

CJBB84

Morfologie a korpus

7.30-9.00 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"/>	Dobry den s kurýrem	, který vynikajícím zákrokem vytěsnil krásnou střelu pod břevno ze	standardky /standardka/NNFS2-----A-----	. Branka : Svízela . ACH TA KONCOVKA ! -
<input type="checkbox"/>	Bílá vlčice	Okamžitě rozpoznal povahu svého zanícení pro tuto neznámou ženu .	Gabriel /gabriel/NNMS1-----A-----	byl náročný muž , potpěl si na dokonalos
<input type="checkbox"/>	Tango Masaj	domana , kterou oblékal jen při slavnostních příležitostech , na	nohou /noha/NNFP6-----A-----	měl kožené sandály a na krku , zápěstích
<input type="checkbox"/>	E15	prodal v roce 2006 Anwilu více než čtyřpětinový podíl ve	Spolaně /spolana/NNFS6-----A-----	za částku přesahující 640 milionů korun .
<input type="checkbox"/>	Hřbitovní kvítí na smetaně	dostatečně inspirativní a pravděpodobně stačilo k výstavbě základní podoby románové	figury /figura/NNFS2-----A-----	. Pro mne osobně je Hardy trochu nudný p
<input type="checkbox"/>	Čtyři kultury ve vzdělávání	uvedených dvou typů reakce na " kritiku modernismu " -	teorie /teorie/NNFS1-----A-----	komunikativní akce a postmoderní dekon:
<input type="checkbox"/>	Století fyzikálních objevů	paprsky vyletující od anody směrem ke katodě a dal jim	název /název/NNIS4-----A-----	anodové nebo kanálové . ● Německý fyzik
<input type="checkbox"/>	Silnice a železnice	možné vyhodnotit garanci systémových parametrů ITS aplikace . VYUŽITÍ GNSS	SIMULÁTORU /simulátor/NNIS2-----A-----	Pro zajištění pilotních i řádných testů a sp
<input type="checkbox"/>	Svět ženy	bylo nám ve třech docela dobře . Ale pak obě	děti /dítě/NNFP1-----A-----	krátce po sobě odešly na studium a za pr
<input type="checkbox"/>	Atlas pískovcových skalnic...	současné době dodává do suťových akumulací další bloky . Kamenná	moře /moře/NNNP1-----A-----	v krystalických horninách (např. rulách)
<input type="checkbox"/>	Britské detektivky	, která se k zločinu připlete mimo službu (Z	příčin /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 [Jdi](#) [Další](#) | [Poslední](#)

doc#0	<code><p></code>	Brod /brod/k1glnSc1qP	porazil Chropyni a ztrácí na ni bod <code></p></code>
doc#0	<code><p></code> Brod porazil	Chropyni /Chropyně/k1gFnSc4	a ztrácí na ni bod <code></p><p></code> Náskok Chropyně
doc#0	<code><p></code> Brod porazil Chropyni a ztrácí na ni	bod /bod/k1glnSc4qP	<code></p><p></code> Náskok Chropyně na čele se snížil
doc#0	porazil Chropyni a ztrácí na ni bod <code></p><p></code>	Náskok /náskok/k1glnSc1qP	Chropyně na čele se snížil. V Brodě prohrála
doc#0	Chropyni a ztrácí na ni bod <code></p><p></code> Náskok	Chropyně /Chropyně/k1gFnSc2	na čele se snížil. V Brodě prohrála a právě
doc#0	ztrácí na ni bod <code></p><p></code> Náskok Chropyně na	čele /čelo/k1gNnSc6qP	se snížil. V Brodě prohrála a právě její
doc#0	<code></p><p></code> Náskok Chropyně na čele se snížil. V	Brodě /brod/k1glnSc6qP	prohrála a právě její dnešní soupeř se
doc#0	dnešní soupeř se za ni dotáhl na nejmenší	rozdíl /rozdíl/k1glnSc4qP	. A to není jediná změna na čele. Sestupný
doc#0	dotáhl na nejmenší rozdíl. A to není jediná	změna /změna/k1gFnSc1qP	na čele. Sestupný kurs po víkendu určitě
doc#0	nejmenší rozdíl. A to není jediná změna na	čele /čelo/k1gNnSc6qP	. Sestupný kurs po víkendu určitě nabere
doc#0	A to není jediná změna na čele. Sestupný	kurs /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. ⁰: 974,65 vztaženo k celému syn2015 | ARF ⁰: 64 433,19 | Výsledek je promíchán

1 / 2 943 ▶▶▶

Výběr řádků: základní (0 vybrané)

