

Korpusy jako zdroje dat pro doplnění obrazu okrajových gramatických jevů

Klára Osolsobě
osolsobe@phil.muni.cz

Značkování korpusů je dobrá věc

ALE

- lingvistické předpoklady : morfologický slovník + tagset
- technická řešení (rozdělení na pozice, různé nelingvistické přístupy k disambiguaci)

Kvalita značkování z lingvistického hlediska

- kvalita zdrojů
- možnosti a meze formalizace
- úplnost výčtu některých okrajových jevů

Futurum

- analytické futurum (slovesa nedokonavá)
- dokonavá slovesa a jejich přítomné tvary
- syntetické futurum (*budu, jít, jet, ...*)

Gramatiky a slovníky

- gramatiky: definují skupinu sloves pohybu výčtem (rozsah odpovídá rozsahu gramatiky).
- praxe slovníků je nejednotná (u některých sloves např. *vézt* se SSJČ a SSČ rozcházejí).

Česká mluvnice (Havránek, Jedlička, 1981)

- hovoří o: ... „*některých slovesech nedokonavých (většinou s významem pohybu, řidčeji stavu)*
- ... *nemají složený tvar času budoucího, leda v jiném významovém odstínu ...*
- možnost tvořit u některých tvarů imperativ ... *s malým významovým rozdílem*, který je na dvou příkladech charakterizován u slovesa *pojd'* jako ... *pohyb k mluvčímu...* u sloves *poběž*, *polet'* jako ...*počátek*
- samostatně se zmiňuje imperativ *poslyš*, který nemá oporu ve tvarech syntetického futura indikativu (*slyšet* - **poslyším* – *slyš* – *poslyš*)

Příruční mluvnice češtiny (Karlík, Nekula, Rusínová, 1996)

- hovoří o ... *zvláštní skupině sloves pohybu a některých jiných, která pohyb implikují*
- dále uvádí, že ... *u nich lze rozeznat rozdíl mezi určeností a neurčeností co do cíle pohybu... přičemž ... oba tyto významy se vyjadřují v rámci sloves nedokonavého vidu...*
- konstatuje, že ... *nejde o dokonavé futurum, od těchto forem nelze vytvořit infinitiv ani préteritum, jsou však možné formy složené ...*

Čeština řeč a jazyk (Čechová a kol., 1996)

- *U některých nedokonavých sloves (sloves pohybu a změny stavu) se tvoří ještě (míní se kromě tvarů budu + infinitiv nedokonavého slovesa) tvary jednoduché s předponou **po-**.*

Mluvnice češtiny 2 (Komárek a kol., 1987)

- ... prefix **po-** (s ojediněle se vyskytující variantou **pů-**): spojuje se s tvary ind. prez. akt. některých nedokonavých sloves, zvláště s významem jednosměrného pohybu v tzv. jednoduché, syntetické futurum ...,
- má tedy gramatickou kategoriální platnost; tato platnost je oslabena ve tvarech imperativu těchto sloves, ..., které jsou synonymní s tvary bez tohoto prefixu.
- slovesa jsou charakterizována jako slovesa jednosměrného pohybu a slovesa trvání nebo změny stavu (jež jsou rovněž pojaty jako jednosměrný pohyb)
- předpona **po-** má zde význam pouze gramatický, nikoli lexikální; nemění ani lexikální význam, ani nedokonavý vid slovesa
- vedle futurálních tvarů jednoduchých je možno tvořit i pravidelné tvary složené futura opisného,

Mluvnice současné češtiny (Cvrček a kol., 2010)

- *Specifická omezená skupina sloves má tvary budoucího času vyjádřené pomocí předpony po/pů : půjdu, pojedu, potrvá, poroste, ponesu, povedu, poběžím, poletíš, polezeš, poteče, poženu, povezeme, poplujete, pohrne, pocestujou, postěhuje, poplazí.*

Syntetické futurum – slovníky a další tištěné zdroje

- SSJČ : 50 sloves (27 navíc, které neuvádí SSČ)
- SSČ : 21 sloves (3 navíc, které neuvádí SSJČ)
- SSSJČ i SSČ : 18 sloves
- Čermák, 1990, s. 106 : 5 dalších sloves, které neuvádí ani SSJČ, ani SSČ
- Šmilauer, 1972 s. 223 : 2 další slovesa
- Kopečný, 1962 s. 46-50 : 115 sloves

vézt

SSJČ Slovník spisovného jazyka českého

véztí

ned. (1. j. vezu, rozk. vez, min. vezl, trp. vezen) (koho, co) *dopravním prostředkem dopravovat*: v nemocného sanitkou do nemocnice; vozy vezou uhlí; v obilí do stodoly, vezl si domů pěkný výdělek
přivážel; bylo mu, jako by ho vezli na popravu (*Wint.*) *odváželi*; --- *véztí se ned.*

1. (~; v čem, čím; na čem) *jet dopravním prostředkem*: někteří šli pěšky, jiní se vezli; ve voze se vezlo pět lidí; vezl se autem; vezli se na valníku; *řídč.* v se v (*řídč.* na Svět.) kočáře
2. (~; po čem) *pohybovat se, šinout se po povrchu něčeho*: v se po skluzavce, po zábradlí *klouzat*; uklouznout a v se po zádech; ruka zpět se veze (*Ner.*)
3. *hovor. expr. (~; s kým, čím) být účasten něčeho (zprav. nepříjemného) bez vlastní viny, zásluhy*: vedoucí udělá chybu a ostatní se s ním vezou *trpí tím*, jsou voláni *k odpovědnosti, trestání*; chytí tě za slovo a už se vezeš (*Zápot.*) (o) předp. *do-, na-, na-se, od- (pood-), po- (popo-), po-se, pro-, pro-se, pře-, pří-, roz-, s-, s-se, u-, u-se, vy-, za-*

vézt

SSC Slovník spisovné češtiny

vézt

ned. (1. j. vezu, rozk. /po/vez, bud. povezu, budu vézt, čin. vezl, trp. vezen; v. po-) *vozidlem přemísťovat* *dopravovat (syno)* vézt nemocného do nemocnice, rodinu autem, vézt zboží,

vézt se ned.

