

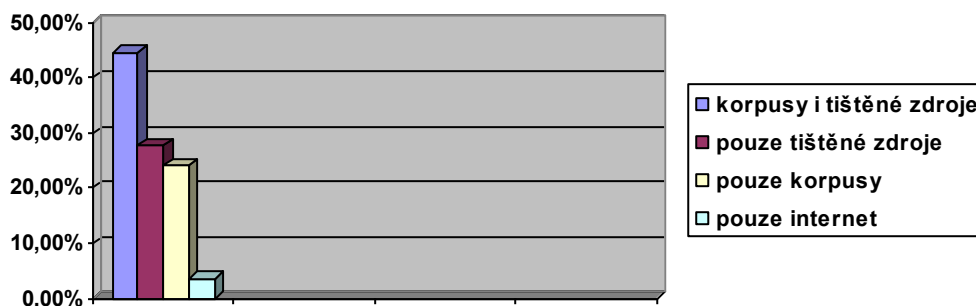
Korpusy jako zdroje dat pro doplnění obrazu okrajových gramatických jevů
Klára Osolsobě
osolsobe@phil.muni.cz
habilitační přednáška
18.10. 2011

Seznam lemmat sloves, která mohou tvořit synteticky tvary futura:

půjde/jít	*pocupe/cupat	postěhuje se/stěhovat se
*pojede/jet	pohasí si to /hasit si to	pocestuje/cestovat
poběží/běžet	*pojechá/jechat	*?potrvá/trvat
*poletí/letět	poklape/klapat	poputuje/putovat
poleze/lézt	poklíčí/klíčit	povane/vanout
popluje/plout	pokulhá/kulhat	*pomaze/mazat
poplave/plavat	pokulí se/kulit se	*poválí (se)/valit (se)
*potáhne se/táhnout	pokvačí/kvačit	pokráčí/kráčet
*poteče/téct	poline se/linout se	*popase se/pást se
*povleče/vléct	poloudá se/loudat se	*pošine se/šinou se
*požene/hnát	popeláší/pelášit	posvíští/svíštět
*ponese/nést	poplaví se/plavit se	*potáhne (se)/táhnout (es)
*povede/vést ?	poplazí se/plazit se	*potrénuje/trénovat
poveze/vézt	poplíží se/plížit se	*popече s kým/péci s kým-
*poroste/růst	porázuje/rázovat	pošupe/šupat
pokvete/kvést	pořine se/řinout se	pobrodí se/brodit se
povládne/vládnout	poštráduje si to/štrádovat	pocrčí/crčet
povrávorá/vrávorat	pověje/vát	pošupajdí/šupajdit
*pošlape/šlapat	povine se/vinout se	poharcuje/harcovat
pomašíruje/maširovat	povíší/viset	pokoučuje/koučovat
pofičí/fičet	*poleje/polije/lít	*povesluje/veslovat
?pokutálí se/ kutálet se	podere se/drát se	poprýští/prýštit
pohrne se/hrnout se	*porazí/razit	pošupačí/šupačit
pochvátá/chvátat	pošíří se/šířit se	pořfrčí/frčet
pokvapí/kvapit	*poprší/pršet	posere se s k/srát se s k
poplyne/plynout	*popálí/pálit	pozdrhá/zdrhat
pocupitá/cupitat	porazuje/rajzovat	potryská/tryskat
pocape/capat	pořítí se/řítit se	potrtá/trtat
pokluše/klusat		*pokrade (se)/krást

* - označuje případy, kde je doložena (slovník/korpusy) tvarová homonymie syntetického futura s tvary od sloves derivovaných prefixem *po-*.

Podíl dokladů sloves, která mohou tvořit futurum synteticky, z tištěných zdrojů a z korpusů:



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.
- 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, 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, PSČ [online]. 2011. Cit. 20. 03. 2011, dostupný z: <<chrome://debdict/content/debdict.xul>>.

