



MORPHOLOGICAL TAGSET “JABLONSKIS”

STANDARD VERSION 2.0

2024-01-13

Erika Rimkutė, Agnė Bielinskienė, Loïc Boizou, Andrius Utkā, Virginijus Dadurkevičius

Institute of Digital Resources and Interdisciplinary Research (SITTI), Vytautas Magnus university

The tagset was firstly used in 2019 at The Centre of Computational Linguistics in the preparation of syntactically annotated corpus ALKSNIS. It was later adapted for annotation of other corpora and in 2023 it was implemented in the morphological analysis tool *Morfuoklis*.

Principles

- The "Jablonskis" standard is based on the abbreviations of parts of speech and other grammatical categories commonly used in Lithuanian linguistic works (e.g., dictionaries, grammars) (see Table 1).
- Tags are connected by a period without space, e.g.: *dkt.tikr.mot.vns.Vt*.
- The first element designates a part of speech. The following categories are treated as parts of speech: punctuation marks (*skyr.*); acronym (*acr.*) - when the abbreviation consists of the first letters of the name, e.g.: *VDU, NATO, BVP*; abbreviation (*sutr.*), consisting of one or more letters of the word, e.g.: *Eur, m., m, gerb.*; foreign language word (*užs.*); other units (*kt.*), e.g.: e-mail, website link, letter names (*vitamin A, group B*), letter numbering (*a, b, c*)...; *i, ii, iii*)...).
- Only a marked category member is marked, when grammatical category is binary. For example, proper nouns are marked (*dkt.tikr.*), while common nouns (*dkt.*) not marked.
- When forming a full tag of morphological categories, the arrangement order must be followed (see Table 2).
- “Jablonskis” can be used in other annotation frameworks, e.g., *Universals dependencies* (CoNLL-U), *SketchEngine* or *NoSketch* (see Table 3).

Table 1. Morphological categories and tags

Category	Elements of the Category (LT)	Elements of the Category (EN)	Tag
POS and other categories	daiktavardis	noun	dkt.
	veiksmažodis	verb	vksm.
	būdvardis	adjective	bdv.
	įvardis	pronoun	įv.
	skaitvardis	numeral	sktv.
	prieveiksmis	adverb	prv.
	jaustukas	interjection	išt.
	ištiktukas	onomatopoeia	jng.
	dalelytė	particle	dll.
	prielinksnis	preposition	prl.
	jungtukas	conjunction	jst.
	akronimas	abbreviation	sutr.
	sutrupinimas	acronym	akr.
	samplaikos ženklas	punctuation	skyr.
	užsienio kalbos žodis	foreign word	užs.
kiti	other	kt.	
Noun type	tikrinis	proper	tikr.
Proper noun subtype	vardas	first name	vrđ.
	pavardė	surname	pvrđ.
	vietovardis	geographical name	vtvrđ.
Reflexiveness	sangražinis	reflexive	sngř.
Verb polarity	neigiamas	negative	neig.
Language standard	substandartinis	substandard	substnd.
	vulgarizmas	vulgar	vulg.
Numeral type	kiekinis pagrindinis	cardinal main	kiek.
	kiekinis kuopinis	cardinal collective	kuopin.
	kiekinis dauginis	cardinal multiple	daugin.
	kelintinis	ordinal	kelint.
Numeral form	arabiškas	Arabic	arab.
	romėniškas	Roman	rom.
	mišrus	mixed numeral (e.g. 2-oji, 50-ies)	mišř.
	žodžiu užrašytas	literal	raid.
Gender	vyriškoji	masculine	vyr.
	moteriškoji	feminine	mot.
	bendroji	common	bendr.
	bevardė	neuter	bev.
Number	vienaskaita	singular	vns.
	daugiskaita	plural	dgs.
	dviskaita	dual	dvisk.

Case	vardininkas	nominative	V.
	kilmininkas	genitive	K.
	naudininkas	dative	N.
	galininkas	accusative	G.
	įnagininkas	instrumental	Įn.
	vietininkas	locative	Vt.
	šauksmininkas	vocative	Š.
	iliatyvas	illative	Il.
Definiteness	įvardžiutinis	pronominal	įvardž.
Possessive pronoun form	savybinis	possessive	savyb.
Degree	aukštesnysis	comparative diminutive	aukštėl.
	aukštesnysis	comparative	aukšt.
	aukščiausiasis	superlative	aukšč.
	nelyginamasis	positive	nelygin.
Verb form	asmenuojamoji	conjugated	asm.
	bendratis	infinitive	bndr.
	dalyvis	participle	dlv.
	padalyvis	gerund	pad.
	pusdalyvis	half particle	pusd.
	būdinys	verbal adverb	būdn.
	siekinys	supine	siekn.
Participle type	veikiamoji	active	veik.
	neveikiamoji	passive	neveik.
	reikiamybės	necessity	reik.
Mood	tiesioginė	indicative	tiesiog.
	liepiamoji	imperative	liep.
	tariamoji	subjective	tar.
	geidžiamoji	optative	geidž.
Tense	esamasis	present	es.
	būtas	past	būt.
	būtas kartinis	past	būt-k.
	būtas dažninis	past frequentative	būt-d.
	būsimasis	future	būs.
Person	pirmasis	I	1.
	antrasis	II	2.
	trečiasis	III	3.
Shortened form	sutrumpintas variantas	shortened	trump.
	ilgasis variantas	long	ilg.
Usage type	reta, nevartotina forma	rare	vrnt.

