

PLIN063_6

Algoritmický popis morfologie

osolsobe@phil.muni.cz

Slovnědruhové přechody a přesahy

- Zájmenná adverbia
- Číslovky

Algoritmický popis adverbíí paradigmaticky tvořených od adjektiv

- Zájmenná adverbia
- Měrové výrazy
- potence a realita - deadverbiální adjektiva
- adverbializace jmenných tvarů adjektiv a problém jejich zachycení a interpretace ve strojových slovnících

Zájmenná adverbia

PŘÍSLOVCE	určitá	neurčitá	záporná	tázací	vztažná	reflexivní	ukazovací
zájmenná	všude, všady, jinde, jinam, onam, druhde	ně()	nikde, nikudy, nikam, odnikud, pranikde	kde, kudy, kam, pokud	kdež, kamž		tu, tady, tam, tudy, tamtudy, odtud, potud, zde, sem, semhle
	pokaždé, jedině, jedinkrát(e), po(v)obakakrát(e), vždy, navždy, jindy, najindy, odjindy, dojindy, druhdy	ně()	nikdy, odnikdy, donikdy, nanikdy, pranikdy	kdy, odkdy, dokdy, nakdy, pokdy, dokavad			ted', tehdy, dotehdy, odtehdy, natehdy,
	jinak, onak	ně()	nijak, pranijak, nikterak	jak	příčemž, načěž, začěž, pročěž		tak, takto

Značkování adverbíí v korpusech řady SYN

- Db
- Dg




	Filtr	tag	Freq	
1	p / n	Db-----	4861873	
2	p / n	Dg---1A--	1821624	
3	p / n	Dg---2A--	283507	
4	p / n	Dg---3A--	73241	
5	p / n	Dg---1N--	51827	
6	p / n	Dg---2A-1	15813	
7	p / n	Db----N--	14496	
8	p / n	Dg---3A-1	3455	
9	p / n	Dg---2N--	1332	
10	p / n	Dg---2A-6	633	
11	p / n	Db----- VB-S--2P-AAI-1	578	
12	p / n	Dg---2A-2	552	
13	p / n	Dg---2A-7	203	
14	p / n	Dg---3A-6	109	
15	p / n	Db-----6	76	
16	p / n	Db-----7	66	
17	p / n	Db-----4	49	
18	p / n	Dg---3A-2	43	
19	p / n	Dg---1N-1	24	
20	p / n	Dg---1A-- VB-S--2P-AAI-1	19	
21	p / n	Dg---3A-7	17	
22	p / n	Db----- Z:----- VB-S--2P-AAI-1	14	
23	p / n	Dg---3N--	12	
24	p / n	Dg---1A-1	6	
25	p / n	Dg---1A-4	5	














Měrové výrazy

- Substantivum, adverbium (*málo*)
- Číslovka, adverbium (*mnoho*)

Slovnědruhov^á platnost výrazu *mnoho* (*mnoha*, *mnohem*, ...)

















Celkem: 3 položky (1 stránka)

	Filtr	lemma_lc [lowercase lemma]	lc [lowercase word]	tag	Freq	
1	p / n	mnoho	mnoho	Ca-1	12219	
2	p / n	mnoho	mnoho	Ca-4	11344	
3	p / n	mnoho	mnoho	Dg-1A	416	


