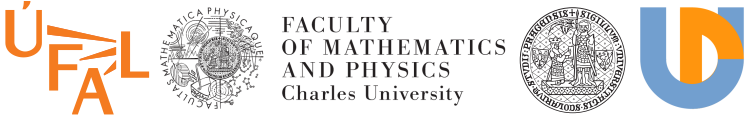
Universal Dependencies: Comparing Languages in Space and Time



Daniel Zeman

zeman@ufal.mff.cuni.cz
https://universaldependencies.org/

Outline

1. Introduction to Universal Dependencies
2. Parsing UD, shared tasks
3. Use in digital humanities
4. Parallel treebanks: Bible, PUD
5. Language change
6. (Closer to meaning: Enhanced and Deep UD)

Universal Dependencies

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

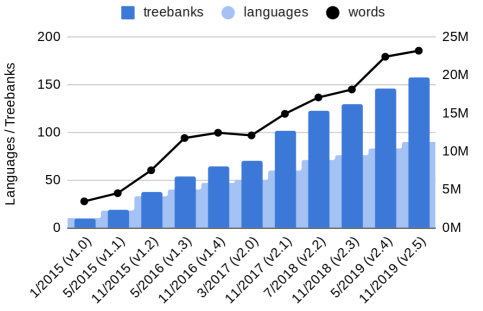
Universal Dependencies

- 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)



