The State of Universal Dependencies

Marie-Catherine de Marneffe, Joakim Nivre, Daniel Zeman

March 16, 2023

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

CoNLL-U Format

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC
1	Le	le	DET	_	_	2	det	_	_
2	chat	chat	NOUN	_	_	3	nsubj	_	_
3	boit	boire	VERB	_	_	0	root	_	_
4-5	du	_	_	_	_	_	_	_	_
4	de	de	ADP	_	_	6	case	_	_
5	le	le	DET	_	_	6	det	_	_
6	lait	lait	NOUN	_	_	3	obj	_	SpaceAfter=No
7	.	.	PUNCT	_	_	3	punct	_	_

- Revised and extended version of CoNLL-X format
- Two-level segmentation and enhanced dependencies

Word Segmentation

Let's go to the sea.

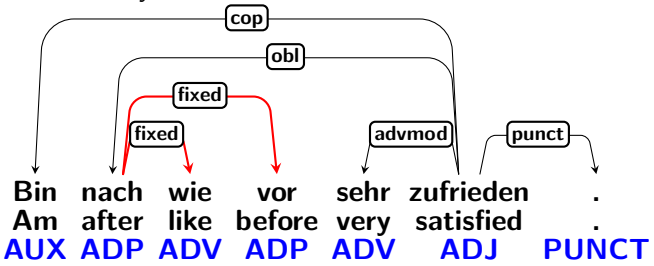
Vámonos	**al**	**mar**	**.**	**Vamos**	**nos**	**a**	**el**	**mar**	**.**
VERB?	**X**	**NOUN**	**PUNCT**	**VERB**	**PRON**	**ADP**	**DET**	**NOUN**	**PUNCT**

- **Syntactic word** vs. orthographic word
- **Multi-word tokens**
- Two-level scheme:
  - Tokenization (low level, punctuation, concatenative)
  - Word segmentation (higher level, not necessarily concatenative)

Fixed Expressions

One syntactic word spans several orthographic words?

I am still very satisfied.



Part-of-Speech Tags

http://universaldependencies.org/u/pos/index.html

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

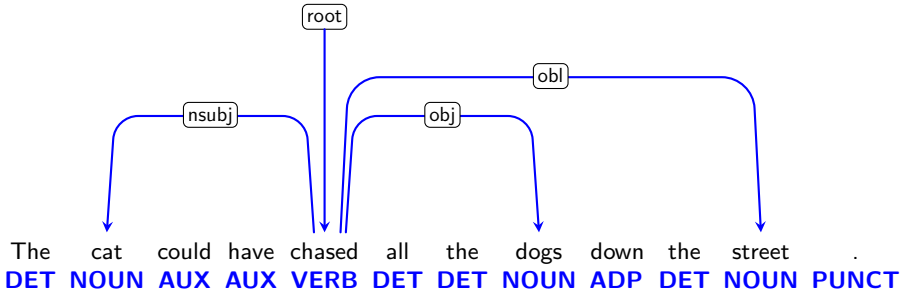
Lexical	Inflectional ("Nominal")	Inflectional ("Verbal, Pronominal")
PronType	Gender	VerbForm
NumType	Animacy	Mood
Poss	NounClass	Tense
Reflect	Number	Aspect
Foreign	Case	Voice
	Definite	Evident
	Degree	Polarity
Abbr		Person
Typo		Polite
		Clusivity

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

Syntax

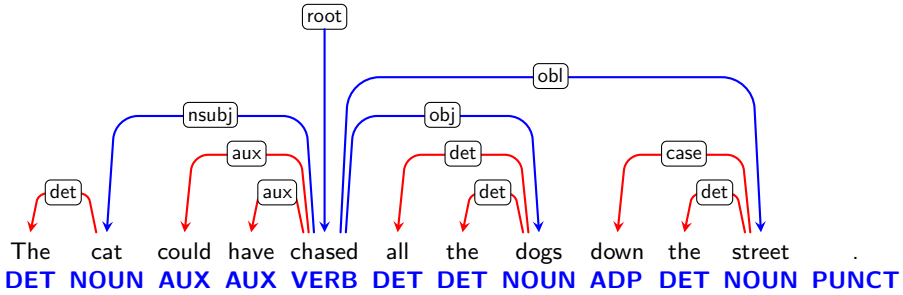
The	cat	could	have	chased	all	the	dogs	down	the	street	.
DET	**NOUN**	**AUX**	**AUX**	**VERB**	**DET**	**DET**	**NOUN**	**ADP**	**DET**	**NOUN**	**PUNCT**