<input type="checkbox"/>		by je zajímala otázka jejich dcery , která bude o	mnoho/mnoho/mnoho/Ca--4 -----	let později svírat tu fotografii v ruce : Jak se
<input type="checkbox"/>		Obsidián byl nejlepší kámen na řezání , na zbraně a	mnoho/mnoho/mnoho/Ca--4 -----	dalších nástrojů , protože je to vlastně sopečné sklo a
<input type="checkbox"/>		ze svého dospívání , případně z dospívání svých dětí .	Mnoho/mnoho/mnoho/Ca--1 -----	dospělých se domnívá , že děti by je měly respektovat
<input type="checkbox"/>		lidského utrpení . Viděla jsem za ty roky smrt na	mnoho/mnoho/mnoho/Ca--4 -----	způsobů , mnohokrát stála u lůžka nemocného při posledním vydechnutí
<input type="checkbox"/>		. Na to Chálid odpověděl , že je jich opravdu	mnoho/mnoho/mnoho/Ca--1 -----	, neboť nejeden muž u nich se pokouší o básnické
<input type="checkbox"/>		. Tím se artefakt stává interpretační úlohou , která má	mnoho/mnoho/mnoho/Ca--4 -----	variant řešení . Vopěnka na jednoduchém příkladu pohledu člověka na
<input type="checkbox"/>		- po lásce . Mějme ji ve svém srdci a	mnoho/mnoho/mnoho/Ca--4 -----	věcí změníme a skrze lásku změníme pisatele anonymu , "
<input type="checkbox"/>		bratrstvím . To není jako v kostele , kde se	mnoho/mnoho/mnoho/Ca--1 -----	lidí tváří dokonale . Přednášející manželský pár nemluví jen o
<input type="checkbox"/>		je jízda s ním neuvěřitelně komfortní a posádce dokáže ušetřit	mnoho/mnoho/mnoho/Ca--4 -----	sil , což na Dakaru 2020 určitě ocení . .
<input type="checkbox"/>		se superprognostiků týká , je stále v plenkách . Existuje	mnoho/mnoho/mnoho/Ca--1 -----	otázek , které jsme teprve začali zkoumat . Jedna z
<input type="checkbox"/>		připouštět , že kolečko klidně mohlo v Evropě existovat o	mnoho/mnoho/mnoho/Dg-----1A ----	dříve a že jej možná běžně používali staří Řekové .
<input type="checkbox"/>		, u kterých byly ubytovány dvě au-pair . To nebylo	mnoho/mnoho/mnoho/Ca--1 -----	, někdo by to mohl nazvat přímo fiaskem a plýtváním
<input type="checkbox"/>		z minulého století nosí těm hodným vajíčka a sladkosti .	Mnoho/mnoho/mnoho/Ca--1 -----	lidí se skutečně věnuje pokání , rozjímání a půstům či
<input type="checkbox"/>		znamenal to všechno . A když už bylo lásky příliš	mnoho/mnoho/mnoho/Ca--1 -----	a on se ke mně tiskl zezadu s rukama na

slovnědruhová platnost výrazu *málo* (*mála, málem*)

Celkem: 16 položek (1 stránka)

	Filtr	sublemma_lc [lowercase sublemma]	lc [lowercase word]	tag	Freq	
1	p / n	méně	méně	Dg—2A—	20802	
2	p / n	málo	málo	Dg—1A—	15529	
3	p / n	méně	nejméně	Dg—3A—	6811	
4	p / n	míň	míň	Dg—2A—	1673	
5	p / n	málo	mála	NNNS2—A—	1580	
6	p / n	neméně	neméně	Dg—2N—	1128	
7	p / n	míň	nejmíň	Dg—3A—	834	
8	p / n	nemálo	nemálo	Dg—1N—	338	
9	p / n	málo	málo	NNNS1—A—	171	
10	p / n	málo	málo	NNNS4—A—	129	
11	p / n	málo	málem	NNNS7—A—	117	
12	p / n	sebeméně	sebeméně	Db—	46	
13	p / n	málo	málu	NNNS3—A—	16	
14	p / n	málo	málu	NNNS6—A—	13	
15	p / n	sebemíň	sebemíň	Db—	11	
16	p / n	málo	mále	NNNS6—A—1	7	

Celkem: 2 položky (1 stránka)

	Filtr	lc [lowercase word]	lemma_lc [lowercase lemma]	tag	Freq	
1	p / n	málem	málem	Db—	5130	

potence a realita - deadverbiální adjektiva

- Dg.*
- Pravidelné rozgenerování a možnost stupňovat
- Deadverbiální adjektiva a problém stupňování
- Adverbiální platnost denumerálií *prve, prvně, druhdy*

adverbializace jmenných tvarů adjektiv a problém jejich zachycení a interpretace ve strojových slovnících

- Predikativa – syntaktická klasifikace
- Jmenné tvary
- Substantivizace a adverbializace

Algoritmický popis číslovek

- číslovka jak sémanticky definovaný slovní druh
- druhy číslovek v českých mluvnicích a značkování číslovek v korpusech
- nepřítomnost slovnědruhového značkování číslovek jako překážka dalšího využití automatických nástrojů v lingvistickém výzkumu
- subklasifikace číslovek a zájmen jako morfologická deskripce a pomůcka pro syntaktickou analýzu
- slovnědruhová přesažnost číslovek, adverbii a zájmen, číslovky neurčité