1. jet vozidlem, vézt se autem, na voze,
2. *hovor. expr.* účastnit se něčeho nepříjemného bez vlastní viny, byl při té výtržnosti, veze se s ostatními,

lézt

ssjc Slovník spisovného jazyka českého

lézti

ned. (1. j. lezu, rozk. lez, mín. lezl, přech. přit. leza, bud. polezu)

- 1.** *pohybovat se plazivě n. šplhavě s celým tělem přiblíženým k podkladu, po němž se pohyb děje (o člověku za pomoci nohou i rukou):* had leze z díry, hlemýžď leze pomalu; moucha leze po listu, po stropě; mravenci jim lezli po nohou; opička leze po stromě, na strom; dítě leze po podlaze po všech čtyřech; voják lezl po břiše ze zákopu; lezl ztěžka vysokou trávou; l. do stráně; l. po žebříku do okna; horolezci lezli na strmou skálu *šplhali*; přen. víno leze po zdi *plazivě roste*; [x] ob. mravenci mu lezou po zádech *mrzí ho*; expr. l. ke křížku, ke kříži *pokořovat se, uznávat svou vinu*; leze (část. plazí se) před ním po břiše *podlézá mu*; zř. neos. jako by po mně lezlo (Něm.) *mrzí mě*
- 2.** ob. a expr. *namáhavě, obtížně se pohybovat, postupovat; vléci se:* koně s těžkým nákladem sotva lezou; nemocný lezl namáhavě do schodů; vlak (povoz) sotva leze do svahu; jídlo mu leze špatně do krku *nechutná mu*; přen. odpoledne leze k nepřechkání *táhne se*
- 3.** ob. (kam) *dostávat se, vstupovat, vnikat, zprav. pomalu, namáhavě n. úporně:* leze do vozu po schůdkách; l. do šatů, do kabátu *obtížně se oblékat*; l. do pasti, přen. *vydávat se v nebezpečí*; l. někomu do stolu, do skříně *neoprávněně je otvírat a probírat se v nich*; ten člověk rád leze do soudů; leze do manželství brzo; [x] ob. leze mi do zelí *přebírá mi někoho (zprav. dívku), něco*; neví, do čeho leze *co ho čeká, řídě*. l. někomu na krk *obtěžovat ho*
- 4.** ob. expr. (kam) *přicházet nevhod, nevitán; cpát se, dolézat:* ty děti jí pořád lezou do kuchyně; klid se, co sem lezeš!; nemám na něho čas a pořád k nám leze; [x] bez peněz do hospody nelez (pořek)
- 5.** ob. expr. (s čím na koho, za kým) *přicházet se žádostí; dolézat, dotírat (na koho), obtěžovat (koho):* leze na přítele se svým úkolem; se vším na něho lezou; leze na představeného, aby ho povýšili; l. někomu za ženou *ucházet se o ni, nabízet se jí, scházet se s ní*
- 6.** (odkuď) *zvolna se objevovat, stávat se viditelným, zjevným, tlačit se ven; vylézat, vyvstávat, ukazovat se:* sněženyky lezou ze země; děti lezou první zoubky; dospívá, lezou mu vousy; leze mu palec z punčochy; z pohovky už lezou žíně; [x] expr. oči mu lezou z důlků *chtivě n. s hrůzou něco pozoruje, diví se; zhrub.* učení mu leze z krku, krkem *je mu protivné, omrzelo ho, znechutilo se mu; zhrub.* teď to leze, proč ujel *vychází najevo; hanl.* leze to z něho jako z chlupeaté deky *mluví pomalu, těžkopádně, rozpačité*
- 7.** (kam) (o věci) *dostávat se někam, ocítat se někde; vnikat, promíkat:* vlasy jí lezly do čela i do očí; sníh leze za krk; přen. expr. (o tělesných pocitech, stavech ap.) *znenáhla se projevoval, působit někde, zasahovat někoho, něco:* zima (mráz) leze do kostí, za nehty, mlha leze do těla; [x] leze na něj chřípka *jde, začíná se u něho*; leze na ni dřímota *chce se jí spát*; víno mu leze do hlavy *omamuje ho, je opojné*; léta na ni lezou *stárne*; neos. lezlo mu za nehty *záblo ho*; leze to na hlavu, na rozum *zbavuje soudnosti; dělá starosti*; to leze do peněz, do kapsy (ob.) *je drahé*
- 8.** řidě. (o věci) *slézat, ztrácet se, mizet:* po nemoci často lezou vlasy *padají*; sníh leze *taje*
- 9.** ob. ve spoj. l. nahoru, l. dolů *zvyšovat n. snižovat hodnotu; stoupat; klesat:* teploměr, tlakoměr leze výš; teplota, tlak vzduchu leze nahoru, dolů; ceny (akcie) lezou hned nahoru, hned zas dolů (o) předp. *do-, na-, o-, ob-, od-, po- (popo-), pod-, pro-, pře-, při-, roz- se, s-, s- se, u-, v-, vy- (povy-), z-, za-; naléztí (vyna-); --> nás. *lézati, *lézávati*

lézt

SSC Slovník spisovné češtiny

lézt

ned. (1. j. lezu, rozk. lez, u význ. 1-3 i po-, bud. polezu, budu lézt, čin. lezl)

1. *pohybovat se plazivě n. šplhavě* hlemýžď leze pomalu, had leze z díry, *pláží se*; dítě už leze, *dovede lézt*; lézt po kolenou, lézt na strom, *šplhat*; lézt po žebříku (nahoru, dolů), *vystupovat, sestupovat*; tlakoměr leze nahoru, dolů, *stoupá, klesá*
2. expr. *s námahou n. pomalu se pohybovat, jít vléci se (syno)*, 2 sotva leze do schodů, na kopec,
3. *dostávat se dovnitř vnikat (syno)* lézt do vody, lézt do kalhot, *oblékat si je*; lézt někomu do (jeho) věci, *neoprávněně si je prohlížet, přen. zasahovat*
4. hovor. *přicházet nevhod obtěžovat (syno)*, 1 leze k nám denně,
5. *objevovat se (syno) ukazovat se (syno)* dítěti už lezou zoubky, sněženky lezou ze země, *vyrůstají*; hřebík leze z prkna, *vyčnívá*
6. *vnikat (syno) promikat (syno)* vlasy lezou za límec, leze mi za nehty, *zebe mě [x]* leze na mě chřipka, *začínám být nemocný chřipkou*; neví, do čeho leze, *co ho čeká*; oči mu lezou z důlků, *chtivě, s hrůzou pozoruje něco*

