

Automatická morfoloická analýza z hlediska pokrytí a nepokrytí morfoloických variant

Klára Osolsobě, Hana Žižková

osolsobe@phil.muni.cz

h.zizkova@email.cz

Rezervy automatické morfologické analýzy

- Kde se berou chyby v automatické morfologické analýze?
- Jaká úskalí mohou přinést pokusy o odstranění těchto chyb?

Kde se berou chyby v automatické morfologické analýze?

- **Tvar je** ve slovníku a **je chybně desambiguován** v důsledku **tvárové homonymie**
- Ve slovníku automatického analyzátoru **chybí adekvátní interpretace**, **tvar/word** tudíž **nemůže** být správně interpretován (**podgenerování na rovině slovníku**)

Tvar je ve slovníku

Výskytů: 2 | i.p.m. θ : 0,02 vztaženo k celému syn2015 | ARF θ : 1,07 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

- Orchestron v hlavě** Dnes mi psala máma Chlapče nebud' fláma Pana hejtmana si **važ** /vázat/Vi-S---2--A----I Spi na levým uchu Drž si nohy v suchu Franzbrantweinem
- Vypravěči amerického Jihu** přišla dolů ? Asi tě slyšela hrát . Toho si **važ** /vázat/Vi-S---2--A----I . Normálně jen vysedává u sebe v křesle a čte

a je chybně desambiguován

Výskytů: 10 | i.p.m. ⁰: 0,08 vztaheno k celému syn2015 | ARF ⁰: 6,75 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	Stopy paměti	. O pár smyček později ze sebe vypravil : „	Nevaž /vážit/Vi-S---2--N----I	mě, boty míň . “ Ta slova musela něco
<input type="checkbox"/>	Sedmička	. . . usínám . Rada Veroniky : Sabi ,	nevaž /vážit/Vi-S---2--N----I	se , odvaž se ! Rada Agáty : Nebát se
<input type="checkbox"/>	Reflex	, každý záběr musí posouvat děj . PAEGAS – „	NEVAŽ /vážit/Vi-S---2--N----I	SE, ODVAŽ SE “ Spot nestojí za řeč –
<input type="checkbox"/>	Lidové noviny	nevěra , nárůst rozvodovosti , experimentující mládež v duchu „	nevaž /vážit/Vi-S---2--N----I	se , odvaž se “ . A protože platí ,
<input type="checkbox"/>	Etika a život	neuživil . Stačí se jen dívat a číst : “	Nevaž /vážit/Vi-S---2--N----I	se , odvaž se ! “ Dej si doušek
<input type="checkbox"/>	Ekonomie rodiny v proměná...	její pobočky neměly zapotřebí útočit na občany svody typu „	Nevaž /vážit/Vi-S---2--N----I	se , odvaž se “ , tak rodina jako klient
<input type="checkbox"/>	Mobility	určené mladším . Ke značce Paegas Twist neodmyslitelně patřil slogan	Nevaž /vážit/Vi-S---2--N----I	se , odvaž se a reklamy plné hudby a bezstarostn
<input type="checkbox"/>	Sedmička	můžete třeba zajímavou aplikací . (F&F) Sabino ,	nevaž /vážit/Vi-S---2--N----I	se , odvaž se ! Módní policie Agáty Prachařové &
<input type="checkbox"/>	Dobry den s kurýrem	hledali uplynulou neděli v popovickém kulturáku Popovskou hvězdu s podtitulem	Nevaž /vážit/Vi-S---2--N----I	se , odvaž se ! Soutěžilo se ve zpěvu ,
<input type="checkbox"/>	Ekonomie rodiny v proměná...	systemu tendenci tvrdit . Tak i dnešní reklamní slogan „	Nevaž /vážit/Vi-S---2--N----I	se , odvaž se ! “ slouží , jak známo