Syntax



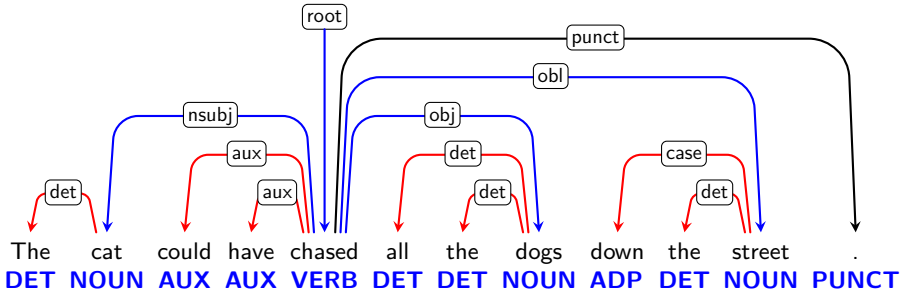
- Content words are related by dependency relations

Syntax



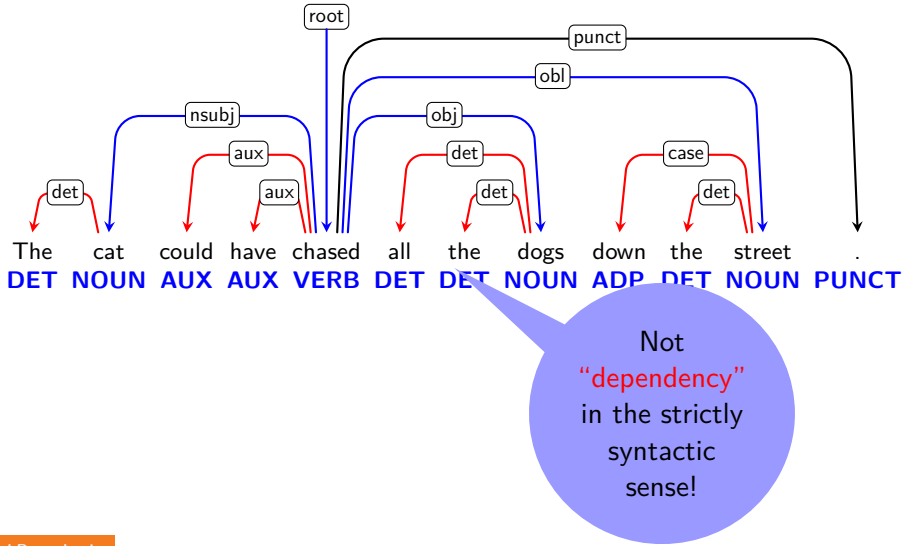
- Content words are related by dependency relations
- Function words attach to closest content words