Zachycení na úrovni morfologické značky (tagu)

- **Pozice 2 - Detailní určení slovního druhu**

B sloveso, tvar přítomného nebo budoucího času

- **Pozice 9 - Čas**

- 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)*

SYN

VB.....F.*

- tvary *budu, budeš, ...*
- několik sloves, která tvoří synteticky futurum prefixem *po-/pů-*

Frekvenční seznam lemmat

The screenshot shows a window titled "Frekvenční distribuce" with a subtitle "Výsledek frekvenční distribuce". At the top, there are controls for "Limit: 0" and "Řádků k zobrazení: 22". Below this, there are radio buttons for "zobrazit" (selected), "součet", and "schovat". The main area displays a list of lemmas and their frequencies, with the word "lemma:" and the frequency column header "##" in blue. The list includes: 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), and kvést (4). At the bottom, there are "Zavřít" and "Uložit" buttons.

| lemma: | ## |
|----------|---------|
| 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 |

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 |

Mohou korpusy přispět k poznání tohoto jevu ?

- a) Výčet získaný pomocí výše uvedeného tagu je úplný/**neúplný**.
- b) Další doklady se **vyskytují**/nevyskytují v korpusech.
- c) **Existuje**/neexistuje nějaký postup, jak je z korpusů získat/**získávat**.

Hledání kandidátů syntetického futura v lemmatizovaných korpusech

- lemma=„po.*“
- P-filtr tag=„(VB.....P.*)|(X.*)“
- N-filtrem lze odstranit lemmata na *pod-, popo-, pona-*, ..., kdy může jít o jiný prefix, kombinaci více prefixů atd.
- frekvenční seznam lemmat
- ruční výběr kandidátů
- ověření předpokladu

Frekvenční seznam lemmat

The screenshot shows a window titled "Frekvenční distribuce" with a subtitle "Výsledek frekvenční distribuce". It features a search bar with "lemma: 0<KWIC" and a "Limit: 0" field. The number of rows to display is set to "Řádků k zobrazení: 45729". There are three radio buttons for display options: "zobrazit" (selected), "součet", and "schovat". The main area contains a table with two columns: "lemma:" and "##".

| lemma: | ## |
|------------|--------|
| potřebovat | 250646 |
| považovat | 238762 |
| počítat | 184238 |
| pořádat | 136100 |
| pokračovat | 127372 |
| používat | 106252 |
| pohybovat | 103491 |
| pomoci | 97821 |
| podařit | 96442 |
| pomáhat | 86202 |
| pocházet | 85285 |
| povést | 74546 |
| podporovat | 70009 |
| poskytovat | 68891 |
| potvrzovat | 68211 |
| požadovat | 61543 |
| popisovat | 59675 |
| pokusit | 48529 |
| postavit | 46085 |
| podílet | 45868 |
| poznat | 39668 |
| podávat | 36492 |
| podpořit | 32507 |

At the bottom of the window are two buttons: "Zavřít" and "Uložit".

Příklady chybné lemmatizace

at , když ho nějaký chuligán náhodou docela líto . Jen si to vezměte , Slavík ji přeje asi dostihly , kdo ví , kudy se do mého psa ! " křičela . " Nebo vás je tím se na Křupana Dicka vrhne a tlapy . . . jak s tím může žít ? !) a a , představila jsem si , že matku teď ozek . Zítřka , ano , zítřka . Chytím ji a které mu přinesli , jej vylil , a brzy se ni ve starém tunelu ! Možná že se tu dnu . Do něčeho vrazím . Někdo mě ale protože je zakovaná v řetězech ,

| | |
|--|--|
| povál /povalit/VB-S---3P-AA---P | na zem , nebo mu nějaká mánička : |
| povál /povalit/VB-S---3P-AA---P | na přechodu , rozbije jí vajíčka , pol |
| povál /povalit/VB-S---3P-AA---P | . Běžel nahoru . Najednou třeskl vy |
| povál /povalit/VB-S---3P-AA---P | na zem . Klidte se od něj , jinak se |
| povál /povalit/VB-S---3P-AA---P | ho , a nechtěl , aby se Anna dívala |
| povál /povalit/VB-S---3P-AA---P | mě na záda . Seberu v sobě vešker |
| povalej /povalit/VB-P---3P-AA--6P | na zem a potáhnou tu singrovku sc |
| poválím /povalit/VB-S---1P-AA---P | na pohovku , na podlahu , na stůl , s |
| povál /povalit/VB-S---3P-AA---P | jako záplava ; jeho proud je však ol |
| povál /povalit/VB-S---3P-AA---P | voda - a chvíli to bude pořádně špi |
| povál /povalit/VB-S---3P-AA---P | . " " Ne , " řekl Ron . " Uvidíš koler |
| poválíš /povalit/VB-S---2P-AA---P | ji prostě na zem a . . . - Ale já žádn |

Zobrazeno: 1+50/323 (15%) Řádek: 10 Vybráno: 3

Heslo *trvat*

ssjc Slovník spisovného jazyka českého

trvati

(+tráti Čel) ned.

1. *stále být (ve význ. I l), nepřestávat, existovat*: bouřka ještě trvá; platnost zákona trvá; pojmy, jimiž trvá svět (Olb.); expr. kdo už ani nežije, kdo jenom trvá (Pujm.) *živoři*, [x] nic netrvá na světě věčně (pořek); práv. trvajicí delikt *kt. se nepromlčuje*
2. *být, konat se po urč. čas*: t. hodinu, nejistota trvala jen krátce; netrvalo dlouho a vrátil se; to jí to trvá!; dlouho trvajicí potlesk
3. (v čem; na čem; při čem; kde) *setrávat, prodlévat l, zůstávat (po urč. čas v jistém stavu, poloze n. na urč. místě)*: t. v mlčení (Svob.); t. na modlitbách (V. Mrš.); t. při svém mírném chování (Svět.); t. v bojových zástupech anebo na hradě (Tyl)
4. (na čem) *vytrvale se přidržovat něj, názoru, neustupovat od něčeho; stát l0*: t. na svém, na svém výroku *setrávat*; na tom netrván; t. na plnění dodávek; t. na ujednaných podmínkách
5. *trvám poněk. zast. v platnosti přisl. (jako ustálenou vsuvku není třeba oddělovat čárkami) jsem přesvědčen, myslím, soudím*: knihovna obsahuje dosud (.) t. (.) jen asi dvanáct svazků (o) předp. **do-**, **po-**, **pro-**, **pře-**, **se-**, **vy-**; --> nás. **trvávati** (o) předp. v. trvati kromě **do-**, **pro-**

SSC Slovník spisovné češtiny

trvat

ned.

