

Morfologická anotácia textov Slovenského národného korpusu

Slovenský národný korpus obsahuje dve verzie morfolologickej anotácie:

- v ručne morfolologicky anotovanom podkorpuse mak (jeho rozsah sa priebežne dopĺňa) sa dôsledne uplatňuje súbor nasledujúcich značiek a pravidiel vrátane pravidiel lematizácie,
- v automatizovane morfolologicky anotovanom celom korpuse sa použil ten istý súbor značiek a pravidiel s výnimkou niektorých pravidiel lematizácie (na to ďalej osobitne upozorníme).

Morfologickej anotácii podliehajú všetky textové jednotky – tokeny, teda reťazce znakov medzi dvoma medzerami, ako aj znaky interpunkcie, pred ktoré sa pri spracúvaní textov v korpuse medzery umelo pridávajú. Je to nevyhnutný predpoklad pre efektívne vyhľadávanie absolútnych výskytov slov – v prípade neoddelenia interpunkčného znamienka medzerou by sa museli vyhľadávať osobitne "čisté" slová/tvary a osobitne tie isté slová/tvary s akýmkoľvek možným interpunkčným znamienkom pred alebo za sebou.

Každému tokenu sa pri ďalšom spracovaní textu priradujú atribúty lema a tag.

Lema je v podstate základný, "slovníkový" tvar tokenu. Pri ručnej anotácii sa nedodržiava dištinkcia malého a veľkého začiatočného písmena, t. j. všetky lemy majú malé začiatočné písmeno a informácia o propriálnosti sa označuje osobitným spôsobom. Pri automatizovanej anotácii sa táto dištinkcia uplatňuje, pretože automatizovaná lematizácia sa zatiaľ realizuje nástrojom H&H, ktorý má niektoré pravidlá lematizácie a anotácie iné. S veľkým začiatočným písmenom sú ním však lematizované všetky, aj nepropriálne tvary stojace na začiatku vety alebo majúce osobitné štylistické či grafické postavenie, na čo treba pamätať pri vyhľadávaní. Negované tvary sloviess sa v ručnej anotácii lematizujú negovaným tvarom infinitívu, afirmácia a negácia sa označujú aj na úrovni tagu. V automatizovanej lematizácii sa všetky negované tvary lematizujú bez morfémy *ne-*.

Morfologické značky sa zapisujú pomocou písmen latinskej abecedy, čísiel a matematických symbolov. Súbor jednotlivých znakov tvorí jeden tag k jednému tokenu. Tag vyjadruje hodnoty formálnych kategórií, ktoré sú pre daný token relevantné. V Slovenskom národnom korpuse sa používajú tagy s variabilným počtom znakov, ich poradie v tagu je však záväzné. Na prvom mieste vždy stojí informácia o príslušnosti k slovnému druhu (podľa zaužívanej desaťčlennej slovnodruhovej typológie), resp. k slovnej triede (sem patria špecifické textové jednotky vrátane interpunkcie a neslovných elementov vyskytujúcich sa v bežnom texte). Nasledujú značky pre príslušné gramatické kategórie (záväzne), resp. značky pre špeciálne skupiny (nezáväzne – stoja na konci tagu po dvojbodke a označujú vlastné mená a chybné zápisy). Všetky značky uvádzame v nasledujúcich tabuľkách, príkladoch a ukážkach, ktoré sú presne prevzaté z ručne anotovaného korpusu.

Podrobnejšie informácie o tokenizácii, lematizácii a morfolologickej anotácii sú tu.

