This page pertains to UD version 2.

Universal Dependencies

Universal Dependencies (UD) is a framework for consistent annotation of grammar (parts of speech, morphological features, and syntactic dependencies) across different human languages. UD is an open community effort with over 600 contributors producing over 200 treebanks in over 150 languages (see the bottom of this page for updated numbers from the latest release). If you are new to UD, you should start by reading the first part of the Short Introduction and then browsing the annotation guidelines.
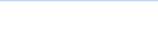
- Short introduction to UD
- UD annotation guidelines
- More information on UD:
  - How to contribute to UD
  - Tools for working with UD
  - Changes to the UD guidelines
  - UD-related events
  - Projects related to UD
- Query UD treebanks online:
  - PML Tree Query maintained by the Charles University in Prague
  - TEITOK maintained by the Charles University in Prague
  - Grew-match maintained by Inria in Nancy
  - INESS maintained by the University of Bergen
- Download UD treebanks
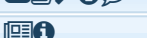
If you want to receive news about Universal Dependencies, you can subscribe to the UD mailing list. If you want to discuss individual annotation questions, use the Github issue tracker.

Current UD Languages

Information about language families (and genera for families with multiple branches) is mostly taken from WALS Online (IE = Indo-European).

| ▶ | 🚩 | Abaza | 1 | <1K | 💬 | Northwest Caucasian |

Language		#	Size		Family
▸	Abkhaz	1	6K		Northwest Caucasian
▸	Afrikaans	1	49K		IE, Germanic
▸	Akkadian	2	25K		Afro-Asiatic, Semitic
▸	Akuntsu	1	1K		Tupian, Tupari
▸	Albanian	2	4K		IE, Albanian
▸	Amharic	1	10K		Afro-Asiatic, Semitic
▸	Ancient Greek	3	456K		IE, Greek
▸	Ancient Hebrew	1	39K		Afro-Asiatic, Semitic
▸	Apurina	1	<1K		Arawakan
▸	Arabic	3	1,042K		Afro-Asiatic, Semitic
▸	Armenian	2	94K		IE, Armenian
▸	Assyrian	1	<1K		Afro-Asiatic, Semitic
▸	Azerbaijani	1	<1K		Turkic, Southwestern
▸	Bambara	1	13K		Mande
▸	Basque	1	121K		Basque
▸	Bavarian	1	15K		IE, Germanic
▸	Beja	1	11K		Afro-Asiatic, Cushitic
▸	Belarusian	1	305K		IE, Slavic
▸	Bengali	1	<1K		IE, Indic
▸	Bhojpuri	1	6K		IE, Indic
▸	Bororo	1	6K		Bororoan
▸	Breton	1	10K		IE, Celtic
▸	Bulgarian	1	156K		IE, Slavic
▸	Buryat	1	10K		Mongolic
▸	Cantonese	1	13K		Sino-Tibetan, Chinese
▸	Cappadocian	2	4K		IE, Greek
▸	Catalan	1	553K		IE, Romance
▸	Cebuano	1	1K		Austronesian, Central Philippine
▸	Chinese	7	309K		Sino-Tibetan, Chinese
▸	Chukchi	1	6K		Chukotko-Kamchatkan
▸	Classical Armenian	1	88K		IE, Armenian
▸	Classical Chinese	2	433K		Sino-Tibetan, Chinese
▸	Coptic	1	57K		Afro-Asiatic, Egyptian
▸	Croatian	1	199K		IE, Slavic
▸	Czech	6	2,252K		IE, Slavic
▸	Danish	1	100K		IE, Germanic
▸	Dutch	2	506K		IE, Germanic
▸	Egyptian	1	14K		Afro-Asiatic, Egyptian
▸	English	11	760K		IE, Germanic