1. *stále být existovat (symo), l nepřestávat (symo)* platnost zákona trvá,
2. *konat se po urč. dobu probíhat (symo)* trvalo to hodinu, to ti to trvá (dlouho), netrvalo dlouho a vrátil se,
3. *neustupovat od něčeho stát@l (symo), l0* trvat na svém, na splnění slibu, [x] nic netrvá věčně, (pořek.);

trvání -í s

trvajicí příd. dlouho trvajicí potlesk,

Heslo *potrvat*

SSJc Slovník spisovného jazyka českého

potrvati

dok.

1. *něj. dobu trvat*: prohlídka zámku potrvala až do večera; potrvalo chvíli, než se vzpamatoval

+

2. *něj. čas setrvat*: potrvám tu několik let (Svět); --- **potrvá** ned. bud. *k trvati* (v. po- III 1) (*vedle* bude trvat)

SSC Slovník spisovné češtiny

potrvat

dok. *něj. dobu trvat* pěkné počasí ještě potrvá,

SYN2000

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 showing "jméno:" and "syn2000". The main area displays a concordance search for the word "potrvá". The results are shown in a table with three columns: the original text, the word "potrvá" in red, and its grammatical analysis "/potrvat/VB-S---3P-AA---". The text on the left is partially visible, showing a paragraph about a man and his son. The text on the right is also partially visible, showing a paragraph about a man and a woman. At the bottom, there is a status bar that reads "Zobrazeno: 1+50/3511 (1%) Řádek: 1".

| Text | Word | Analysis |
|---|--------|--------------------------|
| nohy . " Asi to ještě nějakou dobu | potrvá | /potrvat/VB-S---3P-AA--- |
| hodinu najít někde volný stolec (a | potrvá | /potrvat/VB-S---3P-AA--- |
| ychlíku . Ale jak dlouho ta možnost | potrvá | /potrvat/VB-S---3P-AA--- |
| ami , ale jen tak dlouho , jak dlouho | potrvá | /potrvat/VB-S---3P-AA--- |
| á tak to Rudolf řekl . Jak dlouho to | potrvá | /potrvat/VB-S---3P-AA--- |
| tec se dopustil jedné velké chyby a | potrvá | /potrvat/VB-S---3P-AA--- |
| ouř , " rozdělím si svůj život , dokud | potrvá | /potrvat/VB-S---3P-AA--- |
| otec . Synovi to také nějakou dobu | potrvá | /potrvat/VB-S---3P-AA--- |
| kl Billy , " můžu ti to dát . Chvilí to | potrvá | /potrvat/VB-S---3P-AA--- |
| ně otevíral oči . Cožpak tahle bída | potrvá | /potrvat/VB-S---3P-AA--- |
| avení , ten ji učiní šťastnou , tak to | potrvá | /potrvat/VB-S---3P-AA--- |
| ovor trapný ; budil však naději , že | potrvá | /potrvat/VB-S---3P-AA--- |

SYN2005

Bonito

Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

Nový dotaz jméno: syn2005

| | | |
|--|---|-----------------------------------|
| řířeno z tohoto světa . Mnoho dní to už | nepotrvá /potrvat/VB-S---3P-NA---- | . Maso mu na kostech taje p |
| poň než se přežene ta bouřka . Dlouho | nepotrvá /potrvat/VB-S---3P-NA---- | . " Ne , nemohl čekat . Doje |
| lu a doprovodem smyčců . " Dlouho to | nepotrvá /potrvat/VB-S---3P-NA---- | . " " Nemají žádné peníze . " |
| mě , když prohlásil , že mu celý zákrok | nepotrvá /potrvat/VB-S---3P-NA---- | déle než hodinu . " " Pochop |
| " Jak jsem vyrozuměl , celá procedura | nepotrvá /potrvat/VB-S---3P-NA---- | dlouho . Nebo se mýlím ? " " |
| řovala ho teta . " V každém případě to | nepotrvá /potrvat/VB-S---3P-NA---- | dlouho a budeme ve Španěls |
| bych si , aby to mohlo ještě nějaký čas | potrvat /potrvat/Vf-----A----- | , " řikala , " vím však , že to r |
| o příjemné bydlení ; nu , také to dlouho | nepotrvá /potrvat/VB-S---3P-NA---- | , ale nebrali by všechny ty ne |
| t . Nemysli na to , co bude . Dlouho to | nepotrvá /potrvat/VB-S---3P-NA---- | a ze západu dorazí sem na h |
| s opisovala celé odpoledne a ono mi to | potrvalo /potrvat/VpNS---XR-AA---- | déle , než já myslela . Měla j |
| zítra o deváté . " " Tvoje nová funkce | nepotrvá /potrvat/VB-S---3P-NA---- | dlouho , " vysvětlil Sam Celi |
| řl se po ostatních . " Domnívám se , že | nepotrvá /potrvat/VB-S---3P-NA---- | dlouho . " " Není důvodů , pr |

Zobrazeno: vše/44 Řádek: 1 Vybráno: 2

SYN2010

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 the text "syn2010".

