

Přílohy

Příloha 1: Vnětextová kategorizace korpusu Koditex

MODE	DIVISION	SUPERCLASS	CLASS	Tokeny	Vzorky	
spo (mluvená komunikace)	int (interaktivní)		bru (nepřipravené veřejné / vysílané rozhovory)	221 812	90	
			eli (formální rozhovor)	201 690	82	
			inf (neformální rozhovor)	208 565	86	
	nin (neinteraktivní)		wbs (připravený/čtený projev)	213 201	71	
web (internetová komunikace)	mul (mnohospměrná komunikace)		dis (internetové diskuse)	197 948	87	
			fcb (facebookové statusy)	199 418	91	
			for (webová fóra)	200 104	85	
	uni (jednosměrná komunikace)		blo (blogy)	204 356	74	
			wik (wikipedie)	201 691	84	
wri (psaná komunikace)	fic (beletrie)	nov (romány)	crm (detektivky)	190 026	68	
			fan (fantasy)	189 432	69	
			gen (bez bližšího určení)	193 667	67	
			lov (milostné)	189 893	70	
			scf (sci-fi)	188 703	68	
			col (povídky)	195 595	70	
			scr (scénáře a dramata)	182 689	76	
			ver (poezie a písně)	205 837	76	
	nfc (oborová literatura)	pop (populárně-naučná)	fts (formální a technické vědy)	207 607	68	
			hum (humanitní vědy)	204 837	74	
			nat (přírodní vědy)	204 751	71	
			ssc (společenské vědy)	203 698	68	
			pro (profesní literatura)	fts (formální a technické vědy)	210 010	71
				hum (humanitní vědy)	207 916	69
				nat (přírodní vědy)	209 580	70
				ssc (společenské vědy)	209 385	72
		sci (vědecká literatura)	fts (formální a technické vědy)	202 932	67	
			hum (humanitní vědy)	204 300	71	
			nat (přírodní vědy)	206 716	72	
			ssc (společenské vědy)	205 358	67	

			adm (administrativa)	203 542	82
			enc (encyklopedie)	203 957	73
			mem (auto-/biografie)	203 390	71
nmg (noviny a časopisy)	lei (volnočasová publicistika)	hou (bydlení, zahrada, hobby)	207 499	68	
		int (zajímavosti ze světa)	209 232	69	
		lif (životní styl)	203 124	72	
		mix (víkendové přílohy)	205 310	75	
		sct (bulvár)	201 417	73	
		spo (sport)	199 238	70	
		new (tradiční publicistika)	com (komentáře)	205 372	68
			cul (kultura)	205 690	68
	eco (ekonomika)		211 481	70	
	fre (volnočasové aktivity)		208 532	71	
	pol (politika)		206 893	70	
	rep (reportáže)		206 377	70	
	pri (soukromá komunikace)		cor (dopisy)	96 366	68
	Celkem				9 039 137