Sémantická klasifikace druhů číslovek - počet čeho vyjadřují

- napříč klasifikací – určité/neurčité (vyjadřují určitý/neurčitý počet)
- prostý počet - číslovky základní
- pořadí - číslovky řadové
- počet souborů nebo druhů - číslovky úhrnné/čísllovky druhové/
čísllovky souborné
- četnost - číslovky násobné
- podíl - substantiva z číslovek (*polovina, třetina, ..., tisícina*) - číslovky dílové
- počet skupin - substantiva z číslovek (*dvojka, pětko, ...*) - číslovky skupinové

Subklasifikace slovního druhu ve značkách ČNK: C.

2) Detailní určení slovního druhu

[16]

- a - číslovka neurčitá ("mnoho", "málo", "tolik", "několik", "kdovíkolik", ...)
- d - číslovka druhová, adjektivní skloňování ("jedny", "dvoji", "desaterý", ...)
- h - číslovka druhová "jedny" a "nejedny"
- k - číslovka druhová ≥ 4 , adjektivní postavení, krátký tvar ("čtvery", ...)
- j - číslovka druhová ≥ 4 , substantivní postavení ("čtvero", "desatero", ...)
- l - číslovka základní 1-4, "půl", ...; "sto" a "tisíc" v nesubstantivním skloňování
- n - číslovky základní ≥ 5
- w - číslovka neurčitá s adjektivním skloňováním ("nejeden", "tolikátý", "několikátý" ...)
- y - zlomek zakončený na "-ina" (značkováno jako číslovka)
- ? - číslovka "kolik"
- = - číslo psané číslicemi (značkováno jako číslovka)
- o - číslovka násobná neurčitá ("-krát": "mnohokrát", "tolikrát", ...)
- r - číslovka řadová
- u - číslovka tázací násobná "kolikrát"
- v - číslovka násobná ("-krát": "pětkrát", "poprvé" ...)
- z - číslovka tázací řadová "kolikátý"

Subklasifikace slovního druhu ve značkách analyzátoru *ajka*

4 k4 - Číslovka

x	Druh číslovky (x)
C	Základní
O	Řadová
R	Druhová
G	Gramatika
H	Gramatika

y	Druh číslovky (y)
N	Záporná
I	Neurčitá

Vzory pro číslovky

+stot

<ř,A> 4F,4G1

<ři,A> 6KRAT

<řet,B> PRM

<řetí,B> 6KRAT

<roj,C> PRM,6NASOB,6MO

<roj,k,K> V7

<rojek,K> V78X

<roj,c,K> V7A

<řetin,L> V7,V7B,V78X

<rojic,N> V8,V8A,V78X

+stočt

<yř,A> 4F,4G2,6NASOB,6MO

<:yr,A> 4FQ

<yři,A> 6KRAT

<:yry,A> 6KRAT

<yr,A> 6NASOB

<vrt,B> PRT1,PRT2

Slovnědruhové přehodnocení některých skupin

Číslovky

(<https://wiki.korpus.cz/doku.php/cnk:syn2020:tag>)

Značení číslovek (C) bylo rozšířeno na úkor jiných slovních druhů. Jako číslovky jsou oproti dřívějším korpusům řady SYN značena:

- **původní substantiva**: základní číslovky substantivně skloňované (*sto, tisíc, milion, miliarda* atd.); několik číslovek dílových (*polovina, půl, polovic, čtvrt*); všechny úhrnné číslovky (*čtvero, patero, desatero* atd.; *dvé, tré*)
- **původní adjektiva**: násobné číslovky adjektivně skloňované (např. *dvojitý, čtyřnásobný, mnohonásobný*)
- **původní adverbia**: násobné číslovky adverbialní (*osminásobně, dvojnásob, několikanásobně*), některé adverbialní spřežky (*nadvakrát, zatřetí, naponěkolikáté*)
- Dochází tak k většímu souladu značkování číslovek v korpusu s tradičním dělením číslovek ve slovnících a gramatikách.