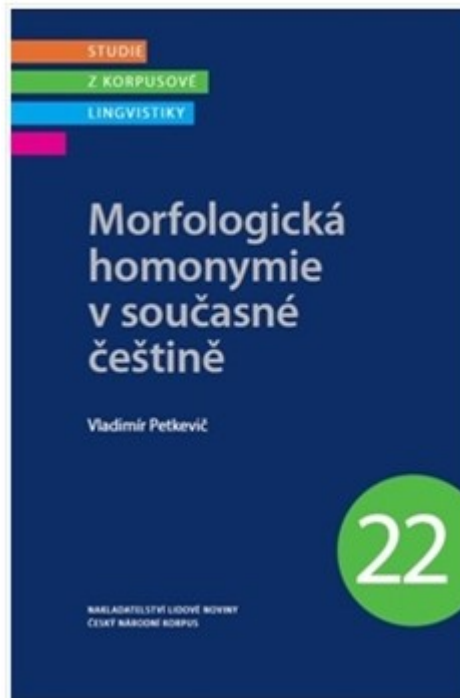
v důsledku tvarové homonymie

Výskytů: 2 | i.p.m. ⁰: 0,02 vztaženo k celému syn2015 | ARF ⁰: 1,16 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	Vedlejší nebe	Moc si toho vážím . . . Hvězda : Nic	nevažte /vážit/Vi-P---2--N----I	ani neměřte . Jde o to , jestli byste do
<input type="checkbox"/>	Co je tvoje, je i moje	nich důvěrně . „ Radši se na to malé moc	nevažte /vážit/Vi-P---2--N----I	, dokud nebude větší jistota , že to zvládne .

Popis homonym



Ve slovníku automatického analyzátoru **chybí adekvátní interpretace**, tvar/word tudíž **nemůže** být správně interpretován (**podgenerování na rovině slovníku**)

Výskytů: 6 | i.p.m. θ : 0,05 vztaheno k celému syn2015 | ARF θ : 4,24 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

- Jíst slušně** nejspíš zvolila pránickou výživu . " Později jsem si to **vygooglovala** /vygooglovat/VpFS----R-AA---P : v hinduismu označuje prána něco jako životní energii ,
- Reflex** , protože jí dodala špatné informace . Nebohá Lauren si **vygooglovala** /vygooglovat/VpFS----R-AA---P na svém telefonu Blackberry trasu ke své procházce . Ta
- Tak trochu šílení** dokumentu na A&E ; spousta lidí ten pořad viděla , **vygooglovala** /vygooglovat/VpFS----R-AA---P si mě a koupila si knihu . " Rhyme měl
- Náš tragický vesmír** jsem si ke stolu u okna , otevřela notebook a **vygooglovala** /vygooglovat/VpFS----R-AA---P Kelseyho Newmana . Našla jsem pár rozhovorů a zastaralou webovou
- Na čem záleží** si vezme na sebe . A jestli si zmíněnou prostitutku **vygooglovala** /vygooglovat/VpFS----R-AA---P a prohlížela si její fotografie v bulváru . Co asi
- Hospodářské noviny** , jde asi o trojnásobek . Paní Hrstková si buď **vygooglovala** /vygooglovat/VpFS----R-AA---P špatná čísla , nebo prospala hodiny matematiky . Růst ceny

Výskytů: 2 | i.p.m. θ : 0,02 vztaheno k celému syn2015 | ARF θ : 1,30 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

- Respekt** dobrovolníků neznala , stačilo si ale během několika minut organizaci **progooglovat** /progooglovat/X@----- , a bylo by jasné , že má co do
- Marianne** a Walt Disney Studios Park , proto se vyplatí dopředu **progooglovat** /progooglovat/X@----- , jaké atrakce vás zajímají , abyste se stejně jako

Možná řešení a jejich úskalí

- Guessery
- Doplnování slovníku

Tvar/varianta není ve slovníku, přestože jiné tvary/varianty ve slovníku jsou, tudíž nemůže být správně desambiguován