The main area of the interface displays a concordance search for the word "potrvat". The search results are shown in a table with three columns: the original text, the word form, and the context. The word forms are highlighted in red. The context is shown in a light blue background.

| | | |
|---|------------------|--|
| by se tím stejně neřešily . " " Ty přece | nepotrvaj | ' věčně , " řekl jsem a vzal ji kolem ramen |
| keré v tomto světě jsou od věků . A navždy | potrvám | . Zanechte naděje . . . Lasciate ogni speranza |
| hanby sen v pekle jeho života . Ten sen | potrvá | , to žena dobře věděla , tak dlouho , dokud |
| ! " Spi klidně , starý Vyšehrade - dokud | potrváš | ty , nad Vltavou osazen , hlaholiti bude |
| zbytky této starobylé svatyně v hradbách a | potrváš | až do dnešních dob . Když byl zastavěn |
| přemáhat velikou únavu a ospalost , která | potrvaly | tak dlouho , dokud se opět nerozednívalo |
| Procházkou se vyvinul i milenecký poměr , který | potrvala | dlouhá léta , až do doby , kdy se Kamila |
| měla dojem ponurý , který ostatně u mne | potrval | až do jeho pohřbu . Krátce před jeho smrtí |
| i aktivními účastníky , u mnohých z nás | potrval | dlouho , aspoň ne tak dlouho , jak jsem |
| byla to pravá Jarní bouře , ale jen aby to | nepotrval | , aby z toho taky vzešly činy . Trochu |
| podal stručný nástin vlastního života , | potrvalo | po celou dobu Trkalova pobytu v Rusku , |
| kousíček dal na to . Takto cesta Ježíšova | potrval | celou Zemí , celou Nocí . Když byl skončil |
| konečku ocasu až po škleb ; ten chvíli ještě | potrvala | , když už ostatek zmizel . " Kočku bez |
| propaganda potrvá v Galerii U Křížovníků | potrval | do 30 . března . Reklama na absurditu Do |
| | potrvá | |

Počet výskytů: 16
> Query : "potrvat"

Zobrazeno: vše/16 Řádek: 13 Vybráno: 9

Výsledky získané z korpusů (20) (ne SSJČ a SSČ)

- *pofrčí/frčet* (**pověje/vát*)
- **pomaže/mazat* **poválí (se)/valit (se)*
- *pokrácí/kračet* **popase se/pást se*
- **pošine se/šinou se* *posviští/svištět*
- **potáhne/táhnout se* **potrénuje/trénovat*
- **popeče s kým/péci* *pošupe/šupat*
- *pobrodí se/brodit se* *pocrčí/crčet*
- *pošupajdí/šupajdit* *poharcuje/harcovat*
- *pokoučuje/koučovat* **povesluje/veslovat*
- *poprýští/prýštit* *pošupačí/šupačit*

Výsledky získané z internetu (3) (ne SSJČ a SSC)

- *potryská/tryskat*
- *potrtá/trtat*
- *pozdrhá/zdrhat*

pořčet ? pořčel.* ? (ani jeden výskyt v SYN)

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The search bar contains "P filtr" and "jméno:". The search criteria are set to "První" (First), "Pozic" (Position), and "KWIC". The search results are displayed in a table with three columns: the original text, the word and its morphological analysis, and the context. The word "pořčet" is highlighted in red in the original text. The morphological analysis for "pořčet" is shown as "pořčim/pořčet/VB-S---1P-AA---P" or "pořči/pořčet/VB-P---3P-AA---P". The context is a snippet of text from a document.

| Text | Word | Context |
|---|---|--------------------------------------|
| íkajíc lepší start . No , Quinne , tak já | pořčim /pořčet/VB-S---1P-AA---P | . " " Počkej , " řekl Quinn naléhavě |
| bslužných . Myslím , že tyhle podniky | pořči /pořčet/VB-P---3P-AA---P | , co ty na to , Gwen ? " Gwen neda |
| zůstaneme tady , spravíme auto a pak | pořčime /pořčet/VB-P---1P-AA---P | ve dne v noci za váma , až vás doh |
| enal Jack . " Teď mi to podepište a já | pořčim /pořčet/VB-S---1P-AA---P | . " Muž napsal své jméno . Přestož |
| pečně dopředu a budu si vyprávět a | pořčime /pořčet/VB-P---1P-AA---P | do New Yorku . Vedle mě si sedu |
| se bojíš řídit , vím , že si ta věčička | pořči /pořčet/VB-S---3P-AA---P | dál sama od sebe a že nesletíš do p |
| ně dohlížet . " " Já vám podám flintu a | pořčim /pořčet/VB-S---1P-AA---P | na ty brambory . Kluci přece rádi t |
| smál se kapitán . " Prostě za pár let | pořčime /pořčet/VB-P---1P-AA---P | v luxusních autech vyschlým řečišt |
| řeci nepojede a když , tak nezastaví a | pořči /pořčet/VB-S---3P-AA---P | si to až do Vídně . Éra prominentní |
| pracovat do školství . Tak zítra zase | pořčim /pořčet/VB-S---1P-AA---P | do Prostějova . Začínám si uvědom |
| řičkářům - Potom si koupí mercedes a | pořči /pořčet/VB-S---3P-AA---P | si to do Mexico City po dálnici El-F |
| kej dojem , že ke konci příštího týdne | pořčim /pořčet/VB-S---1P-AA---P | do Černý Hory , aby mi nezlenivěly |

> Query : "po.*"
> P-filter: "(VB.....P.*)"|(X.*")
> P-filter: "pořčet"

Zobrazeno: 1+50/106 (47%) Řádek: 1

povanout ? povanul ?

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The search bar contains "Nový dotaz" and "jméno:" with a dropdown menu. The search results are displayed in a table with three columns: the original text, the search term, and the context. The search term is "povane/povane/X@-----" for most entries and "povanou/povanou/X@-----" for one entry. The context is a paragraph of text in Czech. At the bottom, the status bar shows "Zobrazeno: vše/42 Řádek: 1".

| Text | Search Term | Context |
|--|-------------------------|---------------------------------------|
| í hodin na kostelní věži . Pokud vítr | povane/povane/X@----- | směřem od kostela , uslyší je doce |
| e sněhu , " řekl Aragorn . " Do rána | povane/povane/X@----- | od východu , " řekl Legolas . " Od |
| ře , bude foukat západní vítr a jestli | povane/povane/X@----- | stejným směrem několik dní . . . N |
| odle předpovědi z Lincolnova letiště | povane/povane/X@----- | v nárazech ze severozápadu , nyn: |
| slední ? " " Záleží to na větru , který | povane/povane/X@----- | zítra . Rozhodně zde nemohu kotv |
| ět polní , tak kvete . Jakž vítr na něj | povane/povane/X@----- | , an' ho není , aniž ho již více pozr |
| přes den od ledu se zvedne a v noci | povane/povane/X@----- | zpět dolů rychlostí 160 km/h . Žád |
| foukat , mýnil , že na dráze nakonec | povane/povane/X@----- | nikoli toužebně očekávaný protivít |
| ni prozradil , odkud a jak silné větry | povanou/povanou/X@----- | , a tak nebylo těžké se přizpůsobit |
| e vítr změň " , netušil , že stejný vítr | povane/povane/X@----- | i daleko za břehy " černého kontin |
| Vše závisí na tom , odkud převážně | povane/povane/X@----- | vítr . Podle dlouhodobých záznamí |
| ven a samozřejmě Ostrava . Pokud | povane/povane/X@----- | pro kapelu příznivý vítr , mohli byc |

Počet výskytů: 42
> Query : "povane|povanou"

Zobrazeno: vše/42 Řádek: 1

povát ? povál ?