Pri práci s korpusovým manažérom Manatee a klientom Bonito sa dajú textové

jednotky vyhľadávať ako konkrétne tvary slova, ako lemy, ale aj pomocou tagu alebo jeho časti. V pravom hornom rohu nastavíme meno korpusu, ktorý chceme používať. Ak máme záujem o ručne morfológicky anotovaný korpus, z ponuky vyberieme verziu mak. Potom v hornom ponukovom riadku otvoríme položku KORPUS, v nej klikneme na položku IMPLICITNÝ ATRIBÚT a vyznačíme podľa predmetu vyhľadávania WORD, LEMMA alebo TAG. Do vyhľadávacieho okienka napíšeme príslušný tvar slova pre word, základný tvar slova pre lemu, tag alebo jeho časť (pomocou regulárnych výrazov) pre vyhľadávanie všetkých možných slov s konkrétnymi gramatickými charakteristikami. Ak chceme príslušné charakteristiky aj vidieť, nezabudneme v položke ZOBRAZENIE v hornom ponukovom riadku nastaviť zobrazenie pre ATRIBÚTY, a to konkrétne LEMY alebo TAGU pre kľúčové slovo alebo pre všetky pozície.

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	S	substantívum	slovo, ryba, ústav, muž
2. paradigma	S	substantívna	chlap, žena, srdce
	A	adjektívna	hlavný, vedúci, Mastný, starká, Slaná, vstupné
	F	zmiešaná	kuli, gazdina
	U	neúplná	kanoe, kupé
3. rod	m	mužský životný	hrdina, hlavný, Mastný
	i	mužský neživotný	strom, rýľ
	f	ženský	ulica, pani, vedúca, Slaná, hradská
	n	stredný	mesto, vysvedčenie, dievča, mlád'a
4. číslo	s	jednotné	slovo, ryba, ústav, muž
	p	množné	slová, ryby, ústavy, muži/mužovia
5. pád	1	nominatív	pán, vedúci, matka, Slaná, more, mlád'a
	2	genitív	pána, vedúceho, matky, Slanej, mora, mlád'at'a
	3	datív	pánovi, vedúcemu, matke, Slanej, moru, mlád'at'u
	4	akuzatív	pána, vedúceho, matku, Slanú, more, mlád'a
	5	vokatív	pane, mami, Táni, oci
	6	lokál	pánovi, mame, Slanej, mori, mlád'ati
	7	inštrumentál	pánom, vedúcim, matkou, Slanou, morom, mlád'at'om

Ukážka:

Prípravili ohnisko/SSns4 na grilovanie/SSns4 a o chvíľu/SSfs4 sa už po lúke/SSfs6 niesla lahodná vôňa/SSfs1, ktorá prilákala ďalších stravníkov/SSmp4.

V domnení/SSns6, že tak zachovajú v Európe/SSfs6:r mier/SSis4, podpísali zástupcovia/SSmp1 západných veľmocí/SSfp2 v septembri/SSis6 1938 Mníchovskú dohodu/SSfs4, ktorou Nemecko/SSns1:r získalo české Sudety/SSip4:r (a časti/SSfp4 dnešnej Bratislavy/SSfs2:r).

Pozícia	Znak	Hodnota	Príklad
---------	------	---------	---------

1. slovný druh	A	adjektívum	<i>milý, svieži, priateľkin, psi</i>
2. paradigma	A	adjektívna	<i>pekny, cudzi, pávi</i>
	F	zmiešaná	<i>otcov, matkin</i>
	U	neúplná	<i>super, nanič, hoden, rád, rada</i>
3. kongruencia v rode	m	mužský životný	<i>láskavý (otec), svieži (muž), matkin (známy)</i>
	i	mužský neživotný	<i>láskavý (pohl'ad), svieži (vietor), matkin (plášť)</i>
	f	ženský	<i>prázdna (ulica), dlhá (kampaň), otcova (košel'a)</i>
	n	stredný	<i>čisté (mesto), super (vysvedčenie), biele (mlád'a)</i>
4. kongruencia v čísle	s	jednotné	<i>láskavý (otec), láskavý (pohl'ad), otcova (košel'a), super (vysvedčenie), biele (mlád'a)</i>
	p	množné	<i>láskavi (otcovia), láskavé (pohl'ady), otcove (košele), super (vysvedčenia), biele (mlád'atá)</i>
5. kongruencia v páde	1	nominatív	<i>vzácný (pán), drahá (mama)</i>
	2	genitív	<i>vzácného (pána), drahej (mamy)</i>
	3	datív	<i>vzácnemu (pánovi), drahej (mame)</i>
	4	akuzatív	<i>vzácného (pána), drahú (mamu)</i>
	5	vokatív	<i>vzácný (pane), drahá (mami)</i>
	6	lokál	<i>vzácnom (pánovi), drahej (mame)</i>
	7	inštrumentál	<i>vzácnym (pánom), drahou (mamou)</i>
6. stupeň	x	pozitív	<i>vzácný, drahá, otcov, psi, strešný</i>
	y	komparatív	<i>vzácnejší, drahší, drevenejší (tanečník)</i>
	z	superlatív	<i>najvzácnejší, najdrahší, najdrevenejší (tanečník)</i>

