Assessing Annotation Consistency in Universal Dependencies Corpora

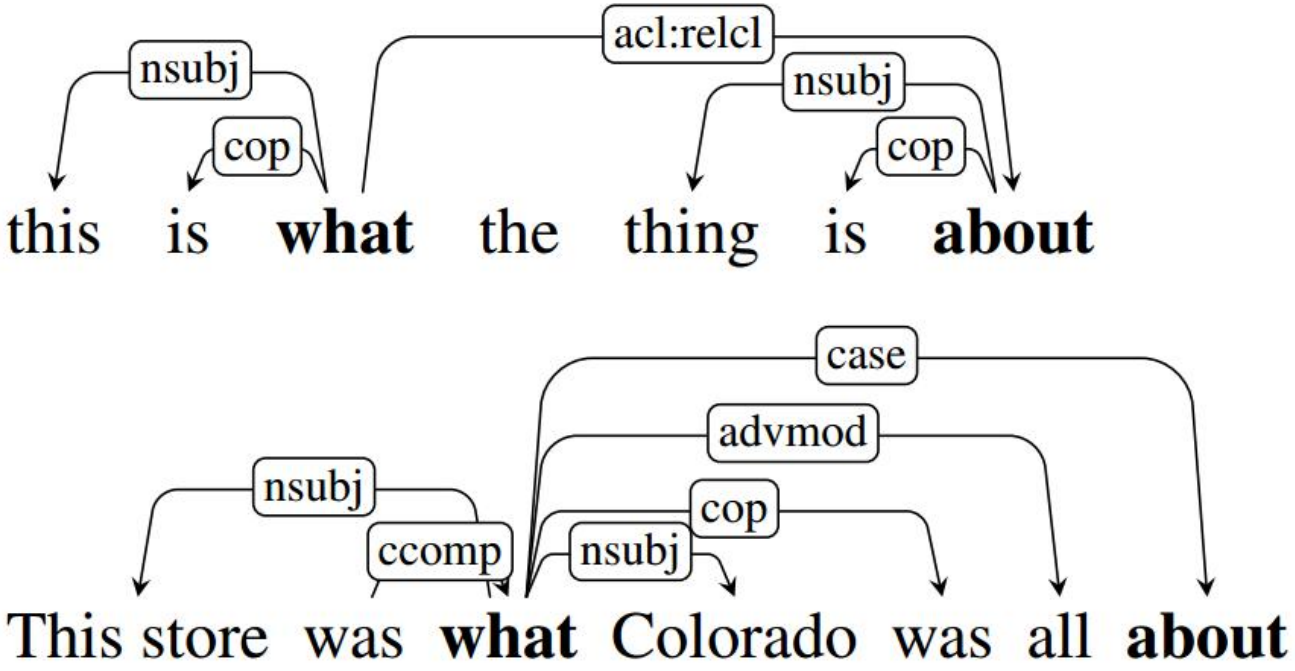
Marie-Catherine de Marneffe, Matias Grioni

Jenna Kanerva, Filip Ginter

THE OHIO STATE UNIVERSITY

Turun yliopisto
University of Turku

Current Inconsistencies

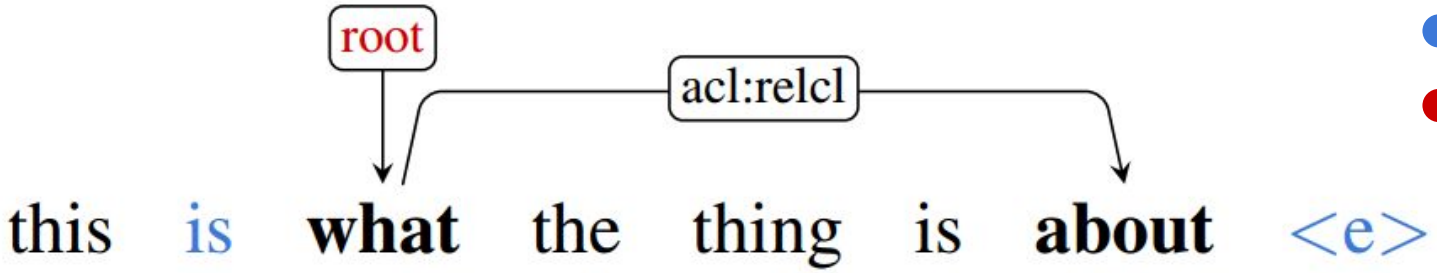


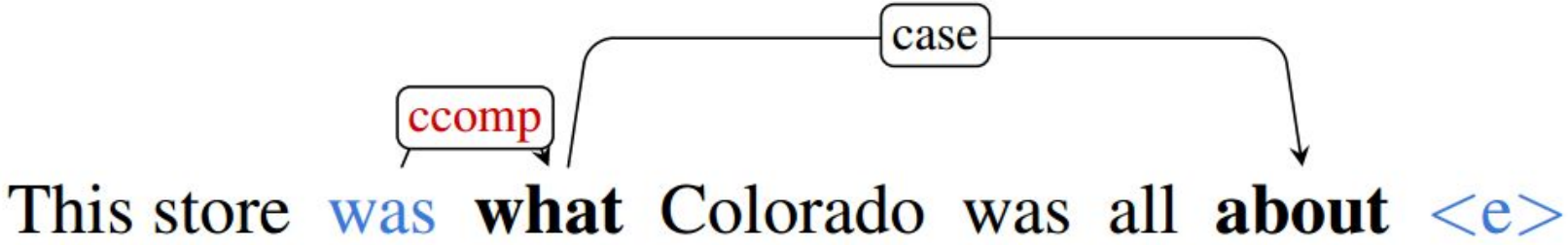
Variation Nuclei for Dependency Annotation

