

PLIN063_2

Algoritmický popis morfologie

osolsobe@phil.muni.cz

Strojové slovníky

- slovoforma/systémové slovo/interpretace – základní jednotka strojového slovníku
- seznam slovoforem
- tvorba seznamu
- interpretace na rovině systémového slova a gramatických významů
- gramatické kategorie a slovotvorné kategorie
- praxe tagsetů v českém prostředí

Hajič – definice morfologického slovníku

- Morfologický slovník obsahuje ke každému lemmatu informaci o **kmeni** slova (v našem případě, kvůli sloučení nejnižších rovin popisu jazyka do jedné, je za kmen slova považována ta část slova, která se při ohýbání nemění), a o přípustných koncovkách.
- Množina koncovek tvoří **vzor**. U každé koncovky je navíc informace o tom, které značky (kombinace hodnot morfologických kategorií) jí pro daný vzor odpovídají.

Tvarosloví a co pod ně zahrnout

- Systémové a textové slovo (slovoforma)
- Slovníkový tvar
- Hranice derivační a flektivní morfologie
- Různost jazyků a různost přístupů

Systemové a textové slovo

- Systemové slovo – lemma
- Textové slovo – word
- Lemmatizace a varianty

Slovníkový tvar

- Heslová slova (SSJČ, korpusy – nerozpoznaná slova)
- Hnízdování (rozgenerování tvarů, tvary zahrnuté pod morfologické vzory)
- Lemmatizace a stupňování
- Lemmatizace a negace

Hranice derivační a flektivní morfologie

- Lemmatizace a slovnědruhová transpozice
- Lemmatizace a jednoslovná morfologie (reflexiva tantum, příslovečné spřežky, víceslovné spojky, ...)

Různost jazyků a různost přístupů

133828	()	, vrávorat' (tacket' sa), húlit' (fajčit'/fajčit') atď . Z hľadiska jazykovej kultúry
348162	()	! (Čičmany) ; Takí šuseň , a už	fajčí/fajčit'	! (Jablonové) ; To šicke jeho dzeci
1003561	()	inú farbu pokožky ako bielu : V Amerike	fajčia/fajčit'	len asociáli a farební . LT 1998 ; Sadla
1474150	()	na vás ! Ste podvyživený ? Pribrať !	Fajčíte/fajčit'	? Okamžite prestaňte ! Nešportujete
1487979	()	všetkých poistencov , teda aj tých , ktorí	nefajčia/nefajčit'	? Alibi pána premiéra a jeho podpredsedu
1756277	()	každý štvrtý má menej ako 15 rokov ,	fajčili/fajčit'	marihuanu minimálne raz . Tento fenomén
1764214	()	a ten má sám asi 40 rokov , marihuanu	fajčil/fajčit'	už jeho otec . Nuž , niekto by povedal
1764252	()	slúžiť ako výzva k tomu , aby ste marihuanu	fajčili/fajčit'	alebo požívali v akejkol'vek inej forme
1930546	()	bezpečnostný pás ? Mal by som prestať	fajčit'/fajčit'	? Je moja štvrť dost' bezpečná ? “
2007463	()	kvantá silno korenených jedál , veľa	fajčil/fajčit'	a pil . Mal preto stále ťažkosti s obličkami

Různost jazyků a různost přístupů

- <https://www.sketchengine.eu/penn-treebank-tagset/>
- <https://www.sketchengine.eu/German-rftagger-part-of-speech-tagset/>

Praxe tagsetů v českém prostředí

- <https://www.sketchengine.eu/french-treetagger-part-of-speech-tagset/>
- <https://www.sketchengine.eu/tagset-reference-for-czech/>
- <https://wiki.korpus.cz/doku.php/en:pojmy:tag>

Poziční systém

- Pozice a hodnota
- Nepřehlednost a heterogennost druhé pozice
- Sdružené hodnoty jako obcházení desambiguace
- Čas a slovesný rod
- Stylistické hodnocení
- Nejednotnost v přístupu k variantám

System atribut/hodnota

- Jednoduchost
- Atribut *m*
- Atribut *d*
- Subklasifikace zájmen
- Subklasifikace číslovek
- Subklasifikace adverbí
- Atribut *w*
- Práce se sdruženými hodnotami pro neohebné slovní druhy

Společné rysy a rozdíly

- Převody dat
- Bez ztrát informací
- Se ztrátami informací
- Stejné informace, ale různé slovníky (data)
- Stejné informace, ale různá pojetí

Slovní druhy

- Interpunkce jako slovní druh
- Slovnědruhov \acute{e} přechody
- Slovnědruhov \acute{a} platnost agreg \acute{a} tů
- Slovnědruhov \acute{a} platnost cizích slov
- Slovnědruhov \acute{a} platnost afixových segmentů