Ukážka:

V domnení, že tak zachovajú v Európe mier, podpísali zástupcovia západných/AAfp2x veľmocí v septembri 1938 Mníchovskú/AAfs4x:r dohodu, ktorou Nemecko získalo české/AAip4x Sudety (a časti dnešnej/AAfs2x Bratislavy).

Vybranú knižku podľa vášho gusta si môžete veľmi rýchlo obstarat' aj on – line/AUis7x nákupom cez internet – priame/AAfp1x linky vás navedú do kníhkupectva Dunaj.

Pozícia	Znak	Hodnota	Príklad	
1. slovný druh	P	pronominum	<i>akýkoľvek, onen, jeho, kadiaľ</i>	
2. paradigma	S	substantívna	<i>koľkátka, všetučko</i>	
	A	adjektívna	<i>aký, ktorá, inakšie, samý</i>	
	P	zámenná	<i>ja, ty, my, vy, seba, sebe</i>	
	F	zmiešaná	<i>on, ona, ono, kto, ten, môj, všetok, čo, žiaden</i>	
	U	neúplná	<i>koľko, jeho, jej, ich</i>	
	D	príslovková	<i>ako, kam, kde, kade, vtedy, začo</i>	
3. rod, resp. kongruencia	m	mužský životný	<i>oni, moji, tí, nejakí (známi)</i>	
	i	mužský neživotný	<i>ony, moje, tie, nejaké (kruhy)</i>	
	f	ženský	<i>ona, moja, tá, nejaká (priateľka)</i>	
	n	stredný	<i>ono, moje, to, nejaké (dieťa)</i>	
4. číslo, resp. kongruencia	h	všeobecný	<i>ja, ty, my, vy, seba</i>	
	s	jednotné	<i>ja, ktorý, ten, nikto, nejaký</i>	
	p	množné	<i>my, ktorí, tí, všetci, nijakí</i>	
	5. pád, resp. kongruencia	1	nominatív	<i>ja, svoj, ktorý, nič, on</i>
		2	genitív	<i>mňa, svojho, ktorého, ničoho, neho</i>
		3	datív	<i>mne, svojmu, ktorému, ničomu, nemu</i>
		4	akuzatív	<i>mňa, svojho, ktorého, nič, oňho</i>
5		vokatív	<i>môj (pane, priateľu), naša (mami)</i>	
6		lokál	<i>mne, svojom, ktorom, ničom, ňom</i>	
7	inštrumentál	<i>mnou, svojím, ktorým, ničím, ním</i>		
6. aglutinovanosť	g	aglutinované	<i>preňho, naňho, oň, zaň, doň</i>	

Ukážka:

Vybranú knižku podľa vášho/PFns2 gusta si môžete veľmi rýchlo obstarat' aj on – line nákupom cez internet – priame linky vás/PPhp4 navedú do kníhkupectva Dunaj .

Určite sa uňho/PFms2g nájde práca pre takého/PAms4 bystrého a schopného mladíka , ako som ja/PPhs1 .