Výskytů: 2 | i.p.m. θ : 0,02 vztaheno k celému syn2015 | ARF θ : 1,45 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	Hliněné dny	k nim a praštil o zem nacpaným pytle . „	neház /nehaž/X@-----	s tím tak ! " křikl Koza přisně . „
<input type="checkbox"/>	Škvírou v žaluzii	Svíci máš - li s lebkou na svém stole ,	neház /nehaž/X@-----	si mašli , ani když máš hlavu dole . St

Výskytů: 44 | i.p.m. θ : 0,36 vztaheno k celému syn2015 | ARF θ : 18,97 | Výsledek je promíchán

1 / 2 ▶▶

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	Otázky a odpovědi pro kluky	dávno prožili a znají i způsob tvého vnímání . Proto	neházej /házet/Vi-S---2--N---I	jejich rady lehkomyšlně za hlavu . Může se stát mou
<input type="checkbox"/>	Magický čtverec Albrechta ...	už jsme spolu celé čtyři roky . " „ Ještě	neházej /házet/Vi-S---2--N---I	flintu do žita . Uvidíš , už brzy dozajista porodíš
<input type="checkbox"/>	Nabíječ	sedadlo a hodila po něm moučník . „ Šilíš ?	neházej /házet/Vi-S---2--N---I	s ní , vždyť ji rozdrobíš ! " vykřikl a
<input type="checkbox"/>	Svatojánská noc	nechat , když tu odpoledne byl . B Jasně H	NEHÁZEJ /házet/Vi-S---2--N---I	S NÍM B Co ? H Promiň – jenom ,
<input type="checkbox"/>	Kmotr se mstí	podobně , to nevydejcháš . " „ To když tím	házej /házet/VB-P---3P-AA--6I	týpci , co nejsou jako my . " „ No
<input type="checkbox"/>	Charakter	tejden bude sucho . Tady bysme těm debilům , co	házej /házet/VB-P---3P-AA--6I	pětikačky do automatu , ani nemohli nalejt , až vyhrajou
<input type="checkbox"/>	Milý Johne	hlavou . „ Tak to ne , na mě to	neházej /házet/Vi-S---2--N---I	! Já jsem nic špatného neudělala . Já nezačala !
<input type="checkbox"/>	Týden	příklad s dítětem . Když řeknete : „ Maruško ,	neházej /házet/Vi-S---2--N---I	písek po kamarádovi , to se nedělá , " ale
<input type="checkbox"/>	Láska je slípka	žije , kde se broskve jen nakusujou a pak se	házej /házet/VB-P---3P-AA--6I	za sebe ! Ale on to neudělal . Možná ,
<input type="checkbox"/>	Rok perel	se zapichují do fotografie oběti špendlíčky nebo se do ní	házej /házet/Vi-S---2--A---I	šípky . Jenže já neměla u sebe už ani jednu

Nedostatečné **pokrytí** má vliv na **nepřesnou** analýzu

- Řešení: **DOPLNĚNÍ TVARŮ DO SLOVNÍKU = ZLEPŠENÍ VÝSLEDKŮ AUTOMATICKÉ ANALÝZY**
- **PROBLÉMY**

Pokrytí slovníku je v pořádku, ale výsledky analýzy nejsou **přesné**

- *Tvar/varianta není ve slovníku, přestože jiné tvary/varianty ve slovníku jsou, je ovšem homonymní s tvarem, který ve slovníku je, a tak desambiguace je technicky vzato v pořádku, nicméně neodpovídá skutečnosti*