Boyd et al., 2008

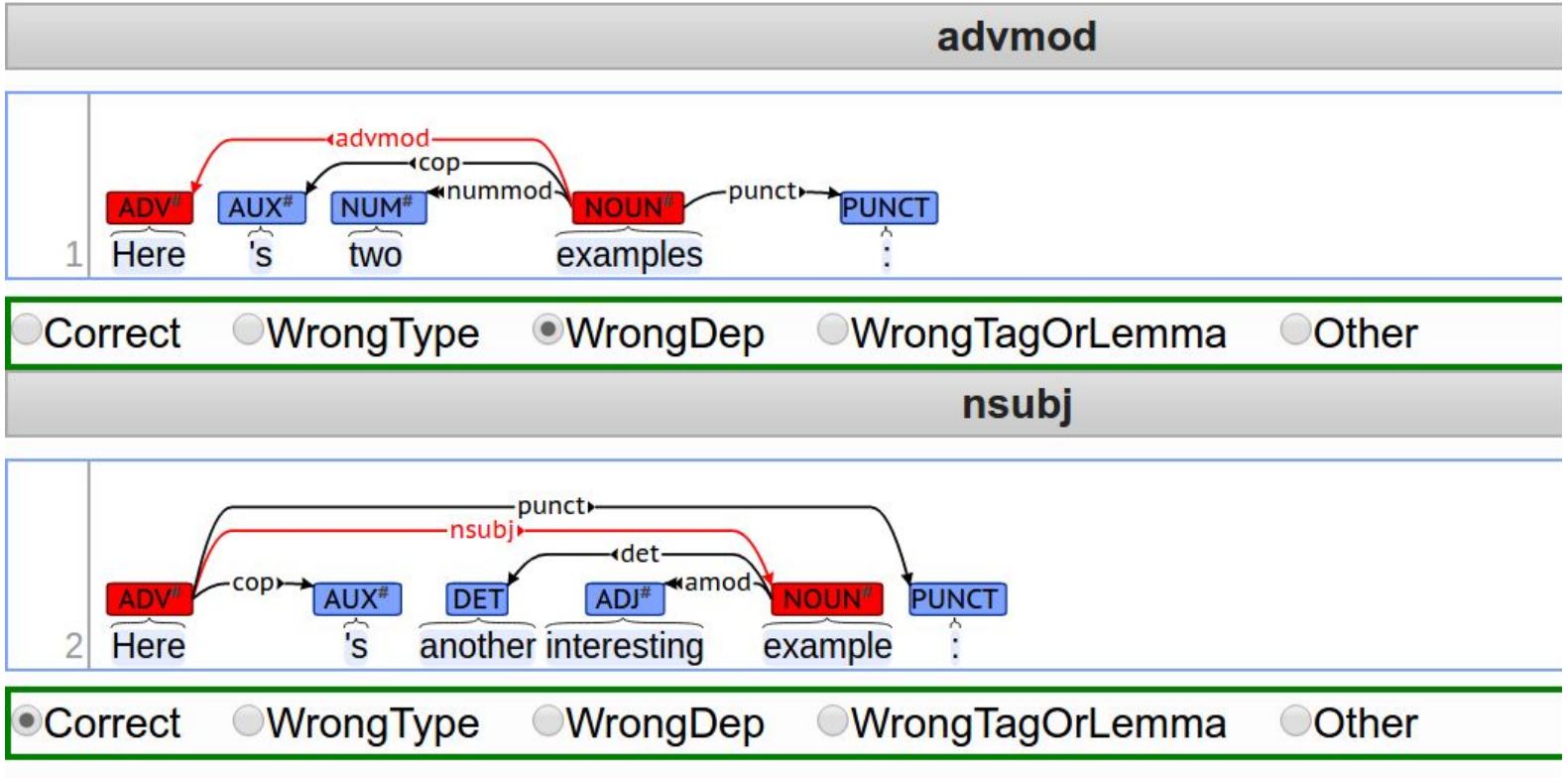
Non-fringe heuristic

Dependency Heuristic





Annotation Tool



Balancing Precision and Recall for Rich Morphology

	LEMMAS			WORDS	
	"Non-fringe"	"Dependency context"		"Non-fringe"	
	Precision (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)
English	62	76	66	72	79
French	65	64	47	76	73
Finnish	19	21	81	72	75

Leveraging Big Data to Ground Correct Annotation

- CoNLL ‘17 Shared Task data (Ginter et al., 2017)
- Several billion words of automatically parsed web data

	# tree errors	“Non-fringe” Precision	% in Boyd
English	54	41%	38%
French	74	57%	36%
Finnish	10	16%	30%

universaldependencies.org/fixit

- Introduce approach to greater UD community
- Intuitive web-based tool to flag inconsistencies
- Inconsistency output for 80+ languages already online