

# Universal Dependencies, UDPipe and UD Morphology



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004984



# UDPipe

A Multilingual tagger, parser, lemmatizer



## UDPipe

[About](#) [Run](#) [REST API Documentation](#)

UDPipe is a trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLL-U files. UDPipe is language-agnostic and can process data in CoNLL-U format. Trained models are provided for nearly all UD treebanks. UDPipe is available as a binary for Linux/Windows/OS X, as a Docker container, and as a web service. [Third-party R CRAN package](#) also exists.

UDPipe is a free software distributed under the [Mozilla Public License 2.0](#) and the linguistic models are free for non-commercial use and distributed under the [CC BY-NC-SA](#) license, although for some models the original data used to create the model may impose additional licensing conditions. UDPipe is versioned using Semantic Versioning.

Copyright 2017 by Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Czech Republic.

Description of the available methods is available in the [API Documentation](#) and the models are described in the [UDPipe 2 models list](#) and [UDPipe 1 models list](#).

## Service

The service is freely available for testing. Respect the [CC BY-NC-SA](#) licence of the models – **explicit written permission of the authors is required for any commercial exploitation of the system**. If you use the service, you agree that data obtained by us during such use can be used for further improvements of the system. Comments and reactions are welcome.

Model:  UD 2.6 (description)  EvaLatin20 (description)Actions:  Tag and Lemmatize  Parse

## Advanced Options

A Input Text

input file

A Output

↓ Process Input ↓

A Output Text

Show Table

Show Trees

Model:  UD 2.6 (description)  EvaLatin20 (description)

Actions:

Options

- 
- arabic-padt-ud-2.6-200830
- armenian-armtdp-ud-2.6-200830
- basque-bdt-ud-2.6-200830
- belarusian-hse-ud-2.6-200830
- bulgarian-btb-ud-2.6-200830
- catalan-ancora-ud-2.6-200830
- chinese-gdsimp-ud-2.6-200830
- chinese-gsd-ud-2.6-200830
- classical\_chinese-kyoto-ud-2.6-200830
- coptic-scriptorium-ud-2.6-200830
- croatian-set-ud-2.6-200830
- czech-pdt-ud-2.6-200830
- czech-cac-ud-2.6-200830
- czech-fictree-ud-2.6-200830
- czech-cltt-ud-2.6-200830
- danish-ddt-ud-2.6-200830
- dutch-alpino-ud-2.6-200830
- dutch-lassysmall-ud-2.6-200830
- english-ewt-ud-2.6-200830
- english-gum-ud-2.6-200830
- english-lines-ud-2.6-200830
- english-partut-ud-2.6-200830

**Model:**  UD 2.6 (description)  EvaLatin20 (description)

 czech-pdt-ud-2.6-200830

**Actions:**  Tag and Lemmatize  Parse

▼ Advanced Options

 Input Text

 Input File

I have heard her reported to be a woman of an invincible spirit.

