

We annotate morphology and syntax of 90 languages in a uniform way.

Universal Dependencies

Daniel Zeman

zeman@ufal.mff.cuni.cz

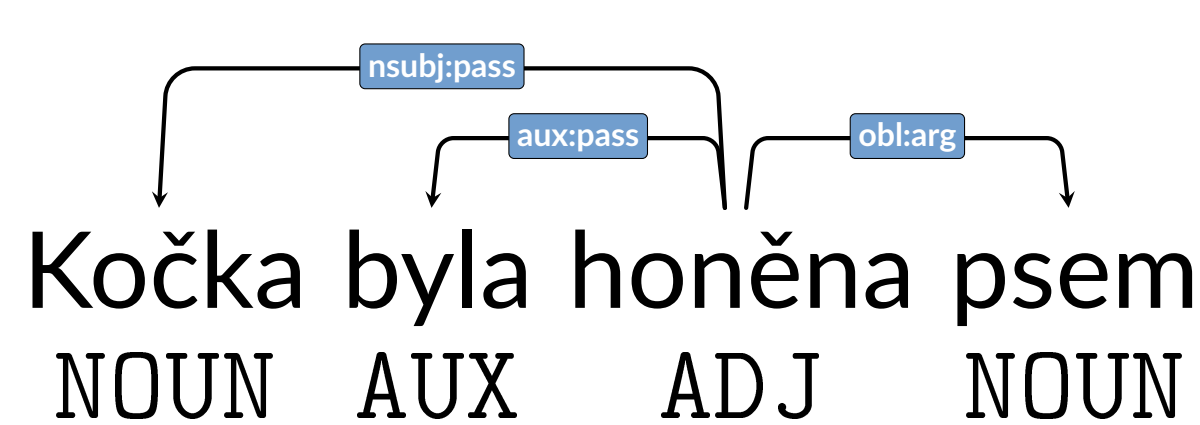
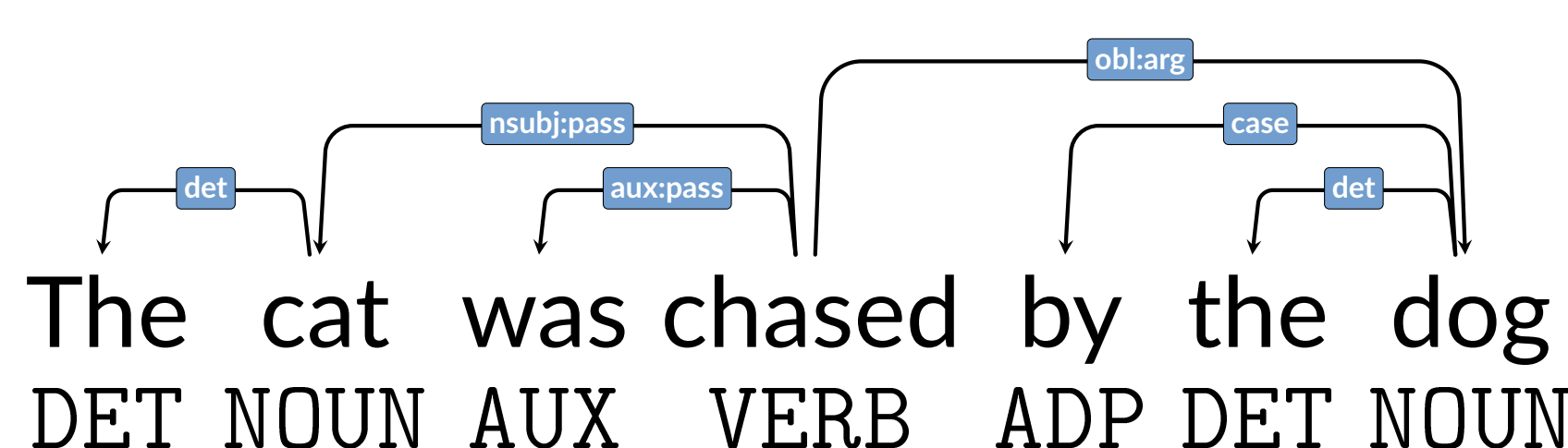
Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Prague, Czechia

Introduction

Universal Dependencies is an open community effort to create cross-linguistically consistent treebank annotation for many languages within a dependency-based lexicalist framework. The annotation consists in a linguistically motivated word segmentation; a morphological layer comprising lemmas, universal part-of-speech tags, and standardized morphological features; and a syntactic layer focusing on syntactic relations between predicates, arguments and modifiers.

