

Enhanced Universal Dependencies: The Current State and Outlook



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University



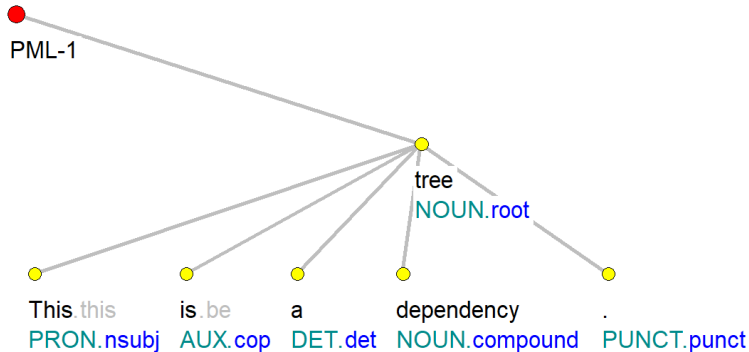
Daniel Zeman

zeman@ufal.mff.cuni.cz
<http://universaldependencies.org/>

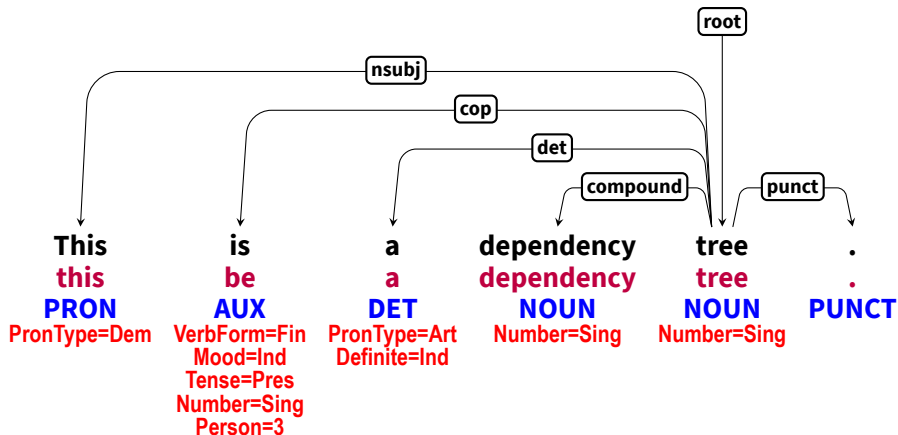
Universal Dependencies



Syntactic Trees



Syntactic Trees



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

455 Contributors

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielé Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Bilge Nas Arican, 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, Kristín 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 Čepľo, 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 Marnette, Valeria de Paiva, Mehmet Ögüz Derin, Elvís de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Jaja Dobrovoljc, Timothy Dozat, Kira Drostanova, Puneet Dwivedi, Hanne Eckhoff, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wograin 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 Gironi, 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, Olájidé Ishola, Kaoru Ito, Tomáš Jelinek, 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 Măranduc, 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 Moskalevskiy, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhle, Juan Ignacio Navarro Horňáček, 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, Adedayo Oluókun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Övrelid, 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-Lapińska, 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 Pretkalnina, Sophie

Manning's Law

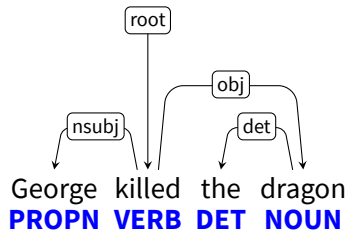
The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:



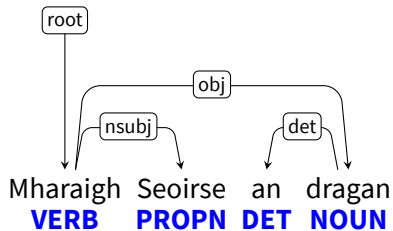
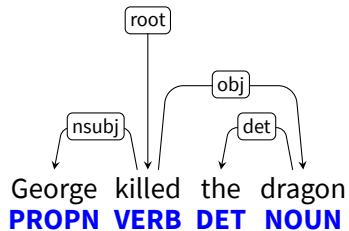
- 1 UD must be satisfactory on linguistic analysis grounds for **individual languages**.
- 2 UD must be good for linguistic **typology**, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
- 3 UD must be suitable for **rapid, consistent annotation** by a human annotator.
- 4 UD must be easily comprehended and used by a **non-linguist**, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor **traditional grammar** notions and terminology.
- 5 UD must be suitable for **computer parsing** with high accuracy.
- 6 UD must support well **downstream language understanding tasks** (relation extraction, reading comprehension, machine translation, ...)

