

The authors of morphological dictionaries of Serbian are Cvetana Krstev and Duško Vitas

The articles you should cite if using this dictionaries are:

Cvetana Krstev, Duško Vitas, "Corpus and Lexicon - Mutual Incompleteness ", in Proceedings of the Corpus Linguistics Conference, 14-17 July 2005, Birmingham, eds. Pernilla Danielsson and Martijn Wagenmakers, ISSN 1747-9398, <http://www.corpus.bham.ac.uk/PCLC/>, 2005.

Duško Vitas, Cvetana Krstev, Ivan Obradović, Ljubomir Popović, Gordana Pavlović-Lažetić", "An Processing Serbian Written Texts: An Overview of Resources and Basic Tools ", in Workshop on Balkan Language Resources and Tools, 21 Novembar 2003, Thessaloniki, Greece, eds, S. Piperidis and V. Karkaletsis, pp. 97-104, 2003.

The more comprehensive bibliography you can find at:

[www.matf.bg.ac.yu/~cvetana](http://www.matf.bg.ac.yu/~cvetana)

	<b>DELAS</b>	<b>DELAF</b>	<b>DELAF/DELAS</b>
<b>Nouns</b>	37357	235975	6.32
<b>Adjectives</b>	22546	401899	17.82
<b>Verbs</b>	15350	451857	29.44
<b>Adverbs</b>	3033	3033	1.00
<b>Other</b>	815	3099	3.80
<b>Derivation</b>	2376	29506	
<b>Total</b>	<b>81477</b>	<b>1125369</b>	<b>13.81</b>
<b>Toponimi</b>	<b>3626</b>	<b>36701</b>	<b>10.12</b>
<b>Imena</b>	3309	16147	
<b>Prezimana</b>	17165	114629	
<b>Imena,engleska</b>	895	3267	
<b>Prezimana,engl</b>	3503	16804	
<b>Imena, poznati</b>	200	1217	
<b>Total</b>	<b>25072</b>	<b>152064</b>	<b>6.07</b>
<b>TOTAL</b>	<b>110175</b>	<b>1314134</b>	<b>11.93</b>

Izvestaj dana 13. VI 2006.

PoS	category	codes	PoS	category	codes
noun – N	gender	m - masculine	adjective - A	gender	As for nouns
		f - feminine		case	
		n - neutre		number	
				animatness	
	case	1 - nominative		degree	a - positive
		2 - genitive			b - comparative
		3 - dative			c - superlative
		4 - accusative		definitness	d - definite
		5 - vocative			k - indefinite
		6 - instrumental			
	number	7 – locative	verb - V	form	W - infinitive
		s – singular			P - present
		p – plural w - paukal			A - aorist
	animatness	v – animate			I - imperfect
		q - non-animate			Y - imperative
		g - no consequence			G - active PP
preposition - PREP	/	/			T - pasive PP
conjunction - CONJ	/	/			F - future
particle - PAR	/	/			S - present gerund
Interjection - INT	/	/			X - past gerund
Numerals - NUM	case			person	x - first
	number				y - second
	gender				z - third
	animatness			negation	h - negated clitic
pronouns - PRO	case				i - postive clitic
	number			number	As for nouns
	gender				
	animatness				
	clitic	i – clitic r – no clitic			

Monday, July 17, 2006

Cvetana Krstev, Dusko Vitas

The complete description of syntactic and semantic markers that are used in Serbian morphological e-dictionary

## **N noun**

+Hum	human (e.g. devojka)
+Zool	animal (e.g. aligator)
+Bot	plant (e.g. badem)
	if none of this the noun refers to the non-living object
+VN	verbal noun (e.g. restrukturiranje)
+Mas	e.g. dečurlija
+Neg	negation (contains the negative affix) (e.g. nepostojanje)
+NumN	numerical noun (e.g. hiljada; dvadesetorica)
+PT	pluralia tantum (e.g. nosila)
+Coll	collective (e.g. kestenje)
+MG	masculine sex (or masculine natural gender) (e.g. bitanga; petorica)
+FG	feminine sex (or feminine natural gender)
+NG	neutral natural gender (e.g. deca)
+Pl	paradigm for singular; denoting more than one object (e.g. deca; sedmorica)
+Dem	diminutive (e.g. stazica)
+Aug	augmentative (e.g. panjina)
+Hip	hipocoristic (e.g. maca)
+Pej	pejorative (e.g. bradetina)
+NProp	proper name (e.g. Litvanija)
+Myth	mythology (e.g. minotaur)
+Relig	religious (e.g. šejtan)
+DerLast	derived from the last name (e.g. reganizam)

