

Universal Description of Morphology and Syntax of Natural Languages



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University



Daniel Zeman

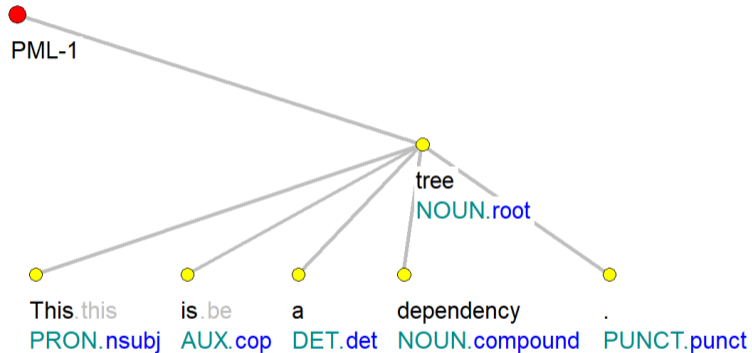
zeman@ufal.mff.cuni.cz

<https://universaldependencies.org/>

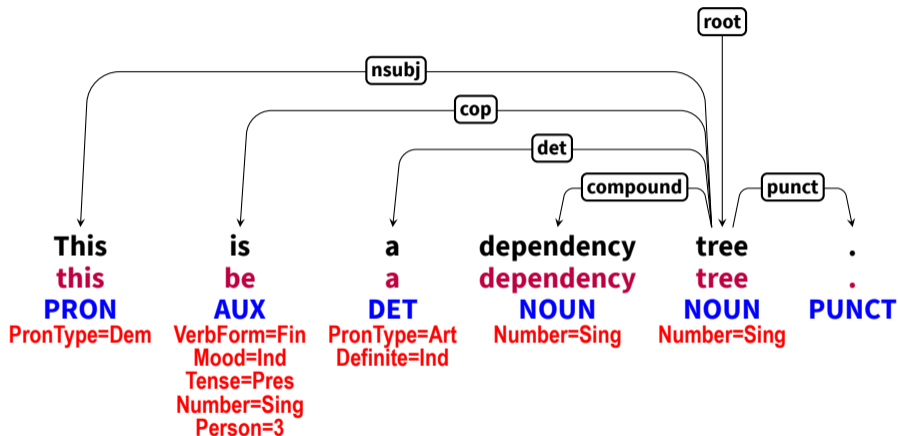
Universal Dependencies



Syntactic Trees



Syntactic Trees



Users (Simplified)



LT engineers

Users (Simplified)



LT engineers



& linguists

Outline

- 1 Introduction to Universal Dependencies
- 2 Parsing UD, shared tasks
- 3 Use in digital humanities
- 4 Use in language technology
- 5 Closer to meaning: Enhanced and Deep UD

Universal Dependencies

- Set of annotation guidelines (+ underlying theory)
- Set of annotated corpora (treebanks)

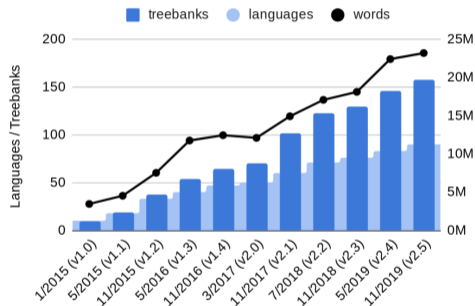
Universal Dependencies

- Set of annotation guidelines (+ underlying theory)
- Set of annotated corpora (treebanks)

- Same things annotated same way across languages...
- ... while highlighting different **coding strategies**

Universal Dependencies: History

- 2014 June: kick-off meeting
- 2015 January: data release 1.0 (10 treebanks, 10 languages)
- 2015 May: release 1.1; since then, **new release every half-a-year**
- 2016 December: v2 guidelines
- 2017 May: first UD workshop in Gothenburg
- 2017 July: first CoNLL shared task in UD parsing
- 2017 November: data release 2.1 (102 treebanks, 60 languages)
- 2020 July: IWPT shared task in Enhanced UD parsing
- 2021 May: release 2.8 (202 treebanks, 114 languages)



Design Principles

- Dependency
 - ▶ Widely used in practical NLP systems
 - ▶ Available in treebanks for many languages

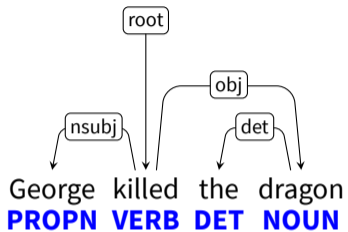
Design Principles

- Dependency
 - ▶ Widely used in practical NLP systems
 - ▶ Available in treebanks for many languages
- Lexicalism
 - ▶ Basic annotation units are words
 - ★ But what is a word? – syntactic words
 - ▶ Words have morphological properties
 - ▶ Words enter into syntactic relations

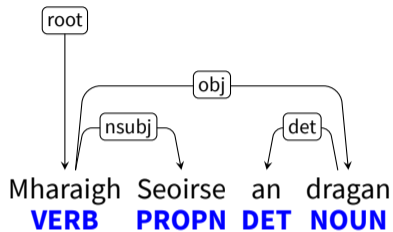
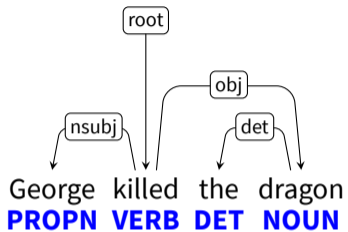
Design Principles

- Dependency
 - ▶ Widely used in practical NLP systems
 - ▶ Available in treebanks for many languages
- Lexicalism
 - ▶ Basic annotation units are words
 - ★ But what is a word? – **syntactic words**
 - ▶ Words have morphological properties
 - ▶ Words enter into syntactic relations
- Recoverability
 - ▶ Transparent mapping from input text to word segmentation

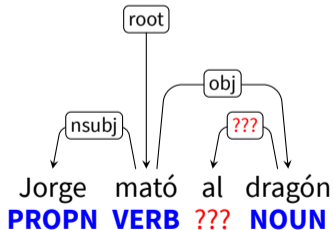
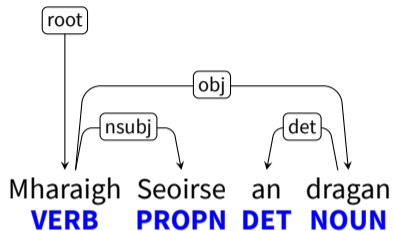
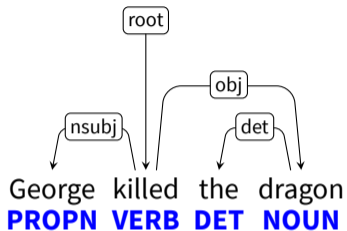
Same Thing Same Way