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

Design Principles

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

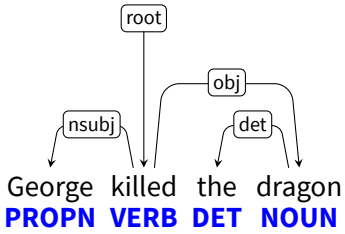
Design Principles

- 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

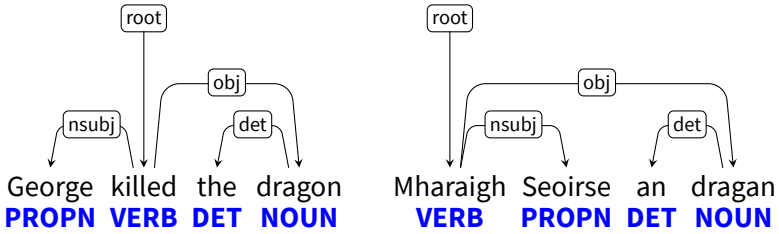
Design Principles

- 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

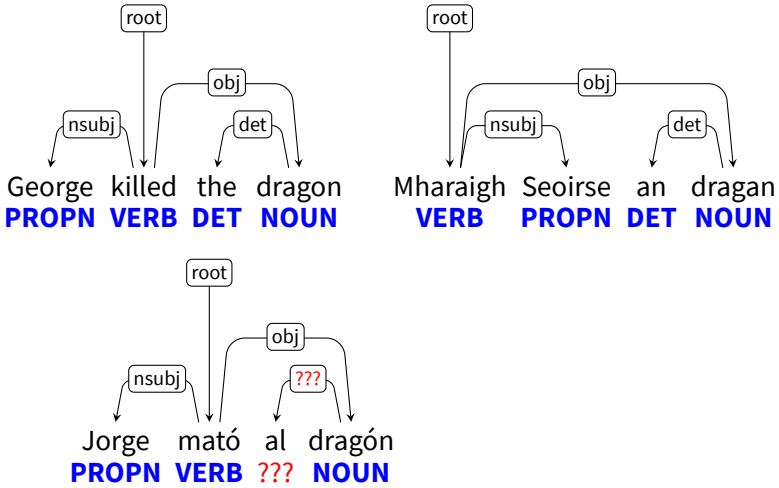
Same Thing Same Way



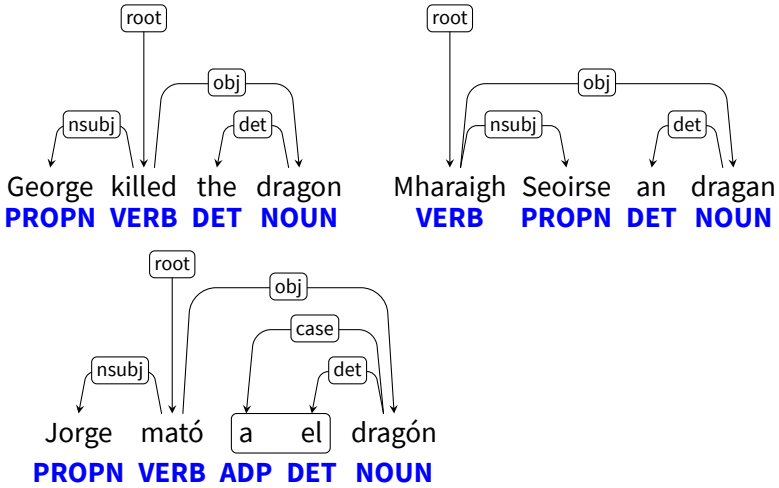
Same Thing Same Way



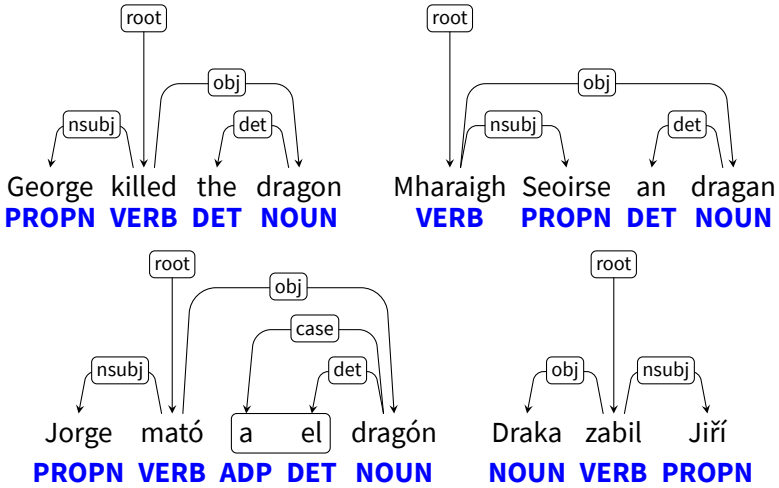
Same Thing Same Way



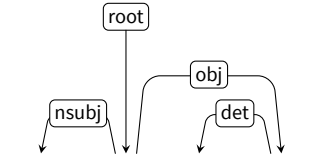
Same Thing Same Way



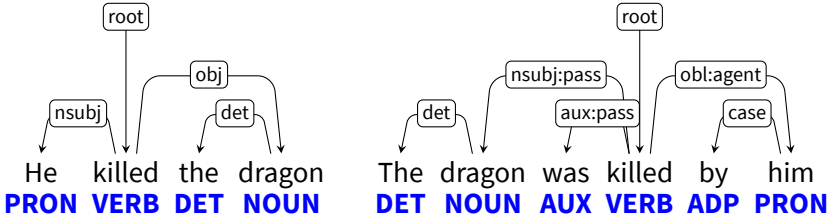
Same Thing Same Way



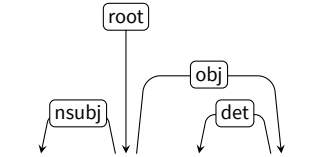
Same Meaning ≠ Same Construction!

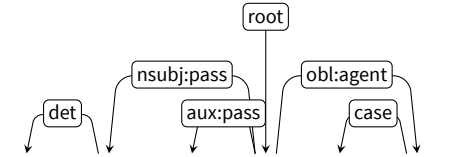


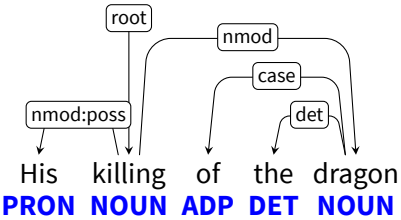
Same Meaning ≠ Same Construction!



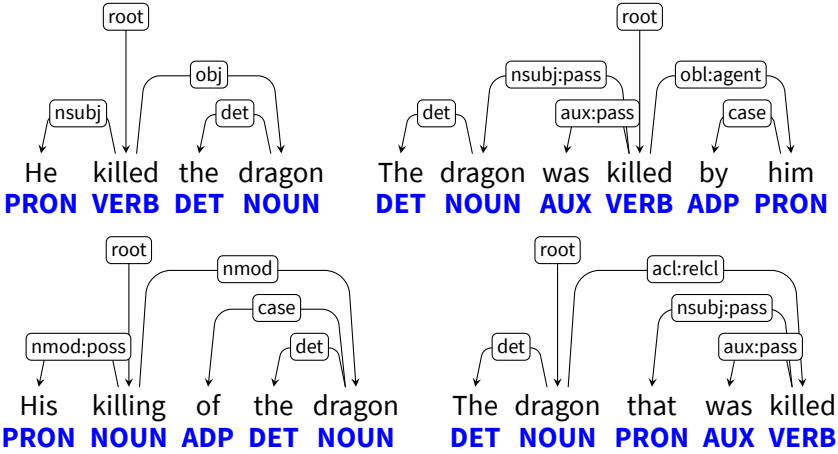
Same Meaning ≠ Same Construction!



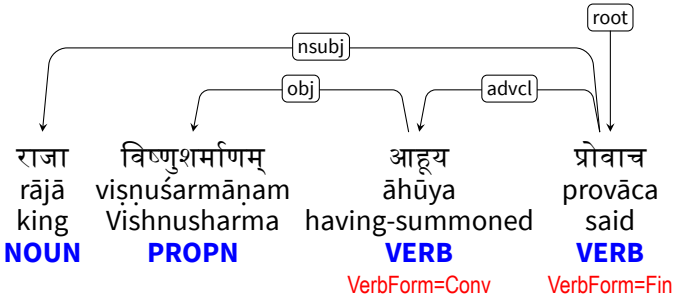




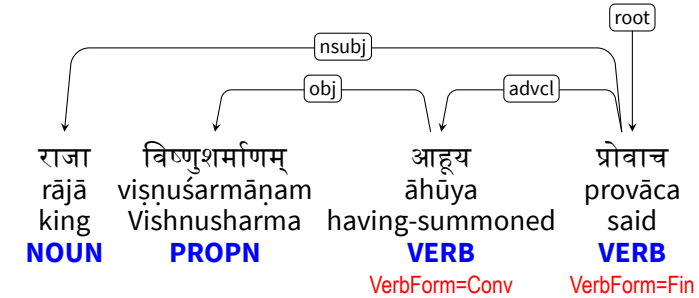
Same Meaning ≠ Same Construction!



Language-specific Preferences



Language-specific Preferences



A Tour through UD

Morphology

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

Morphology

Některé	dívky	si	nicméně	pochvalovaly	zmrzlinu	.
Some	*girls*		*nevertheless*	*praised*	*ice-cream*	*.*
některý	dívka	se	nicméně	pochvalovat	zmrzlina	.

