

# CJBB84

# Morfologie a korpus

**10.00-11.30 G13**

**I.**

# Morfologické značkování českých korpusů

- Pražský systém : tagset\_poziční
- Brněnský systém : tagset\_atributový

# Pražský systém

- Poziční užívaný v korpusech ČNK
- 16 pozic – každá pozice odpovídá nějaké kategorii známé z klasické gramatiky
- Např. na 1. pozici každé značky je uveden slovní druh podle klasické desetičlenné klasifikace slovních druhů. Na stejné pozici je i značka X pro jednotky, které nejsou rozpoznány automatickou morfologickou analýzou a značka Z pro interpunkční znaky.
- Tag=N.\* čteme: značka=substantivum (.\*/.+ je regulární výraz, který čteme libovolný počet opakování libovolného znaku, což značí, že další pozice značky jsou pro nás nezajímavé).

# [tag="N.\*"]

Korpus: [syn2015](#) | Dotaz: [N.\\*](#) (29 549 285 výskytů) ▶ Promíchat: ✓

Výskytů: 29 549 285 | i.p.m. [0](#): 244 717,18 vztaženo k celému syn2015 | ARF [0](#): 20 964 914,34 | Výsledek je promíchán

1 / 738 733 ▶▶▶


Výběr řádků: základní (0 vybrané)

<input type="checkbox"/>	<a href="#">Dobry den s kurýrem</a>	, který vynikajícím zákrokem vytěsnil krásnou střelu pod břevno ze	<a href="#">standardky</a> /standardka/NNFS2-----A-----	. Branka : Svízela . ACH TA KONCOVKA ! -
<input type="checkbox"/>	<a href="#">Bílá vlčice</a>	Okamžitě rozpoznal povahu svého zanícení pro tuto neznámou ženu .	<a href="#">Gabriel</a> /gabriel/NNMS1-----A-----	byl náročný muž , potpěl si na dokonalos
<input type="checkbox"/>	<a href="#">Tango Masaj</a>	domana , kterou oblékal jen při slavnostních příležitostech , na	<a href="#">nohou</a> /noha/NNFP6-----A-----	měl kožené sandály a na krku , zápěstích
<input type="checkbox"/>	<a href="#">E15</a>	prodal v roce 2006 Anwilu více než čtyřpětinový podíl ve	<a href="#">Spolaně</a> /spolana/NNFS6-----A-----	za částku přesahující 640 milionů korun .
<input type="checkbox"/>	<a href="#">Hřbitovní kvítí na smetaně</a>	dostatečně inspirativní a pravděpodobně stačilo k výstavbě základní podoby románové	<a href="#">figury</a> /figura/NNFS2-----A-----	. Pro mne osobně je Hardy trochu nudný p
<input type="checkbox"/>	<a href="#">Čtyři kultury ve vzdělávání</a>	uvedených dvou typů reakce na " kritiku modernismu " -	<a href="#">teorie</a> /teorie/NNFS1-----A-----	komunikativní akce a postmoderní dekon:
<input type="checkbox"/>	<a href="#">Století fyzikálních objevů</a>	paprsky vyletující od anody směrem ke katodě a dal jim	<a href="#">název</a> /název/NNIS4-----A-----	anodové nebo kanálové . ● Německý fyzik
<input type="checkbox"/>	<a href="#">Silnice a železnice</a>	možné vyhodnotit garanci systémových parametrů ITS aplikace . VYUŽITÍ GNSS	<a href="#">SIMULÁTORU</a> /simulátor/NNIS2-----A-----	Pro zajištění pilotních i řádných testů a sp
<input type="checkbox"/>	<a href="#">Svět ženy</a>	bylo nám ve třech docela dobře . Ale pak obě	<a href="#">děti</a> /dítě/NNFP1-----A-----	krátce po sobě odešly na studium a za pr
<input type="checkbox"/>	<a href="#">Atlas pískovcových skalnic...</a>	současné době dodává do suťových akumulací další bloky . Kamenná	<a href="#">moře</a> /moře/NNNP1-----A-----	v krystalických horninách ( např. rulách )
<input type="checkbox"/>	<a href="#">Britské detektivky</a>	, která se k zločinu připlete mimo službu ( Z	<a href="#">příčin</a> /příčina/NNFP2-----A-----	nikoliv přirozených , Černá věž , Plány a t