<input type="checkbox"/>	Nikdy jsem nechtěla zabít	vyzývavé pohledy , jako by jenom čekala , až si	je /oni/PPFP4--3-----	někam odvedou . Trvalo mi nějakou dobu , než jsem
<input type="checkbox"/>	ABC	posměškům trápila , vztekala nebo se urážela . To by	je /oni/PPFP4--3-----	jen vyprovokovalo , aby se ti smáli ještě víc .
<input type="checkbox"/>	Dům u jezera	. Z úplně obyčasných mladíků dělá jiné lidi . Zlomí	je /oni/PPFP4--3-----	. " Pomyslela jsem na Alfreda tu noc na schodech
<input type="checkbox"/>	Hana Benešová	, asi 70 kilometrů severně od New Yorku . Doprovází	je /oni/PPFP4--3-----	Mary Stevensonová , sekretářka Hamiltona Fishe Armstronga , zakladatele a
<input type="checkbox"/>	Právo	nich začal shánět , obžalovaný mu řekl , že mu	je /oni/PPFP4--3-----	vrátí , ale zadarmo to prý nebude . „ Chtěl
<input type="checkbox"/>	Reflex	levicové kavárny . Oficiálně Praha 7 sobě . Nevolil jsem	je /oni/PPFP4--3-----	. Z přesvědčení a asi i z předsudku . Ve
<input type="checkbox"/>	Jihlavské listy	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"/>	Navždy tvůj	ale ona to odmítla . " – „ No to	je /oni/PPFP4--3-----	teda prosím pěkně děs běs , " zhodnotila Bianca .
<input type="checkbox"/>	Děbelská zubařka	„ Okamžitě nasedni , Alfredě ! " V té chvíli	je /oni/PPFP4--3-----	dohnali i pan Snood a studenti dramatické výchovy . Nahruli
<input type="checkbox"/>	Krásné zlo	se rozhodli , že ti pořídíme nové . Dnes už	je /oni/PPFP4--3-----	vyrábějí mnohem bezpečnější , s airbagy a podobně . Peníze
<input type="checkbox"/>	Páté roční období	. " Malin víří hlavou tisíce myšlenek . Snaží se	je /oni/PPFP4--3-----	uspořádat , zachytit , vybavují se jí však jen tváře
<input type="checkbox"/>	Největší rytíř	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"/>	Ona dnes	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. θ : 0,02 vztaženo k celému syn2015 | ARF θ : 1,07 | Výsledek je promíchán

Výběr řádků: základní ▾ (0 vybrané)

- | | | |
|--------------------------|----------------------------------|---|
| <input type="checkbox"/> | Orchestrion v hlavě | Dnes mi psala máma Chlapče nebuď fláma Pana hejtmana si važ /vázat/Vi-S---2--A----I Spi na levém uchu Drž si nohy v suchu Franzbrantweinem |
| <input type="checkbox"/> | 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 vztaženo k celému syn2015 | ARF⁰: 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"/>	Dobrý 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...	systému tendenci tvrdit . Tak i dnešní reklamní slogan „	Nevaž /vážit/Vi-S---2--N----I	se , odvaž se ! “ slouží , jak známo

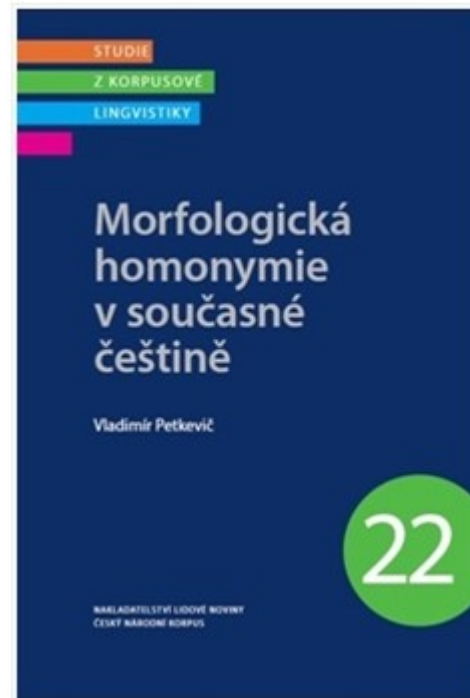
v důsledku tvarové homonymie

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

Výběr řádků: základní ▾ (0 vybrané)

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

Výskytů: 44 | i.p.m. ⁰: 0,36 vztaženo k celému syn2015 | ARF ⁰: 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 . „ Šílíš ?	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 nakusují 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	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. ⁰: 0,17 vztaženo k celému syn2015 | ARF⁰: 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 sepišu
<input type="checkbox"/>	Týdeník Květy	zařve : „ Mámo , ať už ty parchanti kurva	spěj /spět/Vi-S---2--A----I ! ” Nebo rodič , vláčejíci tělesně i duševně vyčerpané
<input type="checkbox"/>	Kloktat dehet	druhej den vlka loví . Po večeri dou spát .	Spěj /spět/Vi-S---2--A----I všichni v jedný seknicí , 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 jednorozců	. „ 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 smrký
<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 břicha ukazováku po vyleštěné plošce . Mezera Všichni tři	spěj /spět/Vi-S---2--A----I na hromadě bágků . Hlídám je , plním kartu svého
<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	, písmáček , stařeček , 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	voní 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

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