It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.

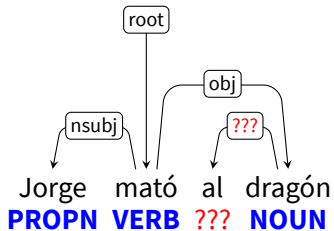
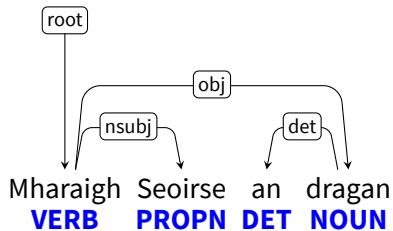
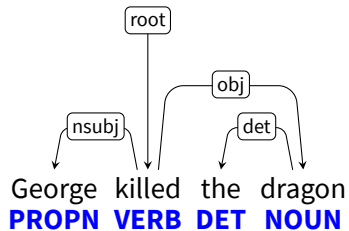
Same Thing Same Way



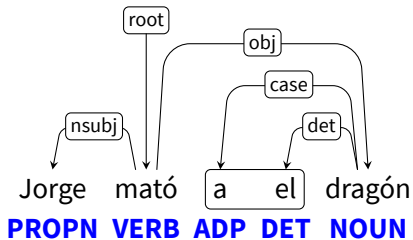
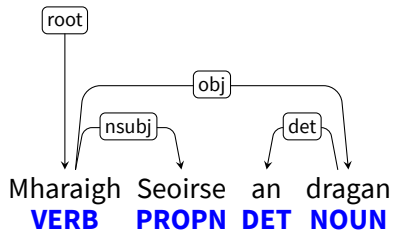
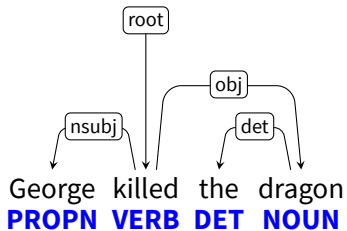
Same Thing Same Way



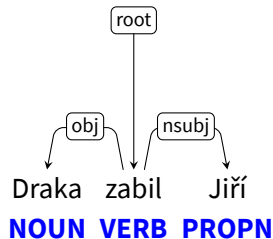
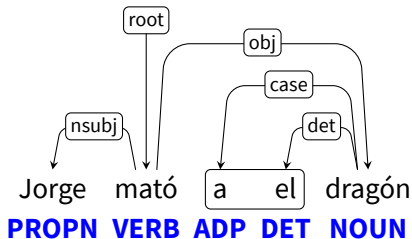
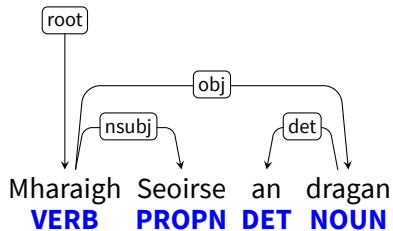
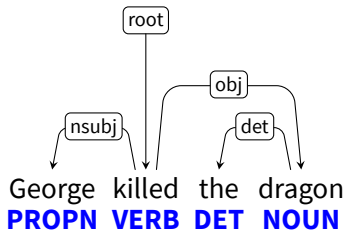
Same Thing Same Way