# Brněnský systém

- Atributový užívaný v korpusech vytvářených na FI MU
- Dvojice atribut\_hodnota odpovídá rovněž gramatickým kategoriím a hodnotám, které nabývají.
- Např. atribut **k** signalizuje slovní druh, je následován číslicí 1,2,3, ...,0, které označují hodnotu klasických deseti slovních druhů ve školním řazení.

# [tag="k1.\*"]

Dotaz **k1.\*** 210,467,197 (počítá se ...) | použito prvních 10,000,000 řádků (použít náhodných). 

Strana  ze 492,591   | [Poslední](#)

<a href="#">doc#0</a>	<code>&lt;p&gt;</code>	<b>Brod</b> /brod/k1glnSc1qP	porazil Chropyni a ztrácí na ni bod <code>&lt;/p&gt;</code>
<a href="#">doc#0</a>	<code>&lt;p&gt;</code> Brod porazil	<b>Chropyni</b> /Chropyně/k1gFnSc4	a ztrácí na ni bod <code>&lt;/p&gt;&lt;p&gt;</code> Náskok Chropyně
<a href="#">doc#0</a>	<code>&lt;p&gt;</code> Brod porazil Chropyni a ztrácí na ni	<b>bod</b> /bod/k1glnSc4qP	<code>&lt;/p&gt;&lt;p&gt;</code> Náskok Chropyně na čele se snížil
<a href="#">doc#0</a>	porazil Chropyni a ztrácí na ni bod <code>&lt;/p&gt;&lt;p&gt;</code>	<b>Náskok</b> /náskok/k1glnSc1qP	Chropyně na čele se snížil. V Brodě prohrála
<a href="#">doc#0</a>	Chropyni a ztrácí na ni bod <code>&lt;/p&gt;&lt;p&gt;</code> Náskok	<b>Chropyně</b> /Chropyně/k1gFnSc2	na čele se snížil. V Brodě prohrála a právě
<a href="#">doc#0</a>	ztrácí na ni bod <code>&lt;/p&gt;&lt;p&gt;</code> Náskok Chropyně na	<b>čele</b> /čelo/k1gNnSc6qP	se snížil. V Brodě prohrála a právě její
<a href="#">doc#0</a>	<code>&lt;/p&gt;&lt;p&gt;</code> Náskok Chropyně na čele se snížil. V	<b>Brodě</b> /brod/k1glnSc6qP	prohrála a právě její dnešní soupeř se
<a href="#">doc#0</a>	dnešní soupeř se za ni dotáhl na nejmenší	<b>rozdíl</b> /rozdíl/k1glnSc4qP	. A to není jediná změna na čele. Sestupný
<a href="#">doc#0</a>	dotáhl na nejmenší rozdíl. A to není jediná	<b>změna</b> /změna/k1gFnSc1qP	na čele. Sestupný kurs po víkendu určitě
<a href="#">doc#0</a>	nejmenší rozdíl. A to není jediná změna na	<b>čele</b> /čelo/k1gNnSc6qP	. Sestupný kurs po víkendu určitě nabere
<a href="#">doc#0</a>	A to není jediná změna na čele. Sestupný	<b>kurs</b> /kurs/k1glnSc1	po víkendu určitě nabere Spytihněv. Rožnov

# Jak se interpretaci ve formě lemmat a tagů dostaly do korpusu?

- Automatická morfologická analýza
- Morfologický slovník
- Automatický nástroj
- Tokenizátor
- Lemmatizace/značkování
- Diasambiguace

# *Jí je špatně.*

- Jí – jíst/VB-P---3P-AA---I
- Jí – on/PPFS3—3-----
- Je – být/VB-S---3P-AA---I
- Je – on/PPFP4—3-----
- Je – on/PNSP4—3-----
- Je – on/PPIP4—3-----
- Je – on/PPMP4—3-----
- Špatně – špatně/Dg-----1A----



Jí je špatně.