↓ Process Input ↓

 Output Text

 Show Table

 Show Trees

Input Text

Input File

I have heard her reported to be a woman of an invincible spirit. |

Process Input

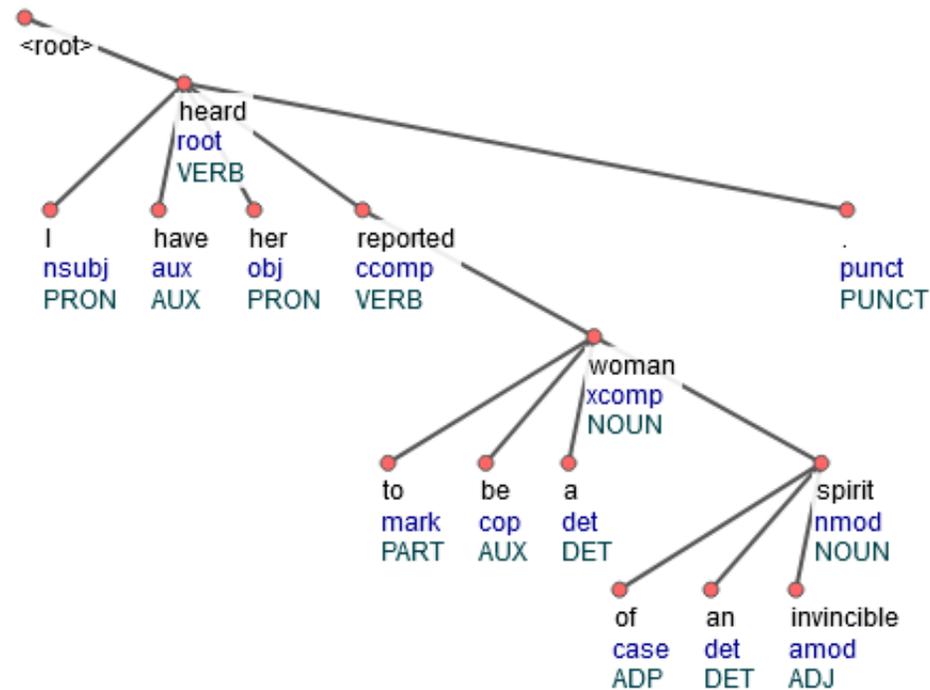
Output Text

Show Table

Show Trees

Save Tree as SVG

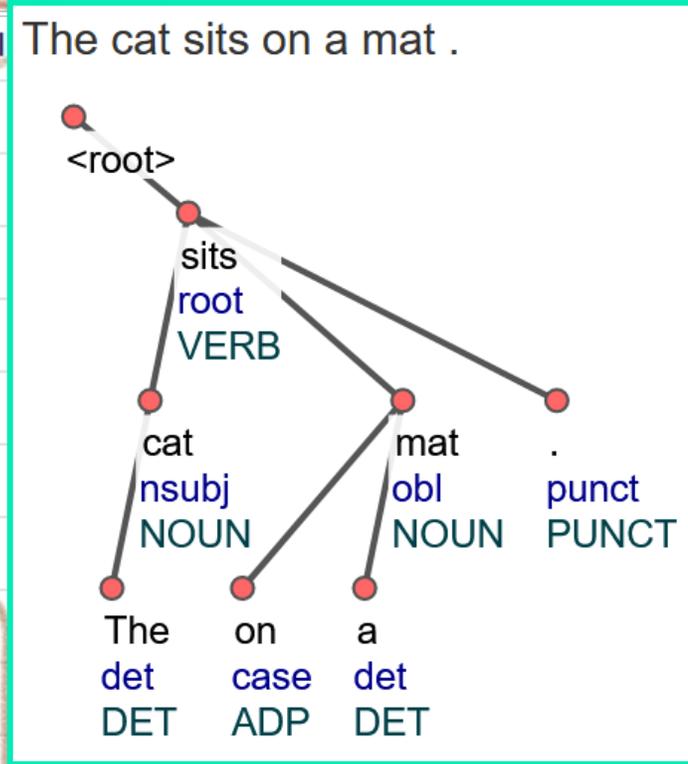
I have heard her reported to be a woman of an invincible spirit .



Id	Form	Lemma	UPostag	XPostag	Feats	Head	DepRel	Deps	Misc
# generator = UDPipe 2, <a href="https://lindat.mff.cuni.cz/services/udpipe">https://lindat.mff.cuni.cz/services/udpipe</a>									
# udpipe_model = english-ewt-ud-2.6-200830									
# udpipe_model_licence = CC BY-NC-SA									
# newdoc									
# newpar									
# sent_id = 1									
# text = I have heard her reported to be a woman of an invincible spirit.									
1	I	I	PRON	PRP	Case=Nom Number=Sing Person=1 PronType=Prs	3	nsubj	_	TokenRange=0:1
2	have	have	AUX	VBP	Mood=Ind Tense=Pres VerbForm=Fin	3	aux	_	TokenRange=2:6
3	heard	hear	VERB	VBN	Tense=Past VerbForm=Part	0	root	_	TokenRange=7:12
4	her	she	PRON	PRP	Case=Acc Gender=Fem Number=Sing Person=3 PronType=Prs	3	obj	_	TokenRange=13:16
5	reported	report	VERB	VBD	Mood=Ind Tense=Past VerbForm=Fin	3	ccomp	_	TokenRange=17:25
6	to	to	PART	TO	_	9	mark	_	TokenRange=26:28
7	be	be	AUX	VB	VerbForm=Inf	9	cop	_	TokenRange=29:31
8	a	a	DET	DT	Definite=Ind PronType=Art	9	det	_	TokenRange=32:33
9	woman	woman	NOUN	NN	Number=Sing	5	xcomp	_	TokenRange=34:39
10	of	of	ADP	IN	_	13	case	_	TokenRange=40:42
11	an	a	DET	DT	Definite=Ind PronType=Art	13	det	_	TokenRange=43:45
12	invincible	invincible	ADJ	JJ	Degree=Pos	13	amod	_	TokenRange=46:56
13	spirit	spirit	NOUN	NN	Number=Sing	9	nmod	_	SpaceAfter=No TokenRange=57:63
14	.	.	PUNCT	.	_	3	punct	_	TokenRange=63:64