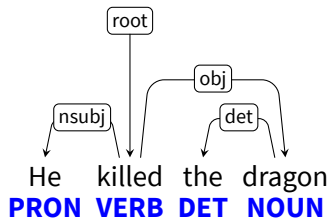
Same Thing Same Way



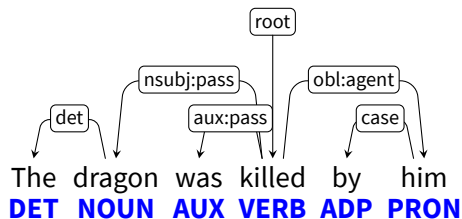
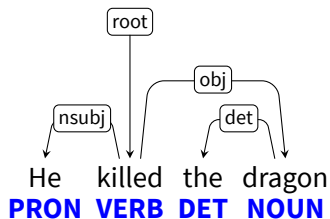
Same Thing Same Way



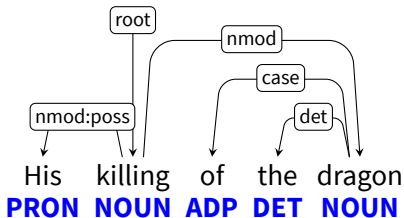
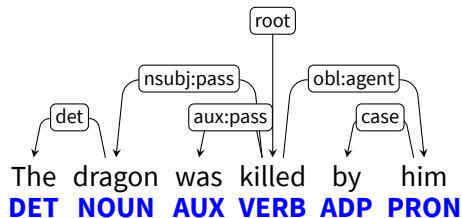
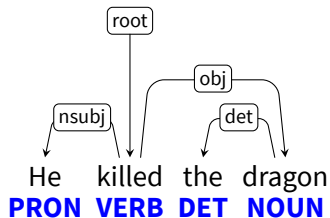
Same Meaning $\neq$ Same Construction!



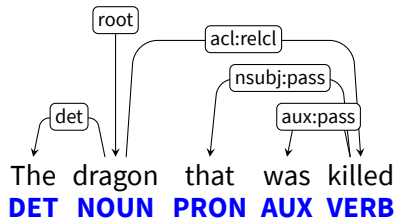
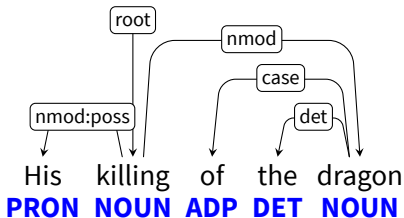
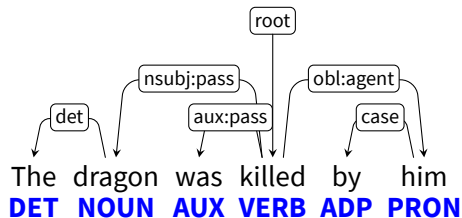
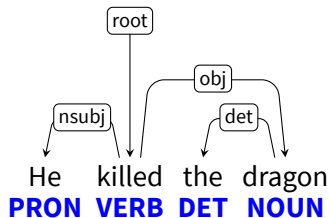
Same Meaning $\neq$ Same Construction!

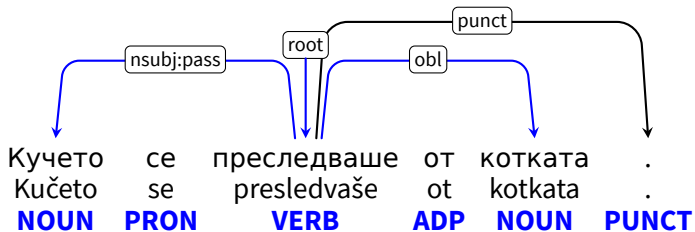
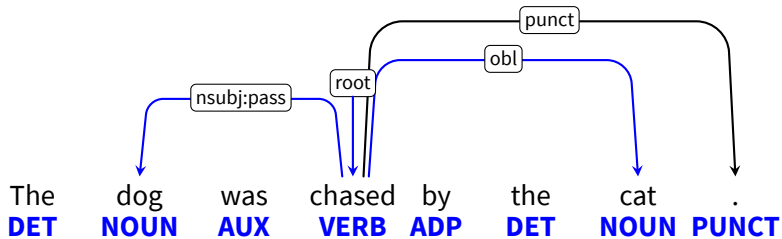