- Lemma representing the semantic content of the word

Morphology

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

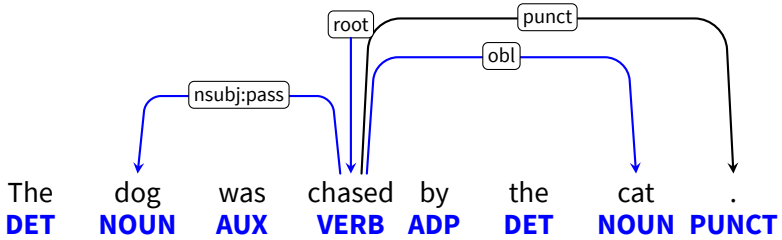
Open		Closed		Other	
NOUN	common noun	**PRON**	pronoun	**PUNCT**	punctuation
PROPN	proper noun	**DET**	determiner	**SYM**	symbol
VERB	verb	**AUX**	auxiliary	**X**	unknown
ADJ	adjective	**NUM**	numeral		
ADV	adverb	**ADP**	adposition		
INTJ	interjection	**SCONJ**	subordinator		
		CCONJ	coordinator		
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

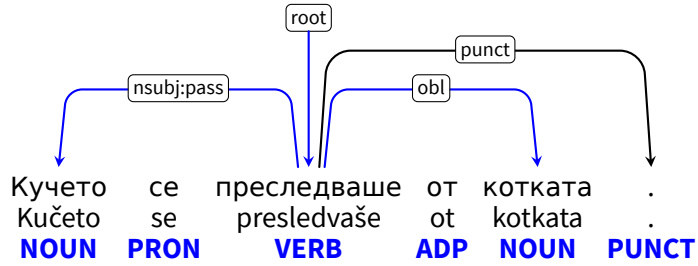
- Taxonomy of 17 universal POS tags
- All languages use the same inventory
  - Not all tags have to be used by all languages
  - Need extensions? Use features!

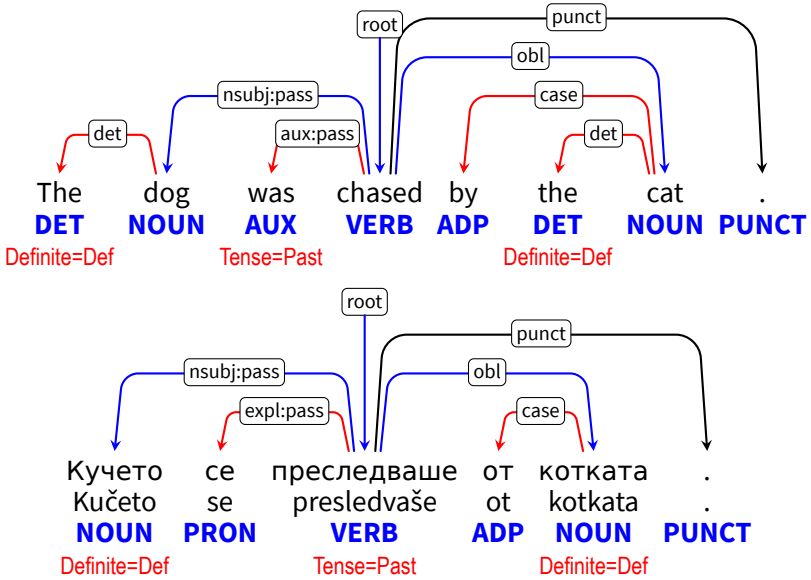
Features

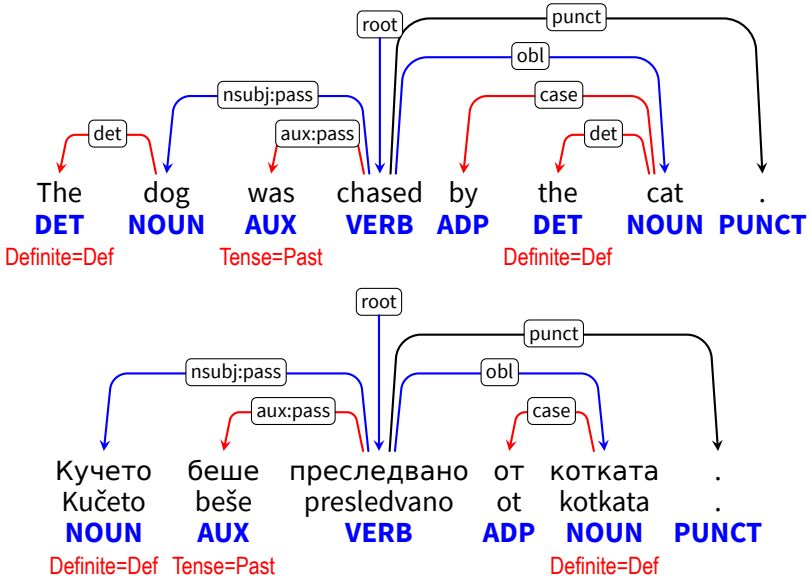
Lexical	Nominal	Pronominal	Verbal
PronType	Gender	Person	VerbForm
NumType	Animacy	Clusivity	Mood
Poss(essive)	NounClass	Polite(ness)	Tense
Reflex(ive)	Number		Aspect
Foreign	Case		Voice
Abbr	Definite(ness)		Evident(iality)
Typo	Degree		Polarity

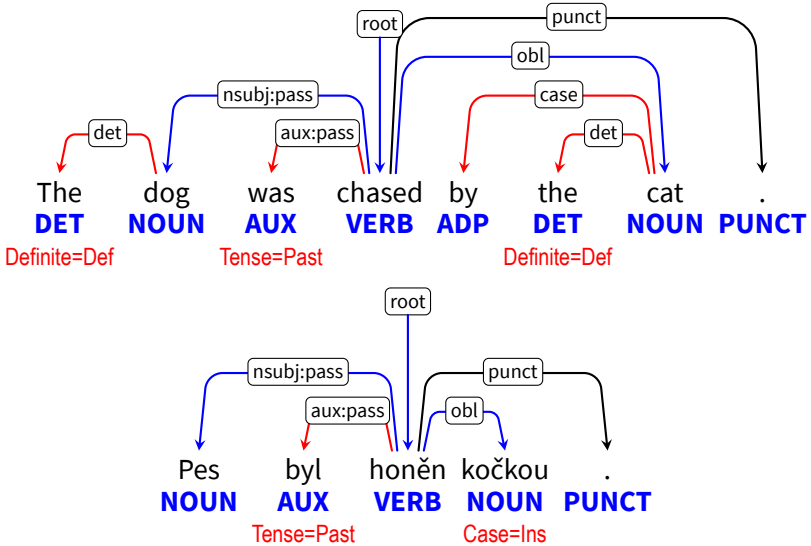
- 24 features, each with a number of possible *values*
- Languages select relevant features
- May add language-specific features or values