## **A adjective**

+Pos	possessive (e.g. carev)
+PosQ	relational (e.g. volonterski)
+PP	derived from the passive past participle, it cannot be easily disambiguated in nominative forms from passive past participles (e.g. kljucan)
+Col	designates colours (e.g. zelenkastosiv)
+Mat	designates materials (e.g. plehan)
+Neg	negation (contains the negative affix) (e.g. neizvodljiv)
+El	elision (e.g. prebogat)
+NProp or	
+NPropre	proper name (e.g. saudijski)
+Ord	ordinal number (e.g. prvi)
+APP	derived from the active past participle (e.g. ogleveo)
+PGA	derived from past gerund active (e.g. zakriljujući)
+DerLast	derived from the last name (e.g. robinhudovski)
+Incorr	incorrect (e.g. plesnjiv)
+Indef	indefinite adjective pronoun (e.g. ijedan)

## **Semantic tags for proper names and adjectives derived from it:**

+Top	toponym (e.g. Valjevo and valjevski)
+Hyd	hydronym (e.g. Dunav and dunavski)
+Oro	oronym (e.g. Jahorina and jahorinski)
+Lang	language (e.g. gelski)
+PDrz	state (e.g. Egipat and egipatski)
+Ppust	desert (e.g. Sahara and sahariski)

+PRav	plain (e.g. Panonija and panonski)
+PFjd	fiord (e.g. Aleksandra and aleksandrijski)
+PAut	autonomous region (e.g. Kosovo and kosovski)
+Preg	region (e.g. Banat and banatski)
+PGgr	capital of a state (e.g. Kairo and kairski)
+PGr4	a very big city (e.g. Atina and atinski)
+PGr3	a big city (e.g. Sevilja and seviljski)
+PGr2	not so small city (e.g. Firenca and firentinski)
+PGr1	a small city (e.g. Apatin and apatinski)
+PDgr	a city quarter (e.g. Palilula and palilulski)
+POps	a county (e.g. Zemun and zemunski)

## **V verb**

+Imperf	imperfective (e.g. cediti)
+Perf	perfective (e.g. procediti)
+Tr	transitive (e.g. celivati)
+It	intransitive (e.g. plakati)
+Ref	reflexive (e.g. smejati se)
+Iref	ireflexive (e.g. cijukati)
+Aux	auxiliary (e.g. jesam)

Note: In a lexicon, lemma can be described both as imperfective and perfectives, as well both as reflexive and ireflexive. In text it is realized only as imeprefective or perfective, and as reflexive and ireflexive. However, such cases have not been disambiguated in a lemmatized corpus.

## **ADV adverb**

+Adj	derived from an adjective (e.g. bistro)
+Comp	in comparative form (e.g. verovatnije)
+Sup	in superlative form (e.g. najjezovitije)
+El	in elative form (e.g. prebogato)
+Neg	negation (contains the negative affix) (e.g. nesigurno)
+Noun	derived from a noun (e.g. mesecima)
+Num	derived from a numeral (e.g. pedesetak)
+DerOvanoIrano	e.g. civilizovano vs.
+DerIranoOvano	civilizirano

## **PREP preposition**

+p2	requests genitive case (e.g. iza)
+p3	requests dative case (e.g. uprkos)
+p4	requests accusative case (e.g. uz)
+p6	requests instrumental case (e.g. za)
+p7	requests locative case (e.g. prema)

## **CONJ conjunction** (e.g. ali)

## **ABB abbreviation**

+Mes	unit of measure (e.g. kg)
------	---------------------------

## **PREF prefix** (e.g.

## **NUMnumeral** (e.g. jedan)

+v1	value =1
+v2	value=2
+v3	value=3
+v4	value=4
+v5	value>=5
+Coll	collective numeral (e.g. dvadestoro)

**INT interjection** (e.g. ah)

**PAR particles** (e.g. se)

**PRO pronoun**

+ProN	nominal pronoun (e.g. gdeko)
+ProA	adjective pronoun (e.g. onaj)
+PrsJB	personal (ja i ti)
+PrsMB	personal (mi i vi)
+PrsJG	personal (on, ona i ono)
+PrsMG	personal (e.g. oni, one i ona)
+Pos	possessive (e.g. njegov)
+Ref	reflexive (e.g. sebe)
+Int	interrogative (e.g. ko)
+Rel	relational (e.g. koji)
+Indef	indefinite (e.g. pokoji)
+Gen	general (e.g. svako)
+Demon	demonstrative (e.g. taj)
+Neg	negative (e.g. niko)

**The tags that apply to all PoS:**

+Ek	ekavian pronunciation (e.g. brzinomer)
+Ijk	ijekavian pronunciation (e.g. brzinomjer)
+Ik	ikavian pronunciation (e.g. misto)
+Sr	specific to Serbian language (e.g. fudbal)
+Cr	specific to Croatian language (e.g. nogomet)