Jí je špatně.



# Homonymie – problém disambiguace

- Statistické metody
- Pravidlové metody

# Chyby

Výskytů: 117 688 | i.p.m. <sup>0</sup>: 974,65 vztaženo k celému syn2015 | ARF <sup>0</sup>: 64 433,19 | Výsledek je promíchán

1 / 2 943 ▶▶▶

Výběr řádků: základní (0 vybrané)

<input type="checkbox"/>	<a href="#">Nikdy jsem nechtěla zabít</a>	vyzývavé pohledy , jako by jenom čekala , až si	je /oni/PPFP4--3-----	někam odvedou . Trvalo mi nějakou dobu , než jsem
<input type="checkbox"/>	<a href="#">ABC</a>	posměškům trápila , vztekala nebo se urážela . To by	je /oni/PPFP4--3-----	jen vyprovokovalo , aby se ti smáli ještě víc .
<input type="checkbox"/>	<a href="#">Dům u jezera</a>	. Z úplně obyčasných mladíků dělá jiné lidi . Zlomí	je /oni/PPFP4--3-----	. " Pomyslela jsem na Alfreda tu noc na schodech
<input type="checkbox"/>	<a href="#">Hana Benešová</a>	, asi 70 kilometrů severně od New Yorku . Doprovází	je /oni/PPFP4--3-----	Mary Stevensonová , sekretářka Hamiltona Fishe Armstronga , zakladatele a
<input type="checkbox"/>	<a href="#">Právo</a>	nich začal shánět , obžalovaný mu řekl , že mu	je /oni/PPFP4--3-----	vrátí , ale zadarmo to prý nebude . „ Chtěl
<input type="checkbox"/>	<a href="#">Reflex</a>	levicové kavárny . Oficiálně Praha 7 sobě . Nevilil jsem	je /oni/PPFP4--3-----	. Z přesvědčení a asi i z předsudku . Ve
<input type="checkbox"/>	<a href="#">Jihlavské listy</a>	lidé věnují méně sportu a jiným aktivitám , které by	je /oni/PPFP4--3-----	udržely v dobré fyzické kondici . Pozor na držení těla
<input type="checkbox"/>	<a href="#">Navždy tvůj</a>	ale ona to odmítla . " – „ No to	je /oni/PPFP4--3-----	teda prosím pěkně děs běs , " zhodnotila Bianca .
<input type="checkbox"/>	<a href="#">Děbelská zubařka</a>	„ Okamžitě nasedni , Alfredě ! " V té chvíli	je /oni/PPFP4--3-----	dohnali i pan Snood a studenti dramatické výchovy . Nahruli
<input type="checkbox"/>	<a href="#">Krásné zlo</a>	se rozhodli , že ti pořídíme nové . Dnes už	je /oni/PPFP4--3-----	vyrábějí mnohem bezpečnější , s airbagy a podobně . Peníze
<input type="checkbox"/>	<a href="#">Páté roční období</a>	. " Malin víří hlavou tisíce myšlenek . Snaží se	je /oni/PPFP4--3-----	uspořádat , zachytit , vybavují se jí však jen tváře
<input type="checkbox"/>	<a href="#">Největší rytíř</a>	členů její rodiny jí v mysli vybledly , jako kdyby	je /oni/PPFP4--3-----	halily stále nové a nové vrstvy mlhy . Když se
<input type="checkbox"/>	<a href="#">Ona dnes</a>	tím , že jste zapoměli zvážít rajčata . „ To	je /oni/PPFP4--3-----	opravdu potřebujete a budete zdržovat ostatní ? " kývne našťvaně

