

# 07 Universal Dependencies

Universal Dependencies (UD) je mednarodno usklajena označevalna shema, ki si prizadeva za poenoteno oblikoslovno in skladenjsko označevanje besedil v vseh svetovnih jezikih, da bi pospešila razvoj večjezičnih jezikovnih tehnologij in kontrastivnih jezikoslovnih raziskav. Znotraj sheme, ki temelji na načelih odvisnostne slovnice, je bil vzpostavljen univerzalni nabor slovničnih kategorij (besednih vrst, oblikoslovnih lastnosti in odvisnostnih skladenjskih relacij) in smernic za njihovo pripisovanje, obenem pa zasnova sheme dovoljuje tudi vpeljevanje jezikovnospecifičnih oznak, če je to potrebno. Doslej je bilo s to shemo ročno označenih že več kot 250 korpusov po vsem svetu, med katerimi sta tudi univerzalni odvisnostni drevesnici pisne (SSJ) in govorjene (SST) slovenščine.

- [Predstavitev oznak](#)
- [Označevalne smernice](#)
- [Reference in povezave](#)

# Predstavitev oznak

Schema Universal Dependencies določa univerzalni nabor oznak za besedne vrste, oblikoslovne lastnosti in odvisnostne skladenjske relacije, ki se lahko prenesejo na drevesnice posameznih jezikov ali pa dopolnijo z novimi oblikoslovnimi oznakami oz. izpeljavami jedrnih relacij, če je to potrebno. Konkretno je bilo na slovenske podatke preneseno vseh 17 besednih vrst (Tabela 1), 22 oblikoslovnih lastnosti z 62 različnimi vrednostmi (Tabela 2) in 35 odvisnostnih relacij (Tabela 3).

Oznaka	Opis
ADJ	pridevnik
ADP	adpozicija (predlog)
ADV	prislov
AUX	pomožnik
CCONJ	priredni veznik
DET	določilnik
INTJ	medmet
NOUN	samostalnik
NUM	števnik
PART	členek
PRON	zaimek
PROPN	lastnoimenski samostalnik
PUNCT	ločilo
SCONJ	podredni veznik
SYM	simbol
VERB	glagol
X	drugo

**Tabela 1:** Na slovenskih besedilih uporabljene oznake za besedne vrste.

Lastnost	Vrednosti	Opis
Abbr	Yes	okrajšanost

Lastnost	Vrednosti	Opis
<b>Animacy</b>	Anim, Inanim	živost
<b>Aspect</b>	Imp, Perf	vid
<b>Case</b>	Nom, Gen, Dat, Acc, Loc, Ins	sklon
<b>Definite</b>	Ind, Def	določnost
<b>Degree</b>	Pos, Cmp, Sup	stopnja
<b>Foreign</b>	Yes	tujejezičnost
<b>Gender</b>	Masc, Fem, Neut	spol
<b>Gender[psor]</b>	Masc. Fem, Neut	spol svojine
<b>Mood</b>	Ind, Imp, Cnd	naklon
<b>Number</b>	Sing, Dual, Plur	število
<b>Number[psor]</b>	Sing, Dual, Plur	število svojine
<b>NumForm</b>	Word, Digit, Roman	oblika števnika
<b>NumType</b>	Card, Ord, Mult, Sets	vrste števnika
<b>Person</b>	1, 2, 3	oseba
<b>Polarity</b>	Neg, Pos	nikalnost
<b>Poss</b>	Yes	svojina
<b>PronType</b>	Prs, Int, Rel, Dem, Tot, Neg, Ind	vrsta zimka
<b>Reflex</b>	Yes	povratnost
<b>Tense</b>	Pres, Fut	čas
<b>Variant</b>	Bound, Short	oblika zaimka
<b>VerbForm</b>	Fin, Inf, Sup, Part, Conv	vrsta glagola

**Tabela 2:** Na slovenskih besedilih uporabljene oznake za oblikoslovne lastnosti. V korpusu so navedene v obliki parov lastnosti in vrednosti (npr. Tense=Pres).

Oznaka	Angleški opis	Slovenski opis
<b>acl</b>	clausal modifier of noun	stavčni prilastki
<b>advcl</b>	adverbial clause modifier	prislovni odvisniki
<b>advmod</b>	adverbial modifier	prislovna določila (v širšem smislu)
<b>amod</b>	adjectival modifier	pridevniški prilastki
<b>appos</b>	appositional modifier	pristavčna določila

Oznaka	Angleški opis	Slovenski opis
<b>aux</b>	auxiliary verb	pomožni glagoli
<b>case</b>	case marking preposition	predlogi
<b>cc</b>	coordinating conjunction	priredni vezniki
<b>ccomp</b>	clausal complement	stavčna dopolnila (predmetni odvisniki)
<b>conj</b>	conjunct	priredno zloženi elementi
<b>cop</b>	copula verb	vezni glagoli
<b>csbj</b>	clausal subject	osebkovi odvisniki
<b>dep</b>	unspecified dependency	nedoločena povezava
<b>det</b>	determiner	določilniki
<b>discourse</b>	discourse element	diskurzni členki
<b>dislocated</b>	dislocated element	dislocirani elementi
<b>expl</b>	expletive	ekspletivne besede
<b>fixed</b>	fixed multi-word expression	funkcijske zveze
<b>flat</b>	flat multi word-expression	eksocentrične zveze
<b>goeswith</b>	disjointed token	razdruženi deli besed
<b>iobj</b>	indirect object	nepremi predmeti
<b>list</b>	list	seznami
<b>mark</b>	marker (subordinating conjunction)	podredni vezniki
<b>nmod</b>	nominal modifier	samostalniški prilastki
<b>nsbj</b>	nominal subject	samostalniški osebki
<b>nummod</b>	numeric modifier	številčna določila
<b>obj</b>	(direct) object	premi predmeti
<b>obl</b>	oblique nominal (adjunct)	odvisne samostalniške zveze
<b>orphan</b>	dependent of missing parent	elementi v eliptičnih strukturah
<b>parataxis</b>	parataxis	stavčna sovedja
<b>punct</b>	punctuation symbol	ločila
<b>reparandum</b>	overriden disfluency	samopopravljanja
<b>root</b>	root element	koren povedi
<b>vocative</b>	vocative	ogovori
<b>xcomp</b>	open clausal complement	odprta stavčna dopolnila