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	N	numerále	jeden, dva, raz, sto, prvý, dvojmo
2. paradigma	S	substantívna	nula, milión, státisíce, raz
	A	adjektívna	jediný, prvý, dvojité, mnohonásobný, obojaký
	N	číslovková	dva, dvaja, oba, obaja, obidva, tri, štyri
	F	zmiešaná	jeden, jedna, jedno
	U	neúplná	sto, tisíc, päť, šesť, dvanásť, šesťdesiat, dvoje, tisícoro, byť u, veľa, plno, málo, pol, trištvrte
	D	príslovková	prvýkrát, sedemkrát, dvojmo, neraz, mnohorako
3. rod, resp. kongruencia	m	mužský životný	jeden (muž), jedni, piati (muži/mužovia)
	i	mužský neživotný	jeden (oblúk), jedny (okuliare), dva (stromy)
	f	ženský	jedna (láska), jedny (dvere), dve (ženy), piatim (ženám)
	n	stredný	jedno (oko), jedny (pl'úca), dve (polia), päť (srdce), (pracovalo) päť (mužov)
4. číslo, resp. kongruencia	s	jednotné	jeden (muž), druhá (osoba), trojaké (víno)
	p	množné	jedny (osoby), druhé (miesta), trojaké (vína)
5. pád, resp. kongruencia	1	nominatív	jeden (pán), druhá (osoba), trojaké (víno)
	2	genitív	jedného (pána), druhej (osoby), trojakého (vína)
	3	datív	jednému (pánovi), druhej (osobe), trojakému
	4	akuzatív	jedného (pána), druhú (osobu), trojaké (víno)
	5	vokatív	(človeče) jeden!
	6	lokál	jednom (pánovi), druhej (osobe), trojakom (víne)
	7	inštrumentál	jedným (pánom), druhou (osobou), trojakým

Ukážka:

Prvým/NAms7 držiteľom bol Darryl F. Zanuck , ktorý si toto ocenenie odniesol neskôr ešte dva/NNip4 razy/NSip4 .

Na opravu jedného/NFis2 bytu treba priemerne od 200 - do 400 - tisíc/NUns4 korún , a práve tu štát štátnou prémieou iniciuje občanov k sporiteľským aktivitám .

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	V	verbum	klásť, čítať, vidieť, činiť
2. slovesná forma	I	infinitív	byť, hriať, volať, viesť, hovoriť
	K	prézent (indikatív)	je, hreje, volá, vedie, hovorí
	M	imperatív	bud'!, hrej!, volajte!, ved'te!, hovor!
	H	prechodník	súc, hrejúc, volajúc, vedúc, hovoriac
	L	l-ové prídavné	bol, hrialo, volali, viedla, hovorili
	B	futúrum byť	budem, budeš, bude, budeme, budete, budú, poletím, povedú
3. vid	d	dokonavý	zohrejem, zavolám, povieme
	e	nedokonavý	budem hriať, volala som, bola by hovorila
	j	obojvidové sloveso	aplikovať, počuť

4. číslo	s	jednotné	<i>je, hrialo, bude, poviem</i>
	p	množné	<i>sú, volali, budeme, hovorili</i>
5. osoba	a	prvá	<i>som, sme, hrejme!, volali sme, budem hovoriť</i>
	b	druhá	<i>si, ste, hriali ste, volajte!, budeš viesť, hovoril by si</i>
	c	tretia	<i>je, sú, hrejú, volalo, povedie, hovoria</i>
6. kongruencia v rode	m	mužský životný	<i>(otec) prišiel, (priatelia) volali</i>
	i	mužský neživotný	<i>(strom) rozkvitol, zafúkal (vietor)</i>
	f	ženský	<i>(známka) sa prilepila, (dievky) tancovali</i>
	n	stredný	<i>(dielo) vyšlo, (šetrenie) pripísalo (dôležitosť)</i>
	h	všeobecný	<i>(vy ste) prišli</i>
	o	neurčený/neurčiteľný	<i>(chlapci, ženy i deti) sa tešili</i>
7. negácia	+	afirmácia	<i>prichádzať, priateliť sa, rásť</i>
	-	negácia	<i>nebyť, neprichádzať, nepriateliť sa, nebáť sa</i>

Ukážka:

Pri každom jedle sa ho doprosovala/VLescf+, aby nebol/VLescm- sebecký a aby si uvedomil/VLdscm+, že malá sestrička je/VKesc+ chorá a takisto potrebuje/VKesc+ jesť/VIe+, ale márne.