Příloha 2: Seznam jazykových rysů užitých v MDA na českém materiálu

ID	Rys	Rovina	Poznámka/příklad
FEI	úženi <i>é</i> > <i>i</i> v kořeni (typ <i>polívka</i>)	fonologie	
FEIK	úženi <i>é</i> > <i>i</i> v koncovce (typ <i>starý město</i>)	fonologie	
FYEJ	přehláska <i>y</i> > <i>ej</i> v kořeni (typ <i>bejt</i>)	fonologie	
FYEJK	přehláska <i>y</i> > <i>ej</i> v koncovce (typ <i>mladej</i>)	fonologie	
WL	průměrná délka slova v textu (v počtu slabik)	fonologie	pouze tokeny obsahující písmena
AA	adjektiva	lexikon	všechny typy
PAJ	sémanticky vyprázdněná adjektiva	lexikon	víc než 3 významy v SSČ
DEMD	ukazovací adverbia	lexikon	
DB	adverbia	lexikon	všechny typy
DT	adverbia vyjadřující určení času	lexikon	
DP	adverbia vyjadřující určení místa či směru	lexikon	
DI	neurčitá adverbia	lexikon	
RST	restriktory (např. <i>hlavně</i> , <i>alespoň</i> , <i>maximálně</i>)	lexikon	
PAV	sémanticky vyprázdněná adverbia	lexikon	víc než 3 významy v SSČ
INT	citoslovce	lexikon	
FNW	synsémantika (zájmena, číslovky, předložky, spojky a částice)	lexikon	vč. <i>být</i> v roli pomocného slovesa v min. čase
POE	poetismy (podle příznaku v SSČ)	lexikon	bez výrazně homonymních jednotek
TIME	časové výrazy (adverbia, jména a časové entity)	lexikon	včera, pondělí, 30. 4. 2010
CONJI	inventář spojek	lexikon	normalizováno pomocí zTTR
MOD	modální slovesa, adjektiva, adverbia a částice	lexikon	
NUM	numerale	lexikon	
PREI	inventář předložek	lexikon	normalizováno pomocí zTTR
DEM	ukazovací zájmena (bez slovního tvaru "to")	lexikon	
IND	neurčitá zájmena	lexikon	
P1P	zájmena pro 1. os. (osobní i přivlastňovací)	lexikon	vč. mluvených variant
P2P	zájmena pro 2. os. (osobní i přivlastňovací)	lexikon	vč. mluvených variant
P3P	zájmena pro 3. os. (osobní i přivlastňovací)	lexikon	vč. mluvených variant
PP	pronomina	lexikon	všechny typy
PPI	inventář zájmen	lexikon	normalizováno pomocí zTTR
POSP	přivlastňovací zájmena	lexikon	viz i P1P, P2P a P3P
NN	substantiva	lexikon	všechny typy
PSB	sémanticky vyprázdněná substantiva	lexikon	víc než 3 významy v SSČ
VTS	slovesa myšlení	lexikon	
VD	slovesa mluvení	lexikon	
VB	slovesa	lexikon	všechny typy

ID	Rys	Rovina	Poznámka/příklad
PVB	sémanticky vyprázdněná slovesa	lexikon	víc než 3 významy v SSČ
ARCH	knižní a zastaralá synsémantika a vybraná adv.	lexikon	např. <i>rovněž, dle, též, aniž</i> ; na základě SSJČ
GENL	lexikální genderové markery	lexikon	propria a apelativa označující ženy
ADN	jmenné tvary adjektiv	morfologie	
DAT	dativ (substantiv)	morfologie	
GEN	genitiv (substantiv)	morfologie	
INS	instrumentál (substantiv)	morfologie	
LOK	lokál (substantiv)	morfologie	
NAKK	nominativ–akuzativ (substantiv)	morfologie	
CAS	pády realizované s předložkou	morfologie	bez předložkově obligatorního lokálu, tj. gen., dat., ak. a inst.
IEK	morfologická konkurence koncovek <i>-é</i> vs. <i>-í</i>	morfologie	typ <i>policisté</i>
VOC	vokativ (substantiv)	morfologie	
SGPL	množné číslo substantiv	morfologie	
VPE1	1. slovesná osoba	morfologie	
VPE2	2. slovesná osoba	morfologie	
ICI	infinitiv na <i>-ci</i>	morfologie	
VIM	imperativ	morfologie	
VIN	indikativ	morfologie	
VTE3	slovesný čas – budoucí	morfologie	
VTE1	slovesný čas – minulý	morfologie	
VVO	slovesný rod – pasivum	morfologie	
VTE2	slovesný čas – přítomný	morfologie	
NEG1	negace větná	morfologie	slovesa s prefixem <i>ne-</i>
VF	verba finita	morfologie	
BYTS	pomocné "být" ve formě <i>-s</i> (typ <i>přišels</i>)	morfologie	
VAS	slovesný vid – perfektivum	morfologie	
VCO	kondicionál	morfologie	verba
NEG2	negace lexikální	morfologie	substantiva s prefixem <i>ne-</i>
NEG4	negace členská (<i>ne, nikoli</i> + varianty)	morfologie	
DROP	pronoun non-dropping	morfologie	explicitní přítomnost osobního zájmena v 1. a 2. osobě
GENG	morfologické genderové markery	morfologie	slovesná participia s ženským rodem
KONT	kontaktné výrazy	pragmatika	např. <i>hele, aha</i>
VYPW	výplňková slova	pragmatika	např. <i>hm, vlastně</i>
EXP	expresivní částice vyjadřující hodnocení, postoje, modalitu a emoce	pragmatika	např. <i>každopádně, vždyť</i>
AMP	částice zesilující význam (amplifiers/boosters)	pragmatika	např. <i>dokonce, vyloženě</i>
DOWN	částice oslabující význam (downtoners/hedges)	pragmatika	např. <i>poněkud, bezmála</i>
IT	polyfunkční "to"	pragmatika	všechny funkce a významy