Nejde o chybu v desambiguaci, ale i o podgenerovaný slovník

Výskytů: 21 | i.p.m. ⁰: 0,17 vztaženo k celému syn2015 | ARF ⁰: 8,51 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	Cesta k moři	až před barák . Katja Doufám , že už rodiče	spěj /spět/Vi-S---2--A---I . Lehnu si do pokoje k dětem . Tady pozor
<input type="checkbox"/>	35 mrtvých	odpověděla . „ A co vaši ? ” „ Ty	spěj /spět/Vi-S---2--A---I , „ zasmála se a začala si rozepínat knoflíčky halenky
<input type="checkbox"/>	Jako vrána na Beckhama	vím . A zase až teď . Ti dva ještě	spěj /spět/Vi-S---2--A---I , já už mám sbaleno . Ještě si všechno sepíšu
<input type="checkbox"/>	Týdeník Květy	zařve : „ Mámo , ať už ty parchanti kurva	spěj /spět/Vi-S---2--A---I ! “ Nebo rodič , vláčejší tělesně i duševně vyčerpané
<input type="checkbox"/>	Kloktat dehet	druhej den vlka loví . Po večeři dou spát .	Spěj /spět/Vi-S---2--A---I všichni v jedný seknci , akorát putovníkovi dopřejou kumbál ,
<input type="checkbox"/>	Mrazivý chlad	to nestalo , a že stejně v tuhle roční dobu	spěj /spět/Vi-S---2--A---I zimním spánkem . Táhle z automatu si koupil pár tyčinek
<input type="checkbox"/>	Poselství jednorožců	. „ Uslyší tě . ” „ Houby ! Všichni	spěj /spět/Vi-S---2--A---I ! “ Ale Veragin ho pro jistotu shodil vedle sebe
<input type="checkbox"/>	Atlas mraků	vlastně Jasnozřiví jedí a pijou ? Spěj na seně ?	Spěj /spět/Vi-S---2--A---I vůbec ? Šest měsíců ! Máma mě proklínala , že
<input type="checkbox"/>	Panoptikon	Noční vychovatelka je už určitě ve věži , jinak všichni	spěj /spět/Vi-S---2--A---I . Je ticho , v polích šustí tráva a smrky
<input type="checkbox"/>	Jako pták na drátě	se dívají na pořad o Hitlerovi . Holky už kupodivu	spěj /spět/Vi-S---2--A---I . Maruška pod peřinou vypadá jako trpaslík . Spí v
<input type="checkbox"/>	Vyvolení	mám Zowalieny z planety Askol , hrozně smrděj a pořad	spěj /spět/Vi-S---2--A---I . To pak mají co kecat , že ti zvířata
<input type="checkbox"/>	Jako vrána na Beckhama	posouvání bříška ukazováku po vyleštěné ploše . Mezera Všichni tři	spěj /spět/Vi-S---2--A---I na hromadě báglů . Hlídám je , plním kartu svýho
<input type="checkbox"/>	O spánku a mizení	veřejný pochlastávání , ale veřejný spaní . A ještě navíc	spěj /spět/Vi-S---2--A---I v nestoudnejch časech ! Křídový byty sice nejsou žádný přístřeší
<input type="checkbox"/>	Jako pták na drátě	, písmácký , stařecký , nebo padlý . Ležej a	spěj /spět/Vi-S---2--A---I , dokud neshnijou . V pralese se vrátíš zpátky ,
<input type="checkbox"/>	Oněgin byl Rusák	všemu průchozí a mezi tím naším a tím , co	spěj /spět/Vi-S---2--A---I naši , jsou dvoukřídly , prosklený dveře . Takže je
<input type="checkbox"/>	Medvědí tanec	voni medvědi v zimě ? Já myslel , že medvědi	spěj /spět/Vi-S---2--A---I zimním spánkem , učili jsme se to . ” „
<input type="checkbox"/>	Atlas mraků	že jo . Co vlastně Jasnozřiví jedí a pijou ?	Spěj /spět/Vi-S---2--A---I na seně ? Spěj vůbec ? Šest měsíců ! Máma
<input type="checkbox"/>	Kloktat dehet	bylo ticho , všichni chlapi se asi tvářili , že	spěj /spět/Vi-S---2--A---I , že využívají polední klid k hlubokému spánku , a
<input type="checkbox"/>	Škoda lásky	ztiš se ! ” „ Ty nás neslyšej . Ty	spěj /spět/Vi-S---2--A---I na tržišti . . . jak ten princ . ”
<input type="checkbox"/>	Hovězí kostky	, a když si vařím poslední kafe , všichni už	spěj /spět/Vi-S---2--A---I , ale já bych bez toho nezabral . Tak víte
<input type="checkbox"/>	Tělo	? “ „ Co by měli podle tebe dělat ?	Spěj /spět/Vi-S---2--A---I . “ „ Ba ne , podle mýho se na