Sloveso *frčet* (lemmatizováno jako sloveso *?pofrčet* a tedy i chybně označkováno) – doklady syntetického futura

The screenshot shows the Bonito software interface with a search query for 'pofrčet'. The results are displayed in a table format with columns for the lemma, its inflection, and the context. The search results are as follows:

Lemma	Inflection	Context
pofrčím	/pofrčet/VB-S---1P-AA---P	. " Počkej , " řekl Quinn naléhavě .
pofrčí	/pofrčet/VB-P---3P-AA---P	, co ty na to , Gwen ? " Gwen nedal
pofrčíme	/pofrčet/VB-P---1P-AA---P	ve dne v noci za váma , až vás doho
pofrčím	/pofrčet/VB-S---1P-AA---P	. " Muž napsal své jméno . Přestože
pofrčíme	/pofrčet/VB-P---1P-AA---P	do New Yorku . Vedle mě si sedne
pofrčí	/pofrčet/VB-S---3P-AA---P	dál sama od sebe a že nesletíš do př
pofrčím	/pofrčet/VB-S---1P-AA---P	na ty brambory . Kluci přece rádi br
pofrčíme	/pofrčet/VB-P---1P-AA---P	v luxusních autech vyschlým řečiště
pofrčí	/pofrčet/VB-S---3P-AA---P	si to až do Vídně . Éra prominentních
pofrčím	/pofrčet/VB-S---1P-AA---P	do Prostějova . Začínám si uvědomo
pofrčí	/pofrčet/VB-S---3P-AA---P	si to do Mexico City po dálnici El-Pa
pofrčím	/pofrčet/VB-S---1P-AA---P	do Černý Hory , aby mi nezlenivěly .

Below the table, the software shows the search statistics: Počet výskytů: 106, > Query : "pofrč.*". At the bottom, it indicates: Zobrazeno: 1+50/106 (47%) Řádek: 1.

Sloveso *vanout* (lemmatizováno jako X) – doklady syntetického futura

The screenshot shows the Bonito software interface with a search query for 'vanout'. The results are displayed in a table format with columns for the lemma, its inflection, and the context. The search results are as follows:

Lemma	Inflection	Context
povane	/povane/X@-----	směrem od kostela , uslyší je docela z
povane	/povane/X@-----	od východu , " řekl Legolas . " Odpoč
povane	/povane/X@-----	stejným směrem několik dní . . . Nikd
povane	/povane/X@-----	v nárazech ze severozápadu , nyní má
povane	/povane/X@-----	zítra . Rozhodně zde nemohu kotvit dl
povane	/povane/X@-----	, anť ho není , aniž ho již více pozná
povane	/povane/X@-----	zpět dolů rychlostí 160 km/h . Žádný s
povane	/povane/X@-----	nikoli toužebně očekávaný protivitr , a
povanou	/povanou/X@-----	, a tak nebylo těžké se přizpůsobit , "
povane	/povane/X@-----	i daleko za břehy " černého kontinentu
povane	/povane/X@-----	vitr . Podle dlouhodobých záznamů z
povane	/povane/X@-----	i Seven a samozřejmě Ostrava . Pokud

Below the table, the software shows the search statistics: Počet výskytů: 42, > Query : "povane|povanou". At the bottom, it indicates: Zobrazeno: vše/42 Řádek: 1.

Chyby v lemmatizaci a značkování slovesa *vést*

The screenshot shows the Bonito software interface with the following details:

- Menu:** Manažer, Korpus, Dotaz, Konkordance, Zobrazení, Výběr, Nápověda
- Search Filter:** P filtr (empty), jméno: (empty), syn (empty)
- Navigation:** První, Poslední, Od 0, Pozic, od, < KWIC, Do 0, Pozic, od, KWIC >
- Results:** A list of search results for the verb 'povede' and 'povedou'. Each entry includes the word, its lemmatized form (e.g., /pověst/VB-S---3P-AA---P), and a snippet of text. The results are:
 - nepovede/pověst/VB-S---3P-NA---P
 - povedou/pověst/VB-P---3P-AA---P
 - povede/pověst/VB-S---3P-AA---P
 - povede/pověst/VB-S---3P-AA---P
 - povede/pověst/VB-S---3P-AA---P
 - povede/pověst/VB-S---3P-AA---P
 - Povedeme/pověst/VB-P---1P-AA---P
 - nepovede/pověst/VB-S---3P-NA---P
 - povedou/pověst/VB-P---3P-AA---P
 - povede/pověst/VB-S---3P-AA---P
 - povede/pověst/VB-S---3P-AA---P
 - povede/pověst/VB-S---3P-AA---P
- Status Bar:** Zobrazeno: 1+50/74546 (0%) Řádek: 9 Vybráno: 4