Subklasifikace slovních druhů

- Heterogenost (značkování ovlivněno potřebami popisu vyšší – syntaktické roviny) – např. vydělení krátkých tvarů os. zájmen, vztažných slov, která fungují jako spojovací výrazy, zacházení s agregáty spojka + kondicionálový morf, volný morf –s za 2. os., ...
- Revize podle slovních druhů
- Inspirace Brnem (pozitivní i negativní)

Slovesný tvar

- Slovesná subparadigmata
- Jmenné a složené tvary participií
- Doplnění variant do slovníku

Adverbia

- Upuštění od sémantické klasifikace adverbii
- Upuštění od explicitního rozdělení adverbii na stupňovatelná a nestupňovatelná
- Zvýraznění vazby mezi zájmeny a zájmennými adverbii
- Adverbiální spřežky

Spojky

- Agregátová povaha výrazů *.*by.**
- Jednoslovná morfologie a víceslovné spojovací výrazy

Částice

- Potřeba definovat seznamy
- Netřeba zatěžovat systém desambiguace v případě, že nebude úspěšná

Projekt NovaMorf





















- Od r. 2012
- Snaha o jednotný morfologický popis češtiny
- Snaha o nezávislost popisu na aplikacích a pracovištích
- Snaha o zachování tradice s ohledem na uživatele
- Snaha o dodržení zlatého pravidla morfologie

[lemma=".*(ou|í)cí"]

tom vím . Používají se	plovoucí /plovoucí/AAFP1-----1A-----	bariéry , naběráky a chemické
stěna nebo střecha . Poskytuje	vynikající /vynikající/AGFP4-----A-----	technické vlastnosti , jeho použití
celé tělo , má vynikající	uvolňující /uvolňující/AGIP4-----A-----	, zklidňující a drenážní účinky
krystalizace solí (hlavní proces	vedoucí /vedoucí/AGIS1-----A-----	ke vzniku dutých tvarů)
okraji společnosti jako nedůtklivý samotář	jedoucí /jedoucí/AGMS1-----A-----	vlastním rytmem , řídící se
hodnocení vývoje nikoli reálný věk	odvíjející /odvíjející/AGIS4-----A-----	se od data narození dítěte
femuru v přirozené poloze u	stojícího /stojící/AGMS2-----A-----	člověka . M 53 (
percepční vadu sluchu . případě	rostoucí /rostoucí/AGFP4-----A-----	ztráty sluchu v souvislosti s
svém posledním románu , hlasem	překypujícím /překypující/AGNS7-----A-----	nadšením . Poté , co
právě přestavba nadrozměrného bytu ,	nacházejícího /nacházející/AGIS2-----A-----	se v boku budovy městského
fakultě na Harvardu , doporučuje	následující /následující/AGNS1-----A-----	: „ Zkuste určit ,
VOC jsou u Kotle 1	spalujícího /spalující/AGIS2-----A-----	hnědé uhlí a topný olej
nás podmínka nutná , nikoliv	postačující /postačující/AGFS1-----A-----	. Pokud jim to neschválí
pěna , obalující ulice ,	plazící /plazící/AGFS1-----A-----	se po zdech domů a
oslovil ho tiše jeho záhadný	cestující /cestující/NNMS1-----A-----	. „ Promiňte , pane

[lemma=".*(ou|í)cí" & tag!="AG.*"]

Celkem: 751 položek (16 stránek)

	Filter	lemma_lc [lowercase lemma]	pos [part of speech]	Freq	
1	p / n	budoucí	A	8049	
2	p / n	vedoucí	N	4935	
3	p / n	stávající	A	4594	
4	p / n	cestující	N	4384	
5	p / n	žádoucí	A	3592	
6	p / n	pracující	N	1643	
7	p / n	plovoucí	A	944	
8	p / n	smrtící	A	876	
9	p / n	okouzlující	A	843	
10	p / n	vroucí	A	604	
11	p / n	věřící	N	596	
12	p / n	kolemjdoucí	N	572	
13	p / n	živoucí	A	551	
14	p / n	kolemjdoucí	A	472	
15	p / n	spolubydlící	N	387	
16	p / n	dlouhotrvající	A	366	
17	p / n	všemohoucí	A	336	
18	p / n	tisící	C	325	
19	p / n	neutuchající	A	321	
20	p / n	plačící	A	307	