	Language	Count	Speakers	Resources	Family
▸	Erzya	1	20K		Uralic, Mordvin
▸	Estonian	2	529K		Uralic, Finnic
▸	Faroese	2	50K		IE, Germanic
▸	Finnish	4	397K		Uralic, Finnic
▸	French	7	635K		IE, Romance
▸	Frisian Dutch	1	3K		Code switching
▸	Galician	3	187K		IE, Romance
▸	Georgian	1	60K		Kartvelian
▸	German	4	3,810K		IE, Germanic
▸	Gheg	1	15K		IE, Albanian
▸	Gothic	1	55K		IE, Germanic
▸	Greek	2	88K		IE, Greek
▸	Guajajara	1	9K		Tupian, Maweti–Guarani
▸	Guarani	1	<1K		Tupian, Maweti–Guarani
▸	Gujarati	1	1K		IE, Indic
▸	Gwichin	1	1K		Na–Dene
▸	Haitian Creole	1	3K		Creole
▸	Hausa	2	18K		Afro–Asiatic, West Chadic
▸	Hebrew	3	368K		Afro–Asiatic, Semitic
▸	Highland P. Nahuatl	1	10K		Uto–Aztecan
▸	Hindi	2	375K		IE, Indic
▸	Hittite	1	1K		IE, Anatolian
▸	Hungarian	1	42K		Uralic, Ugric
▸	Icelandic	4	1,183K		IE, Germanic
▸	Indonesian	3	169K		Austronesian, Malayo–Sumbawan
▸	Irish	3	168K		IE, Celtic
▸	Italian	10	1,001K		IE, Romance
▸	Japanese	6	2,645K		Japanese
▸	Javanese	1	14K		Austronesian, Javanese
▸	Kaapor	1	<1K		Tupian, Maweti–Guarani
▸	Kangri	1	2K		IE, Indic
▸	Karelian	1	3K		Uralic, Finnic
▸	Karo	1	2K		Tupian, Ramarama
▸	Kazakh	1	10K		Turkic, Northwestern
▸	Khunsari	1	<1K		IE, Iranian
▸	Kiche	1	10K		Mayan
▸	Komi Permyak	1	1K		Uralic, Permic
▸	Komi Zyrian	2	10K		Uralic, Permic
▸	Korean	4	513K		Korean

		Language	Count	Speakers	Resources	Family
▶		Kurmanji	1	10K	📄W	IE, Iranian
▶		Kyrgyz	2	24K	📄🖊📖	Turkic, Northwestern
▶		Latgalian	1	<1K	🖊	IE, Baltic
▶		Latin	6	1,002K	☁✉📄🏹ℹ🎵	IE, Italic
▶		Latvian	2	328K	🎓📄🖊🏹📖💬	IE, Baltic
▶		Ligurian	1	6K	☁📄🖊📖ℹ💬W	IE, Romance
▶		Lithuanian	2	75K	📄🏹📖ℹ	IE, Baltic
▶		Livvi	1	1K	📖ℹ🌐	Uralic, Finnic
▶		Low Saxon	1	22K	📄ℹ	IE, Germanic
▶		Luxembourgish	1	<1K	🖊	IE, Germanic
▶		Macedonian	1	1K	🖊	IE, Slavic
▶		Madi	1	<1K	🖊	Arawan
▶		Maghrebi Arabic French	1	19K	📖ℹ	Code switching
▶		Makurap	1	<1K	📖ℹ	Tupian, Tupari
▶		Malayalam	1	2K	🖊	Dravidian
▶		Maltese	1	44K	📄🏹📖ℹW	Afro-Asiatic, Semitic
▶		Manx	1	20K	☁📅📄📖ℹ📡🌐W	IE, Celtic
▶		Marathi	1	3K	📄W	IE, Indic
▶		Mbya Guarani	2	13K	📄ℹ	Tupian, Maweti-Guarani
▶		Middle French	1	12K	📄ℹ	IE, Romance
▶		Moksha	1	4K	📖ℹ	Uralic, Mordvin
▶		Munduruku	1	1K	📖ℹ	Tupian, Munduruku
▶		Naija	1	140K	💬	Creole
▶		Nayini	1	<1K	🖊💬	IE, Iranian
▶		Neapolitan	1	<1K	🖊	IE, Romance
▶		Nheengatu	1	19K	☁📄🖊ℹ💬	Tupian, Maweti-Guarani
▶		North Sami	1	26K	📖ℹ	Uralic, Sami
▶		Northwest Gbaya	1	2K	💬	Niger-Congo, Gbaya-Manza-Ngbaka
▶		Norwegian	2	611K	📅📖ℹ	IE, Germanic
▶		Old Church Slavonic	1	198K	☁	IE, Slavic
▶		Old East Slavic	4	553K	🏹ℹ	IE, Slavic
▶		Old French	1	237K	🏹ℹ🎵	IE, Romance
▶		Old Irish	2	<1K	🎓☁🖊ℹ🎵	IE, Celtic
▶		Old Turkish	1	<1K	ℹ	Turkic, Northeastern
▶		Ottoman Turkish	2	9K	☁📄 📖ℹ	Turkic, Southwestern
▶		Pashto	1	<1K	🖊ℹ	IE, Iranian
▶		Paumari	1	<1K	🖊	Arawan
▶		Persian	2	654K	🎓📅📄🏹🖊📖ℹ📡💬🌐	IE, Iranian
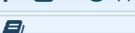
	Language				Classification
	Pesh	1	2K		Chibchan, Pesh
	Phrygian	1	1K		IE, Greek
	Polish	3	499K		IE, Slavic
	Pomak	1	34K		IE, Slavic
	Portuguese	7	1,545K		IE, Romance
	Romanian	5	941K		IE, Romance
	Russian	5	1,896K		IE, Slavic
	Sanskrit	2	208K		IE, Indic
	Scottish Gaelic	1	89K		IE, Celtic
	Serbian	1	97K		IE, Slavic
	Sinhala	1	<1K		IE, Indic
	Skolt Sami	1	2K		Uralic, Sami
	Slovak	1	106K		IE, Slavic
	Slovenian	2	365K		IE, Slavic
	Soi	1	<1K		IE, Iranian
	South Levantine Arabic	1	<1K		Afro-Asiatic, Semitic
	Spanish	4	1,031K		IE, Romance
	Spanish Sign Language	1	1K		Sign Language
	Swedish	3	206K		IE, Germanic
	Swedish Sign Language	1	1K		Sign Language
	Swiss German	1	1K		IE, Germanic
	Tagalog	2	1K		Austronesian, Central Philippine
	Tamil	2	12K		Dravidian
	Tatar	1	2K		Turkic, Northwestern
	Teko	1	2K		Tupian, Maweti-Guarani
	Telugu	1	6K		Dravidian
	Telugu English	1	<1K		Code switching
	Thai	1	22K		Tai-Kadai
	Tswana	1	<1K		Niger-Congo, Bantoid
	Tupinamba	1	4K		Tupian, Maweti-Guarani
	Turkish	9	735K		Turkic, Southwestern
	Turkish German	1	37K		Code switching
	Ukrainian	2	174K		IE, Slavic
	Umbrian	1	<1K		IE, Italic
	Upper Sorbian	1	11K		IE, Slavic
	Urdu	1	138K		IE, Indic
	Uyghur	1	40K		Turkic, Southeastern
	Uzbek	1	5K		Turkic, Southeastern
	Veps	1	1K		Uralic, Finnic