Table 2. Ordering and examples

Part of speech	Ordering of tags	Examples
Noun	noun type, proper noun subtype, reflexiveness, gender, number case, language standard, shortened form, usage type	<i>Lietuvos</i> : dkt.tikr.vtverd.mot.vns.K. <i>Gediminas</i> : dkt.tikr.vrd.vyr.vns.V. <i>Jonaiti</i> : dkt.tikr.pvrd.vyr.vns.Š. <i>vyriausybė</i> : dkt.mot.vns.V. <i>įsipareigojimų</i> : dkt.sngr.vyr.dgs.K. <i>žmonėm</i> : dkt.vyr.dgs.N.trump. <i>šūdas</i> : dkt.vyr.vns.V.substnd. <i>ofisas</i> : dkt.vyr.vns.V.substnd. <i>pidaras</i> : dkt.vyr.vns.V.vulg. <i>kelyj</i> : dkt.vyr.vns.Vt.trump.vrnt.
Adjective	degree, definiteness, gender, number, case, language standard, shortened form, usage type	<i>naudingas</i> : bdv.nelygin.vyr.vns.V. <i>gerojo</i> : bdv.nelygin.įvardž.vyr.vns.K. <i>aukščiausiasis</i> : bdv.aukšč.įvardž.vyr.vns.V. <i>geriem</i> : bdv.nelygin.vyr.dgs.N.trump. <i>šipuotos</i> : bdv.nelygin.mot.dgs.V.substnd.
Pronoun	definiteness, gender, number, case, language standard, shortened form, usage type	<i>kurio</i> : įv.vyr.vns.K. <i>tasai</i> : įv.įvardž.vyr.vns.V. <i>savo</i> (lema <i>savęs</i>): įv.savyb.K. <i>tavim</i> : įv.vns.Įn.trump. <i>mudu</i> : įv.vyr.dvisk.V. <i>kožnu</i> : įv.vyr.vns.Įn.substnd. <i>viens</i> : įv.vyr.vns.V.trump.vrnt.
Numeral	numeral form, numeral type, definiteness, gender, number, case, language standard, shortened form, usage type	<i>penki</i> : sktv.raid.kiek.vyr.dgs.V. <i>penkeri</i> : sktv.raid.daugin.vyr.dgs.dgs.V. <i>penketą</i> : sktv.raid.kuopin.vyr.vns.G. <i>pirmasis</i> : sktv.raid.kelint.įvardž.vyr.vns.V. <i>pirmiem</i> : sktv.raid.kelint.vyr.dgs.N.trump. <i>tūkstantom</i> : sktv.raid.kelint.mot.dgs.Įn.trump.vrnt. <i>kem</i> : sktv.raid.kiek.substnd. 704: sktv.arab. 2-ojo: sktv.mišr.kelint.įvardž.vyr.vns.K. XXI: sktv.rom.
Verb	conjugated forms : verb form (asm.), verb polarity, reflexiveness, mood, tense, number, person, language standard, shortened form, usage type	<i>buvo</i> : vksm.asm.tiesiog.būt-k.3. <i>tebūnie</i> : vksm.asm.geidž. <i>nėra</i> : vksm.asm.neig.tiesiog.es.3. <i>nusikalto</i> : vksm.asm.sngr.tiesiog.būt-k.3. <i>užėikit</i> : vksm.asm.liep.dgs.2. <i>bliuznijo</i> : vksm.asm.tiesiog.būt-k.3.substnd. <i>kalbėtume</i> : vksm.asm.tar.dgs.1.trump. <i>suptute</i> : vksm.asm.tar.dgs.2.trump.vrnt.
	participles : verb form (dlv.), verb polarity, reflexiveness, participle type, tense, degree, definiteness, gender, number, case, language standard, shortened form, usage type	<i>numatyti</i> : vksm.dlv.neveik.būt.vyr.dgs.V. <i>suaugusieji</i> : vksm.dlv.veik.būt-k.įvardž.vyr.dgs.V. <i>buvusį</i> : vksm.dlv.veik.būt-k.vyr.vns.G. <i>pasislėpusi</i> : vksm.dlv.sngr.veik.būt-k.mot.vns.V. <i>mylimiausias</i> : vksm.dlv.neveik.es.aukšč.vyr.vns.V. <i>mylimasai</i> : vksm.dlv.neveik.es.įvardž.vyr.vns.V.vrnt. <i>frizuota</i> : vksm.dlv.neveik.būt.mot.vns.V.substnd. <i>veršią</i> : vksm.dlv.veik.būs.mot.dgs.V.trump.
	gerunds : verb form (pad.), verb polarity, reflexiveness, tense, language standard	<i>neatsisukant</i> : vksm.pad.neig.sngr.es. <i>atburzgiant</i> : vksm.pad.es. <i>bevelijant</i> : vksm.pad.es.substnd.
	half participles : verb form (pusd.), verb polarity, reflexiveness, gender, number, language standard	<i>pradėdami</i> : vksm.pusd.vyr.dgs. <i>nedeldamas</i> : vksm.pusd.neig.vyr.vns. <i>prasimanydamas</i> : vksm.pusd.sngr.vyr.vns.