## Conll-u format

Id	Form	Lemma	UPosTag	XPosTag	Feats	Head	DepRel
# generator = UDPipe 2, <a href="https://lindat.mff.cuni.cz/services/udpipe">https://lindat.mff.cuni.cz/services/udpipe</a>							
# udpipe_model = english-ewt-ud-2.6-200830							
# udpipe_model_licence = CC BY-NC-SA							
# newdoc							
# newpar							
# sent_id = 1							
# text = The cat sits on a mat.							
1	The	the	DET	DT	Definite=Def PronType=Art	2	det
2	cat	cat	NOUN	NN	Number=Sing	3	nsubj
3	sits	sit	VERB	VBZ	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	0	root
4	on	on	ADP	IN	-	6	case
5	a	a	DET	DT	Definite=Ind PronType=Art	6	det
6	mat	mat	NOUN	NN	Number=Sing	3	obl
7	.	.	PUNCT	.	-	3	punct



tokenRange=8:12  
 TokenRange=13:15  
 TokenRange=16:17  
 SpaceAfter=No|TokenRange=18:21  
 TokenRange=21:22

A large, light gray background graphic consisting of several overlapping, stylized shapes that resemble a large, abstract letter 'U' or a series of connected bars, framing the central text.

# Universal Dependencies

[universaldependencies.org](https://universaldependencies.org)

# Consistent grammar annotation across languages

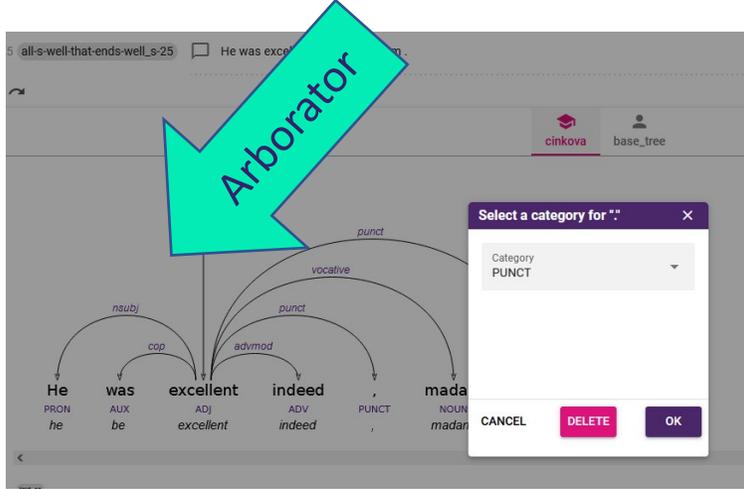


- over 300 contributors
- nearly 200 treebanks (corpora w. syntax annotation)
- over 100 languages
- publicly available

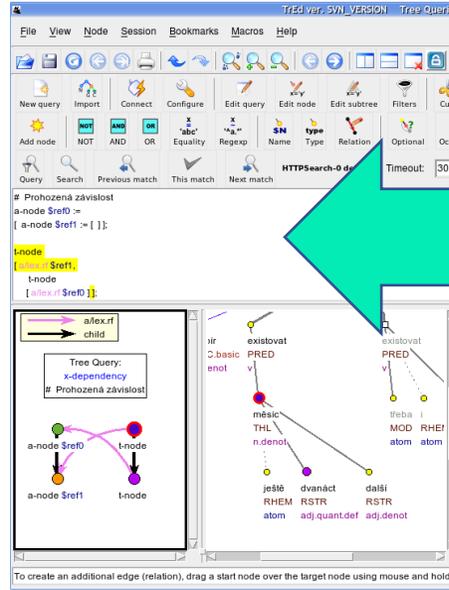
▶		Akkadian	2	25K		Afro-Asiatic, Semitic
▶		Akuntsu	1	1K		Tupian, Tupari
▶		Albanian	1	<1K	W	IE, Albanian
▶		Amharic	1	10K		Afro-Asiatic, Semitic
▶		Ancient Greek	2	416K		IE, Greek
▶		Ancient Hebrew	1	39K		Afro-Asiatic, Semitic
▶		Apurina	1	<1K		Arawakan
▶		Arabic	3	1,042K		Afro-Asiatic, Semitic
▶		Armenian	2	94K		IE, Armenian
▶		Assyrian	1	<1K		Afro-Asiatic, Semitic
▶		Bambara	1	13K		Mande
▶		Basque	1	121K		Basque
▶		Beja	1	<1K		Afro-Asiatic, Cushitic
▶		Belarusian	1	305K		IE, Slavic
▶		Bengali	1	<1K		IE, Indic
▶		Bhojpuri	1	6K		IE, Indic
▶		Breton	1	10K		IE, Celtic
▶		Bulgarian	1	156K		IE, Slavic
▶		Buryat	1	10K		Mongolic
▶		Cantonese	1	13K		Sino-Tibetan
▶		Catalan	1	553K		IE, Romance
▶		Cebuano	1	1K		Austronesian, Central P
▶		Chinese	5	285K		Sino-Tibetan
▶		Chukchi	1	6K		Chukotko-Kamchatkan
▶		Classical Chinese	1	289K		Sino-Tibetan
▶		Coptic	1	52K		Afro-Asiatic, Egyptian
▶		Croatian	1	199K		IE, Slavic
▶		Czech	5	2,227K		IE, Slavic
▶		Danish	1	100K		IE, Germanic
▶		Dutch	2	306K		IE, Germanic
▶		English	9	762K		IE, Germanic