# 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.  $\theta$ : 0,02 vztaženo k celému syn2015 | ARF  $\theta$ : 1,07 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

- |                          |                                  |   |
|--------------------------|----------------------------------|---|
| <input type="checkbox"/> | <b>Orchestrion v hlavě</b>       | Dnes mi psala máma Chlapče nebuď fláma Pana hejtmana si <b>važ</b> /vázat/Vi-S---2--A----I Spi na levém uchu Drž si nohy v suchu Franzbrantweinem |
| <input type="checkbox"/> | <b>Vypravěči amerického Jihu</b> | přišla dolů ? Asi tě slyšela hrát . Toho si <b>važ</b> /vázat/Vi-S---2--A----I . Normálně jen vysedává u sebe v křesle a čte                      |

# a je chybně desambiguován

Výskytů: 10 | i.p.m.<sup>0</sup>: 0,08 vztaženo k celému syn2015 | ARF<sup>0</sup>: 6,75 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

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



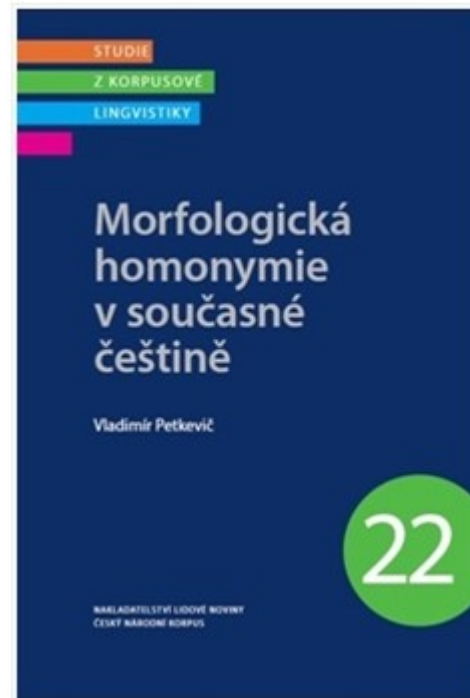
# v důsledku tvarové homonymie

Výskytů: 2 | i.p.m. <sup>0</sup>: 0,02 vztaheno k celému syn2015 | ARF <sup>0</sup>: 1,16 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	<b>Vedlejší nebe</b>	Moc si toho vážím . . . Hvězda : Nic	<b>nevažte</b> /vážit/Vi-P---2--N----I	ani neměřte . Jde o to , jestli byste do
<input type="checkbox"/>	<b>Co je tvoje, je i moje</b>	nich důvěrně . „ Radši se na to malé moc	<b>nevažte</b> /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.  $\theta$ : 0,05 vztaheno k celému syn2015 | ARF  $\theta$ : 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.  $\theta$ : 0,02 vztaheno k celému syn2015 | ARF  $\theta$ : 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. <sup>0</sup>: 0,02 vztaženo k celému syn2015 | ARF <sup>0</sup>: 1,45 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

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

Výskytů: 44 | i.p.m. <sup>0</sup>: 0,36 vztaženo k celému syn2015 | ARF <sup>0</sup>: 18,97 | Výsledek je promíchán

1 / 2 ▶▶▶

Výběr řádků: základní ▾ (0 vybrané)