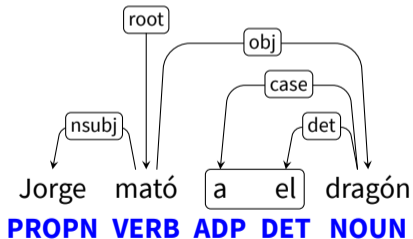
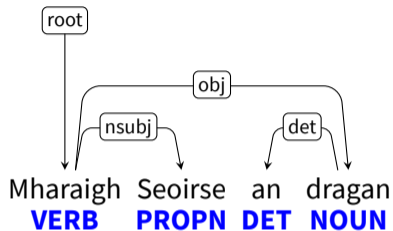
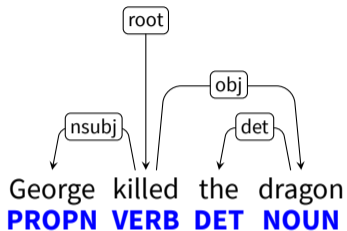
Same Thing Same Way



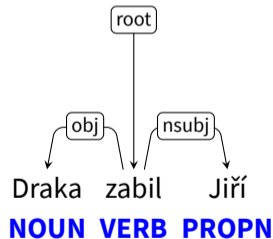
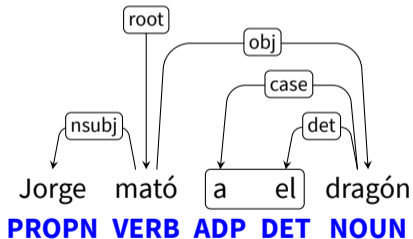
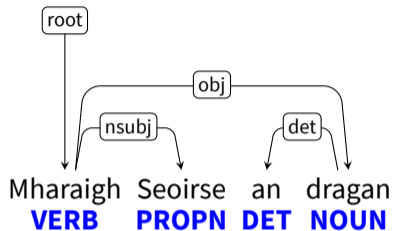
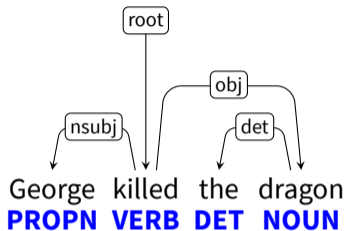
Same Thing Same Way



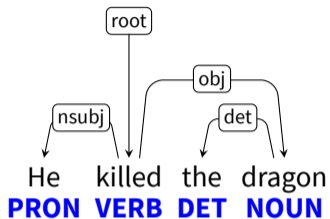
Same Thing Same Way



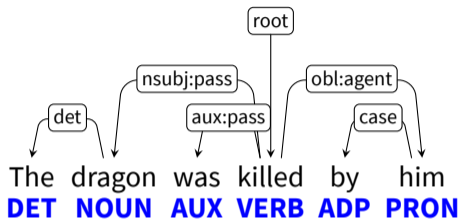
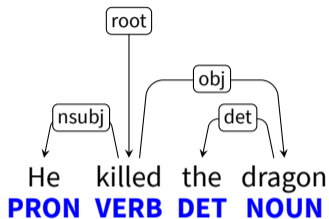
Same Thing Same Way



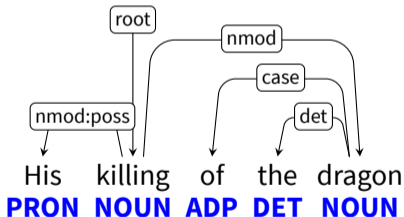
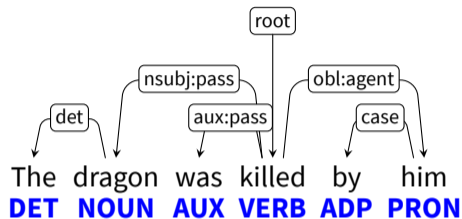
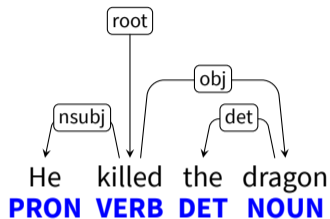
Same Meaning \neq Same Construction!



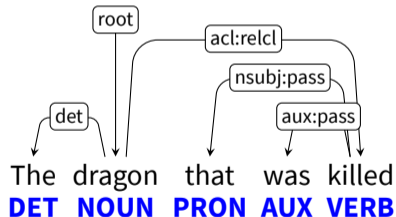
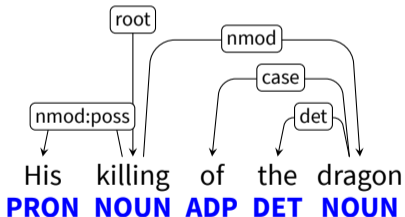
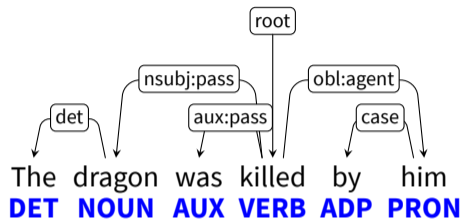
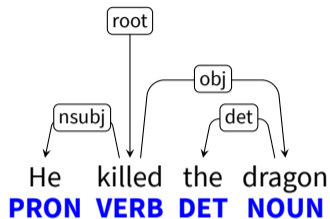
Same Meaning \neq Same Construction!



Same Meaning \neq Same Construction!



Same Meaning \neq Same Construction!



A Tour through UD

Morphology

Některé dívky si nicméně pochvalovaly zmrzlinu .
Some girls nevertheless praised ice-cream .

Morphology

Některé	dívky	si	nicméně	pochvalovaly	zmrzlinu	.
<i>Some</i>	<i>girls</i>		<i>nevertheless</i>	<i>praised</i>	<i>ice-cream</i>	.
některý	dívka	se	nicméně	pochvalovat	zmrzlina	.