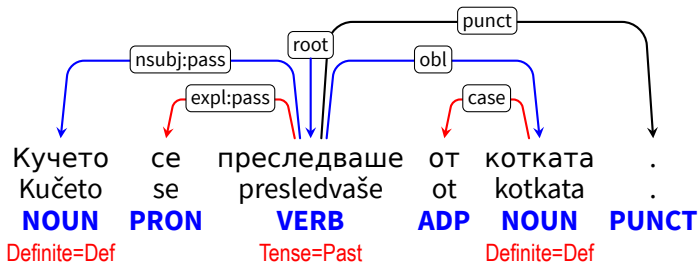
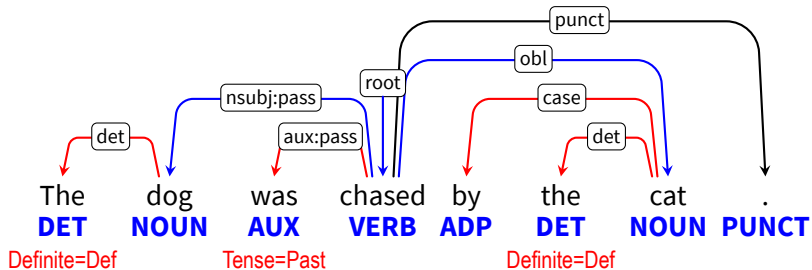
Same Meaning $\neq$ Same Construction!



Same Meaning $\neq$ Same Construction!







Dependents of Clauses (Verbal or Not)

	Nominal	Clausal	Modifier	Function
Core	nsubj	csubj		
Non-Core	obl vocative dislocated expl	advcl	advmod discourse	aux cop mark

Dependents of Verbs, Adjectives and Adverbs

	Nominal	Clausal	Modifier
Core	obj iobj	ccomp xcomp	
Non-Core	obl expl	advcl	advmod

Dependents of Nominals

Nominal	Clausal	Modifier	Function
nmod	acl	amod	det

Noun Phrase

Dependents of Nominals

Nominal

nmod
appos
compound
flat

Clausal

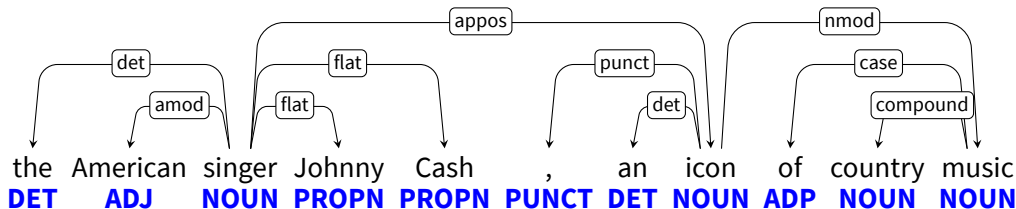
acl

Modifier

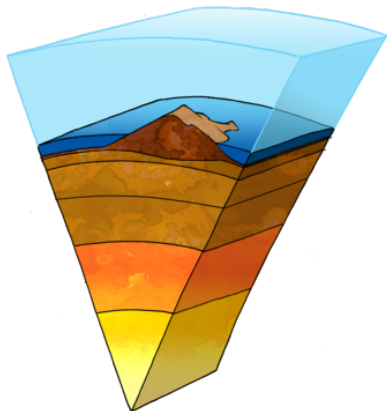
amod
nummod

Function

det
case
clf



Multiple Layers of Dependencies



Form

- Surface syntax
- Deep syntax
- Semantics

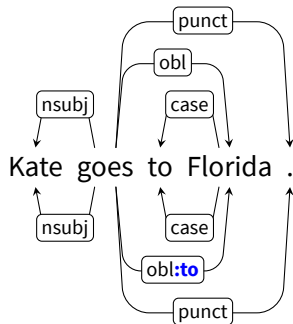
Meaning

Enhanced Universal Dependencies

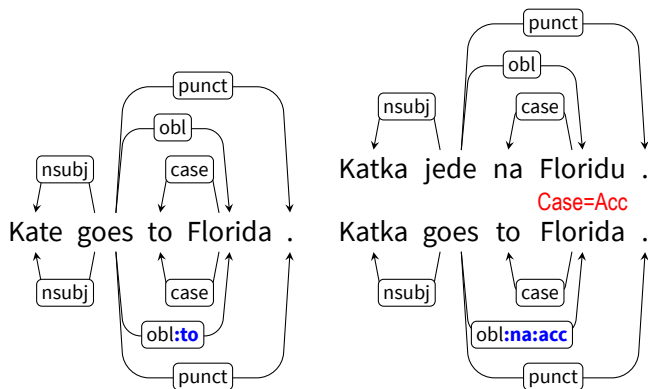
Currently available (at least partially) for 18 languages:

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

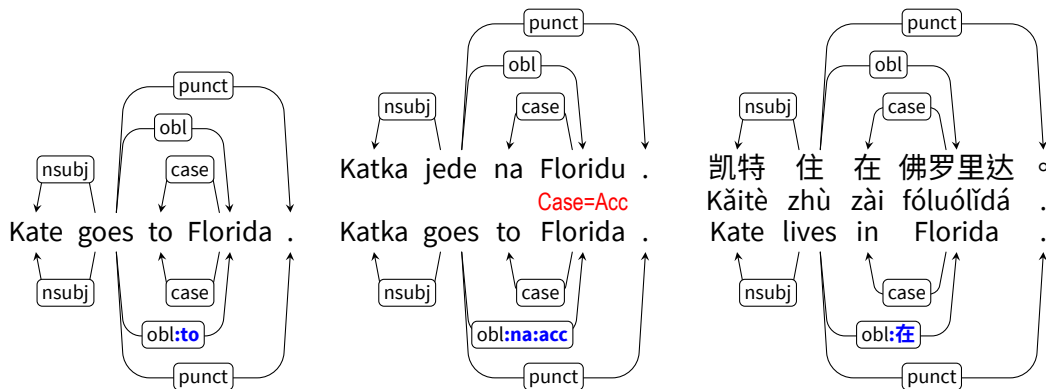
Enhanced UD: Case Information in Dependency Label



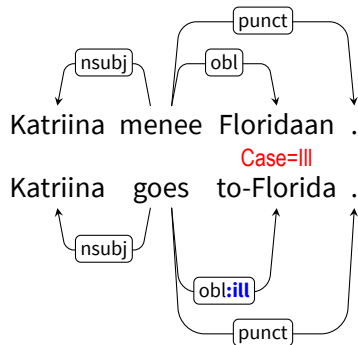
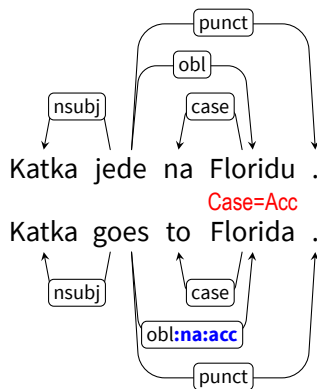
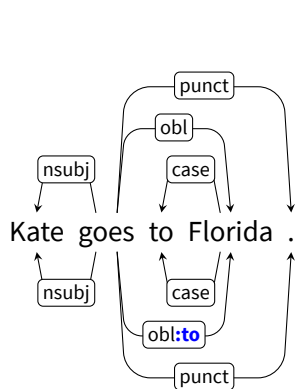
Enhanced UD: Case Information in Dependency Label



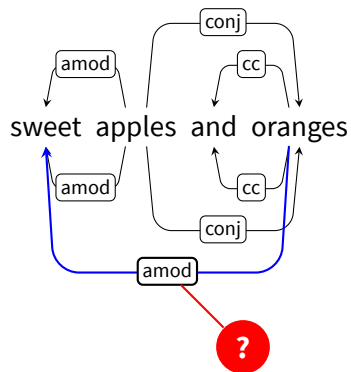
Enhanced UD: Case Information in Dependency Label



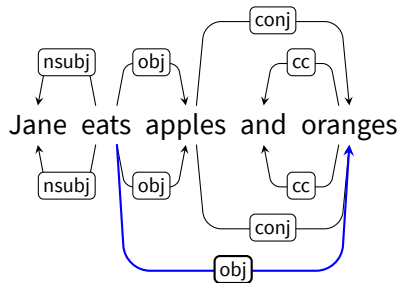
Enhanced UD: Case Information in Dependency Label