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 a dropdown menu set to 'Nový dotaz', a text input field, a 'jméno:' label, another dropdown menu, and a 'syn' button. The main display area shows a search result: 'íc meteorologové hlásí nové vánice , **pověje**/povát/VB-S---3P-AA---P , poprášky a sypanice . Většina sp'. At the bottom, there is a status bar showing 'Počet výskytů: 1' and '> Query : "pověj.*"'. The bottom-most status bar shows 'Zobrazeno: vše/1 Řádek: 1'.

NIKOLI: pověj

ssjc Slovník spisovného jazyka českého

***pověj**

-e ž. *vánice*: byla strašná p. (*Kosm.*); sněhové p-e (*Kar.*)

SSC Slovník spisovné češtiny

psjc Příruční slovník jazyka českého

pověj, -e f. *diaľ. vánice* Byla strašná pověj. *Kosm.* Nastaly dny mrazivé, křišťálově čistého vzduchu, doba to předcházející silné mrazy a sněhové pověje. *Kar.*

pověje, -e f. *zast. oční víčko*. Pavel před Ananiášem stál s očima zavřítými a zavřítými povějemi či víčkami. *Suš. D Zř. dveřní křídlo*. Třeskem rozletěly se pověje nedalekých dveří. *E. Jel.*

pověji *ned. vyjadřuje futurum k slovesu váti*. Výborná vůně ze mstu a vína pověje ti [nymfě] již venku vstříc. *Hanka*

†povějka, -y f. *stuha* Hedvábí nešustilo, povějky vybledly. *Luž.*

ALE: na internetu

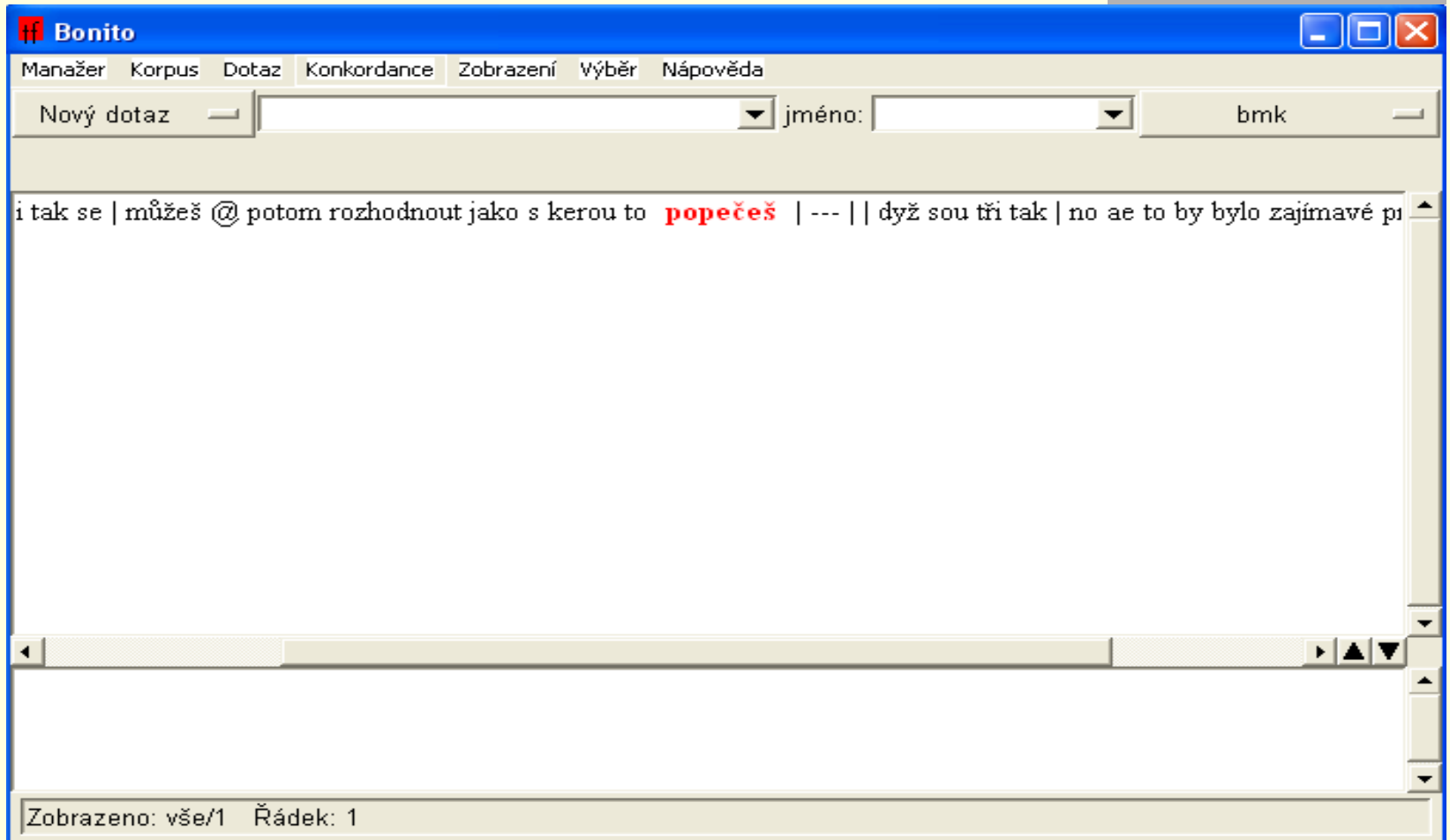
- ... Možná by neškodilo, kdyby vítr, který nám přinese novou vládu, zavál i do těchto koutů. Uvidíme, jakým směrem **pověje** i na našem resortním ministerstvu, které přislíbilo o tomto tématu jednat. ...
- **Pověj**, větříčku, **pověj**,
moju hlavěnku ověj!
Moja hlavěnka mladá
větříčkovi je ráda.
- ... **Pověje-li** vítr od slunce východu, mříti bude dobytek bez plodu. **Pověje-li** vítr od slunce západu, králové, páni nemocní budou, ...