The screenshot shows the LINDAT/CLARIAH-CZ digital library interface. The main content area displays the item 'Universal Dependencies 2.10'. A yellow box contains citation information: 'Please use the following text to cite this item or export to a predefined format: Zeman, Daniel; et al., 2022, *Universal Dependencies 2.10*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-4758>.'. Below this is a share button and a list of authors: 'Zeman, Daniel ; et al. show everyone'. The 'Item identifier' is <http://hdl.handle.net/11234/1-4758>, the 'Project URL' is <http://universaldependencies.org/>, and the 'Referenced by' link is [https://doi.org/10.1162/coli\\_a\\_00402](https://doi.org/10.1162/coli_a_00402). The 'Date issued' is 2022-05-15, the 'Type' is 'corpus, text', and the 'Size' is '29631957 tokens, 30203983 words, 1731173 sentences'. The 'Language(s)' are 'Afrikaans, Akkadian, Akuntsu, Albanian, Amharic, AncientGreek (to 1453), Ancient Hebrew, Apurina, Arabic, Armenian, Assyrian, Bambara, Basque, Beja, Belarusian, Bengali, Bhojpuri, Breton, Bulgarian, Buryat, Cantonese, Catalan, Cebuano, Chinese, Chukchi, Classical Chinese, Coptic, Croatian, Czech, Danish, Dutch, English'.

# Contribute to Universal Dependencies



Arborator



TRED

The screenshot shows the GitHub repository for Universal Dependencies. The repository name is "UniversalDependencies" and it has 306 repositories, 306 projects, 180 packages, and 20 people. The "Pinned" section includes "docs" (Public) and "tools" (Public). The "Repositories" section shows "UD\_Yakut-YKTD" (Public) and "UD\_English-EWT" (Public). The "People" section shows a grid of user avatars. The "Top languages" section shows Python, Shell, Perl, HTML, and SystemVerilog.

The INCEPTION (WebAnno) interface is split into two main sections. The left section, "Active Learning", shows a session with a "Layer" of "Named entity" and a "Text" of "Illinois". It includes a "Recommendation" table with columns for "Text", "Label", "Score", and "Delta". The right section, "Annotation", shows a text snippet: "Barack Hussein Obama II born August 4, 1961) is an American politician who served as the 44th President of the United States from 2009 to 2017. The first African American to assume the presidency, he was previously the junior United States Senator from Illinois from 2005 to 2008. He served in the Illinois State Senate from 1997 until 2004." The text is annotated with various entities and relations, such as "Barack Obama I PER", "date of birth", "occupation", "politician", "position held", "President of the United States of America", and "Illinois River". A dropdown menu for "Identifier" is open, showing options like "illinois", "Illinois Senate", "Illinois River", "Governor of Illinois", "Alton", "Illinois Country", and "Illinois Territory".

INCEPTION (WebAnno)

# Universal Parts of Speech (upos)

UD Morphology

# Morphological categories

- Universal Parts of Speech (**upos**)

- NOUN, PROPN
- VERB, AUX
- ADJ, ADV
- PRON, DET, NUM
- SCONJ, CCONJ, ADP
- PART, INTJ
- PUNCT, SYM, X

- Universal Features (**feats**)

- morphological categories relevant to the given upos







# NOUN vs. PROPN vs. neither

strawberries	NOUN
cat	NOUN
small	neither
Peter	

# NOUN vs. PROPN vs. neither

strawberries	NOUN
cat	NOUN
small	neither
Peter	PROPN
butter	

# NOUN vs. PROPN vs. neither

strawberries	NOUN
cat	NOUN
small	neither
Peter	PROPN
butter	NOUN
beer	

# NOUN vs. PROPN vs. neither

strawberries	NOUN
cat	NOUN
small	neither
Peter	PROPN
butter	NOUN
beer	NOUN
Dutchman	

# NOUN vs. PROPN vs. neither

strawberries	NOUN
cat	NOUN
small	neither
Peter	PROPN
butter	NOUN
beer	NOUN
Dutchman	PROPN
until	

# NOUN vs. PROPN vs. neither

strawberries	NOUN
cat	NOUN
small	neither
Peter	PROPN
butter	NOUN
beer	NOUN
Dutchman	PROPN
until	neither







# VERB vs. AUX vs. neither

are	AUX
can	AUX
(He) did (it)	VERB
Do (you smoke?)	