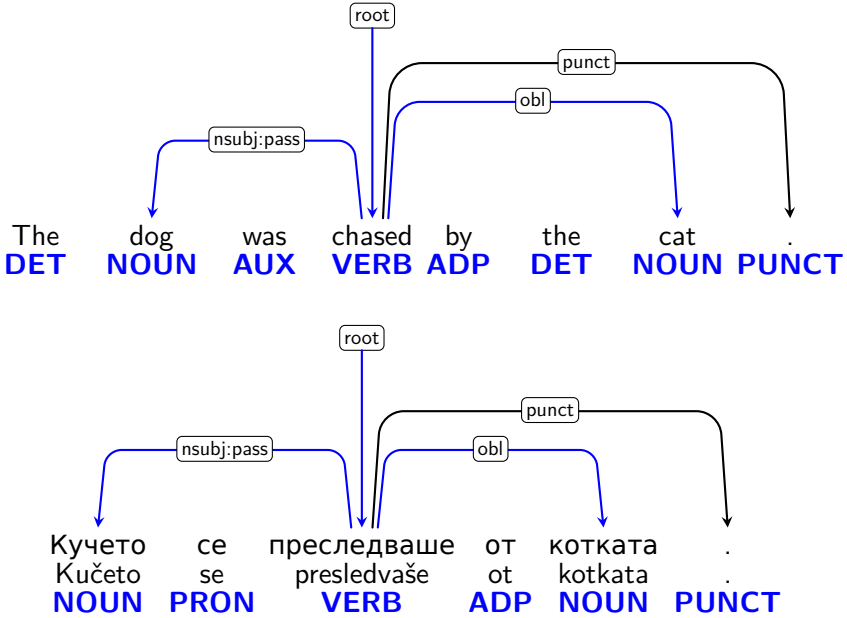
Syntax

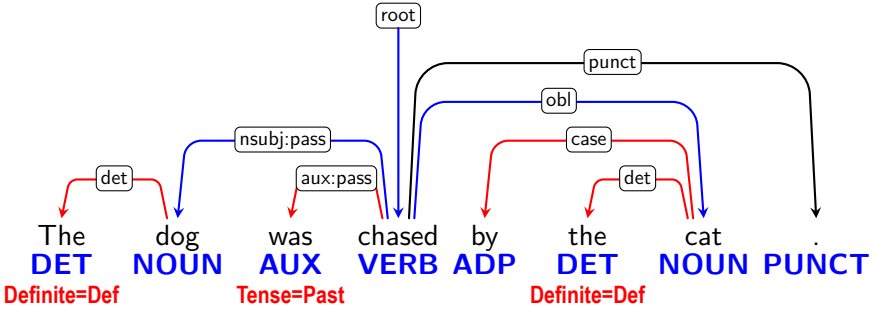


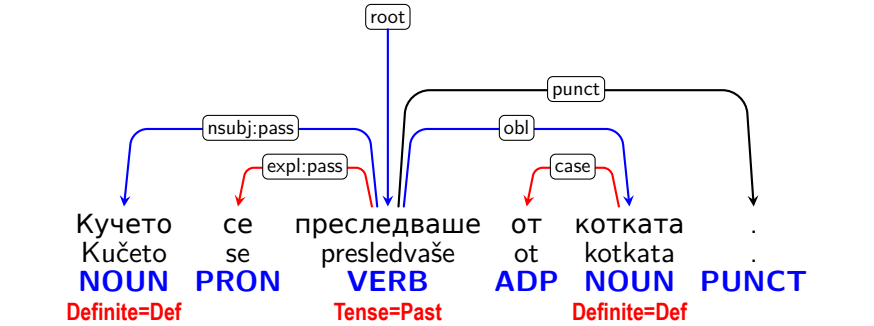
- Content words are related by dependency relations
- Function words attach to closest content words
- Punctuation attach to head of phrase or clause