- Lemma representing the semantic content of the word

Morphology

Některé	dívky	si	nicméně	pochvalovaly	zmrzlinu	.
<i>Some</i>	<i>girls</i>		<i>nevertheless</i>	<i>praised</i>	<i>ice-cream</i>	.
některý	dívka	se	nicméně	pochvalovat	zmrzlina	.
DET	NOUN	PRON	CCONJ	VERB	NOUN	PUNCT

- Lemma representing the semantic content of the word
- Part-of-speech tag representing the abstract lexical category associated with the word

Morphology

Některé	dívky	si	nicméně	pochvalovaly	zmrzlinu	.
<i>Some</i>	<i>girls</i>		<i>nevertheless</i>	<i>praised</i>	<i>ice-cream</i>	.
některý	dívka	se	nicméně	pochvalovat	zmrzlina	.
DET	NOUN	PRON	CCONJ	VERB	NOUN	PUNCT
PronType=Ind Gender=Fem Number=Plur Case=Nom	Gender=Fem Number=Plur Case=Nom	PronType=Prs Reflex=Yes Case=Dat		VerbForm=Part Tense=Past Voice=Act Aspect=Imp Gender=Fem Number=Plur	Gender=Fem Number=Sing Case=Acc	

- Lemma representing the semantic content of the word
- Part-of-speech tag representing the abstract lexical category associated with the word
- Features representing lexical and grammatical properties associated with the lemma or the particular word form

Part-of-Speech Tags

Open		Closed		Other	
NOUN	common noun	PRON	pronoun	PUNCT	punctuation
PROPN	proper noun	DET	determiner	SYM	symbol
VERB	verb	AUX	auxiliary	X	unknown
ADJ	adjective	NUM	numeral		
ADV	adverb	ADP	adposition		
INTJ	interjection	SCONJ	subordinator		
		CCONJ	coordinator		
		PART	particle		

- Taxonomy of 17 universal POS tags
- All languages use the same inventory
 - ▶ Not all tags have to be used by all languages
 - ▶ Need extensions? Use features!

Features

Lexical

PronType

NumType

Poss(essive)

Reflex(ive)

Foreign

Abbr

Typo

Nominal

Gender

Animacy

NounClass

Number

Case

Definite(ness)

Degree

Pronominal

Person

Clusivity

Polite(ness)

Verbal

VerbForm

Mood

Tense

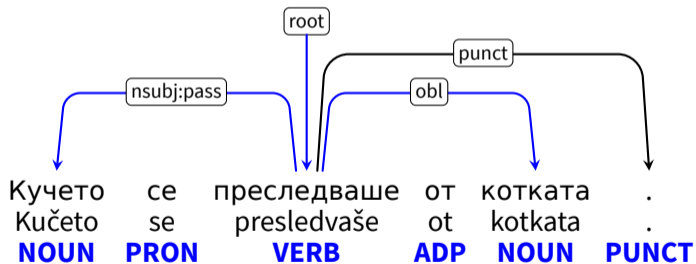
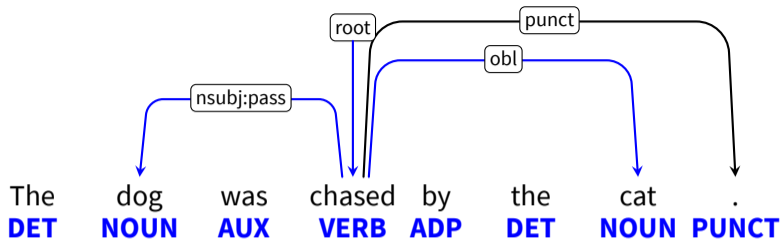
Aspect

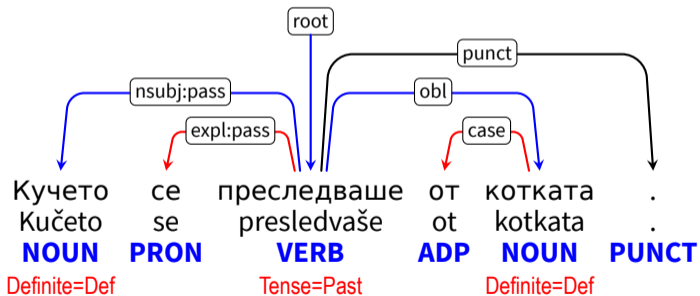
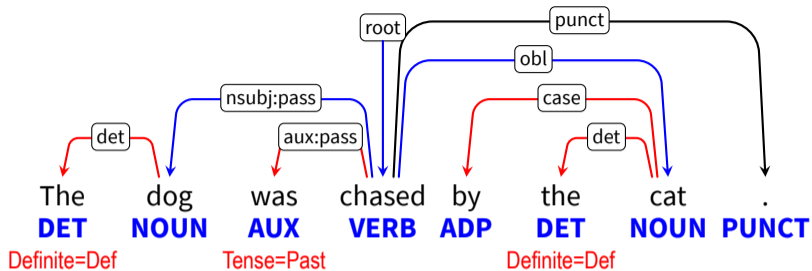
Voice

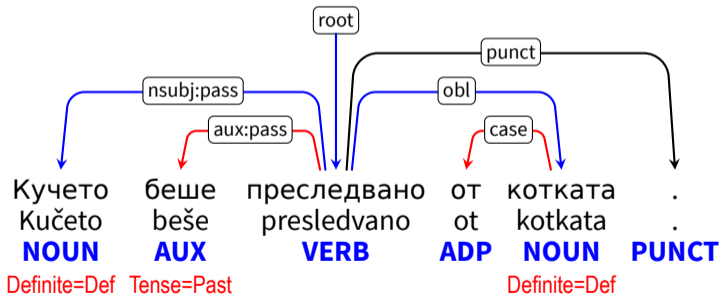
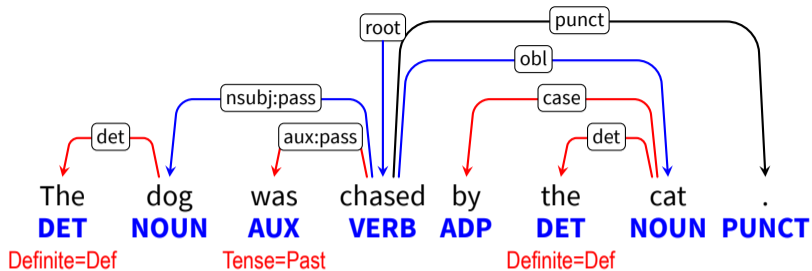
Evident(iality)