Chyby v lemmatizaci a značkování slovesa *pověst se*

The screenshot shows the Bonito software interface with the following details:

- Menu:** Manažer, Korpus, Dotaz, Konkordance, Zobrazení, Výběr, Nápověda
- Search Filter:** P filtr (empty), jméno: (empty), syn (empty)
- Navigation:** První, Poslední, Od 0, Pozic, od, < KWIC, Do 0, Pozic, od, KWIC >
- Results:** A list of search results for the verb 'povedeme' and 'povedete'. Each entry includes the word, its lemmatized form (e.g., /vést/VB-P---1F-AA---I), and a snippet of text. The results are:
 - povedeme/vést/VB-P---1F-AA---I
 - povedete/vést/VB-P---2F-AA---I
 - povedu/vést/VB-S---1F-AA---I
 - povedu/vést/VB-S---1F-AA---I
 - povedu/vést/VB-S---1F-AA---I
 - povedeme/vést/VB-P---1F-AA---I
 - povede/vést/VB-S---3F-AA---I
 - povedou/vést/VB-P---3F-AA---I
 - povedu/vést/VB-S---1F-AA---I
 - povedu/vést/VB-S---1F-AA---I
 - povedu/vést/VB-S---1F-AA---I
 - povede/vést/VB-S---3F-AA---I
- Status Bar:** Zobrazeno: 1+50/2149 (2%) Řádek: 12 Vybráno: 1

Frekvenční distribuce

Výsledek frekvenční distribuce

Limit: 0 Řádků k zobrazení: 18

word: 0<KWIC zobrazit součet schovat

word:	##
Nepokradeš	104
nepokradeš	79
pokradli	10
pokradl	5
Pokradu	4
pokradou	4
nepokradl	3
nepokradou	2
nepokrade	2
pokrad	2
nepokrad	2
NEPOKRADEŠ	2
pokrást	1
NEPOKRADOU	1
Nepokradu	1
pokradeš	1
Pokradeš	1
pokradly	1

Zavřít Uložit

Má sloveso *krást* tvary syntetického futura ? Jak je to s tvarem <nepokradeš> ?

Existuje sloveso derivované od slovesa *krást* sufixem *po-*, tedy *pokrást*. To má význam distributivní (*ukrást + nakrást si*).

Níže uvedené tvary infinitivu a přičestí minulého jsou z větší části tvary slovesa *pokrást* ve významu distributivním.

Nicméně i mezi tvary l-ového participia se najdou příklady slovesa *pokrást* s významem „*krást*“ (ve významu biblického *krást/ukrást* : **myslím, že jsem <nepokradl> a nezabil, ale jinak hlavně to smilstvo**).

Nicméně existují i tvary *pokradu, pokradou*, které by bylo možné interpretovat jako tvary slovesa *krást* (synt. fut.).

Některé následující příklady jsou, alespoň se mi zdá, mimo biblický kontext. Např.:

opus.nazev=Lidové noviny, č. 2/1998,opus.autor=Y

**budou tak nějak sami od sebe slušní a příliš <nepokradou> .
Dosavadní vývoj je však více než poučný**

Bonito Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

N filtr jméno: syn

První Poslední Od 0 Pozic od < KWIC Do 0 Pozic od KWIC >

ukazoval jí všelijaký nože a meče a tesáky , co ! Rovněž nám , pořadatelům , někdo tajně vysoko nad hlavou , v ovoci , jen občas jsi vysoko nad hlavou , v ovoci , jen občas jsi nechali Braniboři v Čechách . Proč by si jeden synku Hynku , doma dli si ! Fendřáci sviňku nechali Braniboři v Čechách . Proč by si jeden Milka (1821.12 : " Kupid s Milkem listky něčí nenechavé ruce prohledaly pokoje a mnoha a mnoha hřichy . Myslím , že jsem zloději v Bagdádu mezi městskými chudými stolíků či sedaček pořezali , pokecali či slepice vybijeme , stejně by nám všechno se neurodily , pomrzly . Něco prý i lidé