Ak nemáte/VKcpb- kontakt na pracovníka STK, počítajte/VKcpb+ s tým, že Vás budú/VBepc+ viacere autá predbiehať/VIe+ a strávite/VKdcpb+ tam viacero hodín.

Pozícia	Znak	Hodnota	Príklad
1. slovná trieda	G	participium	<i>robiaci, sediaci, naložený, zohriaty</i>
2. druh	k	aktívne	<i>pracujúci, visiaci, píšuci, platiaci</i>
	t	pasívne	<i>robený, kosený, obratý, zožatý</i>
3. kongruencia v rode	m	mužský životný	<i>sediaci (otec), pracujúci (muž), vyhodnený (známy)</i>
	i	mužský neživotný	<i>kričiaci (pohľad), pracujúci (stroj), pokrčený (plášť)</i>
	f	ženský	<i>čistená (ulica), strhujúca (kampaň), ušitá (košeľ a)</i>
4. kongruencia v čísle	n	stredný	<i>opustené (mesto), písané (potvrdenie), skáčuce (mača)</i>
	s	jednotné	<i>píšuci (otec), uplakaný (pohľad), vypratá (košeľ a)</i>
	p	množné	<i>píšuci (otcovia), vypraté (košele), skáčuce (mačatá)</i>
5. kongruencia v páde	1	nominatív	<i>kreslený (film), bežiaci (reklama), ušité (sako)</i>
	2	genitív	<i>kresleného (filmu), bežiacej (reklamy), ušitého (saka)</i>
	3	datív	<i>kreslenému (filmu), bežiacej (reklame), ušitému (saku)</i>
	4	akuzatív	<i>kreslený (film), bežiacu (reklamu), ušité (sako)</i>
	5	vokatív	<i>(synu môj) neveriaci!</i>
	6	lokál	<i>kreslenom (filme), bežiacej (reklame), ušitom (saku)</i>
	7	inštrumentál	<i>kresleným (filmom), bežiacou (reklamou), ušitým (sakom)</i>
6. stupeň	x	pozitív	<i>kričiaci, hodena, skáčuce</i>
	y	komparatív	<i>uplakanejší, strhujúcejší</i>
	z	superlatív	<i>najuplakanejší, najstrhujúcejší</i>

Ukážka:

Je to oblačnosť súvisiaca/Gkfs1x so spomínaným/Gtis7x studeným frontom.

Vie sa len, že sochy budú umiestnené/Gtfp1x na terase v Ulici spievajúceho/Gkns2x mora.

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	D	adverbium	<i>prísne, mílo, pravidelne, prázdno</i>
2. stupeň	x	pozitív	<i>draho, vzácne</i>
	y	komparatív	<i>drahšie, vzácnejšie</i>
	v	superlatív	<i>najdrahšie, najvzácnejšie</i>
	z	superlatív	<i>najdrahšie, najvzácnejšie</i>

Ukážka:

Baghíra sa tváril veľmi/Dx vážne/Dx , keď podišiel bližšie/Dy k Balúovi .

A robilo sa tak správne/Dx , lebo sa vyžadovali , najmä na politické účely , krátke , úsečné slová s jednoznačným významom , rýchlo/Dx vysloviteľné a rezonujúce v mysli rečníka čo najkratšie/Dz .

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	E	prepozícia	<i>po, pre, na, do, cez, medzi</i>
2. forma	v	vokalizovaná	<i>so, zo, odo, podo</i>
	u	nevokalizovaná	<i>s, z, od, pod, prostredníctvom</i>
3. spojenie s pádom	2	genitív	<i>od, do, z</i>
	3	datív	<i>k, ku</i>
	4	akuzatív	<i>za, na</i>
	6	lokál	<i>o, na</i>
	7	inštrumentál	<i>s, so, medzi</i>

Ukážka:

Prezident so/Ev7 svojim spoluhračom nad/Eu7 súpermi zvít'azili .