root
punct
obl
nsubj:pass
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT


root
punct
nsubj:pass
obl
Кучето се преследваше от котката .
Kučeto se presledvaše ot kotkata .
NOUN PRON VERB ADP NOUN PUNCT


root
punct
obl
nsubj:pass
case
det
aux:pass
det
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def
root
punct
nsubj:pass
obl
expl:pass
case
Кучето се преследваше от котката .
Kučeto se presledvaše ot kotkata .
NOUN PRON VERB ADP NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def


root
punct
obl
nsubj:pass
case
det
aux:pass
det
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def
root
punct
nsubj:pass
obl
aux:pass
case
Кучето беше преследвано от котката .
Kučeto beše presledvano ot kotkata .
NOUN AUX VERB ADP NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def


root
punct
obl
nsubj:pass
case
det
aux:pass
det
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def
root
nsubj:pass
punct
aux:pass
obl
Pes byl honěn kočkou .
NOUN AUX VERB NOUN PUNCT
Tense=Past
Case=Ins

Parsing UD

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)

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

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

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

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

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

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

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
- HIT-SCIR on Japanese GSD: MLAS = 72.62

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
- HIT-SCIR on Japanese GSD: MLAS = 72.62
  - Sentence segmentation: F = 95.01
  - Word segmentation: F = 94.53

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
- HIT-SCIR on Japanese GSD: MLAS = 72.62
  - Sentence segmentation: F = 95.01
  - Word segmentation: F = 94.53
  - Lemmatization: F = 93.78
  - POS tagging: F = 92.97
  - Feature values: F = 94.52
  - Labeled dependency relations: F = 83.11

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
- HIT-SCIR on Japanese GSD: MLAS = 72.62
  - Sentence segmentation: F = 95.01
  - Word segmentation: F = 94.53
  - Lemmatization: F = 93.78
  - POS tagging: F = 92.97
  - Feature values: F = 94.52
  - Labeled dependency relations: F = 83.11
- Stanford on English EWT: MLAS = 76.33

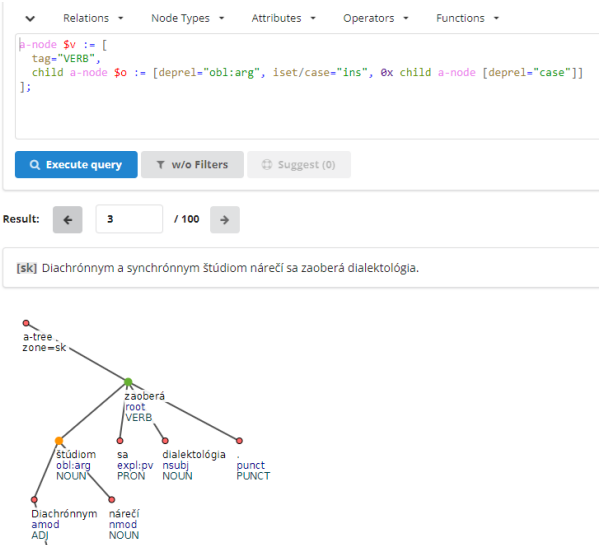
Off-the-Shelf Tools

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

Use in Digital Humanities

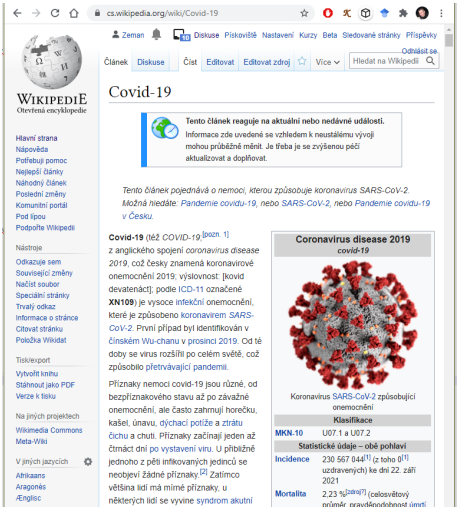
Linguists Can Search Treebanks

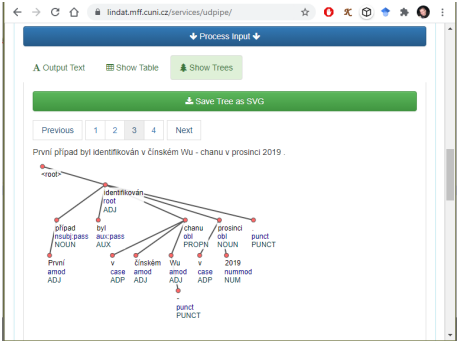
https://lindat.mff.cuni.cz/services/pmltq/



