

Morphological Tagset JABLONSKIS v0.1

A morphological tagset “Jablonskis” has been created in the Centre of Computational Linguistics, Vytautas Magnus university to be used in Lithuanian treebank ALKSNIS v3.0 and in morphologically annotated corpus MATAS v1.0. Human-readable common Lithuanian abbreviations are used for the tagset. This is to ensure readability and to minimize human error. In previous ALKSNIS and MATAS versions, MULTEXT-East standard was used¹.

No.	Category	Attribute (LT)	Attribute (En)	Tag
1	MWE	samplaikos pradžia	start of MWE	sampl.
		nesamplaikinis	non-mwe	(without the tag “sampl.”)
2	POS	skyrybos ženklas	punctuation mark	skyr.
		daiktavardis	noun	dkt.
		būdvardis	adjective	bdv.
		skaitvardis	numeral	sktv.
		įvardis	pronoun	įv.
		veiksmažodis	verb	vksm.
		prieveiksmis	adverb	prv.
		jaustukas	interjection	jst.
		ištiktukas	onomatopoeia	išt.
		dalelytė	particle	dll.
		prielinksnis	preposition	prl.
		jungtukas	conjunction	jng.
		akronimas	acronym	akr.
		sutrumpinimas	abbreviation	sutr.
		samplaikos tęsinys	continuation of MWE	tęs.
		užsienio kalbos žodis	foreign word	užs.
		kiti (pvz., įstatymų numeriai, kuriuos sudaro skaičiai ir raidės, nepilni žodžiai, internetinės svetainės nuoroda, elektroninio pašto adresas)	residual (e.g., law numbers, codes, www links, e-mail addresses)	kita
3	Numeral form	arabiškas skaičius	Arabic	arab.
		romėniškas skaičius	Roman	rom.
		mišrus skaitvardis (pvz., 2-oji, 50-ies)	mixed numeral (e.g. 2-oji, 50-ies)	mišr.
		žodžiu užrašytas skaitvardis	literal	raid.
4	Noun types	tikrinis	proper	tikr.
		bendrini	common	(without the tag “tikr.”)
5	Numeral types	kiekinis	cardinal	kiek.
		kelintinis	ordinal	kelint.
		dauginis	multiple	daugin.

¹ <http://nl.ijs.si/ME/V4/msd/html/index.html>

		kuopinis	collect	kuopin.
6	Verb forms	bendratis	infinitive	bndr.
		dalyvis	participle	dlv.
		padalyvis	adverbial participle	pad.
		pusdalyvis	half participle	pusd.
		būdinys	adverbial participle2	būdn.
		asmenuojama forma	main	asm.
7	Polarity (only for verbs)	neigiamas	negative	neig.
		teigiamas	positive	(without the tag “neig.”)
8	Reflexiveness	sangražinis	reflexive	sngr.
		nesangražinis	non-reflexive	(without the tag “sngr.”)
9	Mood	tiesioginė nuosaka	indicative	tiesiog.
		liepiamoji nuosaka	imperative	liep.
		tariamoji nuosaka	subjunctive	tar.
10	Voice	veikiamoji rūšis	active	veik.
		neveikiamoji rūšis	passive	neveik.
		reikiamybės	necessity	reik.
11	Tense	esamasis laikas	present	es.
		būtas is kartinis laikas	past tense	būt-k
		būtas is dažninis laikas	past frequentative	būt-d.
		būsimasis laikas	future	būs.
		būtas is laikas	simple past	būt.
12	Degree	nelyginamasis laipsnis	positive	nelygin.
		aukštesnysis laipsnis	comparative	aukšt.
		aukščiausias is laipsnis	superlative	aukšč.
13	Definiteness	įvardžiutinis	pronominal	įvardž.
		neįvardžiutinis	non-pronominal	(without the tag “įvardž.”)
14	Gender	moteriškoji giminė	feminine	mot.
		vyriškoji giminė	masculine	vyr.
		bevardė giminė	neuter	bev.
		bendroji giminė	general	bendr.
15	Number	vienaskaita	singular	vns.
		daugiskaita	plural	dgs.
		dviskaita	dual	dvisk.
16	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.
17	Person	pirmasis asmuo	first	1.
		antrasis asmuo	second	2.
		trečiasis asmuo	third	3.

Tag Order

Noun: noun type + reflexiveness + gender + number + case

Adjective: degree + definiteness + gender + number + case

Pronoun: mwe + definiteness + gender + number + case

Numeral: mwe + number form + number type + definiteness + gender + number + case

Verb:

a) <asm.> + polarity + reflexiveness + mood + tense + number + person

b) <dlv.> + polarity + reflexiveness + type + tense + degree + definiteness + gender + number + case

c) <padlv.> + polarity + reflexiveness + tense

d) <psdlv.> + polarity + reflexiveness + gender + number

e) <bndr.>/<bdn.> + polarity + reflexiveness

Adverb: mwe + degree

Preposition: mwe + case

Interjection: mwe

Particle: mwe

Onomatopoeia: mwe

Conjunction: mwe

Abbreviation: –

Acronym: –

Continuation of grammatical mwe: –

Foreign: mwe

Other: –

General Notes on Morphological Annotation. Problematic Cases

- Tags are connected by a period without space, e.g.: dkt.tikr.mot.vns.Vt.
- Besides POS several new top categories are included: punctuation (skyr.), acronym (akr.), abbreviation (sutr.), foreign (užs.), other elements (law numbers, codes, www links, e-mail addresses etc.) (kita.).
- Morphological MWEs are split into two constituents. The annotation of the first word is preceded by „sampl.“ and followed by appropriate POS and categories. The annotation of the second word is „tęs.“, while syntactic tag – AuxL.
- For numerals are marked: Arabic, Roman, mixed or literal.
- 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.