# VERB vs. AUX vs. neither

are	AUX
can	AUX
(He) did (it)	VERB
Do (you smoke?)	AUX
(be) flying	

# VERB vs. AUX vs. neither

are	AUX
can	AUX
(He) did (it)	VERB
Do (you smoke?)	AUX
(be) flying	VERB
(He) used (to swim)	

# VERB vs. AUX vs. neither

are	AUX
can	AUX
(He) did (it)	VERB
Do (you smoke?)	AUX
(be) flying	VERB
(He) used (to swim)	VERB
(She is) going (to win.)	VERB
(You) ought (to smile).	VERB







# VERB vs. AUX vs. neither

(a) winning (strategy)	VERB
(a) rotting (tooth)	VERB
(a) lost (war)	VERB
(a) rotten (tooth)	

# VERB vs. AUX vs. neither

(a) winning (strategy)	VERB
(a) rotting (tooth)	VERB
(a) lost (war)	VERB
(a) rotten (tooth)	neither (adjective)
Let('s dance.)	

# VERB vs. AUX vs. neither

(a) winning (strategy)	VERB
(a) rotting (tooth)	VERB
(a) lost (war)	VERB
(a) rotten (tooth)	neither (adjective)
Let('s dance.)	VERB
(She) wants (food)	

# VERB vs. AUX vs. neither

(a) winning (strategy)	VERB
(a) rotting (tooth)	VERB
(a) lost (war)	VERB
(a) rotten (tooth)	neither (adjective)
Let('s dance.)	VERB
(She) wants (food)	VERB
(She) wants (to win)	VERB
(He) became (professor)	

# VERB vs. AUX vs. neither

(a) winning (strategy)	VERB
(a) rotting (tooth)	VERB
(a) lost (war)	VERB
(a) rotten (tooth)	neither (adjective)
Let('s dance.)	VERB
(She) wants (food)	VERB
(She) wants (to win)	VERB
(He) became (professor)	VERB







# ADJ vs. ADV vs. neither

green	ADJ
happily	ADV
my	neither
many	

# ADJ vs. ADV vs. neither

green	ADJ
happily	ADV
my	neither
many	ADJ
oldest	

# ADJ vs. ADV vs. neither

green	ADJ
happily	ADV
my	neither
many	ADJ
oldest	ADJ
(the) third (year)	

# ADJ vs. ADV vs. neither

green	ADJ
happily	ADV
my	neither
many	ADJ
oldest	ADJ
(the) third (year)	ADJ
(the) poor	

# ADJ vs. ADV vs. neither

green	ADJ
happily	ADV
my	neither
many	ADJ
oldest	ADJ
(the) third (year)	ADJ
(the) poor	ADJ
where	

# ADJ vs. ADV vs. neither

green	ADJ
happily	ADV
my	neither
many	ADJ
oldest	ADJ
(the) third (year)	ADJ
(the) poor	ADJ
where	ADV





# ADJ vs. ADV vs. neither

twice	ADV
(take) off <small>(phrasal verb)</small>	neither
(write) down	ADV
sometime	

# ADJ vs. ADV vs. neither

twice	ADV
(take) off <small>(phrasal verb)</small>	neither
(write) down	ADV
sometime	ADV
yes	

# ADJ vs. ADV vs. neither

twice	ADV
(take) off <small>(phrasal verb)</small>	neither
(write) down	ADV
sometime	ADV
yes	neither
none	

# ADJ vs. ADV vs. neither

twice	ADV
(take) off <small>(phrasal verb)</small>	neither
(write) down	ADV
sometime	ADV
yes	neither
none	neither
how	

# ADJ vs. ADV vs. neither

twice	ADV
(take) off <small>(phrasal verb)</small>	neither
(write) down	ADV
sometime	ADV
yes	neither
none	neither
how	ADV

# ADJ vs. ADV vs. neither

twice	ADV
(take) off <small>(phrasal verb)</small>	neither
(write) down	ADV
sometime	ADV
yes	neither
none	neither
how	ADV
twice	ADV









# SCONJ vs. CCONJ vs. neither

(I hope) that (she will come)	SCONJ
(good) and (bad)	CCONJ
(nobody) but (you)	CCONJ
(this) or (that)	



# SCONJ vs. CCONJ vs. neither

(I hope) that (she will come)	SCONJ
(good) and (bad)	CCONJ
(nobody) but (you)	CCONJ
(this) or (that)	CCONJ
(this or) that	



# SCONJ vs. CCONJ vs. neither

(I hope) that (she will come)	SCONJ
(good) and (bad)	CCONJ
(nobody) but (you)	CCONJ
(this) or (that)	CCONJ
(this or) that	neither
(I know) which (to take)	



# SCONJ vs. CCONJ vs. neither

(I hope) that (she will come)	SCONJ
(good) and (bad)	CCONJ
(nobody) but (you)	CCONJ
(this) or (that)	CCONJ
(this or) that	neither
(I know) which (to take)	neither
(He left,) which (made her sad)	