pokrad slabším . . . jí to ostří lekalo , ale
pokradl v pokladně , zbylo v ní jenom něco na dvě
pokradl , hrozinek ze spížky , pohledů z těl :
pokradl , hrozinek ze spížky , pohledů z těl :
nepokrad ' - jó , v tom je Pepa demokrat a šikulovi
pokradli si . ERBENÁTKY Záhoř Nedaleko metropole
nepokrad ? Jó v tom je Pepa demokrat : vždyť šikulovi
pokradli . . . ") . Právě v této době - a postupně
pokradly , co se dalo . Tyto ruce okradly i hosty
nepokradl a nezabil , ale jinak , hlavně to smilstvo
pokrást , už velvyslanec nevysvětil . " Objektivní
pokradli . Táhl jsem ještě mnohem dále , kameny
pokradli . Z toho , co nás čeká , mám hrůzu . Všichni
pokradli . Majitel vítězné pálenky však neztrácí

> N-filter: ".*eš"
 > Deleted: 16
 > Deleted: 1

Zobrazeno: vše/23 Řádek: 1

Bonito Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

N filtr jméno: syn

První Poslední Od 0 Pozic od < KWIC Do 0 Pozic od KWIC >

mnoha a mnoha hřichy . Myslím , že jsem zloději v Bagdádu mezi městskými chudými stolíků či sedaček pořezali , pokecali či slepice vybijeme , stejně by nám všechno se neurodily , pomrzly . Něco prý i lidé peněženku , doklady a dámskou bundu . Poblý a . Má k tomu důvod - zloději železa totiž Olomouc * Hlíník bezmála za půl milionu korun toho může dohnat mladíka , který předtím olomouckou městskou částí Neředín . Zloději už velikou škodou . Vloupali se do stavení a když praví : nepokradeš ! Jestli předseda Čunek epokradeš ! Jestli předseda Čunek pokradl nebo jich zahubil , některé daroval a zbytek mi

nepokradl a nezabil , ale jinak , hlavně to smilstvo
pokrást , už velvyslanec nevysvětil . " Objektivní
pokradli . Táhl jsem ještě mnohem dále , kameny
pokradli . Z toho , co nás čeká , mám hrůzu . Všichni
pokradli . Majitel vítězné pálenky však neztrácí
pokradl STADICE (krk) - Elektrickou sekačku ,
pokradli po obci například poklopy na kanály právě
pokradli v uplynulých měsících v Olomouci dva bratři
nepokradl ani nezabil , k oně hrůze ? Možná neví
pokradli vybavení za 200 tisíc Jeseník * Škodu kolem
pokradli , často i násilně vyloupili , zmizeli beze
pokradl nebo nepokradl , asi zjistíme až na onom
nepokradl , asi zjistíme až na onom světě , kde předstoupí
pokradli . Bonsaje jsem totiž míval dřív i na zahrádce

> N-filter: ".*eš"
 > Deleted: 16
 > Deleted: 1

Zobrazeno: vše/23 Řádek: 1

Bonito

Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

N filtr jméno: syn

První Poslední Od 0 Pozic od < KWIC Do 0 Pozic od KWIC >

to Boží zákon ? " " Že člověk nezabije , **nepokrade** ani nepožádá manželku bližního svého .
rozprostírá země jen má , nikdo mi ji nikdy **nepokrade** , a jestliže ona krajina za očima zajde
opijou a budou se chtít bit . A taky určitě **pokradou** . " Rudolf pečlivě schovává jízdenky do
budou tak nějak sami od sebe slušní a příliš **nepokradou** . Dosavadní vývoj je však více než poučný

> Query : "pokrást"
> N-filter : ".*eš"
> Deleted : 36

Zobrazeno: vše/4 Řádek: 1

ALE

Z korpusu Araneum

A pak se dle potřeby převlékalo a převlékalo. A kradlo se a kradlo. A dokud bude žít generace Husákových dětí a jejich rodičů, současníci a pamětníci Klause a Zemana, tak se nic nezmění, <pokrade> se a <pokrade> se, nalevo, či napravo, vše jedno, jen když to půjde a ono to půjde, systém je tak nastaven. To jen pár hlupáků si bude vždy myslet, že kradou ti druzí.

Nenalezeny doklady typu <Pokrade> se domů tak, aby ho nikdo neviděl.