mazat NEBO pomazat

The screenshot shows the Bonito software interface. At the top, there are menu items: Manažer, Korpus, Dotaz, Konkordance, Zobrazení, Výběr, and Nápvěda. Below the menu is a search bar with 'P filtr' and a dropdown menu. The search criteria are 'jméno:' and 'syn'. The search results are displayed in a table with columns for the text snippet, the verb form, and the context. The verb forms are highlighted in red. The status bar at the bottom indicates 'Zobrazeno: 1+50/319 (15%) Řádek: 11 Vybráno: 7'.

| Text snippet | Verb form | Context |
|---|---|---------------------------------------|
| .. " Ty si dáš tak akorát zpátečku a vy . . . Ale já vám říkám , že odsud | pomažeš /pomazat/VB-S---2P-AA---P | domů , " zasyčela . " Už jsme kvůli |
| " " Vy a já , " prohlásil šerif , " teď | pomažete /pomazat/VB-P---2P-AA---P | tak rychle , až se vám bude za pata |
| ! Tož si počkej na vdolečky ! Jen je | pomažeme /pomazat/VB-P---1P-AA---P | do hotelu . Máte tu vůz ? " Odvětil |
| i zůstanu vzhůru , ale vy , prckové , | pomažu /pomazat/VB-S---1P-AA---P | povidly a posypu perníkem . " " A t |
| orku . Za ouškem (už zbledlo) se | pomažete /pomazat/VB-P---2P-AA---P | pěkně do postele . " George mu dal |
| " " odpověděli mu neomaleně . " Vy | pomaže /pomazat/VB-S---3P-AA---P | duchy Páté Avenue a je hotova už |
| zbožné sebrance bělocha , který se | pomažete /pomazat/VB-P---2P-AA---P | s Yossarianem - zpátky do služby ! |
| hy , strčím klíče zpátky do kapsy a | pomaže /pomazat/VB-S---3P-AA---P | barvou a cejchuje celý tvůj arzenál |
| evklouzne , a to ani když si předtím | pomažu /pomazat/VB-S---1P-AA---P | co nejrychleji odtud . Však tentokrát |
| ji chytl za paži . " Tak a teď hezky | pomaže /pomazat/VB-S---3P-AA---P | úd ambrou . Občas na Priscillu vzp |
| je a uhníst z nich těsto , kterým pak | pomažeš /pomazat/VB-S---2P-AA---P | zpátky , " řekl , " a hezky s úsměve |
| | pomažu /pomazat/VB-S---1P-AA---P | celé devatenácté století , jako se pa |

bmk



ORAL2008

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The search bar contains "P filtr" and "jméno: oral2008". The search criteria are set to "První" (First), "Od 0", "Pozic" (Positions), "od < KWIC", "Do 0", "Pozic", "od KWIC >". The main display area shows a list of concordance results. The word "posere" is highlighted in red in the original image. The selected entry is: "nás to je jiný , to zajíždí do Biliny a **posere** se ke mně a já budu na myslivně nebo vobraceně". The status bar at the bottom indicates "Zobrazeno: vše/15 Řádek: 7 Vybráno: 1".

| | | |
|--|----------------|---|
| no . před koncem roku , ty jo , já se (smích) sem myslel , že se z toho snad z toho stavebně-historickýho průzkumu tak topit celou noc . hochu , já se asi na pravej pout ? jo . no . já se z toho je budeme mít , tak , tak se Američani | poseru | . je to brutální , no . kopy ti furt do |
| nás to je jiný , to zajíždí do Biliny a | poseru | , jo . hmm . ale jako eště sme spolu nechrápali |
| tamhle na ně čekaj , v zámoří , že se zalehneš , tak já za půl hodiny si to zas si nejdou nakoupit voni , co já se tam nima ? a dybych je tu měla vobě , tak se červený . smřčky , stříbrný . máme . já se ale jen ty sedmy , našťestí . (smích) se nepoblj . neboj , já se poblju , až tě | poseru | , mě to ... já furt normálně všude |
| | poseru | . seš debil , vole ? (smích) bych musel |
| | poseru | , voni si klidně přehoděj vo tejden pouť |
| | poseru | , řeknou , že všichni sou teroristi , |
| | posere | se ke mně a já budu na myslivně nebo vobraceně |
| | poserou | , jen , jen , jen vodevřou dveře , jo |
| | poseru | . no , a proč musíš v půl třetí ? sem |
| | poseru | ? " no , pak eště mi vynadala , protože |
| | poseru | . no a Veruna byla jak pomínutá . k nim |
| | poseru | . hraj , táto , prosim tě . Jindro ! hmm |
| | poseru | . to ani nedělej --- . hlavně se nepoblj |
| | poseru | . pudem domu , jo . to dolíznu a dem . |