Úskalí odstranění podgenerování na rovině slovníku automatického analyzátoru

- Dopad odstranění podgenerování na výsledky automatické morfologické analýzy
- Typy podgenerovaných jednotek

PODGENEROVANÉ JEDNOTKY – tvary sloves (sloveso – klíčové pro další stupně analýzy – pravidlovou desambiguaci)

- Příprava manuálu pro doplnění tvarů sloves nezachycených automatickou analýzou
- Zachycení tvarů sloves uzavřených tříd
- Zachycení tvarů sloves otevřených tříd
- Zachycení variant z hlediska dichotomie kodifikovaný a nekodifikovaný

Konkrétní příklady

- Krácení dlouhých infinitivů
(<http://www.karaoketexty.cz/texty-pisni/lidove-pisne/co-sa-to-supoce-60336>)
- Tvar 3. pl. ind. prez. akt. Na $-[e\check{e}]j$ (viz snímek č. 13), potenciální homonymie s tvarem 2. sg. imp.

Krácení dlouhých infinitivů

Výskytů: 3 | i.p.m. θ : 0,02 vztaženo k celému syn2015 | ARF θ : 1,50 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	Orchestrion v hlavě	hrud klasy Bujný bude květ Kosa bude řinčet Zpěv radostný	znět /znět/X@-----	A ta lípa naše bude zelená Z větví mocných listí
<input type="checkbox"/>	Úvod do fyziky hvězd	se v nich dosud nevznítily . Správná odpověď by mohla	znět /znět/X@-----	: " Hvězdy září , protože je jejich povrch zahřát
<input type="checkbox"/>	Dvě stě let české hudby na Mor...	' Milostné dopisy ' . Myslím , že to bude rozkošně	znět /znět/X@-----	. Vždyť těch našich příhod bylo už dost ! To

PRO POROVNÁNÍ DATA Z KORPUSU SYN

Výskytů: 8 | i.p.m. 0,00 vztaheno k celému syn | ARF 4,31 | Výsledek je promíchán

Výběr řádků: základní (0 vybrané)

<input type="checkbox"/>	opus#14755	květ , / kosa bude řinčet , / zpěv radostný	znět /znět/x@-----	. / / A ta lípa naše / bude zelená
<input type="checkbox"/>	opus#26138	" Hlavní věc je , že to memento zní a	znět /znět/x@-----	musí , " řekl novinářům český prezident Václav Klaus .
<input type="checkbox"/>	opus#43606	věřím , že ještě vydrží . A bude ještě chvílu	znět /znět/x@-----	. Jak vybrat helmu na lyže a snowboard Dočetli jste
<input type="checkbox"/>	opus#28865	, že je mají . " Tak nějak by mohlo	znět /znět/x@-----	motto třináctidílného seriálu z prostředí současné vesnice , jehož hrdina
<input type="checkbox"/>	opus#26138	armády Yisrael Weiss modlitbu . Klaus : memento zní a	znět /znět/x@-----	musí S myšlenkou uspořádat pietní setkání vězňů s osvoboditeli tábora
<input type="checkbox"/>	opus#28292	se pak přemnoží kvasinky a bakterie . Výsledkem je právě	znět /znět/x@-----	: pro kvasinkový je typický bílý výtok , pro bakteriální
<input type="checkbox"/>	opus#28419	, že je mají . " Tak nějak by mohlo	znět /znět/x@-----	motto třináctidílného seriálu z prostředí současné vesnice , jehož hrdina
<input type="checkbox"/>	opus#46145	Po neúspěchu se však v roce 1916 Vlasta Javořická vrací	znět /znět/x@-----	dn Studené . Její prvotinou byla povídka Splněná tužba ,