Linguists Can Parse and Search New Data

`https://lindat.mff.cuni.cz/services/udpipe/`



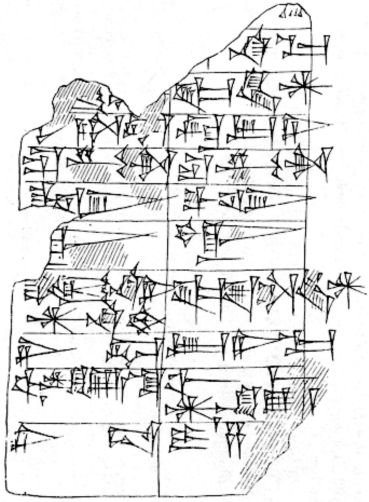


Language Learning

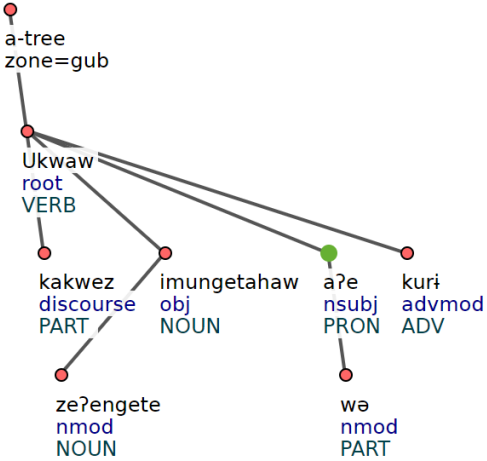
- Check grammar usage in the corpus
- Learner corpora

Historical Linguistics, Classical Languages

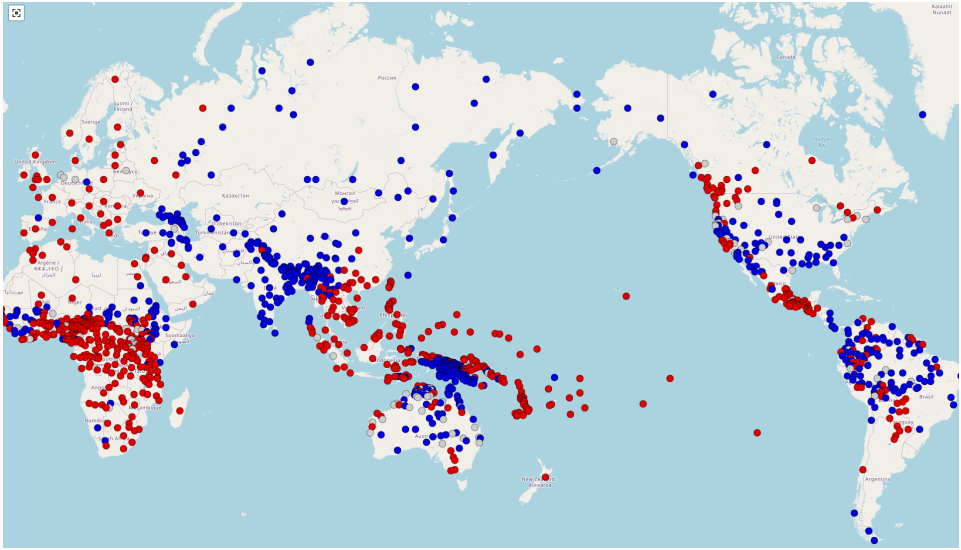
- Old Turkish
- Classical Chinese
- Sanskrit
- Akkadian
- Coptic
- Ancient Greek
- Latin
- Old French
- Gothic
- Old Church Slavonic
- Old East Slavic



Documentation of Endangered Languages



Linguistic Typology



Linguistic Typology

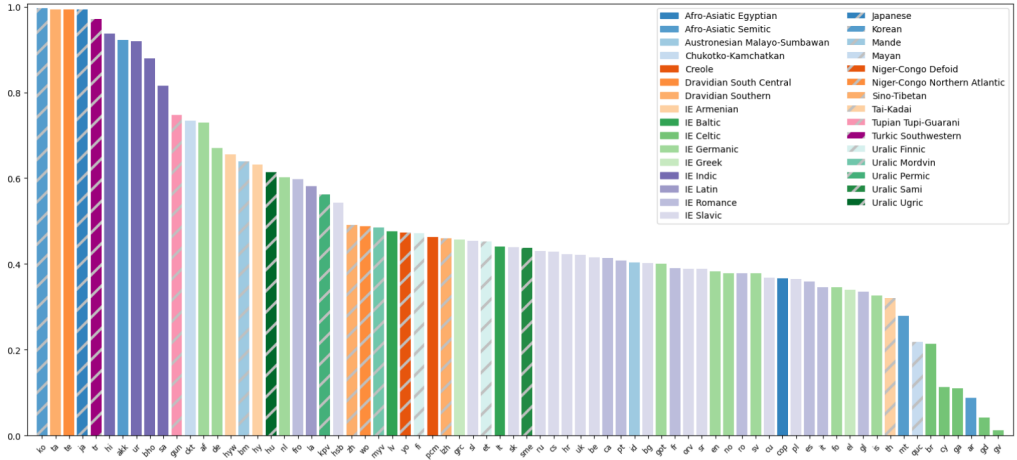
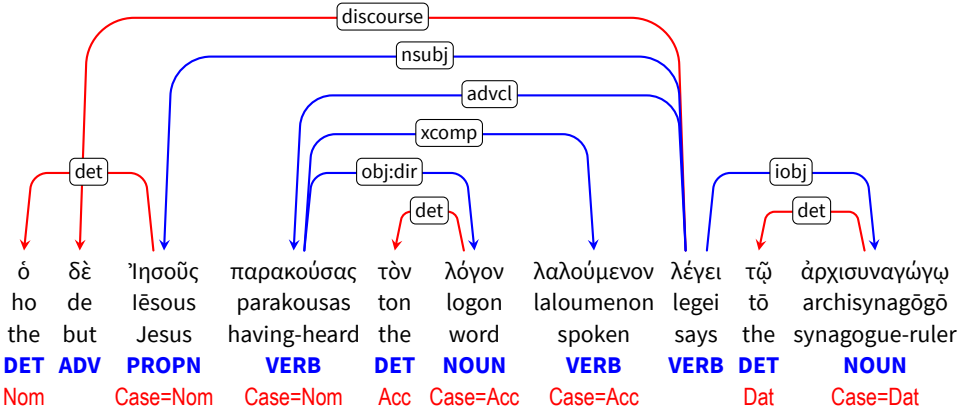


