

# Universal Dependencies and the Slavic Languages



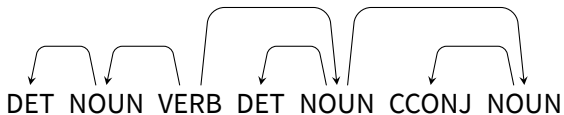
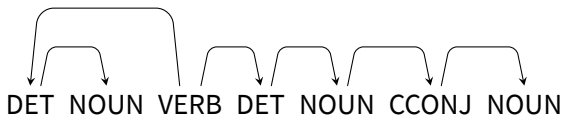
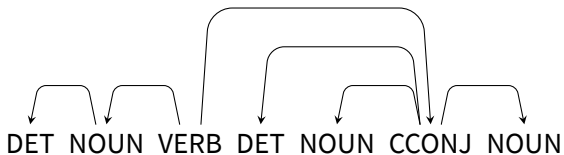
FACULTY  
OF MATHEMATICS  
AND PHYSICS  
Charles University

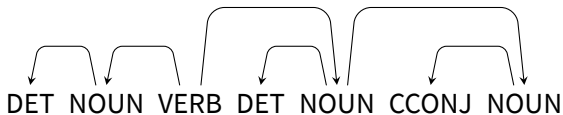
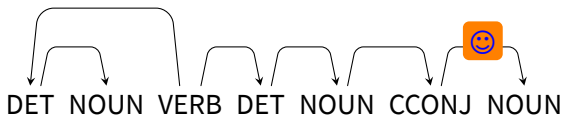
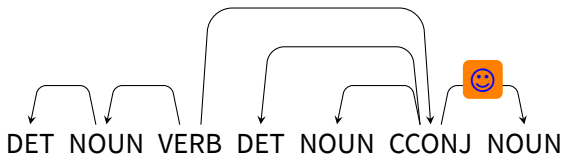


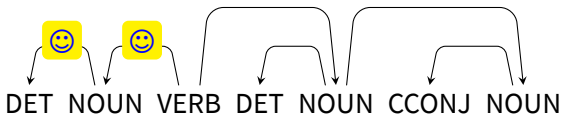
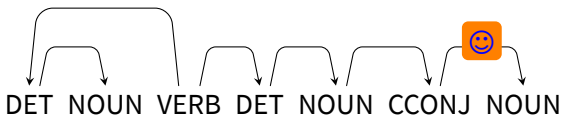
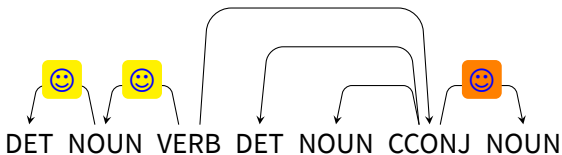
**Daniel Zeman**

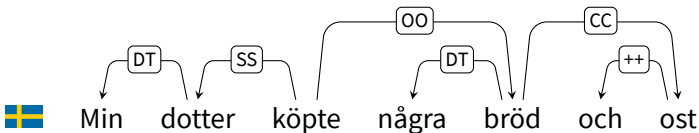
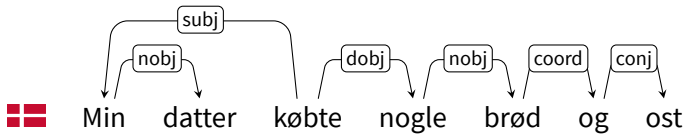
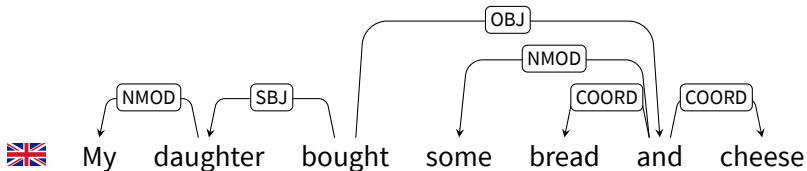
[zeman@ufal.mff.cuni.cz](mailto:zeman@ufal.mff.cuni.cz)

<http://universaldependencies.org/>









# Universal Dependencies

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

# 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**.



*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.*

# 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.



*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.*



# 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.



*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.*

# 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.

*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.*

# 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.

*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.*

# Manning's Law

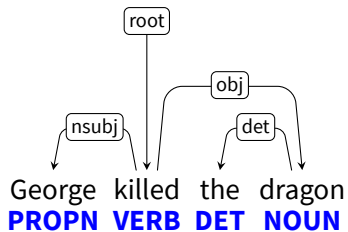
*The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:*



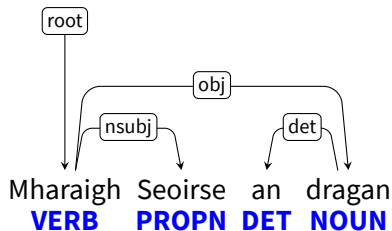
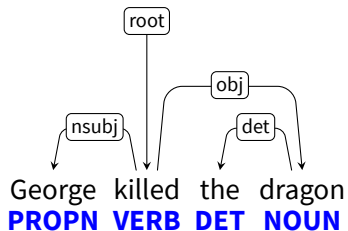
- 1 UD must be satisfactory on linguistic analysis grounds for **individual languages**.
- 2 UD must be good for linguistic **typology**, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
- 3 UD must be suitable for **rapid, consistent 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.*

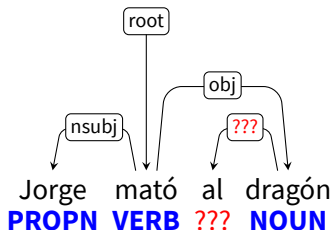
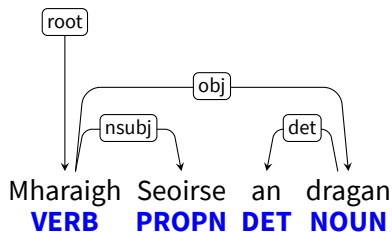
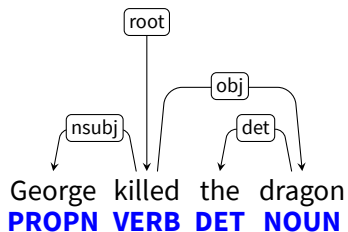
# Same Thing Same Way