Nová subkategorizace číslovek na 2. pozici


(https://wiki.korpus.cz/doku.php/cnk:syn2020:tag:prehled_pozic_1

2)

C	numerál (číslovka, nebo číselný výraz s číslicemi)	l	číslovka základní 1–4 + „nejeden“
		n	číslovka základní 5–99, i pokud je součástí složené číslovky psané dohromady („dvacetpět“, „stotřicet“, „pětapůl“)
		z	číslovka základní se substantivním skloňováním („sto“, „milion“, „nula“ apod.)
		a	číslovka základní neurčitá a tázací („mnoho“, „tolik“, „kolik“)
		y	číslovka dílová („půl“, „polovic“, „polovina“)
		r	číslovka řadová
		w	číslovka řadová neurčitá a tázací
		d	číslovka druhová a souborová („dvojí“, „obojí“, „čtvery“ včetně tvarů „dvoje“, „oboje“, „čtvery“, „obě“, „jedny“)
		h	číslovka druhová a souborová neurčitá a tázací
		j	číslovka úhrnná („čtvero“, „patero“, „devatero“, „dvě“, „tré“)
		k	číslovka úhrnná neurčitá a tázací („několikero“, „tolikero“, „kolikero“)
		u	číslovka násobná (adjektivní typ: „dvojí“, „osminásobný“)
		3	číslovka násobná neurčitá a tázací (adjektivní typ: „mnohonásobný“, „xnásobný“, „kolikanásobný“)
		v	číslovka násobná (adverbiální typ, včetně spřežek: „pětkrát“, „osminásobně“, „trojnásob“, „jednou“)
		o	číslovka násobná neurčitá a tázací (adverbiální typ, včetně spřežek: „mnohokrát“, „několikanásobně“, „pokolikáté“, „naponekolikáté“ ...)
		=	číslo psané arabskými číslicemi
}	číslo psané římskými číslicemi		

Jak je značkováno lemma=první

Celkem: 2 položky (1 stránka)

	Filtr	pos [part of speech]	Freq	
1	p / n	C	121789	
2	p / n	A	8	

Výskytů: 8 | i.p.m.: 0,07 (vztaženo k celému korpusu) | ARF: 3,2 | Výsledek je promíchán

Výběr řádků:

<input type="checkbox"/>		ústních argumentacích) , protože právě ona zprostředkovává ten „	nejprvnější/první/první/AAIS4----3A----	“ dojem a také proto , že podvědomě každý z
<input type="checkbox"/>		– příslušníci letectva , pro něž je obrana vlasti tím	nejprvnějším/první/první/AAIS7----3A----	a nejposvátnějším úkolem , – ulívat někde v zázemí ?
<input type="checkbox"/>		na světě . Abychom ji však mohli prohlásit za “	nejprvnější/první/první/AAFS4----3A----	ve všech ohledech “ , musíme se poohlédnout ještě po
<input type="checkbox"/>		od hladu strápenej , teď pak od některé baby šaty	prvnějšího/první/první/AAMS2----2A----	muže přiděném zdalíž proč jiného přichází , než aby se
<input type="checkbox"/>		na určité místo za vsí o závod . Kdo byl	nejprvnější/první/první/AAMS1----3A----	u cíle , jede nazpět v čele všech ostatních a
<input type="checkbox"/>		vzhled jako takový . Právě ony působí ten úplně „	nejprvnější/první/první/AAIS4----3A----	“ dojem a bude - li vaše písemné podání vypadat
<input type="checkbox"/>		zpívajících mužů , je zachovali : pakli kde co z	prvnější/první/první/AAFS2----2A----	horlivosti ubylo , aby se obnovilo [...] “
<input type="checkbox"/>		se píše , že tito žalmové „ hned od svého	nejprvnějšího/první/první/AANS2----3A----	vydání v jazyku našem do církví našich uvedeni jsou a

Jak je značkováno lemma=druhý (SYN2020 a SYN2010)

Celkem: 1 položka (1 stránka)

	Filtr	pos [part of speech]	Freq
1	p / n	C	106941




Výskytů: 20 172 | i.p.m.: 165,8 (vztaženo k celému korpusu) | ARF: 11 157,8 | Výsledek je promíchán