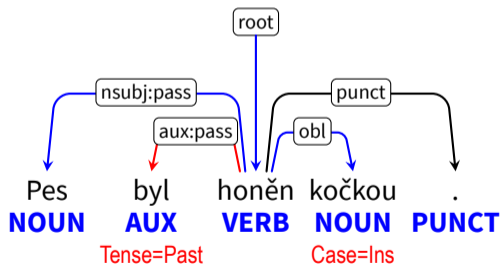
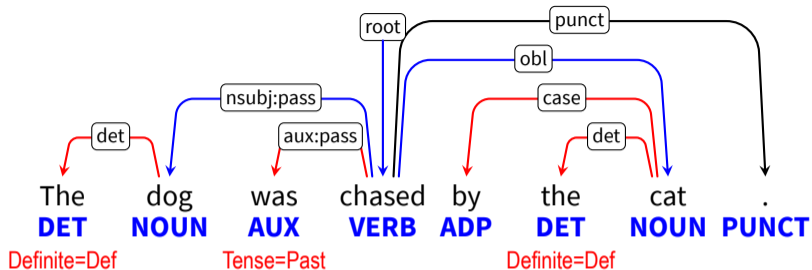
Polarity

- 24 features, each with a number of possible *values*
- Languages select relevant features
- May add language-specific features or values









Basic Universal Dependencies: 114 (112) Languages and Growing

I.-E.: 🇦🇲 Armenian West+East, 🇬🇷 Ancient Greek, Greek, 🇦🇱 Albanian, 🇧🇷 Breton, 🇮🇪 Irish, 🇮🇲 Manx, 🇸🇬 Scottish, 🇨🇾 Welsh, 🇿🇦 Afrikaans, 🇩🇰 Danish, 🇳🇱 Dutch, 🇬🇧 English, 🇫🇷 Faroese, 🇫🇷 Frisian, 🇩🇪 German, 🇬🇴 Gothic, 🇮🇸 Icelandic, 🇩🇪 Low Saxon, 🇳🇴 Norwegian, 🇸🇪 Swedish, 🇨🇭 Swiss German, 🇪🇸 Catalan, 🇫🇷 French, 🇬🇱 Galician, 🇮🇹 Italian, 🇱🇹 Latin, 🇫🇷 Old French, 🇵🇹 Portuguese, 🇷🇴 Romanian, 🇪🇸 Spanish, 🇧🇪 Belarusian, 🇧🇬 Bulgarian, 🇷🇺 Church Slavonic, 🇦🇷 Croatian, 🇨🇪 Czech, 🇷🇺 Old Russian, 🇵🇱 Polish, 🇷🇺 Russian, 🇷🇸 Serbian, 🇸🇰 Slovak, 🇸🇮 Slovenian, 🇺🇦 Ukrainian, 🇸🇮 Upper Sorbian, 🇱🇻 Latvian, 🇱🇹 Lithuanian, 🇰🇷 Kurmanji, 🇮🇷 Persian, Khunsari, Nayini, Soi, 🇮🇳 Hindi, Kangri, Bhojpuri, Marathi, Sanskrit, 🇵🇰 Urdu; **Uralic:** 🇮🇷 Erzya, 🇪🇪 Estonian, 🇫🇮 Finnish, 🇮🇷 Hungarian, 🇰🇷 Karelian, Livvi, 🇸🇮 Komi Permyak+Zyrian, 🇮🇷 Moksha, 🇸🇲 Sámi North+Skolt; **Dravid.:** 🇮🇳 Tamil, Telugu; **Turkic:** 🇰🇿 Kazakh, 🇹🇷 Old Turkish, 🇹🇷 Turkish, 🇺🇾 Uyghur; **Af.-As.:** 🇸🇰 Akkadian, 🇪🇲 Amharic, 🇸🇰 Arabic Modern+Levant, 🇸🇰 Assyrian, 🇸🇰 Beja, 🇸🇰 Coptic, 🇮🇱 Hebrew, 🇲🇹 Maltese; **Sino-Tib.:** 🇨🇳 Cantonese, 🇨🇳 Classical Chinese, 🇨🇳 Chinese; **Tai-Kadai:** 🇹🇭 Thai; **Aus.-As.:** 🇻🇳 Vietnamese; **Austrones.:** 🇮🇳 Indonesian, 🇵🇭 Tagalog; **Tupian:** 🇧🇷 Mundurukú, Akuntsú, Makuráp, Guajajára, Kaapor, Tupinambá, 🇮🇳 Mbyá; **Other:** 🇲🇰 Buryat, 🇯🇵 Japanese, 🇰🇷 Korean, 🇵🇭 Chukchi, 🇺🇾 Yupik, 🇧🇪 Basque, 🇸🇪 Sw. Sign, 🇮🇳 Naija, 🇸🇰 Bambara, 🇸🇰 Wolof, 🇮🇳 Yoruba, 🇸🇰 Warlpiri, 🇸🇰 K'iche', 🇧🇷 Apurinã

Parsing UD

Parsing into Universal Dependencies

- Machine learning:
 - ▶ Train a model on the training part of a UD treebank
 - ▶ Apply it to new data \Rightarrow obtain annotations
 - ▶ **End-to-end:** from raw text to ...
 - ★ Sentence segmentation
 - ★ Tokenization (word segmentation)
 - ★ Morphology (lemmas, POS tags, features)
 - ★ Syntax (relations between words)

Parsing into Universal Dependencies

- Machine learning:

- ▶ Train a model on the training part of a UD treebank
- ▶ Apply it to new data ⇒ obtain annotations
- ▶ **End-to-end:** from raw text to ...
 - ★ Sentence segmentation
 - ★ Tokenization (word segmentation)
 - ★ Morphology (lemmas, POS tags, features)
 - ★ Syntax (relations between words)

- Evaluation: How good is it?