# SCONJ vs. CCONJ vs. neither

(I hope) that (she will come)	SCONJ
(good) and (bad)	CCONJ
(nobody) but (you)	CCONJ
(this) or (that)	CCONJ
(this or) that	neither
(I know) which (to take)	neither
(He left,) which (made her sad)	neither
(Ask) whether (we may leave)	SCONJ









# NUM vs. DET vs. PRON

we	PRON
Which (kids arrived?)	DET
(Say) which (you like)	PRON
myself	



# NUM vs. DET vs. PRON

we	PRON
Which (kids arrived?)	DET
(Say) which (you like)	PRON
myself	PRON
mine	



# NUM vs. DET vs. PRON

we	PRON
Which (kids arrived?)	DET
(Say) which (you like)	PRON
myself	PRON
mine, yours	PRON
my, your, his	



# NUM vs. DET vs. PRON

we	PRON
Which (kids arrived?)	DET
(Say) which (you like)	PRON
myself	PRON
mine, yours	PRON
my, your, his	PRON
every	



# NUM vs. DET vs. PRON

we	PRON
Which (kids arrived?)	DET
(Say) which (you like)	PRON
myself	PRON
mine, yours	PRON
my, your, his	PRON
every	DET
no (man)	

# DET vs. PRON

we	PRON
Which (kids arrived?)	DET
(Say) which (you like)	PRON
myself	PRON
mine, yours	PRON
my, your, his	PRON
every	DET
no (man)	DET











# DET vs. NUM vs. ADJ vs. ADV

many	DET
two	NUM
first (minute)	ADJ
last (minute)	



# DET vs. NUM vs. ADJ vs. ADV

many	DET
two	NUM
first (minute)	ADJ
last (minute)	ADJ
one (man)	



# DET vs. NUM vs. ADJ vs. ADV

many	DET
two	NUM
first (minute)	ADJ
last (minute)	ADJ
one (man)	ADJ
(Charles) IV	



# DET vs. NUM vs. ADJ vs. ADV

many	DET
two	NUM
first (minute)	ADJ
last (minute)	ADJ
one (man)	ADJ
(Charles) IV	NUM
both (men)	



# DET vs. NUM vs. ADJ vs. ADV

many	DET
two	NUM
first (minute)	ADJ
last (minute)	ADJ
one (man)	ADJ
(Charles) IV	NUM
both (men)	DET
twice	



# DET vs. NUM vs. ADJ vs. ADV

many	DET
two	NUM
first (minute)	ADJ
last (minute)	ADJ
one (man)	ADJ
(Charles) IV	NUM
both (men)	DET
twice	ADV







# ADP vs. ADV vs. SCONJ

for (you)	ADP
(forgive me), for (I have done wrong)	SCONJ
ago	ADV
in	

# ADP vs. ADV vs. SCONJ

for (you)	ADP
(forgive me), for (I have done wrong)	SCONJ
ago	ADV
in	ADP
towards	

# ADP vs. ADV vs. SCONJ

for (you)	ADP
(forgive me), for (I have done wrong)	SCONJ
ago	ADV
in	ADP
towards	ADP
upwards	ADV
as/like (a teacher)	

# ADP vs. ADV vs. SCONJ

for (you)	ADP
(forgive me), for (I have done wrong)	SCONJ
ago	ADV
in	ADP
towards	ADP
upwards	ADV
as/like (a teacher)	ADP
(call) as (you go)	

# ADP vs. ADV vs. SCONJ

for (you)	ADP
(forgive me), for (I have done wrong)	SCONJ
ago	ADV
in	ADP
towards	ADP
upwards	ADV
as/like (a teacher)	ADP
(call) as (you go)	SCONJ



# Particles (PART)

- *not, n't*
- *to* (infinitive marker)
- 's (genitive ending)





# Interjections (INTJ)

- yes, no
- please
- well
- hi
- ok, bravo
- like
- lol
- hey
- oh, ouch



# Look it up in the Documentation

- Each treebank has its Documentation
- You get there from the language list at [universaldependencies.org](https://universaldependencies.org)
- Look up the very treebank that was used to train the model you use to parse texts in UDPipe – there are (small) differences
- [https://universaldependencies.org/treebanks/en\\_ewt/index.html](https://universaldependencies.org/treebanks/en_ewt/index.html)

# Universal Features

UD Morphology

# Universal features - feats (English EWT corpus)

- lexical & grammatical properties of words beyond upos tags
- Table: the most common feats, each feature has a set of possible values
- Feature labels should be consistent across languages, but each language can add theirs if not covered
- feats: alphabetically concatenated, separated by | (vertical bar)