A Z KORPUSU ORAL2013

Výskytů: 934 | i.p.m. ^Q: 284,28 vztaženo k celému oral2013 | ARF ^Q: 368,59 | Výsledek je promíchán

1 / 24 ▶▶▶

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	11A084N	. zme měly domluvené že se půjdem podívat na ten byt .] + [mmm] + [jako kde budem a toto ..] + [jenomže ta
<input type="checkbox"/>	09A057N	štyři] + [nebo kolik . jo ?] [no] + [ale mosí to byt prostě ..] + [eee nemusí a finta je v tom .
<input type="checkbox"/>	11A070N	[tak ať vám výde počasí ať ti to nezmokne] + [má byt třinást stupňu .] + [dobře . dobře ..] + [a] [třinást] + [stupňů
<input type="checkbox"/>	11A060N	tu práci ..] + [no] + [takže potom kdyby zas nevyšel ten byt tak m* mám] + [strach jim pak říc (se smíchem) že tam
<input type="checkbox"/>	09M037N	. dali .] + [s tím že by se o ten byt staral] + [a nemuseli by ho hlídat jako co s nim
<input type="checkbox"/>	09A019N	ňáký depresi že že to potřebuje že potřebuje mít] + [nejdřív byt pak potřebuje tadyto a támleto takže] + [zase se to tak
<input type="checkbox"/>	10A094N	své knize to měl] + [takto prezentované jako že to může byt] [jo jo jo] + [jo jak sou třeba i z té
<input type="checkbox"/>	09A134N	tam mají ňáký kamarády . a ty si našli jinej byt .] + [tak bude práznej . a Kačenka s Michalem tam
<input type="checkbox"/>	08A074N	[to si eště poslechnu .. --- ..] + [no tak] + [co byt . dobrý ? ..] + [co ?] + [co byt ? .]
<input type="checkbox"/>	11A028N	bude platit nájem kolik bude] + [platit a hlavně si vybrat byt že on tam má možnost dvou bytů ..] + [jeden s
<input type="checkbox"/>	11A111N	.] + [ty nechceš byt zelená guma ?] + [ne já chcu byt jenom modrá vole nebo jakou maj barvu ti policajti zasraní

Homonymie imperativu 2. sg. a substandardního tvaru 3. pl. ind. prez. akt.

Výskytů: 4 | i.p.m. ⁰: 0,03 vztaženo k celému syn2015 | ARF⁰: 2,48 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	Mánie	na dostihy . Vymyslel naivní systém sázek („ Vždycky	sázej /sázet/Vi-S---2--A----I	na třetího favorita , protože má lepší kurs a první
<input type="checkbox"/>	Vy mně taky!	Na svatou Věruš kušuj a neruš . Na svatého Kravince	sázej /sázet/Vi-S---2--A----I	v České lotynce . Svatá Důra , puč mi búra
<input type="checkbox"/>	Tak trochu šílení	Ale když je počasí na hovno , zůstanou uvnitř a	sázej /sázet/Vi-S---2--A----I	. Jednoduchý jako facka . “ No jo . Tohle
<input type="checkbox"/>	Rodinný dům	za zmínku stojí i stará česká pranostika : „ Brambory	sázej /sázet/Vi-S---2--A----I	na svatého Marka (25 . dubna) , bude

Morfio – nástroj na detekci variant (SYN2015)