- ▶ Align system-produced and gold-standard words
- ▶ For the aligned pairs:
 - ★ Percentage of correctly assigned lemmas, POS tags, feature values
 - ★ Tree structure: correctly assigned parent node + relation type
 - ★ Combined score (morpho+syntax): **MLAS**



Shared Tasks

- Long tradition in natural language processing
- Big tasks in UD parsing: CoNLL 2017 and 2018
 - ▶ <http://universaldependencies.org/conll18/results.html>
- UDPipe-Future on 🇨🇪 Czech PDT: MLAS = 85.10


- Stanford on 🇸🇰 Slovak SNK: MLAS = 75.01
- Stanford on 🇬🇧 English EWT: MLAS = 76.33



Shared Tasks

- Long tradition in natural language processing
- Big tasks in UD parsing: CoNLL 2017 and 2018
 - ▶ <http://universaldependencies.org/conll18/results.html>
- UDPipe-Future on  Czech PDT: MLAS = 85.10
 - ▶ Sentence segmentation: F = 93.41
 - ▶ Word segmentation: F = 99.93


- Stanford on  Slovak SNK: MLAS = 75.01
- Stanford on  English EWT: MLAS = 76.33



Shared Tasks

- Long tradition in natural language processing
- Big tasks in UD parsing: CoNLL 2017 and 2018
 - ▶ <http://universaldependencies.org/conll18/results.html>
- UDPipe-Future on  Czech PDT: MLAS = 85.10
 - ▶ Sentence segmentation: F = 93.41
 - ▶ Word segmentation: F = 99.93
 - ▶ Lemmatization: F = 98.71




- Stanford on  Slovak SNK: MLAS = 75.01
- Stanford on  English EWT: MLAS = 76.33

Shared Tasks




- Long tradition in natural language processing
- Big tasks in UD parsing: CoNLL 2017 and 2018
 - ▶ <http://universaldependencies.org/conll18/results.html>
- UDPipe-Future on  Czech PDT: MLAS = 85.10
 - ▶ Sentence segmentation: F = 93.41
 - ▶ Word segmentation: F = 99.93
 - ▶ Lemmatization: F = 98.71
 - ▶ POS tagging: F = 99.01

- Stanford on  Slovak SNK: MLAS = 75.01
- Stanford on  English EWT: MLAS = 76.33

Shared Tasks

- Long tradition in natural language processing
- Big tasks in UD parsing: CoNLL 2017 and 2018
 - ▶ <http://universaldependencies.org/conll18/results.html>
- UDPipe-Future on  Czech PDT: MLAS = 85.10
 - ▶ Sentence segmentation: F = 93.41
 - ▶ Word segmentation: F = 99.93
 - ▶ Lemmatization: F = 98.71
 - ▶ POS tagging: F = 99.01
 - ▶ Feature values: F = 96.85
- Stanford on  Slovak SNK: MLAS = 75.01
- Stanford on  English EWT: MLAS = 76.33

Shared Tasks

- Long tradition in natural language processing
- Big tasks in UD parsing: CoNLL 2017 and 2018
 - ▶ <http://universaldependencies.org/conll18/results.html>
- UDPipe-Future on  Czech PDT: MLAS = 85.10
 - ▶ Sentence segmentation: F = 93.41
 - ▶ Word segmentation: F = 99.93
 - ▶ Lemmatization: F = 98.71
 - ▶ POS tagging: F = 99.01
 - ▶ Feature values: F = 96.85
 - ▶ Labeled dependency relations: F = 90.32
- Stanford on  Slovak SNK: MLAS = 75.01
- Stanford on  English EWT: MLAS = 76.33

Off-the-Shelf Tools

- UDPipe (<https://lindat.mff.cuni.cz/services/udpipe/>)
 - ▶ <https://ufal.mff.cuni.cz/udpipe>
- Stanza (<https://stanfordnlp.github.io/stanza/>)

Use in Digital Humanities



Linguists Can Search Treebanks

<https://lindat.mff.cuni.cz/services/pmltq/>

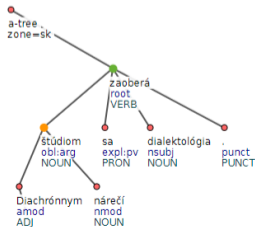
Relations Node Types Attributes Operators Functions

```
a-node $v := [  
  tag="VERB",  
  child a-node $o := [deprel="obl:arg", iset/case="ins", 0x child a-node [deprel="case"]]  
];
```

Execute query w/o Filters Suggest (0)

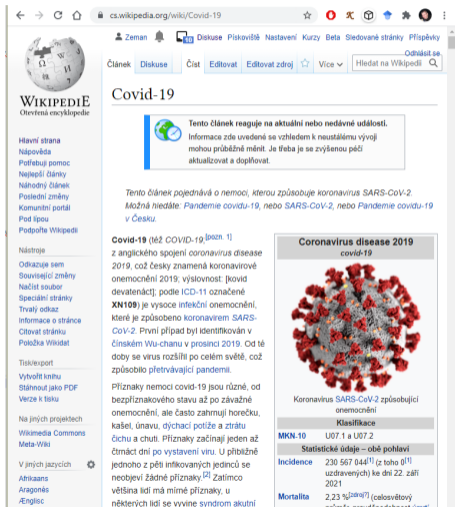
Result: 3 / 100

[sk] Diachrónnym a synchrónnym štúdiom nárečí sa zaoberá dialektológia.

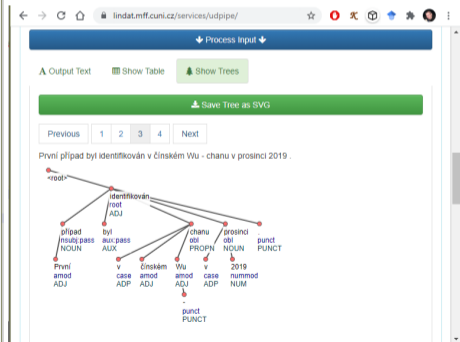


Linguists Can Parse and Search New Data

<https://lindat.mff.cuni.cz/services/udpipe/>



The screenshot shows the Wikipedia page for "Covid-19" in Czech. The page title is "Covid-19". A warning box at the top states: "Tento článek reaguje na aktuální nebo nedávné události. Informace zde uvedené se vzhledem k neustálému vývoji mohou průběžně měnit. Je třeba je se zvýšenou péčí aktualizovat a doplňovat." Below this, a text box says: "Tento článek pojednává o nemoci, kterou způsobuje koronavirus SARS-CoV-2. Možná hledáte: Pandemie covidu-19, nebo SARS-CoV-2, nebo Pandemie covidu-19 v Česku." The main text begins: "Covid-19 (též COVID-19^[pozn. 1] z anglického spojení coronavirus disease 2019, což česky znamená koronavirové onemocnění 2019; výslovnost: [kovid devatenáct]; podle ICD-11 označené **XN109**) je vysoce infekční onemocnění, které je způsobeno koronavirem SARS-CoV-2. První případ byl identifikován v čínském Wu-chanu v prosinci 2019. Od té doby se virus rozšířil po celém světě, což způsobilo přetrvávající pandemii." A sidebar on the right contains a table for "Coronavirus disease 2019 covid-19" with a 3D model of the virus. The table includes classification (Klasifikace), incidence (230 567 044[1] z toho 0[1] uzdravených) ke dni 22. září 2021, and mortality (2,23 %^[zdroj?]) (celosvětový průměr pravděpodobnost úmrtí).