▶	🇻🇳	Vietnamese	2	59K		Austro-Asiatic, Viet-Muong
▶		Warlpiri	1	<1K		Pama-Nyungan, Western
▶	🏴	Welsh	1	52K		IE, Celtic
▶		Western Armenian	1	122K		IE, Armenian
▶	🇲🇽	Western S.P. Nahuatl	1	10K		Uto-Aztecan
▶	🇸🇳	Wolof	1	44K		Niger-Congo, Northern Atlantic
▶	🇧🇷	Xavante	1	1K		Macro-Je
▶	🇨🇳	Xibe	1	15K		Tungusic
▶		Yakut	1	1K		Turkic, Northeastern
▶	🇳🇬	Yoruba	1	8K		Niger-Congo, Defoid
▶		Yupik	1	2K		Eskimo-Aleut
▶	🇳🇬	Zaar	1	17K		Afro-Asiatic, West Chadic

Disclaimer: Our use of flags to symbolise languages is only intended as a visual enhancement of the website and should not be interpreted as a political statement in any way.

Possible Future Extensions

People have expressed interest in providing annotated data for the following languages but no valid data has been provided so far.

▶	🇮🇶	Akkadian	1	117K		Afro-Asiatic, Semitic
▶	🇪🇹	Amharic	2	–		Afro-Asiatic, Semitic
▶		Archaic Irish	1	–		IE, Celtic
▶	🇮🇳	Assamese	1	–		IE, Indic
▶	🇧🇷	Balatipone	1	–		Bororoan
▶	🇧🇩	Bengali	3	–		IE, Indic
▶	🇮🇳	Bhojpuri	1	–		IE, Indic
▶		Classical Nahuatl	1	–		Uto-Aztecan
▶	🇲🇽	Cuicatec	1	–		Oto-Manguean
▶	🇮🇹	Cusco Quechua	1	–		Quechuan
▶	🇨🇿	Czech	1	1,191K		IE, Slavic
▶	🇩🇰	Danish	1	–		IE, Germanic
▶		Dargwa	1	–		Nakh-Daghestanian, Lak-Dargwa
▶	🇬🇧	English	1	–		IE, Germanic
▶		Esperanto	1	–		Constructed
▶	🇫🇷	French	1	–		IE, Romance
▶		Frisian	1	–		IE, Germanic
▶	🇪🇹	Gedeo	1	<1K		Afro-Asiatic, Cushitic
▶	🇬🇪	Georgian	2	1K		Kartvelian