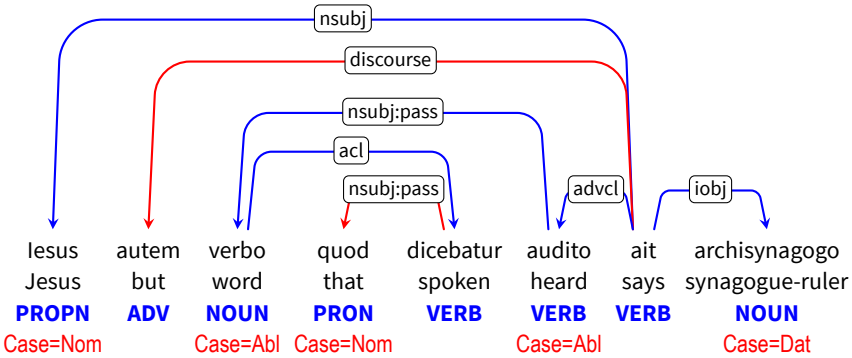
Figure 7 Percentage of head-final dependencies. Each bar is one language.

UD Ancient Greek PROIEL



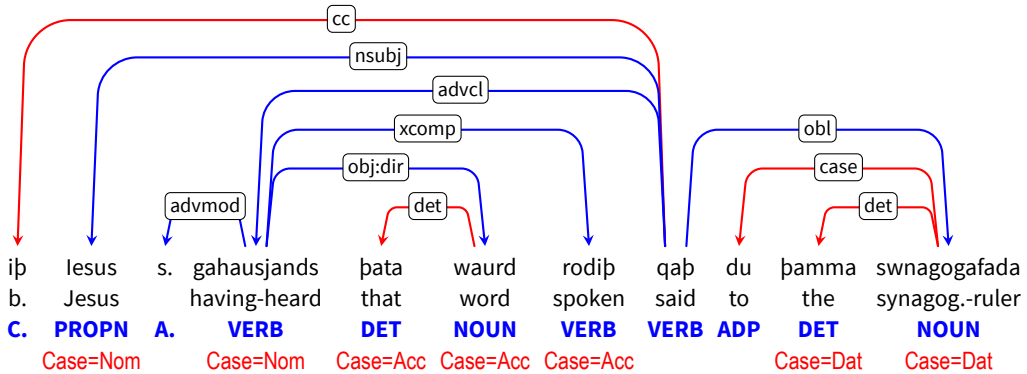
“But overhearing what they said, Jesus said to the ruler of the synagogue” (Mark 5:36)

UD Latin PROIEL



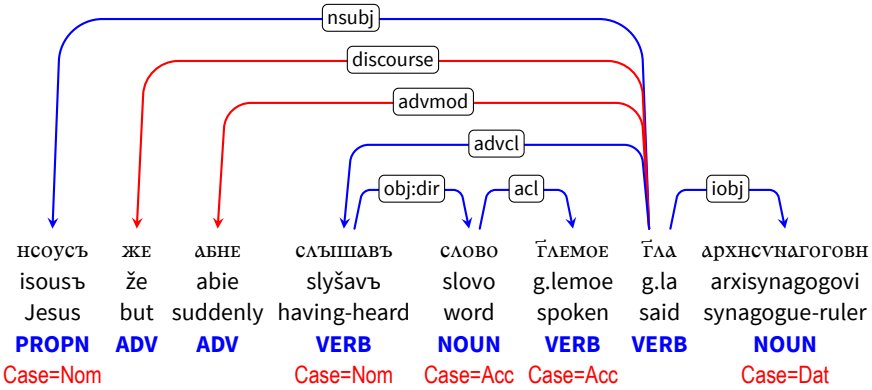
“But overhearing what they said, Jesus said to the ruler of the synagogue” (Mark 5:36)

UD Gothic PROIEL



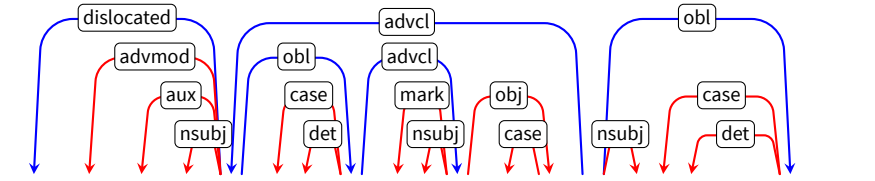
“But overhearing what they said, Jesus said to the ruler of the synagogue” (Mark 5:36)

UD Old Church Slavonic PROIEL



“But overhearing what they said, Jesus said to the ruler of the synagogue” (Mark 5:36)

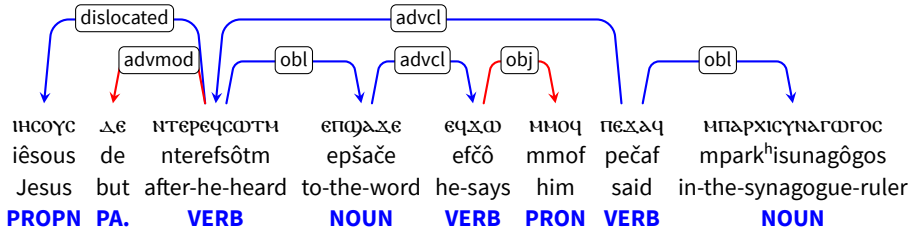
UD Coptic Scriptorium



iêsous de nterefsôtm epšače efčô mmof pečaf mparkʰisunagôgos

“But overhearing what they said, Jesus said to the ruler of the synagogue” (Mark 5:36)

UD Coptic Scriptorium



iêsous de nterefsôtm epšače efčô mmof pečaf mpark^{h}isunagôgos

"But overhearing what they said, Jesus said to the ruler of the synagogue" (Mark 5:36)

Language Change

- Classical Latin → Medieval Latin → Old French → Modern French
  - or Catalan, Spanish, Galician, Portuguese, Italian, Romanian…

Language Change

- Classical Latin → Medieval Latin → Old French → Modern French
  - or Catalan, Spanish, Galician, Portuguese, Italian, Romanian…
- Old Church Slavonic → Old East Slavic → Russian
  - or Ukrainian, Belarusian, Polish, Upper Sorbian, Czech, Slovak, Slovenian, Croatian, Serbian, Bulgarian…

Language Change: The Slavic L-participle

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1st and 2nd persons. Also used in conditional mood.
- L-participle in Slovenian: auxiliary in all persons. Used for past, future and conditional.
- L-participle in old Slavic: perfect (resultative) aspect, and conditional.