ID	Rys	Rovina	Poznámka/příklad
PROPT	toponyma	pragmatika	
PROPA	antroponyma	pragmatika	
EXC	vykřičník	pragmatika	
NGR	výskyt frekventovaných 5-gramů	pragmatika	frekvenční špička ze SYN2015 a ORAL v4
GRAAA	analytické stupňování (typ <i>méně pěkný</i>)	slovotvorba	
ASIM	adjektiva podobnosti (podle přípon)	slovotvorba	adj. na <i>-ovský, -oidní, -ovitý</i>
POS	přívlastňovací adjektiva (typ <i>otcův i matčin</i>)	slovotvorba	
AAV	verbální adjektiva (typ <i>tekoucí i přišedší</i>)	slovotvorba	s vyloučením lexikalizovaných
GRAD	stupňování adjektiv (komparativ + superlativ)	slovotvorba	
GRADD	stupňování adverbíí (komparativ + superlativ)	slovotvorba	
DEMI	deminutiva	slovotvorba	
NNV	verbální substantiva	slovotvorba	
ABS	abstrakta	slovotvorba	
ATA1	přívlastky shodné anteponované	syntax	
ATA21	přívlastky shodné postponované nerozvitě	syntax	
ATA22	adjektivní přívlastky shodné postponované rozvitě	syntax	
CLUAC	klastry pádů adjektiv	syntax	sekvence adj. se stejným pádem
CLUA	klastry adjektiv	syntax	
CLUAD	klastry adverbíí	syntax	
CPD	věty s interog. a vztaž. adv.	syntax	např. <i>kde, kdy, jak</i>
CONJ	konjunkce	syntax	všechny typy
COH1	vícečlenné konektory (typ <i>jednak – jednak</i>)	syntax	
COOR	koordinace	syntax	<i>a, i, nebo, či, ani</i>
PRE	předložky	syntax	všechny typy
PRE2	sekundární předložky	syntax	
REL2	vztažné věty typu <i> který</i>	syntax	
REL3	vztažné věty typu <i>co</i>	syntax	
REL1	vztažné věty typu <i>jenž</i>	syntax	
SER	větná relativita	syntax	<i>což, přičemž</i>
ACM	komplementace adjektiva	syntax	např. <i>je jasné, že...</i>
VCM	komplementace slovesa	syntax	např. <i>myslím, že...</i>
CLUN	klastry substantiv	syntax	
CLUNC	klustry pádů substantiv	syntax	sekvence subst. se stejným pádem
STA4	subst. přívlastky neshodné postponované	syntax	
PA	jmenný přísudek adjektivní	syntax	
PN	jmenný přísudek substantivní	syntax	
CIR3	konektory přípustkové a další adverzativní	syntax	<i>aniž, ačkoli, přesto, třebaže</i>
CIR124	konektory okolnostní	syntax	podmínkové, příčinné

ID	Rys	Rovina	Poznámka/příklad
			a další okolnostní
SL	průměrná délka věty (v počtu slov)	syntax	
NEG3	negace vícenásobná v rámci jedné věty	syntax	
CORR	korelativa (deiktikum + konektor)	syntax	<i>...tak, že... / ...tím, aby...</i>
COH2	částice členící text	text	<i>dále, například</i>
QUE2	otázky (zjišťovací i doplňovací)	text	
QUE1	otázky doplňovací	text	
PHRA	četnost frekventovaných frazémů	text	na základě automatické frazémové anotace
REP	opakování identických slov	text	
LR	Yulův koeficient lexikální repetitivnosti	text	
TC	tematická koncentrace textu	text	
BIG	bigramy	text	zTTR dvojic slov (tvary)
UNG	unigramy	text	zTTR slov (lemmata)

Příloha 3: Distribuce dat jednotlivých subregistrů v první a druhé dimenzi