The screenshot shows the UDPIPE web interface. At the top, there is a "Process Input" button. Below it, there are buttons for "Output Text", "Show Table", and "Show Trees". A "Save Tree as SVG" button is also present. A navigation bar shows "Previous", "1", "2", "3", "4", and "Next". The main content area displays the sentence: "První případ byl identifikován v čínském Wu-chanu v prosinci 2019." Below the sentence is a parse tree diagram. The root node is "<root>". The tree structure is as follows: root (PUNCT) branches into "identifikován" (ADJ) and "prosince" (PUNCT). "identifikován" branches into "případ" (NOUN) and "byl" (AUX). "případ" branches into "První" (ADJ) and "insult" (PASS). "byl" branches into "sus" (PASS) and "v" (ADP). "v" branches into "case" (ADP) and "čínském" (ADJ). "čínském" branches into "Wu" (ADJ) and "chane" (ADJ). "Wu" branches into "amod" (ADJ) and "v" (ADP). "chane" branches into "case" (ADP) and "2019" (NUM). "prosince" branches into "obl" (NOUN) and "2019" (NUM). "obl" branches into "punct" (PUNCT) and "2019" (NUM).

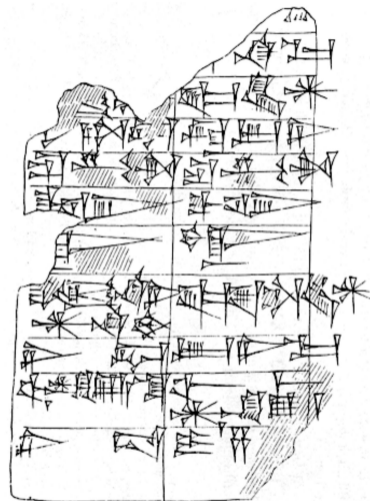
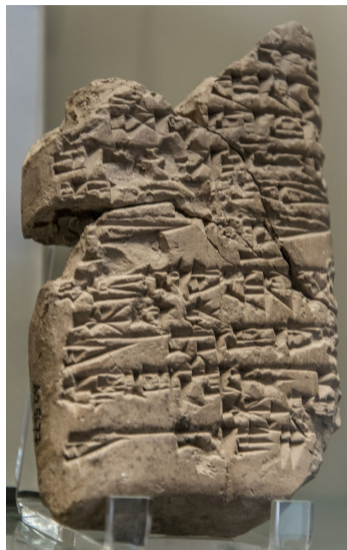
Language Learning

- Check grammar usage in the corpus
- Learner corpora

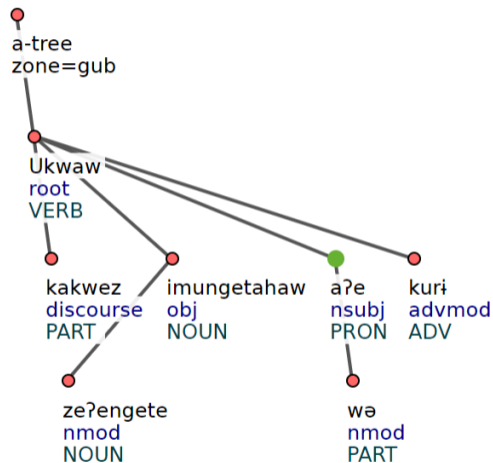


Historical Linguistics, Classical Languages

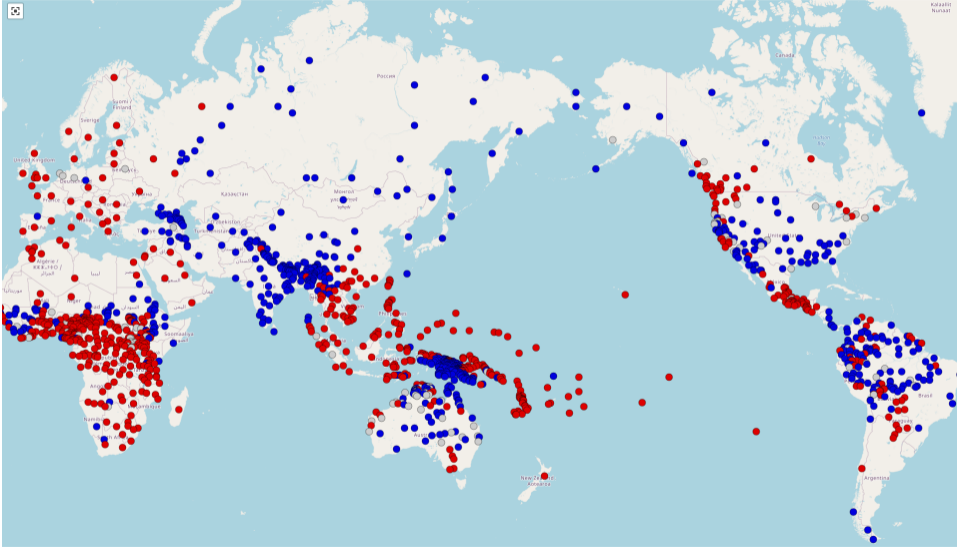
- Old Turkish
- Classical Chinese
- Sanskrit
- Akkadian
- Coptic
- Ancient Greek
- Latin
- Old French
- Gothic
- Old Church Slavonic
- Old East Slavic