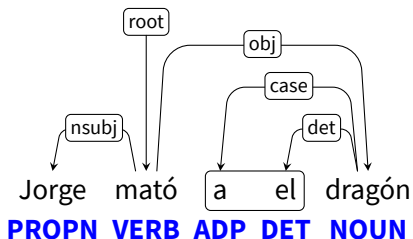
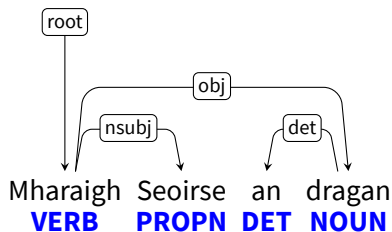
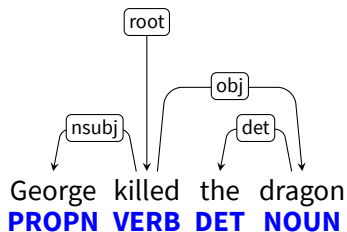
# Same Thing Same Way



# Same Thing Same Way

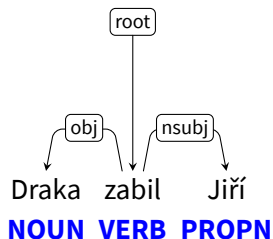
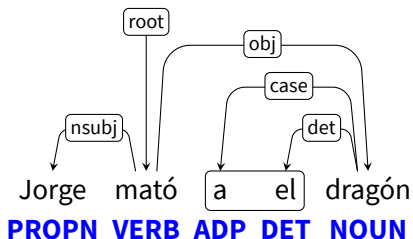
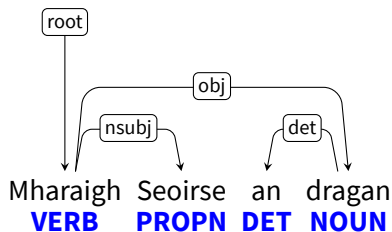
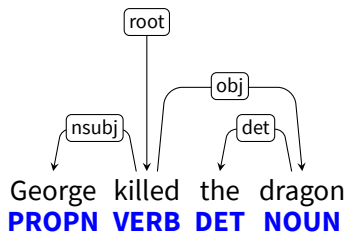


# Same Thing Same Way

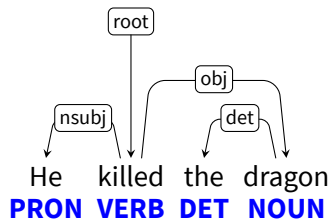




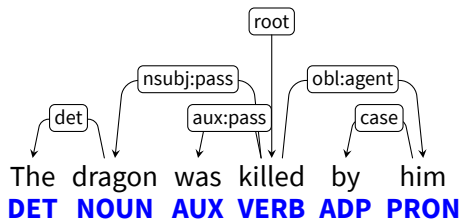
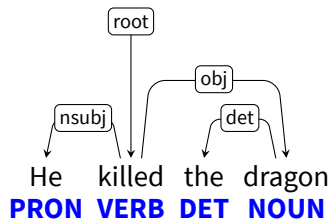
# Same Thing Same Way



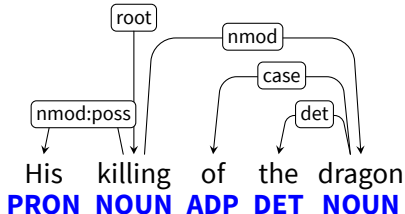
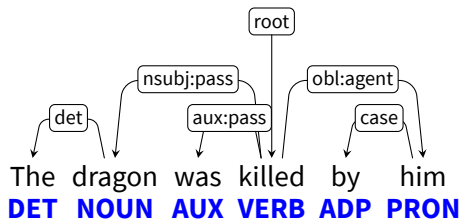
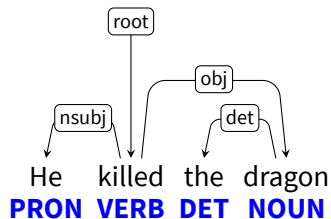
# Same Meaning $\neq$ Same Construction!



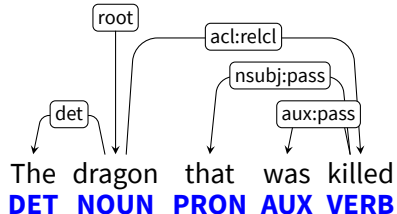
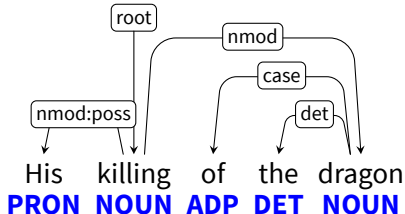
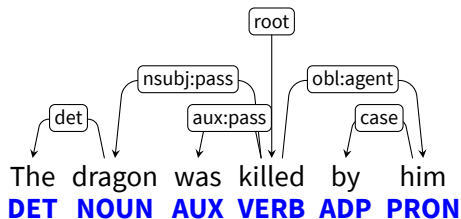
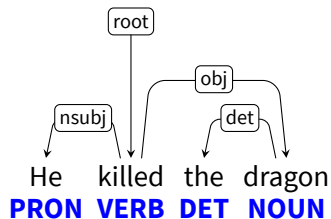
# Same Meaning $\neq$ Same Construction!



