

# 07 Universal Dependencies

Universal Dependencies (UD) is an internationally harmonised annotation framework that aims to standardize the morphological and syntactic tagging of texts across languages in order to foster the development of multilingual language technologies and contrastive linguistics research. Grounded in dependency grammar principles, UD provides a universal set of grammatical categories – including parts of speech (POS), morphological features and syntactic dependency relations – and guidelines for their use. The framework is flexible, incorporating language-specific tags as needed. To date, over 240 corpora worldwide have been manually tagged using the framework, including the universal treebanks of both written and spoken Slovene.

- [Introduction to Tags](#)
- [Annotation Guidelines](#)
- [References and Links](#)

# Introduction to Tags

The Universal Dependencies framework establishes a comprehensive and universal set of tags for parts of speech (POS), morphological features and syntactic dependencies that can be adopted in the treebanks of individual languages, or supplemented with new morphological features or derivations of core relations when necessary. For the Slovene language data, this includes the adoption of all 17 parts of speech (see Table 1), 22 morphological features spanning 62 distinct values (see Table 2), and 35 types of dependency relations (see Table 3).

Tag	Description
ADJ	adjective
ADP	adposition
ADV	adverb
AUX	auxiliary
CCONJ	coordinating conjunction
DET	determiner
INTJ	interjection
NOUN	noun
NUM	numeral
PART	particle
PRON	pronoun
PROPN	proper noun
PUNCT	punctuation
SCONJ	subordinating conjunction
SYM	symbol
VERB	verb
X	other

**Table 1:** Part-of-speech tags used in Slovene texts.

Feature	Value	Description
Abbr	Yes	abbreviation

Feature	Value	Description
<b>Animacy</b>	Anim, Inanim	animacy
<b>Aspect</b>	Imp, Perf	aspect
<b>Case</b>	Nom, Gen, Dat, Acc, Loc, Ins	case
<b>Definite</b>	Ind, Def	definiteness or state
<b>Degree</b>	Pos, Cmp, Sup	degree
<b>Foreign</b>	Yes	is this a foreign word?
<b>Gender</b>	Masc, Fem, Neut	gender
<b>Gender[psor]</b>	Masc. Fem, Neut	possessor's gender
<b>Mood</b>	Ind, Imp, Cnd	mood
<b>Number</b>	Sing, Dual, Plur	number
<b>Number[psor]</b>	Sing, Dual, Plur	possessor's number
<b>NumForm</b>	Word, Digit, Roman	numeral form
<b>NumType</b>	Card, Ord, Mult, Sets	numeral type
<b>Person</b>	1, 2, 3	person
<b>Polarity</b>	Neg, Pos	polarity
<b>Poss</b>	Yes	possessive
<b>PronType</b>	Prs, Int, Rel, Dem, Tot, Neg, Ind	pronominal type
<b>Reflex</b>	Yes	reflexive
<b>Tense</b>	Pres, Fut	tense
<b>Variant</b>	Bound, Short	alternative form of word
<b>VerbForm</b>	Fin, Inf, Sup, Part, Conv	form of verb or deverbative

**Table 2:** Tags for morphological features used in Slovene texts. In the corpus, these are listed in the form of feature and value pairs (e.g., Tense=Pres).

Tag	Description
<b>acl</b>	clausal modifier of noun
<b>advcl</b>	adverbial clause modifier
<b>advmod</b>	adverbial modifier
<b>amod</b>	adjectival modifier
<b>appos</b>	appositional modifier

Tag	Description
<b>aux</b>	auxiliary verb
<b>case</b>	case marking preposition
<b>cc</b>	coordinating conjunction
<b>ccomp</b>	clausal complement
<b>conj</b>	conjunct
<b>cop</b>	copula verb
<b>csbj</b>	clausal subject
<b>dep</b>	unspecified dependency
<b>det</b>	determiner
<b>discourse</b>	discourse element
<b>dislocated</b>	dislocated element
<b>expl</b>	expletive
<b>fixed</b>	fixed multi-word expression
<b>flat</b>	flat multi word-expression
<b>goeswith</b>	disjointed token
<b>iobj</b>	indirect object
<b>list</b>	list
<b>mark</b>	marker (subordinating conjunction)
<b>nmod</b>	nominal modifier
<b>nsbj</b>	nominal subject
<b>nummod</b>	numeric modifier
<b>obj</b>	(direct) object
<b>obl</b>	oblique nominal (adjunct)
<b>orphan</b>	dependent of missing parent
<b>parataxis</b>	parataxis
<b>punct</b>	punctuation symbol
<b>reparandum</b>	overriden disfluency
<b>root</b>	root element
<b>vocative</b>	vocative
<b>xcomp</b>	open clausal complement

**Table 3:** Tags for syntactic dependency relations (without subtypes) used in Slovene texts.



# Annotation Guidelines

This chapter summarizes the annotation guidelines for the Universal Dependencies (UD) morphology and syntax as applied to Slovene texts. The guidelines are arranged from the latest, up-to-date version to the oldest version.

## Version 1.3

### Project **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 govorne slovenščine. [\[DOCX\]](#) [\[PDF\]](#) - only in Slovene

## Version 1.0

### Project **Development of Slovene in a Digital Environment**

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

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

[\[DOCX\]](#) [\[PDF\]](#) - only in Slovene

# References and Links

This chapter compiles relevant references and provides links to projects where the the Universal Dependencies (UD) morphology and syntax have been developed and applied to Slovene texts.

Main website of the Universal Dependencies project: <https://universaldependencies.org/>

General guidelines: <https://universaldependencies.org/guidelines.html>

Slovene guidelines (in English): <https://universaldependencies.org/sl/index.html>

## Corpora containing manually revised UD tags

Slovenian-SSJ written language treebank: [https://github.com/UniversalDependencies/UD\\_Slovenian-SSJ](https://github.com/UniversalDependencies/UD_Slovenian-SSJ)

Slovenian-SST spoken language treebank:

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

Slovenian-SSJ treebank as part of the SUK 1.0 training corpus: Zwitter Vitez, Ana; et al., 2023, Spoken corpus Gos 2.0 (transcriptions), Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1771>.

## References

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] - only in Slovene

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>