

# Co-constructions in UD

how to annotate, validate, visualize?

Arianna Masciolini

Språkbanken, University of Gothenburg

- ❖ PhD student at the University of Gothenburg
- ❖ current main focus: treebanking L2 Swedish **essays**
- ❖ recent side project: annotation of **interactional** L2 Italian data (StraParla)
- ❖ experience with SpLAn-UD: < 1 week

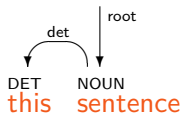
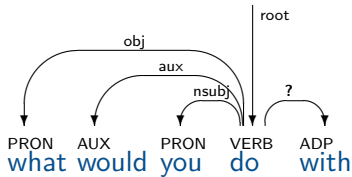
# A first example



*what would you do with this sentence*

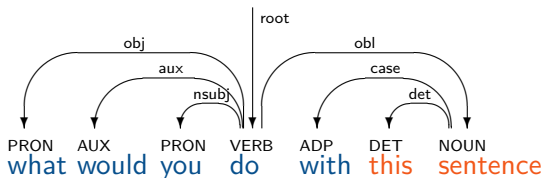
- ❖ Speaker A: blue
- ❖ Speaker B: orange

# Option 1: splitting by speaker



- ❑ how meaningful is this annotation?
- ❑ what to do with "with"?

# Option 2: single sentence



(with speaker IDs in the MISC field)

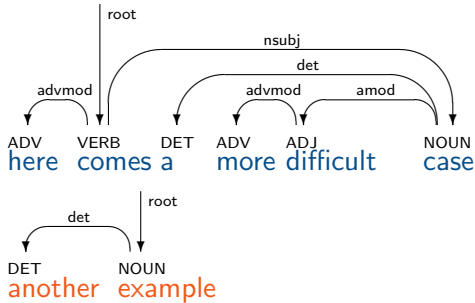
- ❑ complete sentence
- ❑ easy-to-visualize co-construction

*here comes a more difficult case*  
*another example*

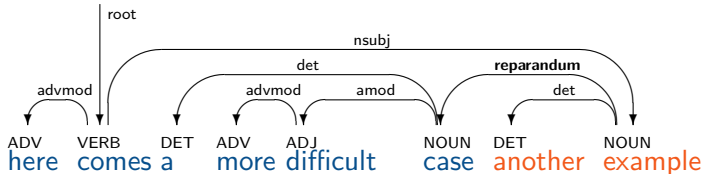
A and B :

- ❖ fill in the same slot with different values
- ❖ typically (but not necessarily) overlap

# Option 1: splitting by speaker



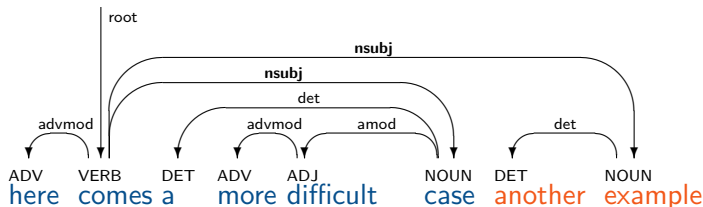
# Option 2: reparandum



- ❖ makes sense if B is correcting A
- ❖ ...but that is not always the case



# Option 3: double subject



❖ intuitive analysis

❖ problem:

[Line 55 Node 2]: [L3 Syntax too-many-subjects]  
Multiple subjects [6, 8] not subtyped as ':outer'.  
Outer subjects are allowed if a clause acts as the  
predicate of another clause.

- ❖ if spoken data is segmented in a speaker-based fashion, some information is inevitably lost or hidden (at least from current UD tools)
- ❖ if multi-speaker (co-constructed) “sentences” are allowed, the validator needs to account for speaker information