Documentation of Endangered Languages



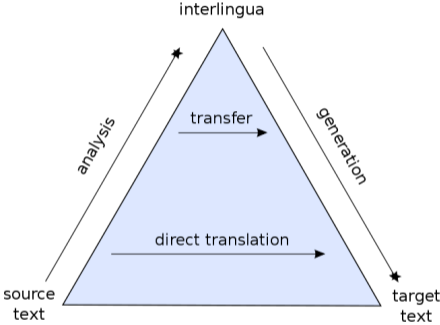
Linguistic Typology



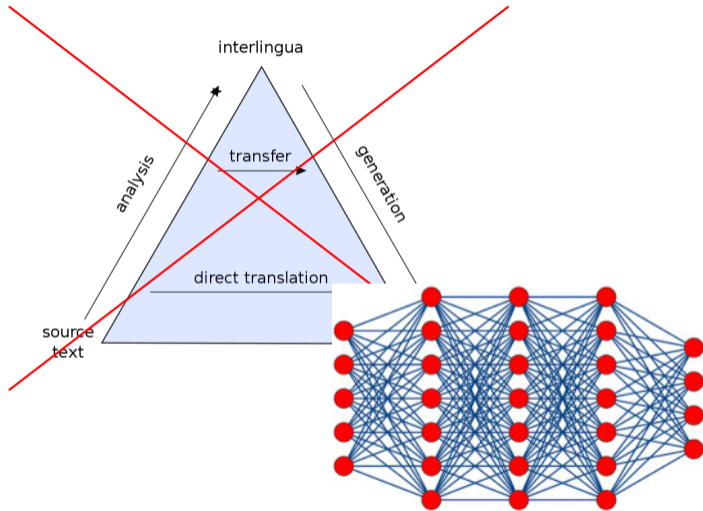
Use in Language Technology



Use in Language Technology?



Use in Language Technology?



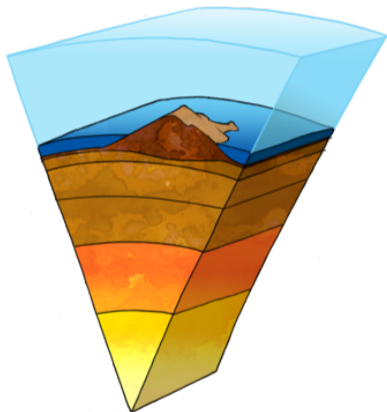
Use in Language Technology

- But neural networks are data-hungry
- We don't always have enough data tailored for a particular task

- Use a neural parser to get the trees
- Then use rules/heuristics to extract information from the trees
 - ▶ *Who did what to whom?*

Closer to Meaning: Enhanced and Deep UD

Multiple Layers of Dependencies



Form

- Surface syntax
- Deep syntax
- Semantics

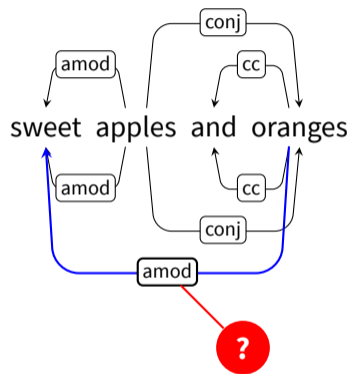
Meaning

Enhanced Universal Dependencies

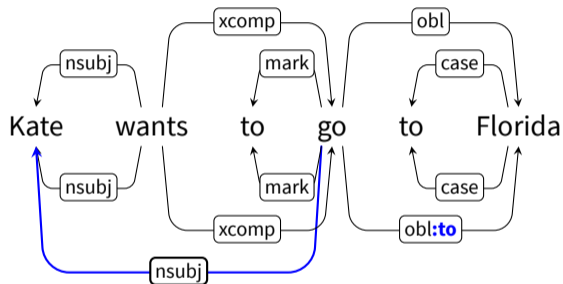
Currently available (at least partially) for 18 languages:

 Arabic,  Belarusian,  Bulgarian,  Chukchi,  Czech,  Dutch,  English,  Estonian,  Finnish,  Italian,  Latvian,  Lithuanian,  Polish,  Russian,  Slovak,  Swedish,  Tamil,  Ukrainian

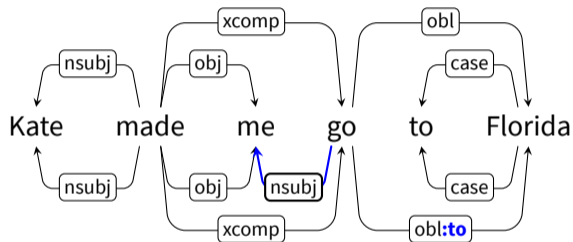
Enhanced UD: Shared Dependent of Coordination



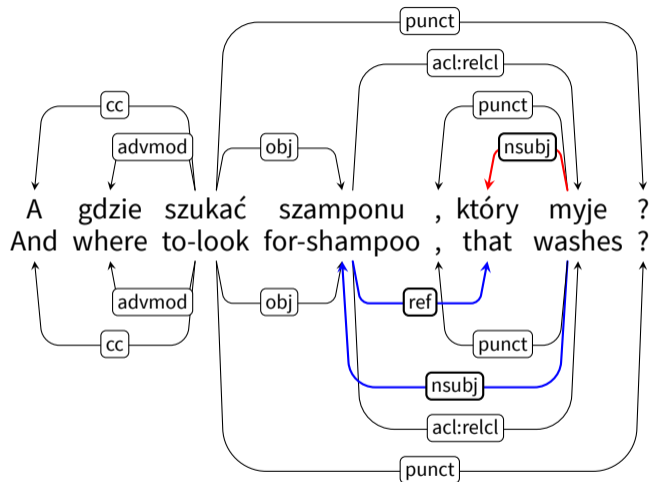
Enhanced UD: External Subject of Controlled Predicate



Enhanced UD: External Subject in Object-Control Construction

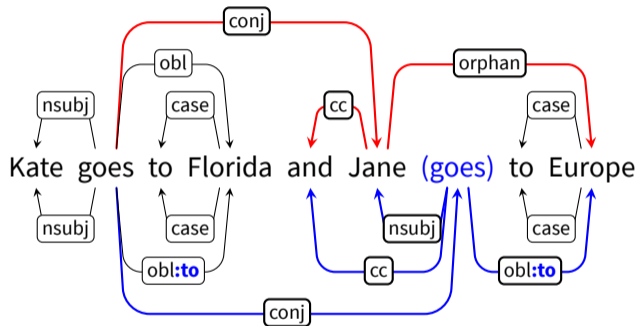


Enhanced UD: Relative Clauses

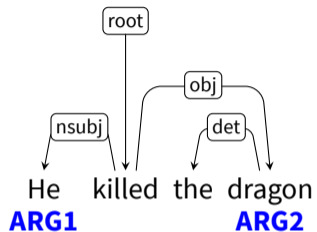


“And where to look for shampoo that works?”