<input type="checkbox"/>	<b>Otázky a odpovědi pro kluky</b>	dávno prožili a znají i způsob tvého vnímání . Proto	<b>neházej</b> /házet/Vi-S---2--N---I	jejich rady lehkomyšlně za hlavu . Může se stát mou
<input type="checkbox"/>	<b>Magický čtverec Albrechta ...</b>	už jsme spolu celé čtyři roky . “ „ Ještě	<b>neházej</b> /házet/Vi-S---2--N---I	flintu do žita . Uvidíš , už brzy dozajista porodíš
<input type="checkbox"/>	<b>Nabíječ</b>	sedadlo a hodila po něm moučník . „ Šílíš ?	<b>Neházej</b> /házet/Vi-S---2--N---I	s ní , vždyť ji rozdrobíš ! “ vykřikl a
<input type="checkbox"/>	<b>Svatojánská noc</b>	nechat , když tu odpoledne byl . B Jasně H	<b>NEHÁZEJ</b> /házet/Vi-S---2--N---I	S NÍM B Co ? H Promiň - jenom ,
<input type="checkbox"/>	<b>Kmotr se mstí</b>	podobně , to nevydejcháš . “ „ To když tím	<b>házej</b> /házet/VB-P---3P-AA--6I	týpci , co nejsou jako my . “ „ No
<input type="checkbox"/>	<b>Charakter</b>	tejden bude sucho . Tady bysme těm debilům , co	<b>házej</b> /házet/VB-P---3P-AA--6I	pětikačky do automatu , ani nemohli nalejt , až vyhrajou
<input type="checkbox"/>	<b>Milý Johne</b>	hlavou . „ Tak to ne , na mě to	<b>neházej</b> /házet/Vi-S---2--N---I	! Já jsem nic špatného neudělala . Já nezačala !
<input type="checkbox"/>	<b>Týden</b>	příklad s dítětem . Když řeknete : „ Maruško ,	<b>neházej</b> /házet/Vi-S---2--N---I	písek po kamarádovi , to se nedělá , “ ale
<input type="checkbox"/>	<b>Láska je slípka</b>	žije , kde se broskve jen nakusují a pak se	<b>házej</b> /házet/VB-P---3P-AA--6I	za sebe ! Ale on to neudělal . Možná ,
<input type="checkbox"/>	<b>Rok perel</b>	se zapichují do fotografie oběti špendlíčky nebo se do ní	<b>házej</b> /házet/Vi-S---2--A---I	šipky . 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. <sup>0</sup>: 0,17 vztaženo k celému syn2015 | ARF<sup>0</sup>: 8,51 | Výsledek je promíchán

Výběr řádků: základní (0 vybrané)

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



# Časté případy podgenerování slovníku

- Cizojazyčné jednotky
- Vlastní jména
- Termíny
- Překlepy
- Řídká slova
- Řídce doložené varianty
- Varianty substandardní (nekodifikované, které nebyly zaneseny do morfologického slovníku)

# [tag="X.\*"]

Výskytů: 1 281 881 | i.p.m. 10 616,10 vztaženo k celému syn2015 | ARF 551 760,28 | Výsledek je promíchán

1 / 32 048 ▶▶▶

Výběr řádků: základní (0 vybrané)

