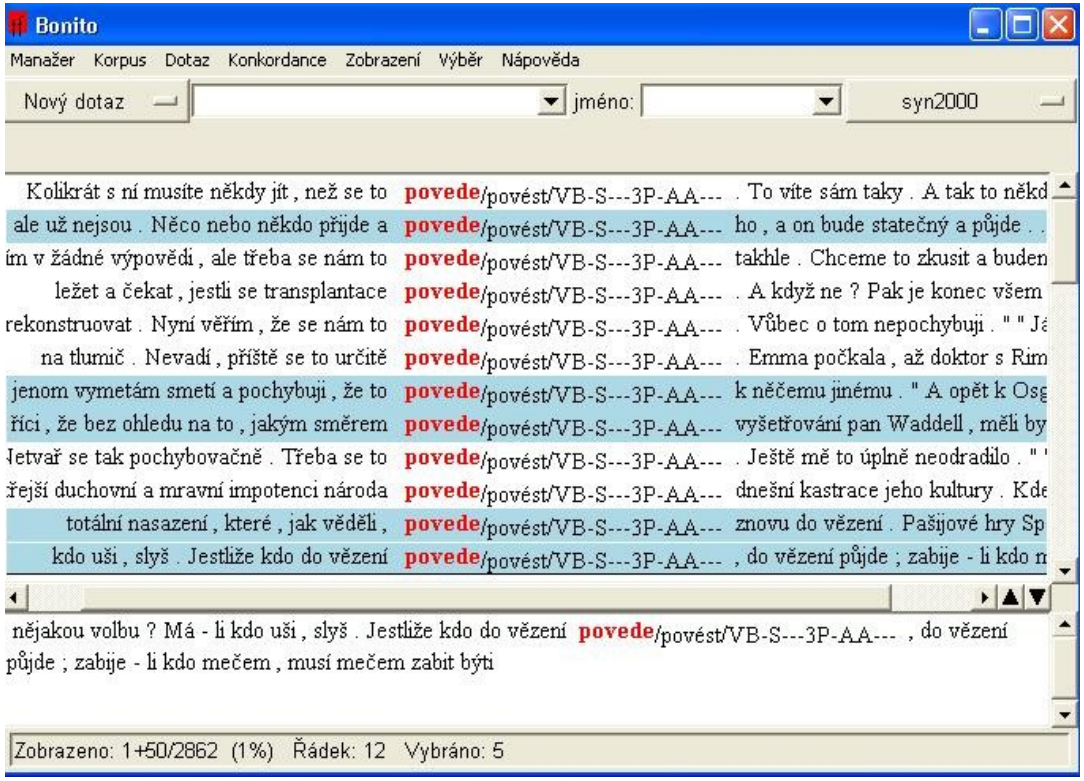
„Zůstaň u nás, budeš nám vařit,
a když všechno budeš dělat,
povede se ti dobře !“

(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



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)/povalí (se)*
- *pást (se)/popase (se)*
- *šinout (se)/pošine (se)*
- *trénovat/potrénuje*
- *trvat/potrvá*
- *péct/popeče (s kým)*

korpusy dokládají (tag!=„V.*“)
9 sloves (doložena též na internetu)

- šupat/pošupe
- brodit se/pobrodí se
- crčet/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, pojede, poběží, poplazí se, pokráčí, 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, pocrčí, 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 paradigmatika 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.