Zo/Ev2 sna t'a prebudí bozk mládenca , ktorý t'a bude nado/Ev4 všetko l'úbiť .

A to najmä v/Eu6 súvislosti/SSfs6 s/Eu7 cenou vstupného .

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	O	konjunkcia	<i>a, ale, alebo, či, pretože, že</i>
2. kondicionálnosť	Y	kondicionálnosť	<i>aby, keby, čoby, žeby</i>

Ukážka:

Najhoršie by si počínal , keby/OY to odkladal .

Kritika je opodstatnená , ale/O nepáči sa mi , že/O prezident vyšiel z koalície , a/O ako/O autorita by mohol viac využívať svoju funkciu na ovplyvnenie jej činnosti .

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	T	partikula	<i>azda, nuž, bodaj, sotva, áno, nie</i>
2. kondicionálnosť	Y	kondicionálnosť	<i>kiežby, žeby</i>

Ukážka:

Ozdravenie bánk je nepochybne/T prvým a zrejme/T rozhodujúcim krokom pri vytváraní priaznivého podnikateľského prostredia .

STV však/T nie/T je v štandardnej situácii a musí bojovať o každého diváka .

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	J	interjekci	<i>fíha, bác, bums, dokelu, ahoj, cveng, plesk</i>
		a	

Ukážka:

Ach/J , prosím vás , nekričte tak !

Deti sa chytili za ruky , urobili okolo Maji s Vilkom kruh a spievali : dzum/J , dzum/J , dzum/J . . .

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	R	reflexívum	sa, si

Ukážka:

Ako ľahko by sa/R dalo uveriť , keby sa/R človek nemusel obzerat' okolo seba , pomyslel si/R Winston ...

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	Y	kondicionálová morféma	by

Ukážka:

To by/Y sa mali naučiť' aj kluby .

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	W	abreviácie, značky	km, kg, atď., H ₂ O, SND

Ukážka:

Na preloma júla a augusta , v období tzv/W . psích dní , u nás leto vrcholí .

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	Z	interpunkcia	., !, (, +

Ukážka:

Na preloma júla a augusta , v období tzv /Z psích dní , u nás leto vrcholí /Z

Deti sa chytili za ruky /Z urobili okolo Maji s Vilkom kruh a spievali :/Z dzum /Z dzum /Z dzum /Z /Z /Z

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	Q	neurčiteľný slovný druh	bielo(-čierny), New (York)

Ukážka:

Otvorene vnímam hodnotenie ťažkej sociálno/Q - ekonomickej situácie väčšiny občanov na Slovensku .

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	#	neslovný element	XXXX, -----

Ukážka:

**/# */# */# Ani dnes v čase internetu ...*

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	%	citátový výraz	<i>šaj pes dovakeras, take it easy!, náměští</i>

Ukážka:

Johansson potvrdil zrušenie letnej súťaže Intertoto/% cup/% , lebo peniaze investované do nej ...

Pozícia	Znak	Hodnota	Príklad
<u>1.</u> slovná trieda	0	číslica	<i>8, 14000, 3 (razy)</i>

Ukážka:

Denná teplota v sobotu dosiahne 20/0 až 24/0 , na severe Slovenska len okolo 18/0 stupňov Celzia .

Pozícia	Znak	Hodnota	Príklad
<u>posledná</u>	:r	vlastné meno	<i>Emil, Molnárová, Vysoké, Tatry, Slovenské (národné divadlo)</i>

Ukážka:

GRIGORIJ/SSms1:r MESEŽNIKOV/SSms1:r , politológ , Inštitút pre verejné otázky : Prezident neprekvapil .

Pozícia	Znak	Hodnota	Príklad
<u>posledná</u>	:q	chybný zápis	<i>papeirníctvo, zhrzený</i>

Ukážka:

Nepridátete/VKdppb-:q sa ku mne , Alia ?