**Various derivational forms:**

+Der0H	e.g. ambar	+DerH0	e.g. hambar
+Der0I	e.g. talijanština	+DerI0	e.g. italijanština
+Der0U	e.g. skladnost	+DerU0	e.g. sukladnost
+DerAHa	e.g. meana	+DerHaA	e.g. mehana
+DerAhuAu	e.g. začahuriti v.s.	+DerAuAhu	e.g. začauriti
+DerArisatiIrati	e.g. komentarisati v.s.	+DerIratiArisati	e.g. komentirati
+DerAtiIrati	e.g. izmiksati	+DerIratiAti	e.g. izmiksirati
+DerAtiOvati	e.g. švercati	+DerOvatiAti	e.g. švercovati
+DerAvatiIvati	e.g. raseljavati	+DerIvatiAvati	e.g. raseljivati
+DerAvatiUvati	e.g. nađinđavati	+DerUvatiAvati	e.g. nađinđuvati
+DerBV	e.g. barbar	+DerVB	e.g. varvar
+DerCijskiTorski	e.g. stabilizacijski	+DerTorskiCijski	e.g. stabilizatorski
+DerCiratiKovati	e.g. deducirati	+DerKovatiCirati	e.g. dedukovati
+DerCK	e.g. cedar	+DerKC	e.g. kedar
+DerCS	e.g. certifikat	+DerSC	e.g. sertifikat
+DerCxatiTati	e.g. shvaćati	+DerTatiCxati	e.g. shvatati
+DerCxivatiTavati	e.g. upropašćivati	+DerTavatiCxivati	e.g. upropaštavati
+DerErisatiIrati	e.g. hohsxtaplerisati	+DerIratiErisati	e.g. hohštaplirati
+DerFV	e.g. kafa	+DerVF	e.g. kava
+DerGH	e.g. astragan	+DerHGe.g.	asgrahan
+DerGK	e.g. garnišna	+DerKGe.g.	karnišna
+DerHaVa	e.g. kuhati	+DerVaHa	e.g. kuvati
+DerHiI	e.g. sahibija	+DerIH	e.g. saibija
+DerHJ	e.g. proha	+DerJH	e.g. proja
+DerHK	e.g. hronološki vs.	+DerKHe.g.	kronološki
+DerHrR	e.g. hrđati	+DerRHr	e.g. rđati
+DerHuU	e.g. čaurast	+DerUHu	e.g. čaurast
+DerHV	e.g. suhoparno vs.	+DerVHe.g.	suvoparno

+DerIJ	e.g. ionizacija	+DerJI	e.g. jonizacija
+DerIratiOvati	e.g. afrikanizirati	+DerOvatiIrati	e.g. afrikanizovati
+DerJV	e.g. proja	+DerVJ	e.g. prova
+DerIsatiOvati	e.g. argumentisati	+DerOvatiIsati	e.g. argumentovati
+DerKovatiZirati	e.g. kritikovati	+DerZiratiKovati	e.g. kritizirati
+DerRatiSati	e.g. afirmirati	+DerSatiRati	e.g. afirmisati
+DerSaSxa	e.g. spasavati	+DerSxSa	e.g. spašavati
+DerSatiTi	e.g. kaparisati	+DerTiSati	e.g. kapariti
+DerSatiZirati	e.g. hipnotisati	+DerZiratiSati	e.g. hipnotizirati
+DerSatiZovati	e.g. karakterisati	+DerZovatiSati	e.g. karakterizovati
+DerSavatiSxavati	e.g. spasavati	+DerSxavatiSavati	e.g. spašavati
+DerShSX	e.g. shematičan	+DerSXSx	e.g. šematičan
+DerSivatiSxavati	e.g. uresivati	+DerSxavatiSivati	e.g. urešavati
+DerSivatiSxivati	e.g. nadvisivati	+DerSxivatiSivati	e.g. nadvišivati
+DerSSX	e.g. spekulativan	+DerSXS	e.g. špekulativan
+DerSxavatiSxivati	e.g. nadvišavati	+DerSxivatiSxavati	e.g. nadvišivati
+DerSZ	e.g. kosmički	+DerZS	e.g. kozmički
+DerZivatiZxavati	e.g. unizivati	+DerZxavatiZivati	e.g. unižavati

### **Dictionary of famous persons** (fictious and non-fictious), organizatios, etc.

+Cel	celebrity
+Sci	scientist e.g. Darwin
+Fict	fictious e.g. Kandid
+Hist	history e.g. Gestapo
+Lit	literature e.g. Bajron
+Art	art e.g. Picasso
+Mus	music e.g. Mocart

### **Dictionary of personal names**

+First	first name	e.g. Miloš
+Last	surname	e.g. Utvić
+Nick	nick name	e.g. Miško