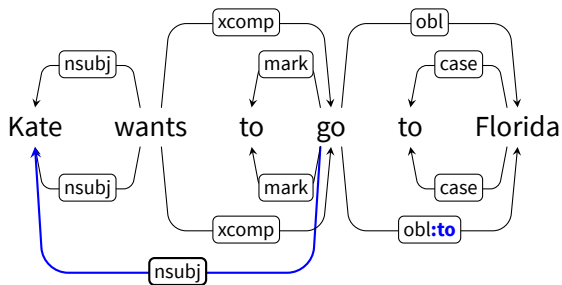
Enhanced UD: Shared Dependent of Coordination



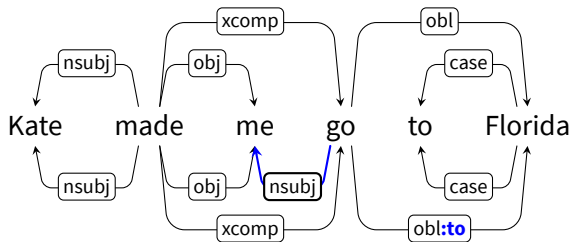
Enhanced UD: Parent of Coordination



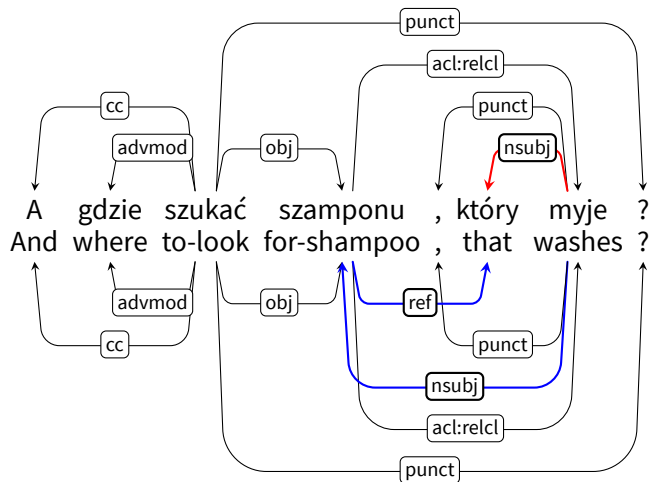
Enhanced UD: External Subject of Controlled Predicate



Enhanced UD: External Subject in Object-Control Construction

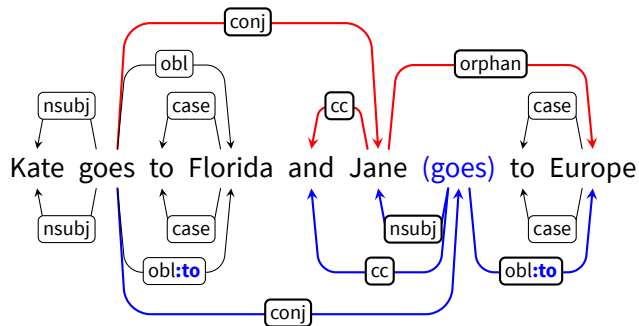


Enhanced UD: Relative Clauses

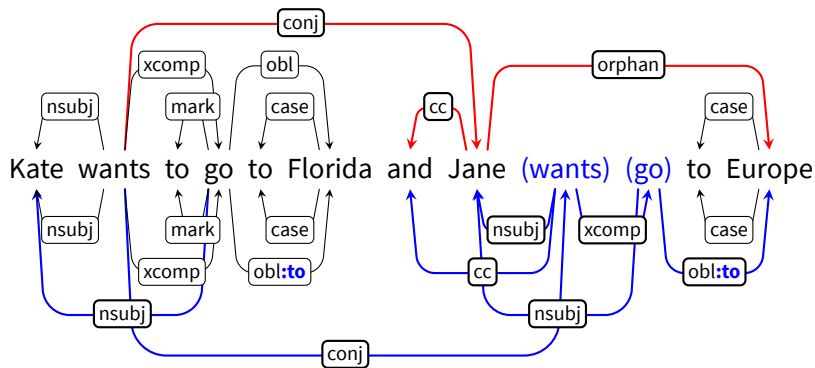


“And where to look for shampoo that works?”