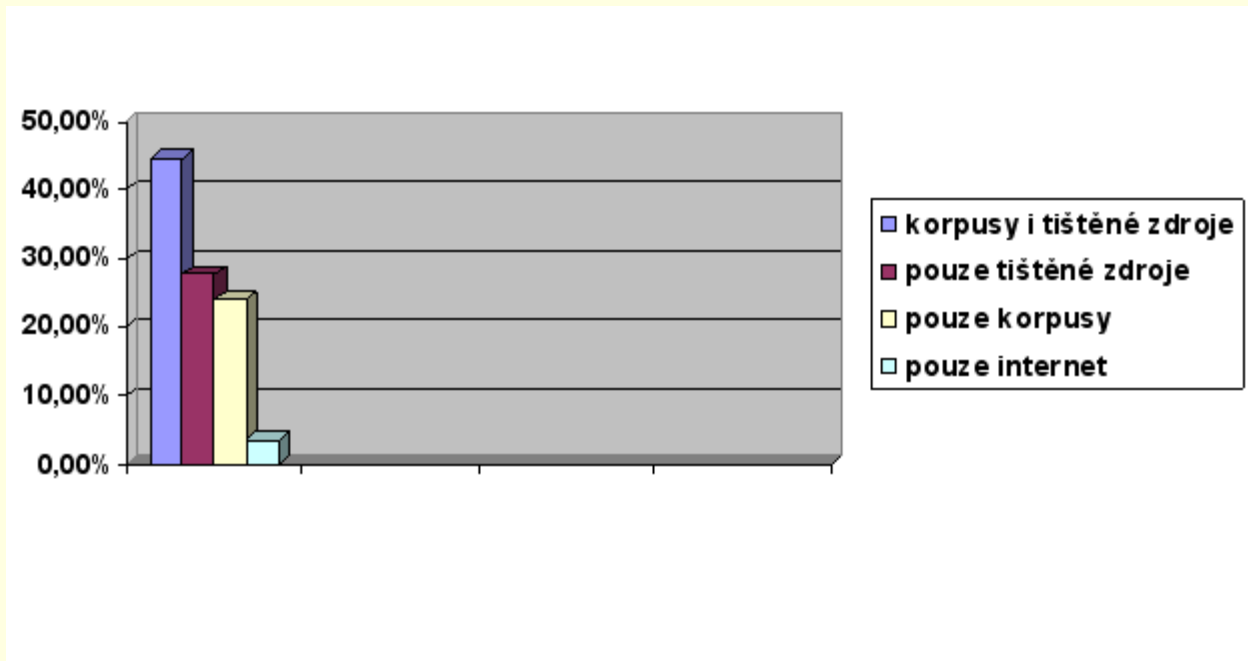
stavíte ... : " ... u nás to je jiný , to zajíždí do Biliny a **posere** se ke mně a já budu na myslivně nebo vobraceně , jo ? to je

Zobrazeno: vše/15 Řádek: 7 Vybráno: 1

Zastoupení lemmat v tištěných zdrojích, korpusech a na internetu

| | | |
|-------------------------------------|-------------|---------|
| slovníky a tištěné zdroje | 53+7 lemmat | 72,29 % |
| pouze korpuse | 20 lemmat | 24,1 % |
| korpuse i slovníky a tištěné zdroje | 37 lemmat | 44,58 % |
| pouze slovníky a tištěné zdroje | 23 lemmat | 27,71 % |
| pouze internet | 3 lemmata | 3,61 % |
| slovníky + tištěné zdroje + korpuse | 83 lemmat | 100 % |

Podíl dokladů z tištěných zdrojů a z korpusů



Hodnocení

- seznam lze nadále doplňovat (slovníky, gramatiky, učebnice)
- vysoká míra homonymie – problémy automatické lemmatizace a následné disambiguace
- další možnosti zkoumat poměr syntetického a analytického vyjádření futura (rozsah korpusů)
- sémantické vymezení skupiny sloves

Upřesnění definice významu sloves schopných tvořit tvary syntetického futura

- **Pohyb v konkrétním slova smyslu** (přemísťování z místa na místo jednosměrně/ v určeném směru) : (*půjde, pojedje, 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, 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*)

Závěr

- otevřená třída sloves
- upřesnění definice
- opravy chybné lemmatizace a značkování

Bibliografie

- CVRČEK A KOL. (2010): *Mluvnice současné češtiny*. Praha : Karolinum.
- ČECHOVÁ A KOL. (1996): *Čeština, řeč a jazyk*. Praha : ISVN.
- ČERMÁK, F. (1990): *Syntagmatika a paradigmatika českého slova*. Praha : Karlova Univerzita.
- HAJIČ J. (1994): *Unification Morphology Grammar*. Praha : MFF UK, (disert. práce).
- HAJIČ J. (2004): *Desambiguation of Rich Inflection (Computational Morphology of Czech)*. Praha : Karolinum.
- HAVRÁNEK, B. A KOL. (1989): *Slovník spisovného jazyka českého (SSJČ)*. Praha : Academia.
- FILIPEC, J. A KOL. (2005): *Slovníku spisovné češtiny pro školu a veřejnost (SSČ)*. Praha : Academia.
- KARLÍK, P., NEKULA, M., RUSÍNOVÁ, Z. (1996): *Příruční mluvnice češtiny*. Praha : Nakladatelství Lidové noviny.
- KARLÍK, P., NEKULA, M., PLESKALOVÁ, J. (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.

Bibliografie

- KOMÁREK, M. a kol. (1986): *Mluvnice češtiny II.*, Praha : Academia.
- KOPEČNÝ, F. (1962): Slovesný vid v češtině. Praha : Nakladatelství československé akademie věd.
- 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, s. 201-218.
- OSOLSOBĚ, K. (2008): 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. (ed.) *Grammar & Corpora / Gramatika a korpus 2007*. Academia : Praha, s. 407-416.
- SEDLÁČEK, R. (2004): *Morphematic analyser for Czech*. Brno : FI MU (disert. práce).
- ŠMILAUER, V. (1972): *Nauka o českém jazyku*. Praha : SPN.

Elektronické zdroje

- SYN: Český národní korpus - SYN. Ústav Českého národního korpusu FF UK, Praha. Cit. 20. 03. 2011, dostupný z WWW: <<http://www.korpus.cz>>.
- Korpusový manažer BONITO [online]. 2011. Cit. 20. 03. 2011, dostupný z <<http://ucnk.ff.cuni.cz/bonito/>>.
- Automatický analyzátor *ajka* [online] 2011. Cit. 20. 03. 2011, dostupný z <<http://nlp.fi.muni.cz/projekty/wwwajka>>.
- Internetový vyhledávač Google [online]. 2011. Cit. 20. 03. 2011, dostupný z <<http://www.google.com/>>.
- DebDict – internetový prohlížeč slovníků umožňující mj. přístup k elektronickým verzím SSJČ, SSČ, PSC [online]. 2011. Cit. 20. 03. 2011, dostupný z: <<chrome://debdict/content/debdict.xul>>.

Děkuji Vám za pozornost !