она	пришла	к	нему
ona	prišla	k	nemu
she	came	to	him
PRON	**VERB**	**ADP**	**PRON**
	VerbForm=Fin		
	Tense=Past		

я	хотел	бы	поблагодарить
ja	hotel	by	poblagodarit’
I	like	would	to-thank
PRON	**VERB**	**AUX**	**VERB**
	VerbForm=Fin		VerbForm=Inf
	Tense=Past		

Language Change: The Slavic L-participle

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1st and 2nd persons. Also used in conditional mood.
- L-participle in Slovenian: auxiliary in all persons. Used for past, future and conditional.
- L-participle in old Slavic: perfect (resultative) aspect, and conditional.

přišel	jsi	nás	zabít
came	you-have	us	to-kill
VERB	**AUX**	**PRON**	**VERB**
VerbForm=Part	VerbForm=Fin		VerbForm=Inf
Tense=Past	Tense=Pres		

přišel	nás	zabít
he-came	us	to-kill
VERB	**PRON**	**VERB**
VerbForm=Part		VerbForm=Inf
Tense=Past		

Language Change: The Slavic L-participle

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1st and 2nd persons. Also used in conditional mood.
- L-participle in Slovenian: auxiliary in all persons. Used for past, future and conditional.
- L-participle in old Slavic: perfect (resultative) aspect, and conditional.

přišel	by	nás	zabít
came	he-would	us	to-kill
VERB	**AUX**	**PRON**	**VERB**
VerbForm=Part	VerbForm=Fin		VerbForm=Inf
Tense=Past	Mood=Cnd		

přišel	nás	zabít
he-came	us	to-kill
VERB	**PRON**	**VERB**
VerbForm=Part		VerbForm=Inf
Tense=Past		

Language Change: The Slavic L-participle

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1st and 2nd persons. Also used in conditional mood.
- L-participle in Slovenian: auxiliary in all persons. Used for past, future and conditional.
- L-participle in old Slavic: perfect (resultative) aspect, and conditional.

daleč	boste	prišli	prišel	je	do	ugotovitve
far	you-will	come	come	he-has	to	conclusion
ADV	**AUX**	**VERB**	**VERB**	**AUX**	**ADP**	**NOUN**
	VerbForm=Fin	VerbForm=Part	VerbForm=Part	VerbForm=Fin		
	Tense=Fut			Tense=Pres		

Language Change: The Slavic L-participle

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1st and 2nd persons. Also used in conditional mood.
- L-participle in Slovenian: auxiliary in all persons. Used for past, future and conditional.
- L-participle in old Slavic: perfect (resultative) aspect, and conditional.

прншелъ	есн	погоубнтъ	насъ	да	не	бн	отъшелъ	отъ	нихъ
prišelъ	jesi	pogubitъ	nasъ	da	ne	bi	otъšelъ	otъ	nichъ
come	you-have	to-kill	us	that	not	would	leave	from	them
VERB	**AUX**	**VERB**	**PRON**	**SCONJ**	**PART**	**AUX**	**VERB**	**ADP**	**PRON**
VerbForm=Part	VerbForm=Fin	VerbForm=Inf				Fin	VerbForm=Part		
Aspect=Res	Tense=Pres					Cnd	Aspect=Res		

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

455 Contributors

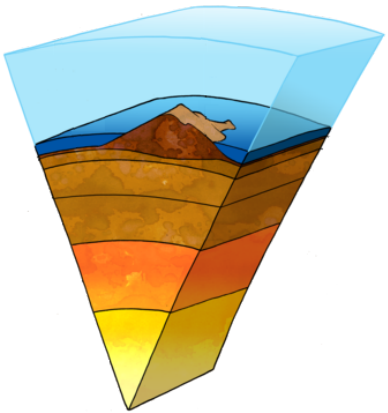
Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielė Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Bilge Nas Arıcan, Þórunn Arnardóttir, 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, Verginica Barbu Mititelu, Starkaður Barkarson, 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, Kristīne Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaar, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Lauren Cassidy, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, 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 Ilarraza, 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, Wograine 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 Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, 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, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Eva Huber, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Ọlájídé 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 Kopacewicz, Timo Korkiakangas, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Oğuzhan Kuyrukçu, 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 Miekka, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Judit Molnár, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Lương Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayọ̀ Olúòkun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özates, 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, Cenel-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 Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, 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 Roșca, Davide Rovati, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Benoît Sagot, Aleksi Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Sanıyar, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibussirri, Dmitry Sichinava, Janine Siewert, Einar Freyr Sigurðsson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Rachele Sprugnoli, Steinþór Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Samson Tella, Isabelle Tellier, Marinella Testori, Guillaume Thomas, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, 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, Maximilan 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. Yavrumyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, Hanzhi Zhu, Anna Zhuravleva, Rayan Ziane