Enhanced UD: Gapping and Stripping



Enhanced UD: Gapping + Control Verb Construction



Enhanced UD: Six Enhancements

- Null nodes for **gapping** (24 treebanks in UD 2.8)
 - Common **parent of coordination** (26 treebanks)
 - Shared **dependents of coordination** (25 treebanks)
 - External subjects of **controlled predicates** (21 treebanks)
 - Cyclic dependencies to/from **relative clauses** (24 treebanks)
 - **Case**-enhanced dependency labels (23 treebanks)
-
- All 6 types: 15 treebanks, 8 languages
 - At least 1 type: 30 treebanks, 18 languages
 - Only basic UD: 172 treebanks

“Enhanced Plus” — Proposals beyond the Official Guidelines

- Enhanced UD help us identify more predicate-argument relations
- But some patterns are still not handled...
- Adverbial **infinitives**
 - ▶ *They will meet **to discuss** a contract.*
 - ▶ But not `ccomp` infinitives: *He recommended to replace the tyres.*

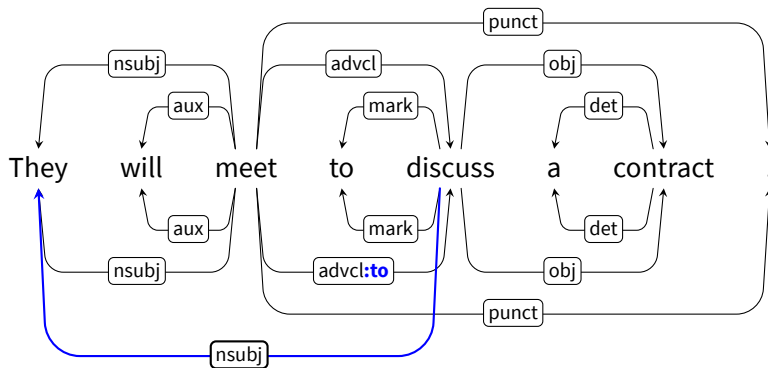
“Enhanced Plus” — Proposals beyond the Official Guidelines

- Enhanced UD help us identify more predicate-argument relations
- But some patterns are still not handled...
- Adverbial **infinitives**
 - ▶ *They will meet **to discuss** a contract.*
 - ▶ But not `ccomp` infinitives: *He recommended to replace the tyres.*
- Adverbial **converbs (gerunds)**
 - ▶ *Terrorists detonated a bomb **killing** five people.*

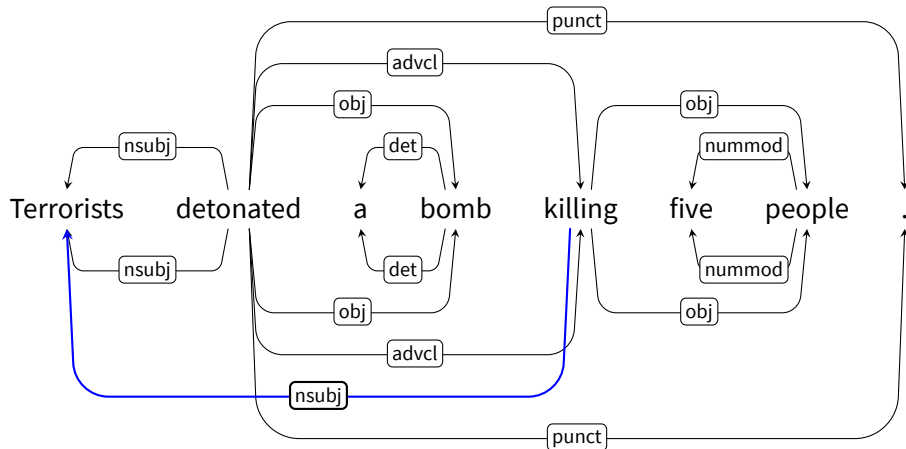
“Enhanced Plus” — Proposals beyond the Official Guidelines