Doklady ze SYNv7

Jihlavské listy	proti konkurentovi svému . Nebo nepfelepiš plakátu jeho . Aníž	pokradeš /krást/VB-S---2F-AA---I	volebního materiálu jeho . Uhradí poplatky dalším tisícům JIHLAVA-
Mladá fronta DNES	tehdy pokoušeli namlavit zdejším lidem , že jim bženci všechno	pokradou /pokrást/VB-P---3P-AA---P	.což je nesmysl , " konstatovala místní mladá žena
Mladá fronta DNES	opijou a budou se chtít bit . A taky určitě	pokradou /pokrást/VB-P---3P-AA---P	" Rudolf pečlivě schovává jízdenky do jediné kasy na
Deník Moravia	Jak Svobodová poznamenala, báli se, že lidé svíčky	pokradou /pokrást/VB-P---3P-AA---P	a papíry poníž . " Nic takového se zatím nestalo
Pisání	! " Ano, a byl tomu tak - ostrov	Pokradu /pokrást/VB-S---1P-AA---P	cyril potažený kopzorou palmovkami a kokosovými olechy . " Víbec
Pisání	Lus si musí zaplátnout novou dřevitou nohu , protože na	Pokradu /krást/VB-S---1F-AA---I	utíkat od krbu . Sten Gunn se vybrní jako mladík
Literárky - můj esot	hlas buěl : Kapusta , kapusta , čo že si	pokradou /pokrást/VB-P---3P-AA---P	/ ej že fa poviazali / do hrubých povrazov
Lidové noviny	, doporučená cena 159 Kč . Pavel Jansa Zabijel ,	pokradeš /pokrást/VB-S---2P-AA---P	, zesmlíní Jednání a osudy hrdinů této napínavé knihy ovládá
Hospodářské noviny	, že krádeží víc ztratí , než získají , neboť	pokradou /pokrást/VB-P---3P-AA---P	- li , budou potrestáni špatným svědomím a spaním na
Deník Bohemia	i násobně větší , než poctivé pracující . Jakmile takový	pokradeš /pokrást/VB-S---3P-AA---P	, okamžitě useknout pazouru , když nic nemusejí dělat a
Regionální týdeník	a tak jsme měli strach , že nám to tady	pokradou /pokrást/VB-P---3P-AA---P	, " tekla unaveným hlasem Zdena Štelcová . " Staří
Ekonom	ženu i muže v žaláři dusí . Když obecní pastvě	pokradou /pokrást/VB-P---3P-AA---P	z husí , Gráží větší volně kráčí , žádná muka
Deník Bohemia	pro méně majetné bude dobře , že si supermarket navzájem	pokradou /pokrást/VB-P---3P-AA---P	zákazníky a budou nuceny snížit ceny . . . Matřkářka
Sport	na Spartě s šálami po Praze , sparťani jim je	pokradou /pokrást/VB-P---3P-AA---P	a pak si je vyvěsí u sebe v kottli .
Pisání	, kerej si říká Sten Gunn a žije po celém	Pokradu /pokrást/VB-S---1P-AA---P	mnoho let , neboť mu krutý kapitál Flint udělil černý
Deník Bohemia	hitlové uvidí ochranku , tak si dvakrát rozmyslí , než	pokradeš /pokrást/VB-S---3P-AA---P	Jitka Machová Myslí si , že situace na hitlovech
Přívě	v desateru . Žádná strana ale přece nehlásá : "	Pokradeš /pokrást/VB-S---3P-AA---P	! " Podstatné není ani tak to , čo kdo
Pisání	společně pějíci jakýs takýs chvalozpěv .) Konec 19 Ostrov	Pokradu /krást/VB-S---1F-AA---I	V malé pobřežní hospodě v Pistolu se odbírají spolek odbanců
Deník Moravia	, aby propojila Mikulčice se Slovenskem , přejdou Romové a	pokradou /pokrást/VB-P---3P-AA---P	v dědině všechno železo do sběru . Doufám , že
Regionální týdeník	, aby propojila Mikulčice se Slovenskem , přejdou Romové ,	pokradou /pokrást/VB-P---3P-AA---P	v dědině všechno železo a zpeněží ho ve sběrných surovinách