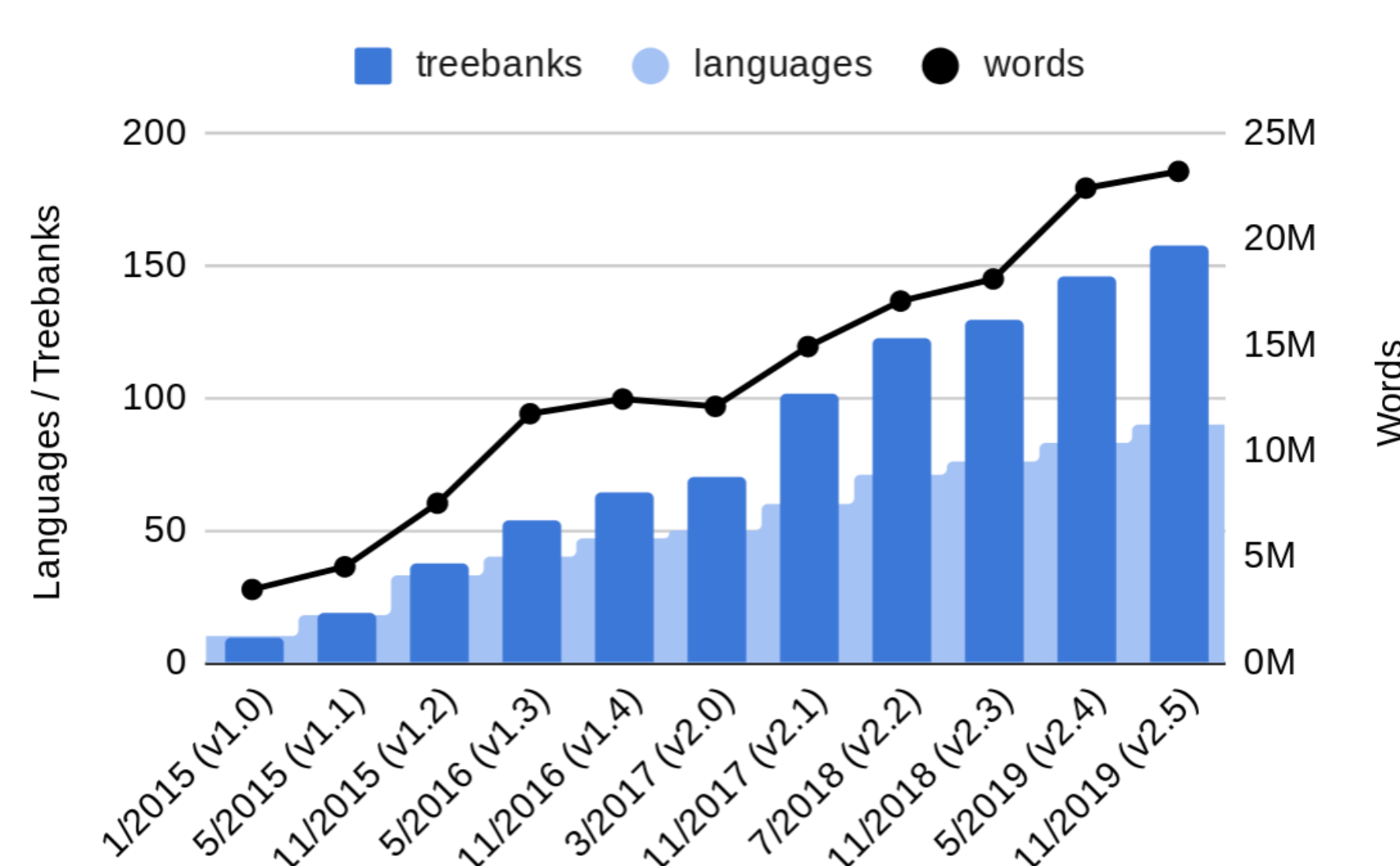
```
# text = Die Katze war vom Hunde gejagt.
# text_en = The cat was chased by the dog.
```