- Enhanced UD help us identify more predicate-argument relations
- But some patterns are still not handled...
- Adverbial **infinitives**
 - ▶ *They will meet **to discuss** a contract.*
 - ▶ But not `ccomp` infinitives: *He recommended to replace the tyres.*
- Adverbial **converbs (gerunds)**
 - ▶ *Terrorists detonated a bomb **killing** five people.*
- Attributive **participles**
 - ▶ *The shares **reflected** on your statement.*

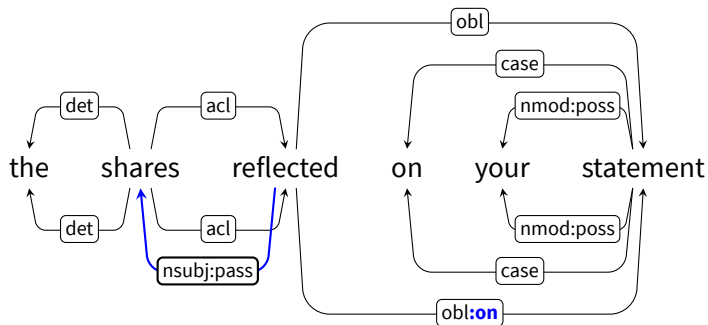
External Subject of Adverbial Infinitive



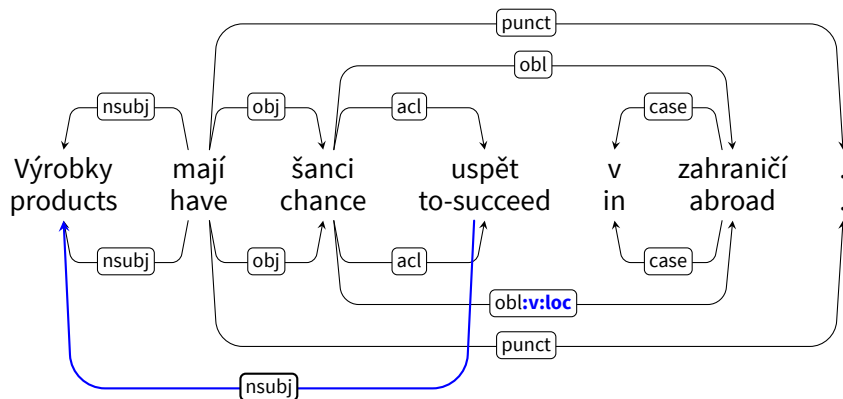
External Subject of Adverbial Gerund



External Subject of Participle

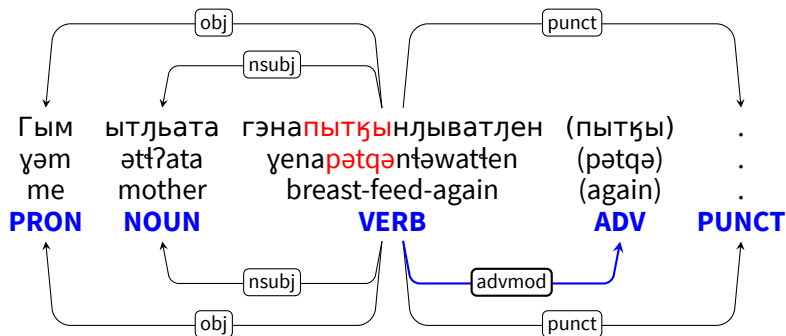


Control by Light-Verb Constructions



“The products have a chance to succeed abroad.”

Empty Nodes for Adverb Incorporation in Chukchi



Summary

- Enhanced UD: additional implicit relations made explicit
- Closer to meaning $\Rightarrow$ easier to discover **who did what to whom**
- UD 2.8: 30 treebanks out of 202
- 18 languages
- Some enhancements can be derived/guessed from basic trees
- “Enhancers”: could generate EUD for remaining treebanks?

Thanks!
Спасибо!

`https://universaldependencies.org/`