Thanks!
ありがとう！

https://universaldependencies.org/

Closer to Meaning: Enhanced and Deep UD

Multiple Layers of Dependencies



Form

- Surface syntax
- Deep syntax
- Semantics

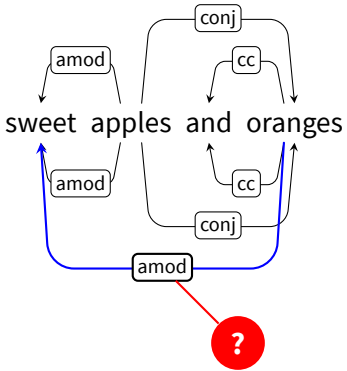
Meaning

Enhanced Universal Dependencies

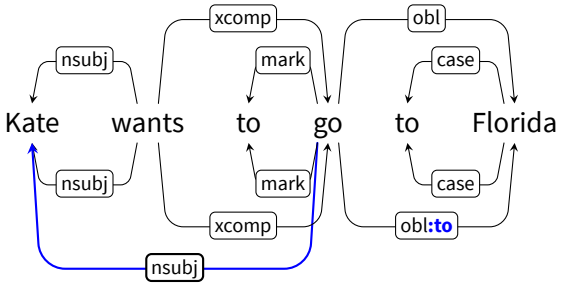
Currently available (at least partially) for 18 languages:

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

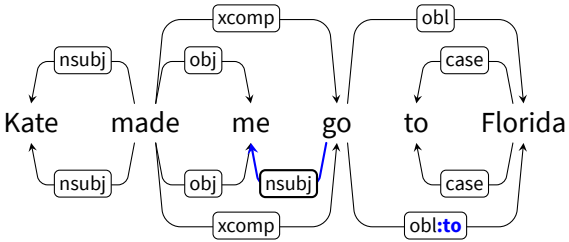
Enhanced UD: Shared Dependent of Coordination



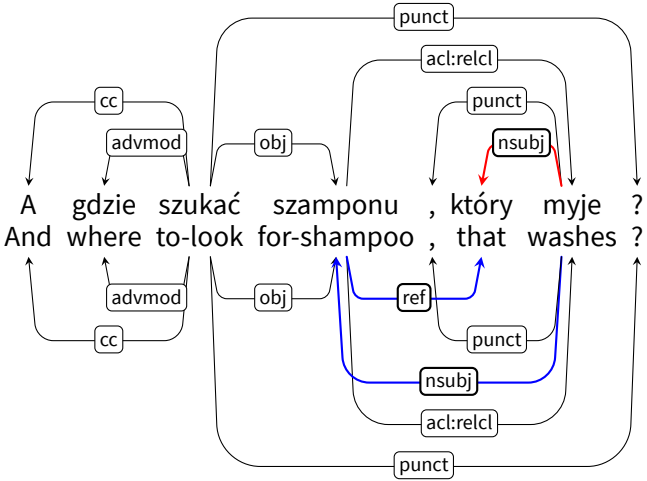
Enhanced UD: External Subject of Controlled Predicate



Enhanced UD: External Subject in Object-Control Construction

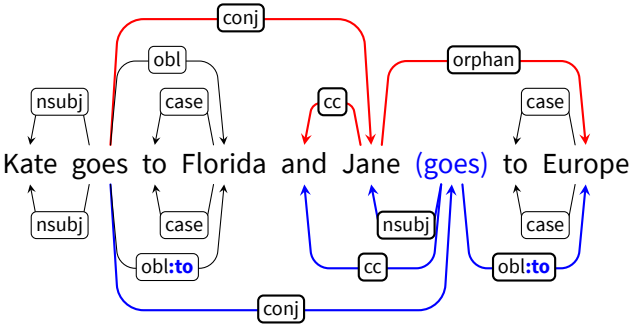


Enhanced UD: Relative Clauses

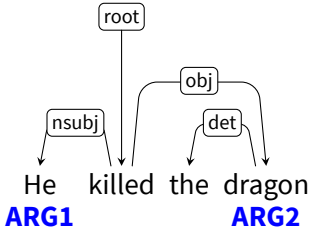


“And where to look for shampoo that works?”

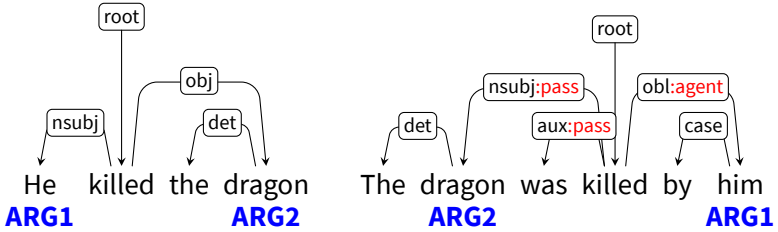
Enhanced UD: Gapping and Stripping



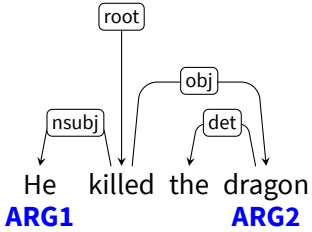
Deep UD: Normalization of Syntactic Alternations

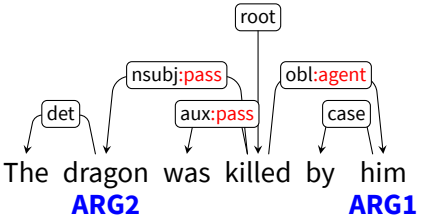


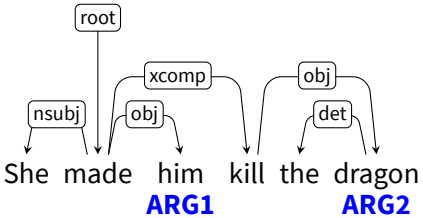
Deep UD: Normalization of Syntactic Alternations



Deep UD: Normalization of Syntactic Alternations







Deep UD: Normalization of Syntactic Alternations