AA.* *plačící*

setníkem pod německým velením ,	plačící /plačící/AAFP1----1A-----	ženy a učenníci – tak
té uranové předdoby jak praví	plačící /plačící/AAFS1----1A-----	ERATÓ a MELPOMENÉ si je
. " Z ložnice vyšla	plačící /plačící/AAMS1----1A-----	Kiška a běžela k matčiným
jak den po maturitě plivla	plačící /plačící/AAFS3----1A-----	Renatě do obličeje a ječela
většinou z nich se vlekly	plačící /plačící/AAFP1----1A-----	děti , matky měly kojence
v lebce . Roztomilá holčička	plačící /plačící/AAFS7----1A-----	bezmocí v jarním slunku .
zřejmě nechtěla , co vdova	plačící /plačící/AAFS2----1A-----	jen sebe zachránit v té
Už žádné boje , žádné	plačící /plačící/AAFP1----1A-----	matky u pohřebních hranic .
našli jsme na schodech hystericky	plačící /plačící/AAFS7----1A-----	Agness . Našla Eleanor zhroucenou
trestů . Mazlení či konejšení	plačícího /plačící/AANS2----1A-----	dítěte je projevem výchovné slabosti
celý , plyn , oko	plačící /plačící/AANS1----1A-----	. – Ach , posteskněme
tragické krásy a spatřili nevěstu	plačící /plačící/AAFS4----1A-----	nad Salimem , v úleku
pro člověka , určitě pro	plačícího /plačící/AAMS4----1A-----	člověka , který je od
chabě se usmívat a označit	plačící /plačící/AAIP4----1A-----	miliony za sentimentální hlupáky utápějící
košíčky jablek , které jeho	plačící /plačící/AAFS1----1A-----	manželka vysypávala do velkých prádelních
Mobile , " vzala jsem	plačící /plačící/AAFS4----1A-----	dívku kolem ramen , shovívavě

AG.* *plakající*

mne . Výborný uklidňoval i	plakající /plakající/AGMP4-----A-----	hráče Hodně emotivní byl závěr
tehdy Černý své dílo .	Plakající /plakající/AGMS1-----A-----	Marťánek na Ještědu Historie vzniku
, ponurá atmosféra , smutné	plakající /plakající/AGIP1-----A-----	obličej přilepené na skla oken
vůbec dovolit . Zažil jsem	plakající /plakající/AGFS4-----A-----	Dánku , která už nemohla
když ji doprovázely záběry roztřesených	plakajících /plakající/AGFP2-----A-----	děti a jejich rodičů .
někdo zmocní trezorového snímku o	plakajícím /plakající/AGMS6-----A-----	klaunovi a konečně jej pustí
skončila . Dnes musím kupovat	plakající /plakající/AGFP4-----A-----	panenky a drahá auta na
je postarat se o pět	plakajících /plakající/AGNP2-----A-----	miminek najednou ? Vždyť ruce
Noční budíčky , kašlající a	plakající /plakající/AGFS2-----A-----	malé berušky a broučci ,
manželka v jedné scéně lituje	plakající /plakající/AGNS1-----A-----	, nově nakoupenou otrokyni ,
Apfelthaler . A kde jsou	plakající /plakající/AGIP1-----A-----	stromy v Českém Krumlově k
Rusko : Holčičky zahynuly hladu	PLAKAJÍCÍ /plakající/AGFP1-----A-----	DĚTI NIKOHO NEZAJÍMALY MOSKVA -
Vlasta Formánková žalobu zamítla .	Plakající /plakající/AGFS1-----A-----	a třesoucí se Pavlova matka
a musíte investovat do panenky	plakající /plakající/AGFS2-----A-----	, čurající , zpívající nebo
" v plném žalu srdečně	plakajícími /plakající/AGMP7-----A-----	" stála v jedné ze
« řekl moudrý hřebec .	Plakající /plakající/AGMS1-----A-----	slepýš zalezl nešťastně do díry

X@.* pláčící

Pomůžu zvukově například na závěr

, a naposled spatřil svou

pláčící /pláčící/x@-----

pláčící /pláčící/x@-----

Marii . Aby se v

matičku . Dokonalý obrat nastal

Ke čtení

- OSOLSOBĚ, Klára a Jaroslava HLAVÁČOVÁ. Chystané změny projektu NovaMorf jako výsledek dlouhého období konvergencí a divergencí ve zpracování české morfologie. In Aleš Horák, Klára Osol sobě, Adam Rambousek, Pavel Rychlý. *Slavonic Natural Language Processing in the 21st Century*. 1. vyd. Brno: Tribun EU, 2019. s. 93-99, 7 s. ISBN 978-80-263-1545-2.
- OSOLSOBĚ, Klára, Vladimír PETKEVIČ, Jaroslava HLAVÁČOVÁ, Josef ŠIMANDL a Martin SVÁŠEK. Parts of Speech in NovaMorf, a New Morphological Annotation of Czech. *Jazykovedný časopis, Versita*, 2019, roč. 70, 2019, č. 2, s. 358-369. ISSN 0021-5597. doi:10.2478/jazcas-2019-0065.
- OSOLSOBĚ, Klára, Jaroslava HLAVÁČOVÁ, Vladimír PETKEVIČ, Josef ŠIMANDL a Martin SVÁŠEK. Nová automatická morfologická analýza češtiny. *Naše řeč*, AV ČR, Ústav pro jazyk český, 2017, roč. 2017, č. 4, s. 225-234. ISSN 0027-8203.