		<i>mozodamas</i> : vksm.pusd.vyr.vns.substnd.
	infinitives : verb form (bndr.), verb polarity, reflexiveness, language standard, shortened form, usage type	<i>išlaikyti</i> : vksm.bndr. <i>pasišnekėti</i> : vksm.bndr.sngr. <i>nepaslysti</i> : vksm.bndr.neig. <i>bėgt</i> : vksm.bndr.trump. <i>arenduoti</i> : vksm.bndr.substnd.
	verbal adverbs : verb forms (būdn.), verb polarity, language standard	<i>bėgte</i> : vksm.būdn. <i>nebėgte</i> : vksm.būdn.neig. <i>perste</i> : vksm.būdn.substnd.
	supines : verb form (siekn.), verb polarity, language standard	<i>darytų</i> : vksm.siekn. <i>smožytų</i> : vksm.siekn.substnd.
Adverb	degree, language standard, shortened form, usage type	<i>dažnai</i> : prv.nelygin. <i>daugiau</i> : prv.aukšt. <i>daugiausia</i> : prv.aukšč. <i>daugiausiai</i> : prv.aukšč.ilg. <i>iš karto</i> : prv.nelygin. <i>ubagiškai</i> : prv.nelygin.substnd.
Preposition	governing case, language standard	<i>link</i> : prl.K. <i>apie</i> : prl.G. <i>su</i> : prl.Įn. <i>apart</i> : prl.K.substnd.

Table 3. Examples of annotated text in different frameworks

Input text	Standard	Annotated text
Šis tekstas anotuotas naudojant „Jablonskio“ standartą.	CoNLL-U	<pre># sent_id = 1 # text = Šis tekstas anotuotas naudojant „Jablonskio“ standartą. 1 Šis šis PRON jv.vyr.vns.V. Case=Nom Gender=Masc Number=Sing _ _ _ _ 2 tekstas tekstas NOUN dkt.vyr.vns.V. Case=Nom Gender=Masc Number=Sing _ _ _ _ 3 anotuotas anotuoti VERB vksm.dlv.neveik.būt.vyr.vns.V. Case=Nom Gender=Masc Number=Sing Tense=Past VerbForm=Part Voice=Pass _ _ _ _ 4 naudojant naudoti VERB vksm.pad.es. Tense=Pres VerbForm=Ger _ _ _ _ 5 „ „ PUNCT skyr. _ _ _ SpaceAfter=No 6 Jablonskio Jablonskis PROP dkt.tikr.pvrd.vyr.vns.K. Case=Gen Gender=Masc NameType=Sur Number=Sing _ _ _ SpaceAfter=No 7 “ “ PUNCT skyr. _ _ _ SpaceAfter=No 8 standartą standartas NOUN dkt.vyr.vns.G. Case=Acc Gender=Masc Number=Sing _ _ _ SpaceAfter=No 9 . . PUNCT skyr. _ _ _ SpaceAfter=No</pre>
	Sketch Engine	<pre><doc> <p> <s> Šis šis jv.vyr.vns.V. šis-p tekstas tekstas dkt.vyr.vns.V. tekstas-n anotuotas anotuoti vksm.dlv.neveik.būt.vyr.vns.V. anotuoti-v naudojant naudoti vksm.pad.es. naudoti-v „ „ skyr. „-t </g> Jablonskio Jablonskis dkt.tikr.pvrd.vyr.vns.K. Jablonskis-n </g> “ “ skyr. “-t standartą standartas dkt.vyr.vns.G. standartas-n </g> . . skyr. .-t </s> </p> </doc></pre>
	NoSketch Engine	<pre><doc> <p> <s> Šis jv.vyr.vns.V. šis tekstas dkt.vyr.vns.V. tekstas anotuotas vksm.dlv.neveik.būt.vyr.vns.V. anotuoti naudojant vksm.pad.es. naudoti „ skyr. „ </g> Jablonskio dkt.tikr.pvrd.vyr.vns.K. Jablonskis </g> “ skyr. “ standartą dkt.vyr.vns.G. standartas </g> . skyr. . </s> </p> </doc></pre>