# Same Meaning $\neq$ Same Construction!



# Same Meaning $\neq$ Same Construction!



# A Tour through UD

# Word Segmentation

- Must be **reproducible** on new data
- Surface tokens vs. syntactic words
- Chinese, Vietnamese etc.: no clues, non-trivial algorithm
- Arabic, Tamil etc.: part of morphological analysis
- Spanish, German etc.: rather limited cases of contractions
- Others: only punctuation (low-level tokenization)

## Word Segmentation

Vámonos al mar .  
VERB+PRON ADP+DET NOUN PUNCT

Vamos nos a el mar .  
VERB PRON ADP DET NOUN PUNCT

# Word Segmentation

- Fusions

- ▶ al = a + el
- ▶ naň = na + něj

- Clitics

- ▶ vámonos = vamos + nos
- ▶ izmenjat'sja = izmenjat' + sja
- ▶ potrafilibyśmy = potrafili + by + jesteśmy

## Word Segmentation

Cośmy jeszcze zaniedbali ?  
PRON+AUX PART VERB PUNCT

Co (jeste)śmy jeszcze zaniedbali ?  
PRON AUX PART VERB PUNCT



# 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 <i>Some</i>	dívky <i>girls</i>	si	nicméně <i>nevertheless</i>	pochvalovaly <i>praised</i>	zmrzlinu <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

ADJ

ADV

INTJ

NOUN

PROPN

VERB

### Closed

ADP

AUX

CCONJ

DET

NUM

PART

PRON

SCONJ

### Other

PUNCT

SYM

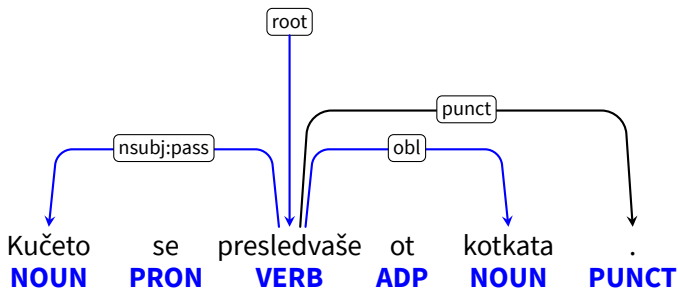
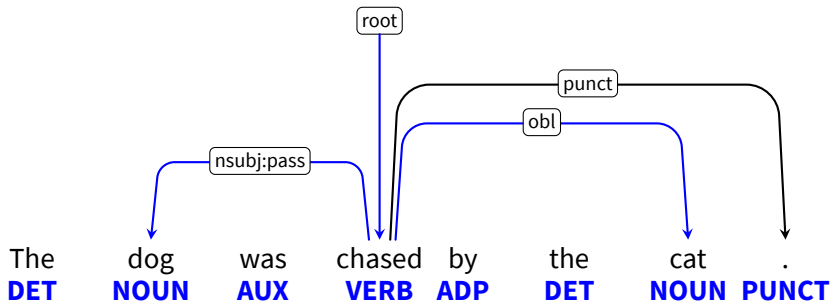
X

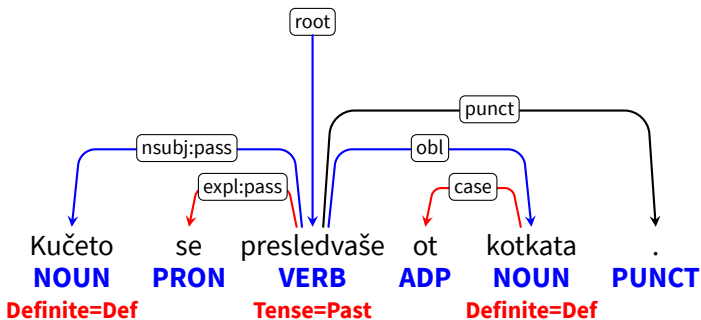
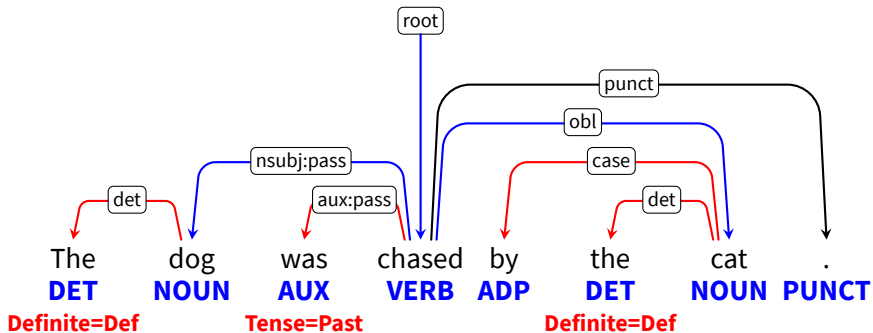
- Taxonomy of 17 universal part-of-speech tags, based on the Google Universal Tagset (Petrov et al., 2012)
- All languages use the same inventory, but not all tags have to be used by all languages

## Features

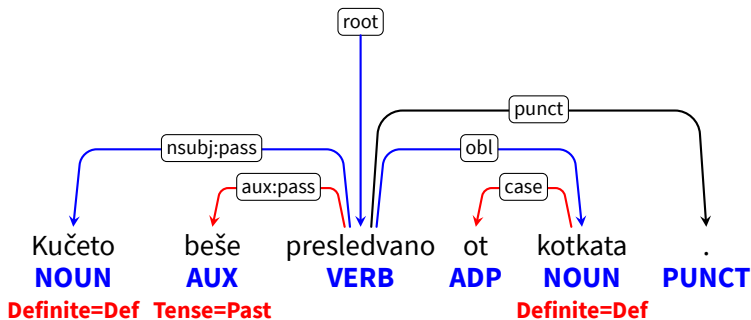
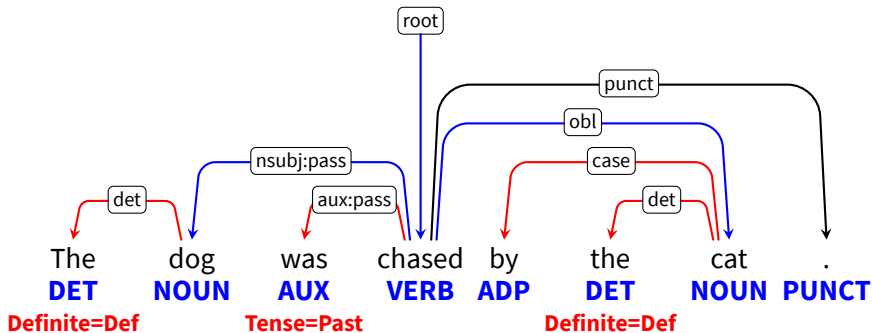
<b>Lexical</b>	<b>Nominal</b>	<b>Pronominal</b>	<b>Verbal</b>
PronType	Gender	Person	VerbForm
NumType	Animacy	Clusivity	Mood
Poss	NounClass	Polite	Tense
Reflex	Number		Aspect
Foreign	Case		Voice
Abbr	Definite		Evident
	Degree		Polarity