**Tabela 3:** Na slovenskih besedilih uporabljene oznake za odvisnostne skladišne relacije (brez podoznaka).

# Označevalne smernice

V tem poglavju so zbrane označevalne smernice za oblikoslovje in skladnjo po sistemu Universal Dependencies (UD).

## Različica 1.7

### projekt [SPOT](#)

DOBROVOLJC, Kaja in TERČON, Luka: 2024. *Universal Dependencies: Smernice za označevanje besedil v slovenščini*. Različica 1.7. Rezultat projekta Na drevesnici temelječ pristop k raziskavam govornjene slovenščine. [\[DOCX\]](#) [\[PDF\]](#)

## Različica 1.3

### projekt [SPOT](#)

DOBROVOLJC, Kaja in TERČON, Luka: 2023. *Universal Dependencies: Smernice za označevanje besedil v slovenščini*. Različica 1.3. Rezultat projekta Na drevesnici temelječ pristop k raziskavam govornjene slovenščine. [\[DOCX\]](#) [\[PDF\]](#)

## Različica 1.0

### projekt [Razvoj slovenščine v digitalnem okolju](#)

DOBROVOLJC, Kaja in TERČON, Luka: 2023. *Universal Dependencies: Smernice za označevanje besedil v slovenščini*. Rezultat projekta Razvoj slovenščine v digitalnem okolju. [\[DOCX\]](#) [\[PDF\]](#)

Priloga k smernicam: Odprta vprašanja pri prenosu označevalne sheme Universal Dependencies na slovenska besedila

[\[DOCX\]](#) [\[PDF\]](#)

# Reference in povezave

V tem poglavju so zbrane relevantne reference in povezave na projekte, v katerih se je označevalni sistem razvijal ter uporabljal.

Krovna spletna stran projekta Universal Dependencies: <https://universaldependencies.org/>

Splošne smernice: <https://universaldependencies.org/guidelines.html>

Slovenske smernice (v angleščini): <https://universaldependencies.org/sl/index.html>

Platforma za diskusijo glede slovenskih smernic in predloge izboljšav (v angleščini):  
[https://github.com/UniversalDependencies/UD\\_Slovenian-SSJ/issues](https://github.com/UniversalDependencies/UD_Slovenian-SSJ/issues)

## Korpusi z ročno pregledanimi oznakami UD

Drevesnica pisne slovenščine SSJ: [https://github.com/UniversalDependencies/UD\\_Slovenian-SSJ](https://github.com/UniversalDependencies/UD_Slovenian-SSJ)

Drevesnica govorne slovenščine SST: [https://github.com/UniversalDependencies/UD\\_Slovenian-SST](https://github.com/UniversalDependencies/UD_Slovenian-SST)

Drevesnica SSJ kot del trenutne različice učnega korpusa SUK: Arhar Holdt, Špela; et al., 2024, Training corpus SUK 1.1, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1959>.

## Reference

Dobrovoljc, K. (2024). Extending the Spoken Slovenian Treebank. Conference on Language Technologies and Digital Humanities (JT-DH-2024), Ljubljana, Slovenia.

<https://doi.org/10.5281/zenodo.13936394>

Dobrovoljc, K., Terčon, L., Ljubešić, N. (2023). Universal Dependencies za slovenščino: nove smernice, ročno označeni podatki in razčlenjevalni model. Slovenščina 2.0, 11(1): 218–246.

<https://doi.org/10.4312/slo2.0.2023.1.218-246> [PDF]

Dobrovoljc, K., Terčon, L., & Ljubešić, N. (2022). Universal Dependencies za slovenščino: nadgradnja smernic, učnih podatkov in razčlenjevalnega modela. In D. Fišer & T. Erjavec (Eds.), Jezikovne tehnologije in digitalna humanistika: zbornik konference (pp. 30–39). Inštitut za novejšo zgodovino. [https://nl.ijs.si/jtdh22/pdf/JTDH2022\\_Dobrovoljc-et-al\\_Universal-Dependencies-za-slovenscino.pdf](https://nl.ijs.si/jtdh22/pdf/JTDH2022_Dobrovoljc-et-al_Universal-Dependencies-za-slovenscino.pdf)

de Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal Dependencies. Computational Linguistics, 47(2), 255–308. [https://doi.org/10.1162/coli\\_a\\_00402](https://doi.org/10.1162/coli_a_00402)

Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F., & Zeman, D. (2020). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. Proceedings of the Twelfth Language Resources and Evaluation Conference, 4034–4043.

<https://aclanthology.org/2020.lrec-1.497>

Dobrovoljc, K., Erjavec, T., & Krek, S. (2017). The Universal Dependencies Treebank for Slovenian. Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing, 33–38.

<https://doi.org/10.18653/v1/W17-1406>

Dobrovoljc, K., & Nivre, J. (2016). The Universal Dependencies Treebank of Spoken Slovenian. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 1566–1573. <https://aclanthology.org/L16-1248>