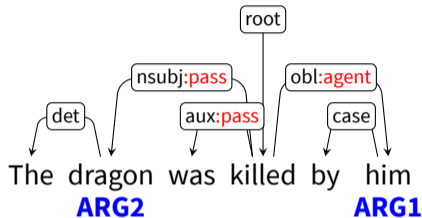
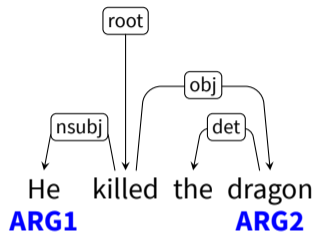
Enhanced UD: Gapping and Stripping



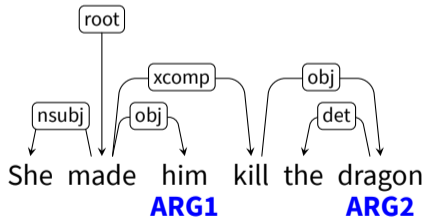
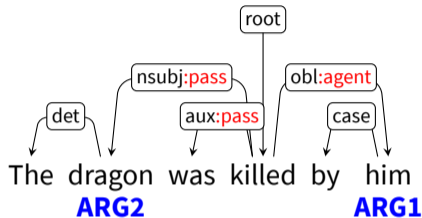
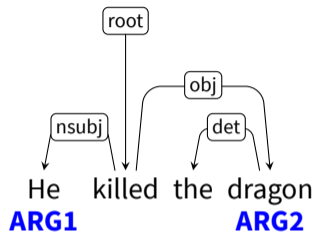
Deep UD: Normalization of Syntactic Alternations



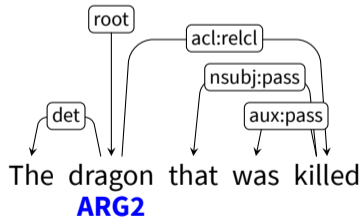
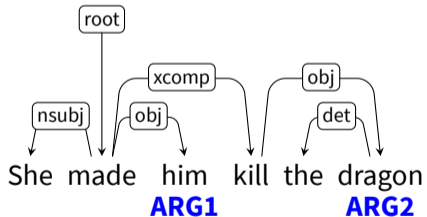
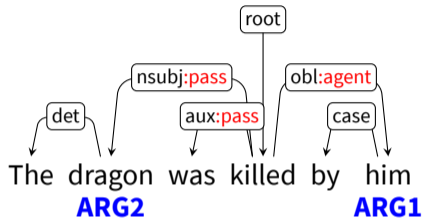
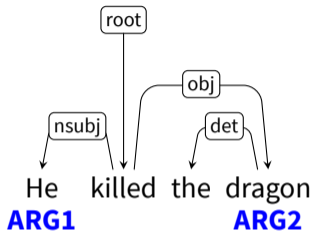
Deep UD: Normalization of Syntactic Alternations



Deep UD: Normalization of Syntactic Alternations



Deep UD: Normalization of Syntactic Alternations



Summary



- Over the last 7 years, UD has become very popular in computational linguistics
- Annotated data now exist for languages for which it was unthinkable before

Summary



- Over the last 7 years, UD has become very popular in computational linguistics
- Annotated data now exist for languages for which it was unthinkable before
- Common annotation scheme for typologically diverse languages
- Easier processing and searching
- Cross-linguistic studies possible

Summary



- Over the last 7 years, UD has become very popular in computational linguistics
- Annotated data now exist for languages for which it was unthinkable before
- Common annotation scheme for typologically diverse languages
- Easier processing and searching
- Cross-linguistic studies possible
- Enhanced UD: additional implicit relations made explicit
- Closer to meaning \Rightarrow easier to discover **who did what to whom**

455 Contributors

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepfli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielė Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aponova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Bilge Nas Arıcan, Póruan Arnadóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Deniz Baran Aslan, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginia Barbu Mititelu, Starkaður Barkaron, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Kepa Bengoetxea, Gözde Berk, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristin Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaa, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Lauren Cassidy, Tatiana Cavalcanti, Gülşen Cebreroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čeplo, Neslihan Cesur, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Shweta Chauhan, Ethan Chi, Taishi Chika, Yongseok Cho, Jinho Choi, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Mihaela Cristescu, Philemon. Daniel, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilaraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wognaire Evelyn, Sidney Facundes, Richárd Farkas, Marília Fernanda, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Fabrício Ferraz Gerardi, Kim Gerdes, Filip Ginter, Gustavo Godoy, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grióni, Loic Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinnsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barбора Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Eva Huber, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Olájdíde Ishola, Kaoru Ito, Tomáš Jelínek, Apoorva Jha, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Sarveswaran K, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Neslihan Kara, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Abdullatif Köksal, Kamil Kopaciewicz, Timo Korhikangas, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Kirayathu, Oğuzhan Kuyrukcu, Aslı Kuzgun, Sookyoung Kwak, Veronika Laippala, Lucia Lam, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phương Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Cheuk Ying Li, Josie Li, Keying Li, Yuan Li, KyungTae Lim, Bruna Lima Padovani, Krister Lindén, Nikola Ljubešić, Olga Loginova, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Büşra Marşan, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Alessandro Mazzei, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Niko Miekkä, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foushani, Judit Molnár, Amiraeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskiy, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisepp, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horňiáček, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Lương Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaïdo, Vitaly Nikolaev, Rattima Nitisaraj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayò Olúòkun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Övrelid, Şaziye Betül Özafe, Merve Özçelik, Arzucan Özgür, Balkız Öztürk Başaran, Hyunji Hayley Park, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cene-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalīna, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyyssalo, Peng Qi, Oriana Rabbis, Alexandre Rademaker, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rasher, Mohammad Sadegh Rasooli, Vinit Ravishanker, Livy Real, Petru Rebeja, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Eiríkur Rögnvaldsson, Mykhailo Romanenko, Rudolf Rosa, Valentin Roşca, Davide Rovati, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Benoît Sagot, Aleksí Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Saniyar, Dage Särg, Baiba Saulite, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddadh, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibussirri, Dmitry Sichinava, Janine Siewert, Einar Frey Sigurðsson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachudubova, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Rachele Sprugnoli, Steinhör Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Aliane Suhr, Yogi Lesmana Sulestio, Umuto Sulubacak, Shingo Suzuki, Zolt Szántó, Doa Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Samson Tella, Isabelle Tellier, Marinella Testori, Guillaume Thomas, Lili Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Uřešová, Larraitz Uriá, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Rob van der Goot, Martine Vanhove, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Natalia Vlasova, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilian Wendt, Paul Widmer, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrummyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, Hanzhi Zhu, Anna Zhuravleva, Rayan Ziane

Thanks!
Ďakujem!

<https://universaldependencies.org/>