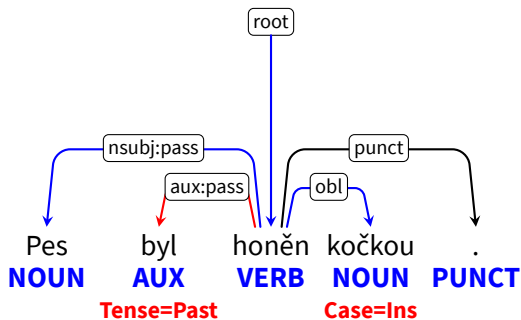
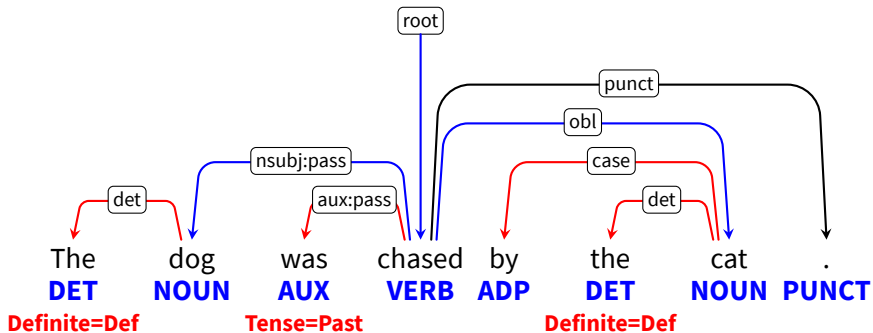
- Standardized inventory of 23 morphological features, based on InterSet (Zeman, 2008)
- Languages select relevant features and can add language-specific features or values with documentation











## Dependents of Clauses (Verbal or Not)

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

## Dependents of Verbs, Adjectives and Adverbs

	<b>Nominal</b>	<b>Clausal</b>	<b>Modifier</b>
<b>Core</b>	<b>obj</b> <b>iobj</b>	ccomp xcomp	
<b>Non-Core</b>	<b>obl</b> expl	advcl	advmod

## Dependents of Nominals

<b>Nominal</b>	<b>Clausal</b>	<b>Modifier</b>	<b>Function</b>
nmod appos	acl	amod nummod	det case

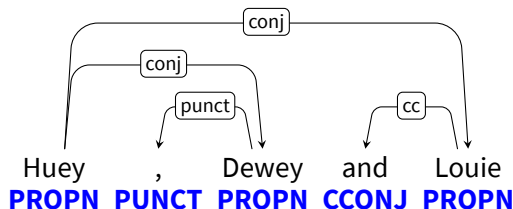
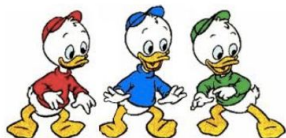
# Language-Specific Relation Subtypes

- Language-specific relations are **subtypes** of universal relations added to capture important phenomena
- Subtyping permits us to “back off” to universal relations

## Language-Specific Relation Subtypes

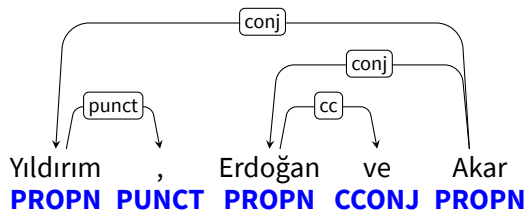
<b>Relation</b>	<b>Explanation</b>
acl:relcl	Relative clause
compound:prt	Verb particle (dress <b>up</b> )
nmod:poss	Possessive nominal ( <b>Mary's</b> book)
obl:agent	Agent in passive (saved <b>by the bell</b> )
cc:preconj	Preconjunction ( <b>both</b> ... and)
det:predet	Predeterminer ( <b>all</b> those ...)

# Coordination



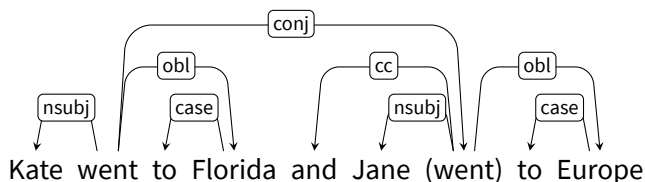
- Coordinate structures are headed by the first conjunct
  - ▶ Subsequent conjuncts depend on it via the **conj** relation
  - ▶ Conjunctions depend on the next conjunct via the **cc** relation
  - ▶ Punctuation marks depend on the next conjunct via the **punct** relation

## But Some Languages Might Prefer the Opposite



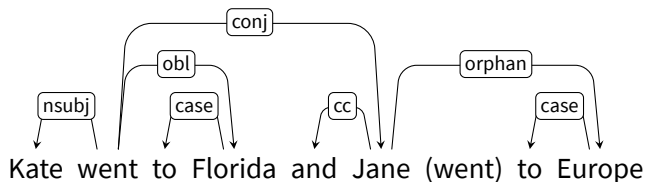
- Coordinate structures would be headed by the last conjunct
  - ▶ Preceding conjuncts would depend on it via the **conj** relation
  - ▶ Conjunctions would depend on the preceding conjunct
  - ▶ Punctuation marks would depend on the preceding conjunct