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	bruslejí (1)	bruslej (1)
2	čučejí (1)	čučeř (7)
3	frejí (1)	frej (32)
4	mějí (1)	mej (1)
5	nejpozdějí (1)	nejpozděj (1)
6	obličejí (1)	obličej (16764)
7	orsejí (1)	orsej (12)
8	pozdějí (3)	pozděj (10)
9	prodejí (1)	prodej (11157)
10	ráčejí (3)	ráčej (1)
11	silnějí (1)	silněř (2)
12	stěžejí (5)	stěžeř (6)
13	terejí (1)	tereř (48)

Nejde vždy o vzor *sázet* a homonymii substandardního tvaru s tvarem standardním (SYN2010)

vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1 častěji (1)	častěj (2)
2 ferveji (1)	fervej (2)
3 frčejí (1)	frčej (2)
4 jednodušejí (1)	jednodušej (1)
5 laskavějí (1)	laskavěj (3)
6 lpějí (24)	lpěj (2)
7 mastějí (1)	mastěj (2)
8 mířejí (1)	mířej (3)
9 mějí (2)	mej (13)
10 nebezpečnějí (1)	nebezpečněj (1)
11 obličejí (1)	obličej (15030)
12 onemocnějí (1)	onemocněj (1)
13 plazejí (1)	plazej (3)
14 pohřbějí (1)	pohřběj (1)
15 pozdějí (4)	pozděj (36)
16 praštějí (2)	praštěj (1)
17 prodejí (1)	prodej (13364)
18 ráčejí (1)	ráčej (1)
19 sprejí (1)	sprej (642)
20 stěžejí (1)	stěžej (11)
21 šejí (1)	šej (3)
22 udržejí (1)	udržej (3)
23 uvolnějí (1)	uvolněj (2)
24 vodrážejí (1)	vodrážej (2)
25 výprodejí (1)	výprodej (481)
26 zasloužejí (1)	zasloužej (6)
27 zmizejí (5)	zmizej (1)

Závěr

- Slovník automatického analyzátoru má rezervy a nedostatky jsou při využití korpusů vidět.
- Doplnění slovníku a zvýšení jeho pokrytí je v zásadě možné a díky korpusům lze zajistit zachycení jak centrálních, tak okrajových variant.
- Zvýšení pokrytí slovníku nemusí znamenat zlepšení přesnosti automatické morfologické analýzy.

Kam dál?

- Chceme novou automatickou morfologickou analýzu, i když víme, že i ona bude obsahovat chyby?
- Ano, ale informujme uživatele o tom, na které úrovni která chyba vzniká!
- Děkuji Vám za pozornost.

Bibliografie

- BABINCOVÁ, Adriana. *Morfologické varianty jako problém strojové analýzy češtiny*. FF MU: Brno. 2015. (Nepublikovaná bakalářská práce pod vedením K. Osolobě).
- HLAVÁČKOVÁ, Dana. Korpusové zpracování korespondenčních textů: morfologické značkování. In Hladká, Zdeňka a kol.. *Soukromá korespondence jako lingvistický pramen*. Vyd. 1. Brno: Masarykova univerzita, 2013. s. 19-31.
- HLAVÁČKOVÁ, Dana a Klára OSOLSOBĚ. Morfologické značkování mluvených korpusů, zkušenosti a otevřené otázky. Kopřivová, Marie, Waclawičová, Martina. In *Čeština v mluveném korpusu*. 1. vyd. Praha: Nakladatelství Lidové noviny/ Ústav Českého národního korpusu, 2008. s. 105-114.
- OSOLSOBĚ, Klára. Slovesné tvary nerozpoznané automatickou morfologickou analýzou. Rkp. 2016.
- PETKEVIČ, Vladimír. Morfologická homonymie v současné češtině. Praha: Nakladatelství Lidové noviny/ Ústav Českého národního korpusu, 2016.