▶	🇬🇷	Greek	3	–	✏️	IE, Greek
▶	🇵🇭	Hiligaynon	1	<1K	✏️	Austronesian, Central Philippine
▶	🇮🇳	Hindi	1	4K	📶	IE, Indic
▶	🇲🇽	Huave	1	–	✏️	Huavean
▶	🇮🇹	Italian	1	–	✏️	IE, Romance
▶	🇯🇵	Japanese	2	–	✏️💬	Japanese
▶	🇩🇿	Kabyle	1	23K	📖ℹ️	Afro-Asiatic, Berber
▶	🇧🇷	Kadiweu	1	–	✏️	Guaicuruan
▶	🇮🇳	Kannada	1	–	✏️	Dravidian
▶	🇳🇦	Khoekhoe	1	–	📖ℹ️	Khoe-Kwadi
▶	🇺🇬	Kiga	1	–	📖ℹ️	Niger-Congo, Bantoid
▶		Komi	1	<1K	✏️	Uralic, Permic
▶	🇰🇷	Korean	2	–	📖ℹ️📄	Korean
▶	🇮🇱	Ladino	1	–	📖ℹ️	IE, Romance
▶	🇹🇷	Laz	1	2K	✏️	Kartvelian
▶	🇮🇳	Magahi	2	7K	📖ℹ️W	IE, Indic
▶	🇮🇳	Mandyali	1	<1K	📖ℹ️	IE, Indic
▶		Mansi	1	–	✏️	Uralic, Ugric
▶	🇮🇳	Marathi	1	205K	📑📑📑	IE, Indic
▶		Middle Irish	2	<1K	📕/📖ℹ️	IE, Celtic
▶	🇲🇳	Mongolian	1	–	📖ℹ️	Mongolic
▶	🇹🇿	Ndengeleko	1	–	📖ℹ️	Niger-Congo, Bantoid
▶		Nenets	1	–	✏️	Uralic, Samoyedic
▶	🇳🇵	Nepali	1	–	📖ℹ️	IE, Indic
▶	🇺🇬	Nkore	1	–	📖ℹ️	Niger-Congo, Bantoid
▶		Occitan	2	–	📖ℹ️	IE, Romance
▶	🇮🇳	Odia	1	<1K	📖ℹ️	IE, Indic
▶		Old English	2	–	📕ℹ️	IE, Germanic
▶	🇯🇵	Old Japanese	1	3K	📕	Japanese
▶	🇨🇴	Palenquero	1	–	✏️	Creole
▶		Papiamento	2	–	✏️	Creole
▶	🇦🇫	Pashto	1	–	✏️	IE, Iranian
▶	🇮🇷	Persian	1	–	✏️	IE, Iranian
▶	🇮🇳	Pnar	1	–	✏️💬	Austro-Asiatic, Khasian
▶		Pontic	1	–	📖ℹ️	IE, Greek
▶	🇵🇹	Portuguese	1	–	W	IE, Romance
▶	🇮🇳	Prakrit	1	<1K	📖ℹ️	IE, Indic
▶	🇮🇳	Punjabi	1	6K	📅📖ℹ️	IE, Indic
▶	🇵🇪	Puno Quechua	1	–	✏️	Quechuan

▶	Romanian	1	–		IE, Romance
▶	Romansh	2	–		IE, Romance
▶	Seri	1	–		Hokan, Seri
▶	Shipibo Konibo	1	–		Pano-Tacanan
▶	Sindhi	2	6K		IE, Indic
▶	Sinhala	1	–		IE, Indic
▶	Somali	1	–		Afro-Asiatic, Cushitic
▶	Sorani	1	–		IE, Iranian
▶	Spanish	1	<1K		IE, Romance
▶	Spanish English	1	–		Code switching
▶	Swahili	1	–		Niger-Congo, Bantoid
▶	Tagabawa	1	<1K		Austronesian, Central Philippine
▶	Tagalog	1	340K		Austronesian, Central Philippine
▶	Tetun	1	–		Austronesian, Malayo-Polynesian
▶	Thai	1	–		Tai-Kadai
▶	Tigrinya	1	–		Afro-Asiatic, Semitic
▶	Tunisian Arabic	1	–		Afro-Asiatic, Semitic
▶	Turkish	1	142K		Turkic, Southwestern
▶	Tuwari	1	–		Sepik
▶	Uspanteko	1	–		Mayan
▶	Uyghur	1	–		Turkic, Southeastern
▶	Uzbek	1	–		Turkic, Southeastern
▶	Western S.P. Nahuatl	1	–		Uto-Aztecan

Disclaimer: Our use of flags to symbolise languages is only intended as a visual enhancement of the website and should not be interpreted as a political statement in any way.