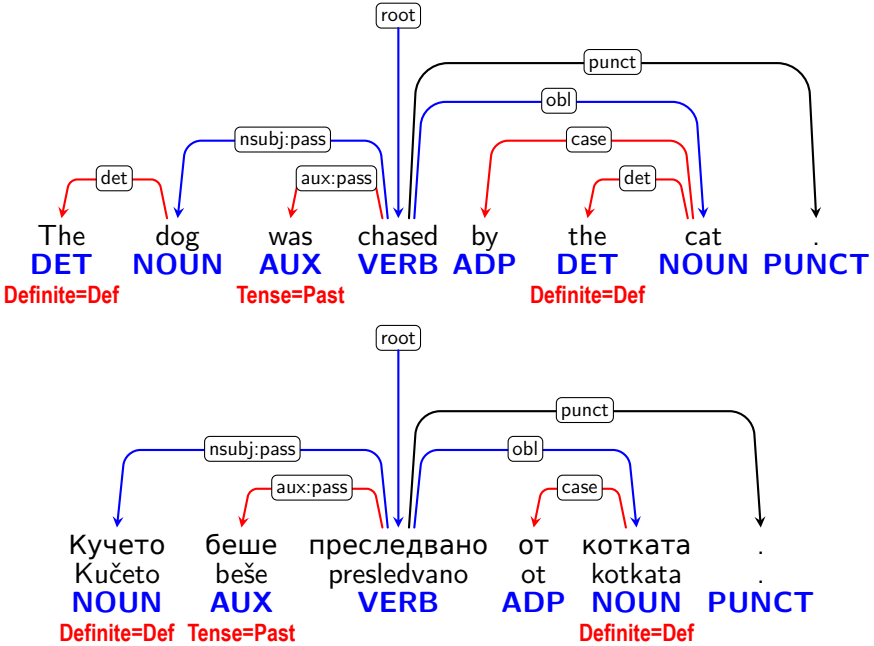
Syntax

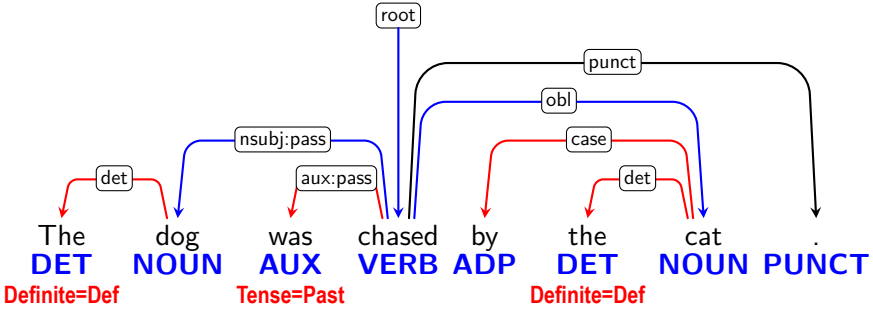

root
punct
obl
nsubj
obj
aux
det
case
det
aux
det
det
The cat could have chased all the dogs down the street .
DET NOUN AUX AUX VERB DET DET NOUN ADP DET NOUN PUNCT
Not "dependency" in the strictly syntactic sense!

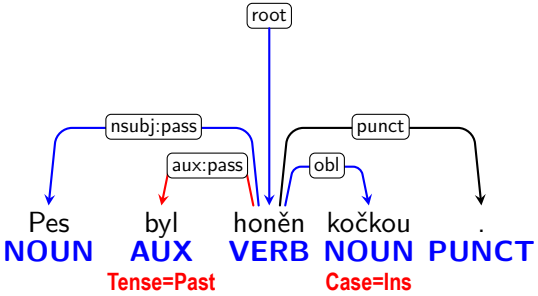

root
punct
obl
nsubj:pass
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT
root
punct
nsubj:pass
obl
Кучето се преследваше от котката .
Kučeto se presledvaše ot kotkata .
NOUN PRON VERB ADP NOUN PUNCT


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

Dependents of Clauses (Verbal or Not)

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

Dependents of Verbs, Adjectives and Adverbs

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

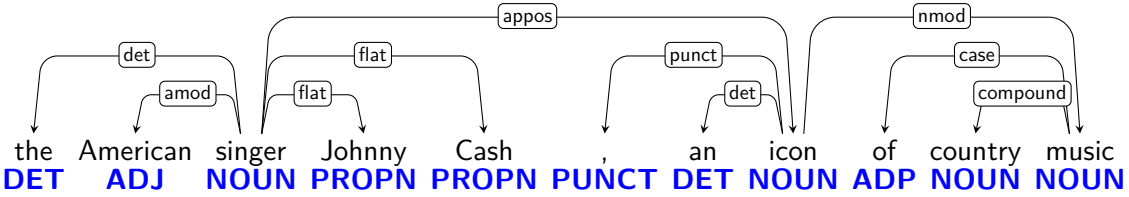
Dependents of Nominals

Nominal	Clausal	Modifier	Function
nmod	acl	amod	det
appos		nummod	case

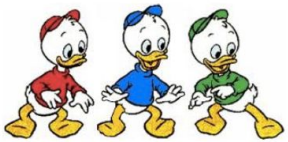
Noun Phrase

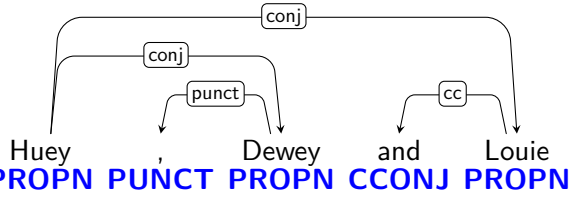
Dependents of Nominals

Nominal	**Clausal**	**Modifier**	**Function**
nmod	acl	amod	det
appos		nummod	case
compound			clf
flat			



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

Multiword Expressions

Relation	Examples
fixed	*as well, by and large, according to, more than*
flat	*president Havel, New York, four thousand*
compound	*phone book, dress up*
goeswith	*notwith standing, with out*

- UD annotation almost does not permit "words with spaces"
  - Multiword expressions are analyzed using special relations
  - The fixed, flat and goeswith relations are always head-initial
  - The compound relation reflects the internal structure
- Words with spaces allowed in exceptional cases:
  - Vietnamese (spaces delimit syllables, not words)
  - Numbers ("1 000 000")
  - Possibly other approved cases, e.g. multi-word abbreviations

Other Relations

Other Relations

Relation	Explanation
parataxis	Loosely linked clauses of same rank
list	Lists without syntactic structure
orphan	Orphans in ellipsis linked together
reparandum	Disfluency linked to (speech) repair
dep	Unspecified dependency
root	The single syntactically independent element of the sentence

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

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