Lexical features*	Inflectional features*	
	Nominal*	Verbal*
<u>PronType</u> ★	<u>Gender</u> ★	<u>VerbForm</u> ★
<u>NumType</u> ★	<u>Animacy</u>	<u>Mood</u> ★
<u>Poss</u> ★	<u>NounClass</u>	<u>Tense</u> ★
<u>Reflex</u> ★	<u>Number</u> ★	<u>Aspect</u>
<u>Foreign</u> ★	<u>Case</u> ★	<u>Voice</u> ★
<u>Abbr</u> ★	<u>Definite</u> ★	<u>Evident</u>
<u>Typo</u> ★	<u>Degree</u> ★	<u>Polarity</u> ★
		<u>Person</u> ★
		<u>Polite</u>
		<u>Clusivity</u>

# Features mostly describe only grammatical categories explicitly indicated by morphemes

- *he writes* Person=3, but *they write* does not have Person!
- *is sleeping* ≠ present progressive tense, but 2 verbs
  - *is*  
Mood=Ind | Number=Sing | Person=3 | Tense=Present | VerbForm=Fin
  - *sleeping* Tense=Pres | VerbForm=Part
- Many inconsistencies:
  - e. g. *be*: parser tries to assign person beside 1<sup>st</sup> and 3<sup>rd</sup> singular present tense, other verbs not so much.

# Case

- Nom, Acc
- with PRON, mostly `PronType=Prs` (Personal pronouns)
  - Nom: *I, they, we, he, she...* but also *you, it*,
  - Acc: *me, them, him, us, her...* but also *it, you, yourself, myself, themselves*

# Gender

- Fem, Masc, Neut
- with PRON, PronType=Prs
- usually also co-occurs with Number, Person, Case, Poss

# Person

- 1, 2, 3
- with VERB and AUX, mostly with VerbForm=Fin, Mood=Ind, Number=Sing, Tense=Pres
- with PRON, mostly with PronType=Pers, Case, Poss, and Number (any values)

# Number

- `Plur, Sing`
- with `NOUN` and `PROPN`
- with `PRON`, mostly with `PronType=Prs, Case, Gender, Poss`
- with `DET`, mostly with `PronType=Dem`

# Tense

- Past, Pres
- with VERB and AUX, mostly with VerbForm=Fin, Mood=Ind, Number, Person
- with SCONJ – Past: *given, based, provided*

# Mood

- Imp, Ind, Sub
- with VERB and AUX, mostly with VerbForm=Fin, Number, Person, Tense

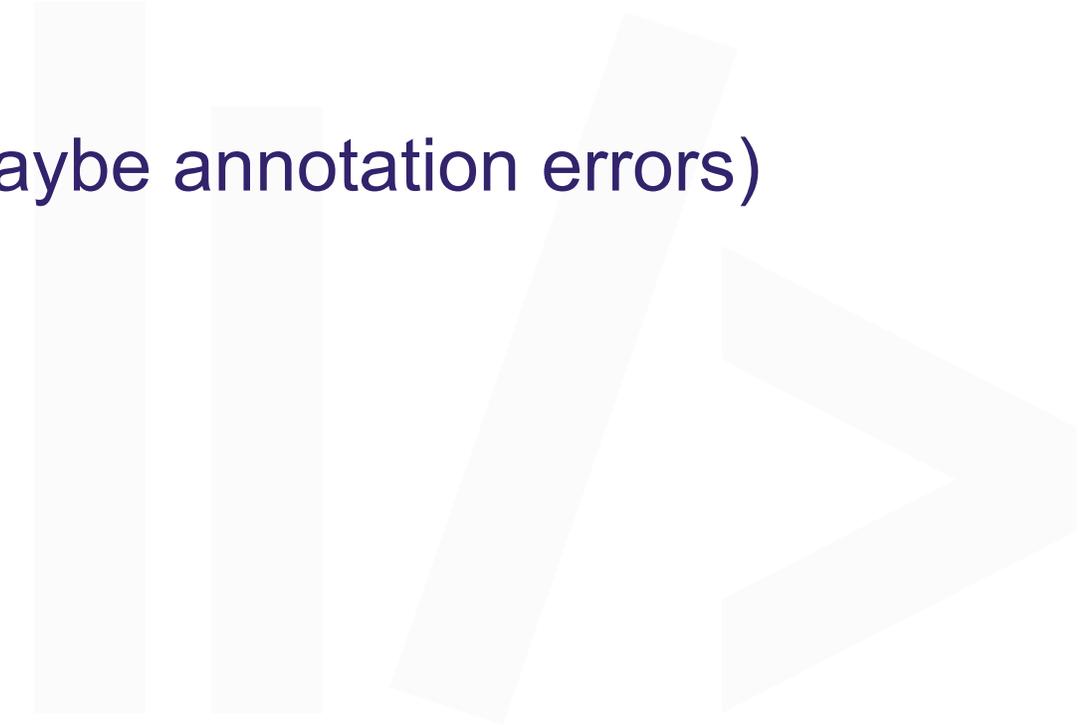


# Voice

- Pass
- with VERB, mostly with `VerbForm=Part, Tense=Past`
- This is quite a weird feature in English. It occurs systematically in past participles, when they are combined with be as AUX (*I was invited*). In this case, it considers the context. Cf. (the invited experts: `Voice=Pass` is not there, just `Tense=Past | VerbForm=Part`).
- Perhaps the parser just decided to do this, based on input from some other data?