Retired Treebanks

The following treebanks have been part of one or more UD releases in the past but they are no longer maintained and they have been excluded from the most recent release.

▶	English	1	97K		IE, Germanic
▶	French	1	573K		IE, Romance
▶	Hindi English	1	26K		Code switching
▶	Japanese	2	204K		Japanese
▶	Norwegian	1	55K		IE, Germanic

Disclaimer: Our use of flags to symbolise languages is only intended as a visual enhancement of the website and should not be interpreted as a political statement in any way.

Download

The data is released through [LINDAT/CLARIAH-CZ](#).

- The next release (v2.16) is scheduled for May 15, 2025 (data freeze on May 1).
- Version 2.15 treebanks are available at http://hdl.handle.net/11234/1-5787. 296 treebanks, 168 languages, released November 15, 2024.
- Version 2.14 treebanks are archived at http://hdl.handle.net/11234/1-5502. 283 treebanks, 161 languages, released May 15, 2024.
- Version 2.13 treebanks are archived at http://hdl.handle.net/11234/1-5287. 259 treebanks, 148 languages, released November 15, 2023.
- Version 2.12 treebanks are archived at http://hdl.handle.net/11234/1-5150. 245 treebanks, 141 languages, released May 15, 2023.
- Version 2.11 treebanks are archived at http://hdl.handle.net/11234/1-4923. 243 treebanks, 138 languages, released November 15, 2022.
- Version 2.10 treebanks are archived at http://hdl.handle.net/11234/1-4758. 228 treebanks, 130 languages, released May 15, 2022.
- Version 2.9 treebanks are archived at http://hdl.handle.net/11234/1-4611. 217 treebanks, 122 languages, released November 15, 2021.
- Version 2.8 treebanks are archived at http://hdl.handle.net/11234/1-3687. 202 treebanks, 114 languages, released May 15, 2021.
- Version 2.7 treebanks are archived at http://hdl.handle.net/11234/1-3424. 183 treebanks, 104 languages, released November 15, 2020.
- Version 2.6 treebanks are archived at http://hdl.handle.net/11234/1-3226. 163 treebanks, 92 languages, released May 15, 2020.
- Version 2.5 treebanks are archived at http://hdl.handle.net/11234/1-3105. 157 treebanks, 90 languages, released November 15, 2019.
- Version 2.4 treebanks are archived at http://hdl.handle.net/11234/1-2988. 146 treebanks, 83 languages, released May 15, 2019.
- Version 2.3 treebanks are archived at http://hdl.handle.net/11234/1-2895. 129 treebanks, 76 languages, released November 15, 2018.
- Version 2.2 treebanks are archived at http://hdl.handle.net/11234/1-2837. 122 treebanks, 71 languages, released July 1, 2018.
- Version 2.1 treebanks are archived at http://hdl.handle.net/11234/1-2515. 102 treebanks, 60 languages, released November 15, 2017.
- Version 2.0 treebanks are archived at http://hdl.handle.net/11234/1-1983. 70 treebanks, 50 languages, released March 1, 2017.
    - Test data 2.0 are archived at http://hdl.handle.net/11234/1-2184. 81 treebanks, 49 languages, released May 18, 2017.
- Version 1.4 treebanks are archived at http://hdl.handle.net/11234/1-1827. 64 treebanks, 47 languages, released November 15, 2016.
- Version 1.3 treebanks are archived at http://hdl.handle.net/11234/1-1699. 54 treebanks, 40 languages, released May 15, 2016.
- Version 1.2 treebanks are archived at http://hdl.handle.net/11234/1-1548. 37 treebanks, 33 languages, released November 15, 2015.
- Version 1.1 treebanks are archived at http://hdl.handle.net/11234/LRT-1478. 19 treebanks, 18 languages, released May 15, 2015.
- Version 1.0 treebanks are archived at http://hdl.handle.net/11234/1-1464. 10 treebanks, 10 languages, released January 15, 2015.
- In general, we intend to have regular treebank releases every six months. The v2.0 and v2.2 releases were brought forward because of their usage in the [CoNLL 2017 and 2018 Multilingual Parsing Shared Tasks](#).