<input type="checkbox"/>	<a href="#">Právní rádce</a>	dne 12 . 7 . 2007 , sp. zn. 26	<a href="#">Cdo /cdo/X@-----</a>	1837/2006 ) . Z § 711 odst. 1 písm. g
<input type="checkbox"/>	<a href="#">Sport</a>	při show a inspiraci pro celý svět sport fashion Barbory	<a href="#">Žehanové /žehanové/X@-----</a>	Šárka může být od medaile po dvacáté třicáté
<input type="checkbox"/>	<a href="#">PRO-ENERGY</a>	pro skladování plynů v plynné fázi ( plynojemy ) ,	<a href="#">TPG /tpg/X@-----</a>	703 01 Průmyslové plynovody ( pro materiály
<input type="checkbox"/>	<a href="#">STEREO &amp; VIDEO</a>	kanálu Voyo Cinema obsahují ještě komediální Voyo Comedy a rodinný	<a href="#">Voyo /voyo/X@-----</a>	Family . Naše Voyo . cz oproti tomu nabízí už
<input type="checkbox"/>	<a href="#">Xantypa</a>	na socialistických pódiih uvidíme skutečně živé The Plastic People Of	<a href="#">The /the/XX-----</a>	Universe ! Píše se rok 1987 a Plastici - byť
<input type="checkbox"/>	<a href="#">Blesk Hobby</a>	k dostání ní d za 145,90 a 319,90 Kč .	<a href="#">9,90Kč /9,90kč/X@-----</a>	? ? Po dobrém obědě oceníte sklenku piva Ma
<input type="checkbox"/>	<a href="#">Maminka</a>	: Sprchový gel , šampon a kondicionér v jednom ,	<a href="#">Caribbean /caribbean/X@-----</a>	Melodies Piña Colada , Hello Kitty ( 80Kč ) 4
<input type="checkbox"/>	<a href="#">Týden</a>	si toho musí být autoři vědomi . Sociální psycholog Uri	<a href="#">Simonsohn /simonsohn/X@-----</a>	z Pensylvánské univerzity nedávno s kolegy p
<input type="checkbox"/>	<a href="#">Při loupání cibule</a>	se ráda smála , hloubavý hrabě Syzzo a melancholický donátor	<a href="#">Thimo /thimo/X@-----</a>	von Küstritz , konečně pak Ekkehard Druhý a j
<input type="checkbox"/>	<a href="#">InStyle</a>	tentokrát útočí rovnou na celou postavu . Rozkaz od značek	<a href="#">Burberry /burberry/X@-----</a>	Prorsum , Lanvin , Michael Kors či Narciso Ro
<input type="checkbox"/>	<a href="#">Sport</a>	se Williamsová . Pokud sm dnes uspěje , bude nejstarší	<a href="#">kud /kud/X@-----</a>	dn grandslamovou šampionkou grandsl za po
<input type="checkbox"/>	<a href="#">Stručný přehled separačnic...</a>	vrstvu pomocí mikročerpadel ( průtok mobilní fáze je > 1	<a href="#">ml/s /ml/s/X@-----</a>	) pro zajištění rovnoměrného toku eluentu tenl
<input type="checkbox"/>	<a href="#">Mrznoucí mrholení</a>	laskavou oblost Ruka poté nevědoucí vědoucí Tváře božen němcoých Tříska	<a href="#">třísečka /třísečka/X@-----</a>	Možná nejsoucí 9 . 2 . 1971 VŠECHNY NAŠE Č
<input type="checkbox"/>	<a href="#">Základní charakteristiky on...</a>	nemocemi . Období středního věku V období středního věku (	<a href="#">Maturus /maturus/X@-----</a>	II ) mezi 45 . – 60 . rokem dochází
<input type="checkbox"/>	<a href="#">Základy biologie a genetiky...</a>	se po několika děleních zygoty podílejí komplexy HOX genů .	<a href="#">HOX /hox/X@-----</a>	geny jsou během embryonálního vývoje odpov
<input type="checkbox"/>	<a href="#">Tajný deník</a>	teď trochu strach , že jsem něco pokazil , paní	<a href="#">McNultyová /mcnultyová/X@-----</a>	. " Byla jsem příliš unavená , abych už zase
<input type="checkbox"/>	<a href="#">Praktická cvičení z klinické...</a>	hlavy , kdy se poškodí osmoreceptory v hypothamu ) .	<a href="#">Hyponatremie /hyponatremie/X@-----</a>	může vzniknout v souvislosti s hyperglykemií

# Závěr

- Výsledky automatické morfologické analýzy - užitečný nástroj
- Vždy bude obsahovat chyby
- Chyby je možné odstraňovat
- Nikdy nebudou odstraněny všechny chyby
- Práce s a bez použití automatické analýzy

# Úkol na příště

- Prostudovat pražský tagset  
(<http://wiki.korpus.cz/doku.php/pojmy:tag?redirect=1#tagset>)
- Prostudovat brněnský tagset  
(<https://nlp.fi.muni.cz/projekty/ajka/tags.pdf>)
- článek \_ IS