# VerbForm

- Fin, ~~Ger~~, Inf, Part
- with VERB and AUX
- with SCONJ (very little cases, maybe annotation errors)



# Playtime!

[https://quizlet.com/\\_bkoupi?x=1jqt&i=c5q4t](https://quizlet.com/_bkoupi?x=1jqt&i=c5q4t)

[https://quizlet.com/\\_bkoq mz?x=1jqt&i=c5q4t](https://quizlet.com/_bkoq mz?x=1jqt&i=c5q4t)

# PronType

- Art, Dem, Emp, Int, Prs, Rel
- with PRON
  - Dem (demonstrative): *this, that, those, these*;
  - Emp (emphatic): *ourselves/yourselves/themselves, him/her/my/your/itself*;
  - Int (interrogative): *what, which, who, whom, whose*
  - Rel (relative): *that, who, which, whom, what, whose, whatever, whoever, whomever*
  - Prs: *I, you, it, they, my, we, he, your, me, them, their*
- with DET
  - Art: *the, a, an*
  - Dem: *this, that, these, those*
  - Int: *what, which, whatever*
  - Rel: *what, which*
  - EMPTY: *all, some, any, no, another, every, each, both, such*

# PronType - continuation

- with ADV
  - Dem: *then, there, here*
  - Int: *how, why, where, when, whenever, however*
  - Rel: *when, where, how, wherein*
  - EMPTY: *so, just, very, also, now, even, only, as, back, well*
- with SCONJ
  - Int: *when, how, where, why, whenever, wherever, who*
  - Rel: *where, when, why*
  - EMPTY: *that, if, as, because, for, of, since, before, like, with*

# Definite

- Def, Ind
- with DET
  - Def: *the*
  - Ind: *an, a*
  - EMPTY: *this, all, some, any, no, that, these, another, every, such*

# NumType

- Card, Frac, Mult, Ord
- with NUM:
  - Card: *one, two, 1,30...*
- with ADJ:
  - Frac: *half*
  - Ord: *first, second, third, 16<sup>th</sup>, ...*
- with ADV:
  - Frac: *half*
  - Mult: *once, twice*

# Degree

- Cmp, Pos, Sup
- with ADJ and ADV:
  - Cmp: *more, better, less, bigger...*
  - Pos: *good, great, new, far, well, soon, late, little, close...*
  - Sup: *best, most, least, worst, cheapest, largest...*

# Poss (is it possessive?)

# Reflex (is it reflexive?)

- Yes
- with PRON, mostly with `PronType=Prs`, `Gender`, `Number`, `Person`



**Foreign (is it in a foreign language?)**

**Typo (is it a typo?)**

**Abbr (is it an abbreviation?)**

- Yes



The background features a large, light gray left-pointing arrow on the left and a large, light gray right-pointing arrow on the right. In the center, there are three vertical bars of varying heights, resembling a bar chart or a stylized graphic element.

# Playtime!

[https://quizlet.com/\\_bo1jkz?x=1jqt&i=c5q4t](https://quizlet.com/_bo1jkz?x=1jqt&i=c5q4t)

The background features several large, light gray decorative elements: a large left-pointing chevron on the left, a large right-pointing chevron on the right, and a central graphic consisting of two vertical bars of different heights and a diagonal bar crossing them.

# Your turn

in groups

# Feats and their values in your languages!



- A mind map of features (mainly of verbs) across languages is here:

[https://www.orgpad.com/o/DfIElyUSIBzY6YTaK-pUDf?token=Dp\\_2WHU1pHFKcAmAsmqLeC&open=all](https://www.orgpad.com/o/DfIElyUSIBzY6YTaK-pUDf?token=Dp_2WHU1pHFKcAmAsmqLeC&open=all)

- The UD documentation page on feats is here:

<https://universaldependencies.org/u/feat/all.html>

- Create groups and set up a list of words from your languages that would combine features and values not present in English.
- Are there word forms with ambiguous upos, such as participles in adjectival positions? Show us!
- You can consult UDPipe: <https://lindat.mff.cuni.cz/services/udpipe/>
  - Select an appropriate language model
  - Create an example sentence with the candidate and check out the markup.
  - If there are several models for your language, do they disagree?