# Deleted Predicates in Coordination



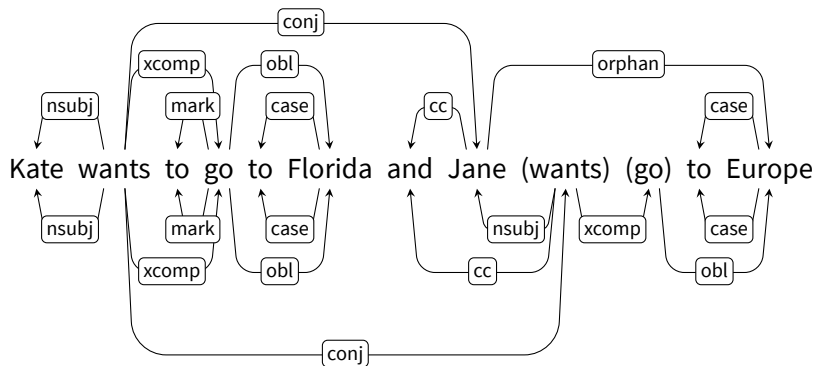
- Some treebanks would use an **empty node** to represent the second *went*.
- UD **enhanced representation** now allows empty nodes!
- But the basic representation sticks with the overt words.

# Ellipsis in Basic UD

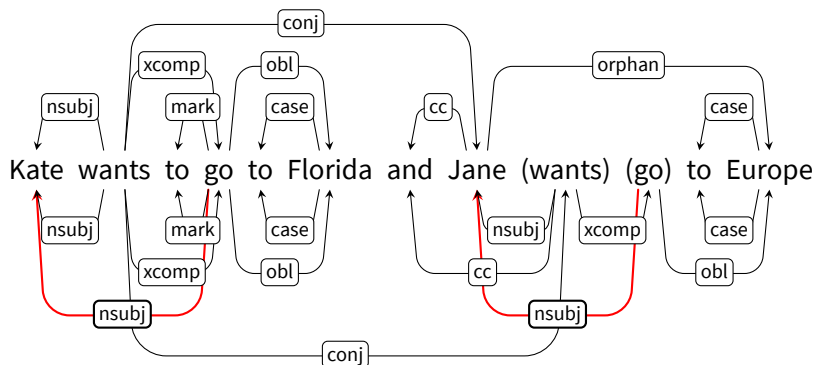




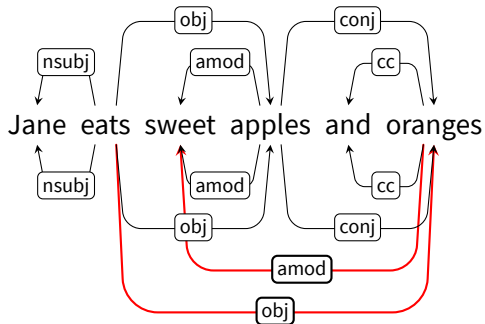
# Enhanced Dependencies: Gapping



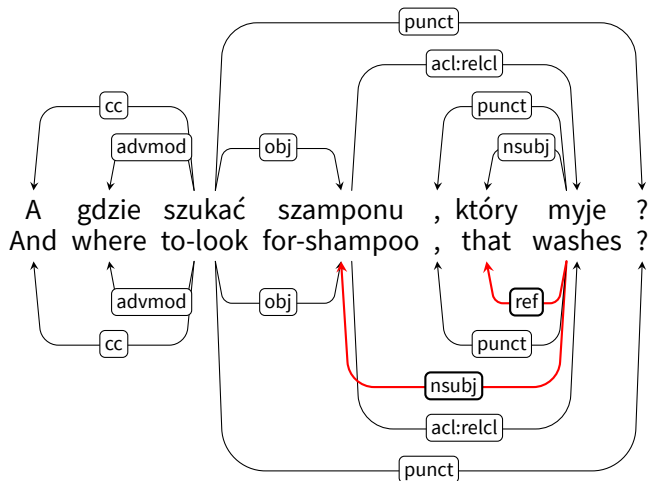
# Enhanced Dependencies: Gapping and Control



# Enhanced Dependencies: Coordination



# Enhanced Dependencies: Relative Clauses



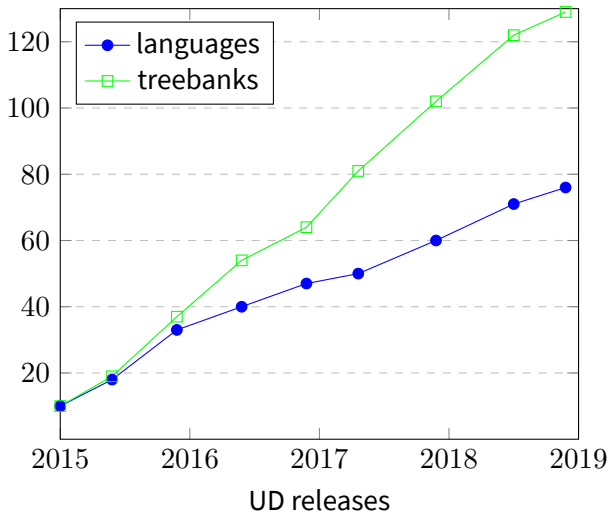
# Where Are We Now?



# Where Are We Now?

- Four and half years
- 9 treebank releases (every 6 months)
- 129 treebanks
- 75 languages (over 50% world's population)
- Over 18M tokens; treebanks range from <1K to 1.5M
- Over 280 contributors
  - ▶ language group consistency SIGs
- Two UD workshops
- Two CoNLL Shared Tasks in parsing UD

# Growing Number of Languages














# A Big Family: 75 Languages and Growing

- I.-E.:  Armenian,  Ancient Greek, Greek,  Breton,  Irish
  - ▶ Germanic:  Afrikaans,  Danish,  Dutch,  English,  Faroese,  German,  Gothic,  Norwegian,  Swedish
  - ▶ Romance:  Catalan,  French,  Galician,  Italian,  Latin,  Old French,  Portuguese,  Romanian,  Spanish
  - ▶ Balto-Slavic:  Belarusian,  Bulgarian,  Croatian,  Czech,  Church Slavonic,  Polish,  Russian,  Serbian,  Slovak,  Slovenian,  Ukrainian,  Sorbian,  Latvian,  Lithuanian
  - ▶ Indo-Ir.:  Kurmanji,  Persian,  Hindi, Marathi, Sanskrit, Urdu
- Uralic:  Erzya,  Estonian,  Finnish,  Hungarian,  Komi,  Sami; Dravidian:  Tamil, Telugu
- Turkic:  Kazakh,  Turkish,  Uyghur
- Afro-Asiatic:  Akkadian,  Amharic,  Arabic,  Coptic,  Hebrew,  Maltese
- Sino-Tibetan:  Cantonese,  Chinese; Aus.-As.:  Vietnamese
- Tai-Kadai:  Thai; Austronesian:  Indonesian,  Tagalog
- Other:  Buryat,  Japanese,  Korean,  Basque,  Sw. Sign,  Naija,  Bambara,  Yoruba,  Warlpiri