Výběr řádků: základní

<input type="checkbox"/>	Y → Lidé a země, č. 7/2006	v nivě Moravy , si nikdo z nich nepoložil .	Druhý/druhý/AAIS1----1A-----	, mikulčický břeh přitom přitahoval archeology dlouhá léta . Od
<input type="checkbox"/>	Y → Nedělní sport, č. 21/2008	Behrami . Švýcaři neustávali v pekelném tempu a brzy přidali	druhý/druhý/CrIS4-----	gól . Po akci Vonlanthena s Behramim poslal míč do
<input type="checkbox"/>	Y → Právo, 14. 9. 2005	nezářil . " Ale jeho čas přijde , " věří	druhý/druhý/CrMS1-----	ze střešovických koučů Petr Harapát . Minimálně čtyři utkání by
<input type="checkbox"/>	Y → Právo, 21. 7. 2007	MS supersportů (18 kol) , 15.30 h :	druhý/druhý/CrIS1-----	závod MS superbiků (20 kol) . Příjmení Haškovi
<input type="checkbox"/>	Herriot, James → Moje nejmilejší kočičí historky	že do osmačtyřiceti hodin bude Olly mrtvý . Bylo to	druhý/druhý/CrIS1-----	den ráno . Helena volala rozrušeně ze zahrady za domem
<input type="checkbox"/>	Venclik, Vlastimil → Nedopsané příběhy a jiné hry	Král : (vyštěkne) A co ten královrah ?	Druhý/druhý/CrMS1-----	úředník : Čistému vše čisté . Král : (ho
<input type="checkbox"/>	Petráňová, Lydia → Domovní znamení staré Prahy	ulici , patřily různým městům : jeden Malé Straně a	druhý/druhý/AAFS3----1A---6-	Hradčanům . Bílé jablko na Strakovském domě , vkomponované do
<input type="checkbox"/>	Budinský, Václav → Gymnázium	improvizovaný , že nebyl do programu zařazen a schválen .	Druhý/druhý/CrIS4-----	den se ráno před vyučováním zastavil třídní Hlasák ve třídě
<input type="checkbox"/>	Callananová, Maggie - Kelleyová, Patricia → Poslední dary	psů hlasitě zafuněl a sedl si mi na nohu a	druhý/druhý/CrMS1-----	si opřel svou chundelatou zlatou hlavu o mou sukni .
<input type="checkbox"/>	Y → Sport, 25. 3. 2008	na bedně , stoprocentně , v normálním závodě bych byl	druhý/druhý/AAMS1----1A-----	, " prohlásil Lewis Hamilton . " Když jsem stál
<input type="checkbox"/>	Kolář, Jan → Biologické hodiny rostlin) výška stonku . Další podrobnosti v textu . Ještě	druhý/druhý/CrIS1-----	příklad : Mendel vybral pro své pokusy mimo jiné i
<input type="checkbox"/>	Hathaway, Charla → Erotická masáž	velkého dne , ale já si užívala přítomný okamžik .	Druhý/druhý/CrIS4-----	den jsme naproti sobě seděli jako muž a žena .
<input type="checkbox"/>	Y → Právní rádce, č. 4/2009) nebo právního řádu (italský , švýcarský) .	Druhý/druhý/CrIS1-----	poukaz se zdá být přesnější - vlastník nemusí být omezen
<input type="checkbox"/>	Knopp, Guido → Vatikán: moc papežů	modlí . Ale ve zpětném pohledu je zřetelné jedno :	Druhý/druhý/CrIS4-----	vatikánský koncil nedokázal naznačit žádnou konkrétní cestu k jedno
<input type="checkbox"/>	Rotrekl, Zdeněk → Světlo přichází potmě	ležel na posteli proti výlevce , Přeražený nos i ten	Druhý/druhý/CrMS1-----	zmizeli , jak se dostal zpátky , nedokázal pochopit ,
<input type="checkbox"/>	Woolrich, Cornell → Noční můry	. " Ten chlap musí být neviný , " poznamenal	druhý/druhý/CrMS1-----	detektiv , když se vrátil . " Ví , že
<input type="checkbox"/>	Burian, Jan → Sága o cestě na Island	TELEVIZE Islandská televize má dva programy , jeden státní a	druhý/druhý/AAIS4----1A-----	soukromý . Za soukromý musíte zaplatit zvlášť . Pokud tak