ID	FORM	LEM	UPOS	XPOS	FEATS	H	DEPREL
1	Die	der	DET	ART	Case=Nom Definite=Def ...	2	det
2	Katze	Katze	NOUN	NN	Case=Nom Gender=Fem ...	7	nsubj:pass
3	war	sein	AUX	VAFIN	Mood=Ind Number=Sing ...	7	aux:pass
4-5	vom	-	-	-	-	-	-
4	von	von	ADP	APPR	-	6	case
5	dem	der	DET	ART	Case=Dat Definite=Def ...	6	det
6	Hunde	Hund	NOUN	NN	Case=Dat Gender=Masc ...	7	obl:agent
7	gejagt	jagen	VERB	VVPP	Tense=Past VerbForm=Part	0	root
8	.	.	PUNCT	\$.	-	7	punct



Coverage

- Indo-European:** Armenian, Ancient Greek, Greek, Breton, Irish, Scottish, Welsh; Afrikaans, Danish, Dutch, English, Faroese, German, Gothic, 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, Urdu, Hindi, Bhojpur, Marathi, Sanskrit
- Uralic:** Erzya, Estonian, Finnish, Hungarian, Karelian, Livvi, Komi Permyak+Zyrian, Moksha, Sámi North+Skolt
- Dravidian:** Tamil, Telugu
- Turkic:** Kazakh, Turkish, Uyghur
- Afro-Asiatic:** Akkadian, Amharic, Arabic, Assyrian, Coptic, Hebrew, Maltese
- Sino-Tibetan:** Cantonese, Classical Chinese, Chinese
- Austro-Asiatic:** Vietnamese; **Tai-Kadai:** Thai
- Austronesian:** Indonesian, Tagalog
- Other:** Buryat, Japanese, Korean, Basque, Swedish Sign, Nijja, Bambara, Wolof, Yoruba, Warlpiri, Mbyá Guaraní



Traditional POS	UPOS	Category
noun	NOUN	Common noun
	PROPN	Proper noun
verb	VERB	Main verb
	AUX	Auxiliary verb or other TAM particle
adjective	ADJ	Adjective
	DET	Determiner
	NUM	Numeral
adverb	ADV	Adverb
pronoun	PRON	Pronoun
preposition	ADP	Adposition (preposition/postposition)
conjunction	CCONJ	Coordinating conjunction
	SCONJ	Subordinating conjunction
interjection	INTJ	Interjection
-	PART	Particle (special single word markers)
-	X	Other (e.g., foreign language)
-	SYM	Non-punctuation symbol
-	PUNCT	Punctuation

Feature	Values
pronominal type	PronType Art Dem Emp Exc Ind Int Neg Prs Rcp Rel Tot
numeral type	NumType Card Dist Frac Mult Ord Range Sets
possessive	Poss Yes
reflexive	Reflex Yes
foreign word	Foreign Yes
abbreviation	Abbr Yes
wrong spelling	Typo Yes
gender	Gender Com Fem Masc Neut
animacy	Animacy Anim Hum Inan Nhum
noun class	NounClass Bantu1-23 Wol1-12 ...
number	Number Coll Count Dual Grpa Grpl Inv Pauc Plur Ptan Sing Tri
case	Case Abs Acc Erg Nom Abe Ben Cau Cmp Cns Com Dat Dis Equ Gen Ins Par Tem Tra Voc Abl Add Ade All Del Ela Ess Ill Ine Lat Loc Per Sub Sup Ter
definiteness	Definiteness Com Cons Def Ind Spec
comparison	Degree Abs Cmp Equ Pos Sup
verbal form	VerbForm Conv Fin Gdv Ger Inf Part Sup Vnoun
mood	Mood Adm Cnd Des Imp Ind Jus Nec Opt Pot Prp Qot Sub
tense	Tense Fut Imp Past Pqp Pres
aspect	Aspect Hab Imp Iter Perf Prog Prosp
voice	Voice Act Antip Cau Dir Inv Mid Pass Rcp
evidentiality	Evident Fh Nfh
polarity	Polarity Neg Pos
person	Person 0 1 2 3 4
politeness	Polite Elev Form Humb Infm
clusivity	Clusivity In Ex

head \ dependent	Nominals	Clauses	Modifier words	Function words
Clausal core arguments	nsubj obj iobj	csubj ccomp xcomp		
Clausal non-core arguments	obl vocative expl dislocated	advcl advmod discourse	aux cop mark	
Nominal dependents	nmod appos nummod	acl amod	det clf case	
Coordination	MWE	Loose Special	Other	
conj	fixed compound	list parataxis reparandum	orphan goeswith	punct root
cc	flat			dep