# Existing Slavic Treebanks

Language	Code	Treebanks
Belarusian 	[be]	HSE 8K
Bulgarian 	[bg]	BTB 156K
Church Slavonic	[cu]	PROIEL 57K
Croatian 	[hr]	SET 197K
Czech 	[cs]	PDT 1506K, CAC 494K, CLTT 35K, PUD 18K, FicTree 167K
Polish 	[pl]	SZ 83K, LFG 130K
Russian 	[ru]	SynTagRus 1107K, GSD 99K, PUD 19K, Taiga 20K
Serbian 	[sr]	SET 86K
Slovak 	[sk]	SNK 106K
Slovenian 	[sl]	SSJ 140K, SST 29K
Ukrainian 	[uk]	IU 116K
Upper Sorbian 	[hsb]	UFAL 11K

# Pronouns and Determiners

- English + Romance languages: **DET** = article or pronominal adjective  
*this, which, every*

# Pronouns and Determiners

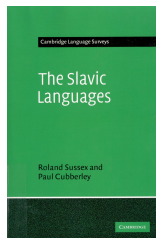
- English + Romance languages: **DET** = article or pronominal adjective  
*this, which, every*
- We don't have this category! (Traditionally → PRON.)

# Pronouns and Determiners

- English + Romance languages: **DET** = article or pronominal adjective  
*this, which, every*
- We don't have this category! (Traditionally → PRON.)
- We have the words (except for articles).

# Pronouns and Determiners

- English + Romance languages: **DET** = article or pronominal adjective  
*this, which, every*
- We don't have this category! (Traditionally → PRON.)
- We have the words (except for articles).
- Some authors do recognize determiners in Slavic!



# Pronouns and Determiners

- English + Romance languages: **DET** = article or pronominal adjective  
*this, which, every*
- We don't have this category! (Traditionally → PRON.)
- We have the words (except for articles).
- Some authors do recognize determiners in Slavic!
- UD v1: functional borderline (but ellipsis?)
  - ▶ *This.DET car is expensive.*
  - ▶ *This.PRON is expensive.*
- Less strict in UD v2.

# Verb Forms

- **Conflicting terminologies** in traditional grammars
- Participle ... verb or adjective?
- Converb / transgressive / gerund ... verb or adverb?
- Verbal noun / gerund ... verb or noun?
- Tags and features apply to **individual words!**

# Conflicting Terminology & Language Change

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1<sup>st</sup> and 2<sup>nd</sup> 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	я	хотел	бы	поблагодарить
ona	prišla	k	nemu	ja	hotel	by	poblagodarit’
she	came	to	him	I	like	would	to-thank
<b>PRON</b>	<b>VERB</b>	<b>ADP</b>	<b>PRON</b>	<b>PRON</b>	<b>VERB</b>	<b>AUX</b>	<b>VERB</b>
	VerbForm=Fin Tense=Past				VerbForm=Fin Tense=Past		VerbForm=Inf



# Conflicting Terminology & Language Change

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1<sup>st</sup> and 2<sup>nd</sup> 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
<b>VERB</b>	<b>AUX</b>	<b>PRON</b>	<b>VERB</b>
VerbForm=Part Tense=Past	VerbForm=Fin Tense=Pres		VerbForm=Inf

přišel	nás	zabít
he-came	us	to-kill
<b>VERB</b>	<b>PRON</b>	<b>VERB</b>
VerbForm=Part Tense=Past		VerbForm=Inf

# Conflicting Terminology & Language Change

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1<sup>st</sup> and 2<sup>nd</sup> 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
<b>VERB</b>	<b>AUX</b>	<b>PRON</b>	<b>VERB</b>
VerbForm=Part Tense=Past	VerbForm=Fin Mood=Cnd		VerbForm=Inf

přišel	nás	zabít
he-came	us	to-kill
<b>VERB</b>	<b>PRON</b>	<b>VERB</b>
VerbForm=Part Tense=Past		VerbForm=Inf

# Conflicting Terminology & Language Change

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1<sup>st</sup> and 2<sup>nd</sup> 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	přišli	přišel	je	do	ugotovitve
far	you-will	come	come	he-has	to	conclusion
<b>ADV</b>	<b>AUX</b>	<b>VERB</b>	<b>VERB</b>	<b>AUX</b>	<b>ADP</b>	<b>NOUN</b>
	VerbForm=Fin Tense=Fut	VerbForm=Part	VerbForm=Part	VerbForm=Fin Tense=Pres		

# Conflicting Terminology & Language Change

- Past tense in modern East Slavic: -l / -la / -lo / -li, “finite” (no auxiliary).
- Past tense / active (“l”-)participle in West Slavic: auxiliary in 1<sup>st</sup> and 2<sup>nd</sup> 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.