Jak je značkován tvar lc=jednou

Celkem: 3 položky (1 stránka)

	Filtr	lemma_lc [lowercase lemma]	tag	Freq	
1	p / n	jednou	Db-----	16806	
2	p / n	jednou	Cv-----	14576	
3	p / n	jeden	CIFS7-----	10283	

Jak jsou značkována lemmata *jinaký*, *jinačí*, *onačí*, *onačejší*

Výskytů: 155 | i.p.m.: 1,27 (vztaženo k celému korpusu) | ARF: 87,43 | Výsledek je promíchán 1 / 4 ▶▶▶

Výběr řádků:

<input type="checkbox"/>		si postupně uvědomovaly , že je doma čeká taková či	onaká/onaký/onaký/AAFS1----1A----	práce . Omlouvaly se a zvolna se trousily pryč .
<input type="checkbox"/>		S díky jsem přijal . Byl to dobrý tabák .	Jinačí/jinačí/jinačí/AAMP1----1A----	, než jaký měl ten žebrák . Takhle chutnalo lepší
<input type="checkbox"/>		Perspektiva Má milá rozmilá , neplakej ! Život už není	jinakej/jinaký/jinaký/AAIS1----1A---6	. Dnes budme ještě veselí na naší bílé posteli !
<input type="checkbox"/>		jinak se napojovaly na kosti , a hlavně měly docela	jinačí/jinačí/jinačí/AAFS4----1A----	pevnost . Zůstal bezmocně stát . Věděl , že teď
<input type="checkbox"/>		, že ve svém věku za sebou mám i věru	onačejší/onaký/onaký/AAFP4----2A----	výpravy . To ano , to opravdu ano . Třeba
<input type="checkbox"/>		. Poslušně odběhl a do sálu se vrátil s něčím	onačejším/onaký/onaký/AAANS7----2A----	. Na stůl mezi ně rozložil původně bělostnou tuniku zbrocenou
<input type="checkbox"/>		čemu je to výzva ! Lidé prchají a zeměkoule bude	jinačí/jinačí/jinačí/AAFS1----1A----	, než byla ... S tím žádný soucit , žádné
<input type="checkbox"/>		. To je fakt výjimečné setkání . Já tu zažil	jinačí/jinačí/jinačí/AAFP4----1A----	věci ! Velmi diskutabilní . S lesem , se stromy
<input type="checkbox"/>		věku nic , to bych musel mít k dispozici úplně	jinačí/jinačí/jinačí/AAIS4----1A----	elixír . Ale když jsem byl mladší , vždýcky jsem
<input type="checkbox"/>		výlet do naší milované psychiatrické léčebny Kaščenko ! Tam jsou	onačejší/onaký/onaký/AAMP1----2A----	pikadoři ! " Vzpomněl jsem si na Ljoňu a Hemingwaye
<input type="checkbox"/>		zlatý důl . Jenže černou dírou je zdravotnictví i v	onačejším/onaký/onaký/AAIS6----2A----	smyslu . Do lepší starosti o pacienty se dá vždy
<input type="checkbox"/>		těch prvních dnů v New Yorku . Všechno byl tak	jinačí/jinačí/jinačí/AAANS1----1A----	, tak hrůzně cizí . Předtím jsem nikdy nevytáhl paty
<input type="checkbox"/>		lepší , anebo horší ? Těžko soudit , proto píšu	jinačí/jinačí/jinačí/AAANS4----1A----	. 19 . února 1984 , neděle , ve 22
<input type="checkbox"/>		a poslouchala nás . Měl velké množství neuvěřitelných informací ,	jinačí/jinačí/jinačí/AAANS4----1A----	než ty otřepané historky našeho vysloužilce . Jako advokát nahlédl
<input type="checkbox"/>		kanceláře . Jejich předchozí nájemci je dávno vyměnili za něco	onačejšího/onaký/onaký/AAANS2----2A----	a vyklidili tak prostor pro penzisty , duševně vyšinutý a

Závěr

- Některé návrhy projektu NovaMorf (např. změny v subklasifikaci číselnek) byly již implementovány v praxi
- Některé problémy souvisí s desambiguací
- Některé problémy se dále řeší