<b>ПРИШЕАЪ</b>	ЕСИ	ПОГОУБИТЪ	НАСЪ	ДА	НЕ	БИ	<b>ОТЪШЕАЪ</b>	ОТЪ	НИХЪ
<b>prišelъ</b>	jesi	pogubitъ	nasъ	da	ne	bi	<b>otъšelъ</b>	otъ	nichъ
come	you-have	to-kill	us	that	not	would	leave	from	them
<b>VERB</b>	<b>AUX</b>	<b>VERB</b>	<b>PRON</b>	<b>SCONJ</b>	<b>PART</b>	<b>AUX</b>	<b>VERB</b>	<b>ADP</b>	<b>PRON</b>
VerbForm=Part Aspect=Res	VerbForm=Fin Tense=Pres	VerbForm=Inf				Fin Cnd	VerbForm=Part Aspect=Res		

# Core Arguments

- Easier cross-linguistically than argument-adjunct?
- **S**ubject of intransitive verb
- **A**gent of transitive verb
- **P**atient (direct object) of transitive verb

# Core vs. Oblique Dependents

- **Core arguments:** what exactly is it?
  - ▶ Subject
  - ▶ Object
  - ▶ (Indirect object)
- English:
  - ▶ *He gave **John** the book.* (iobj)
  - ▶ *He gave the book **to John**.* (obl)
- Spanish:
  - ▶ *Dio el libro **a John**.* (iobj?)
- Slavic languages and German:
  - ▶ Only bare accusative deserves to be obj?
  - ▶ All other “objects” should be oblique arguments?

# Reflexive Pronouns

- Direct or indirect object (obj, iobj):  
*Řízl **se** do prstu / Řízl **ho** do prstu.*
  - ▶ Including reciprocal usage:  
*Políbili **se**. / They kissed **each other**.*

# Reflexive Pronouns

- Direct or indirect object (obj, iobj):  
*Řízl **se** do prstu / Řízl **ho** do prstu.*
  - ▶ Including reciprocal usage:  
*Políbili **se**. / They kissed **each other**.*
- Inherently reflexive verbs: ***smát se**, **bát se** / laugh, fear*
  - ▶ `expl:pv` (pronominal verb; previously compound)



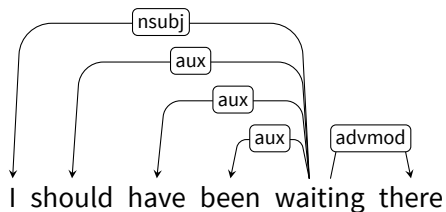
# Reflexive Pronouns

- Direct or indirect object (obj, iobj):  
*Řízl **se** do prstu / Řízl **ho** do prstu.*
  - ▶ Including reciprocal usage:  
*Políbili **se**. / They kissed **each other**.*
- Inherently reflexive verbs: ***smát se**, **bát se** / laugh, fear*
  - ▶ `expl:pv` (pronominal verb; previously compound)
- Reflexive passive:  
*To se **snadněji** řekne než udělá. / That is easier said than done.*
  - ▶ `expl:pass` (previously `auxpass:reflex`)

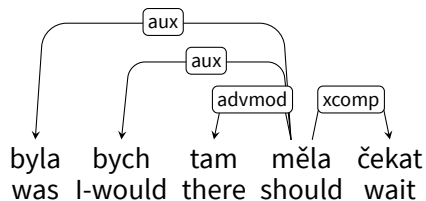
# Reflexive Pronouns

- Direct or indirect object (obj, iobj):  
*Řízl **se** do prstu / Řízl **ho** do prstu.*
  - ▶ Including reciprocal usage:  
*Políbili **se**. / They kissed **each other**.*
- Inherently reflexive verbs: ***smát se**, bát se / laugh, fear*
  - ▶ `expl:pv` (pronominal verb; previously compound)
- Reflexive passive:  
*To se snadněji řekne než udělá. / That is easier said than done.*
  - ▶ `expl:pass` (previously `auxpass:reflex`)
- Impersonal construction (~ passive?):  
*Zde se mluví německy. / German is spoken here.*
  - ▶ `expl:impers`

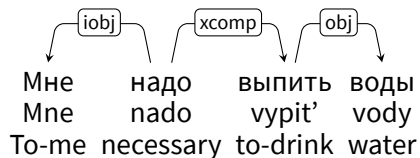
# Modal Auxiliary in English



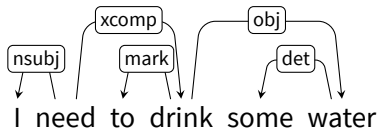
# Modal Verb in Czech



# Modal Predicative in Russian



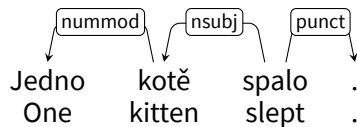
# Modal / Control Verb in English



# Still Have Time?

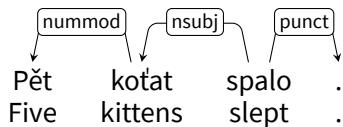


# Quantified Noun Phrase

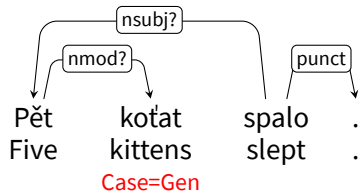




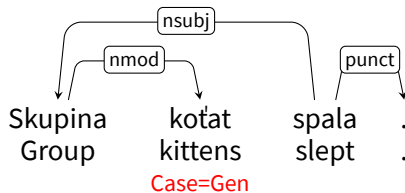
# Quantified Noun Phrase



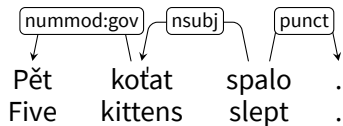
# Quantified Noun Phrase



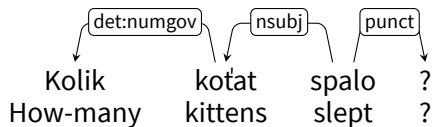
# Quantified Noun Phrase



# Quantified Noun Phrase



# Quantified Noun Phrase

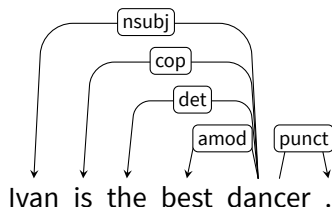


# Language-Specific Labels

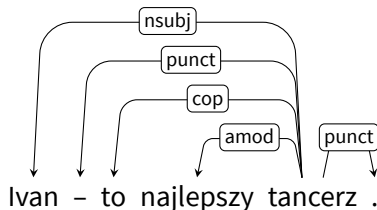
	<b>Numeric</b>	<b>Pronominal</b>
<b>Noun governs</b>	nummod	det:nummod
<b>Numeral governs</b>	nummod:gov	det:numgov

# Nonverbal Predicate and Copula

- Some languages use a copula verb:

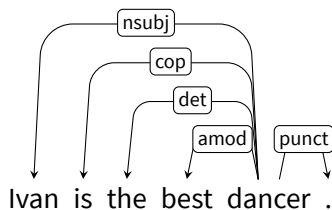


- Some languages use a copula **pronoun**:

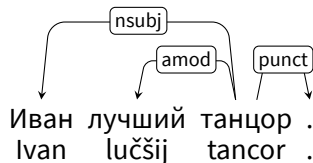


# Nonverbal Predicate and Copula

- Some languages use a copula verb:



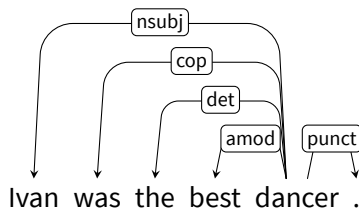
- Some languages **omit the copula**:



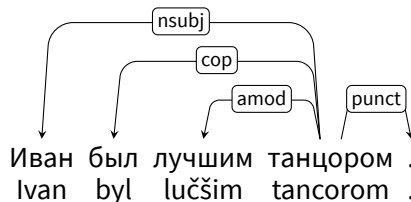


# Nonverbal Predicate and Copula

- Some languages use a copula verb:

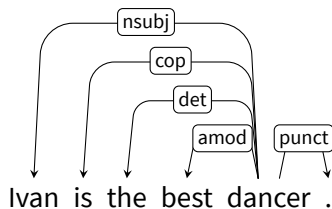


- Some languages use it **only in some tenses**:

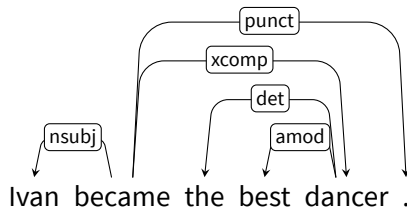


# Copula Verbs: We Are Restrictive!

- *To be* is copula:

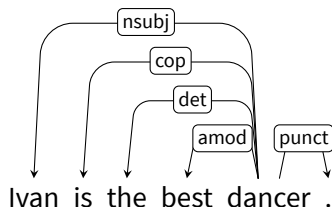


- *To become* is not copula:

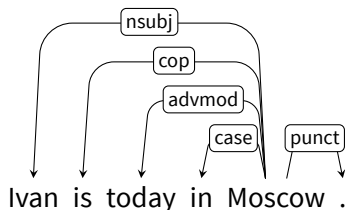


# Once Copula, Always Copula!

- This is parallel with Russian:

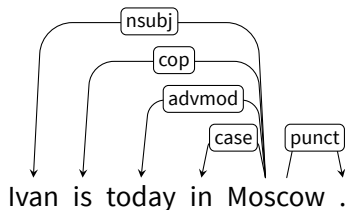


- This is also parallel with Russian:

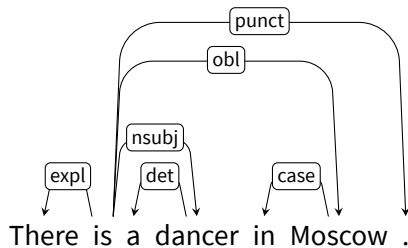


# Well, Almost...

- This is parallel with Russian:

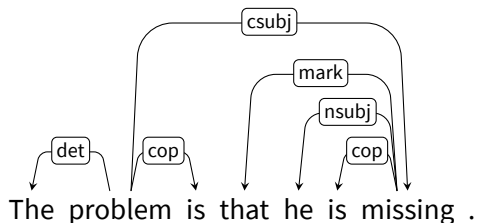


- But not with this in English:

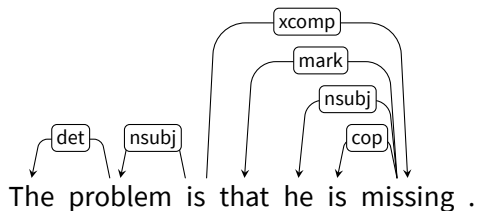


# Clauses and Copula

- A clause can be the subject:



- But it cannot be annotated as the nonverbal predicate:



# Wrapping Up



# Wrapping Up

- UD is now the de-facto standard



# Wrapping Up

- UD is now the de-facto standard
- Slavic languages very well represented
- Consistency matters!





# Wrapping Up

- UD is now the de-facto standard
- Slavic languages very well represented
- Consistency matters!
- Get involved. It's fun!



Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Rogier Blokland, Victoria Bobicev, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Carly Dickerson, Peter Dirix, Kaja Dobrovoljic, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Tomaž Erjavec, Aline Etienne, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökirmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Radu Ion, Elena Irimia, Olájúdí Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanera, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Kamil Kopaciewicz, Natalia Kotsyba, Simon Krek, Sookyung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phương Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Măranduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horfiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lương Nguyễn Thị, Huyền Nguyễn Thị Minh, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adedayo Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Slav Petrov, Jussi Piitulainen, Emily Pitler, Barbara Plank, Thierry Poibeau, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Reemo Pyysalo, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Michael Rießler, Larissa Rinaldi, Laura Rituma, Luísa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roşca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddāh, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Muh Shohibussirri, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umot Sulubacak, Zsolt Szántó, Dima Taji, Yuta Takahashi, Takaaki Tanaka, Isabelle Tellier, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Jing Xian Wang, Jonathan North Washington, Seyi Williams, Mats Wirén, Tsegay Woldemariam, Tak-sum Wong, Chunxiao Yan, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Manying Zhang, Hanzhi Zhu

Džakuju so!  
Prašenja?

Dziękuję!  
Pytania?

Дякую!  
Питання?

Děkuji!  
Otázky?

Thank You!  
Questions?

Ďakujem!  
Otázky?

Hvala!  
Vprašanja?

Hvala!  
Pitanja?

Благодаря!  
Въпроси?