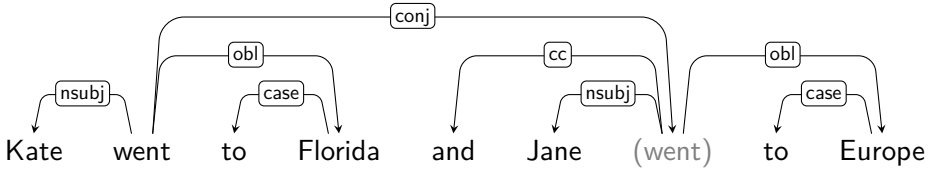
Enhanced Universal Dependencies

Daniel Zeman

April 27, 2023

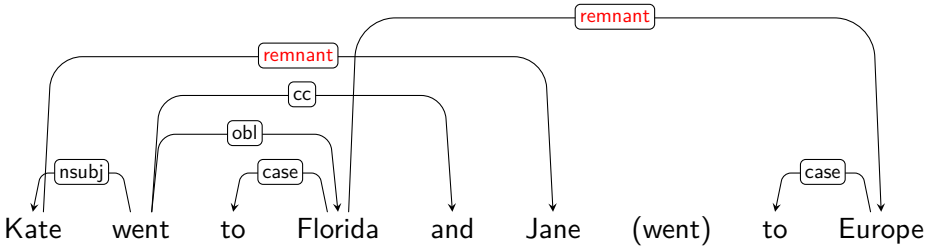
Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

Deleted Predicates in Coordination

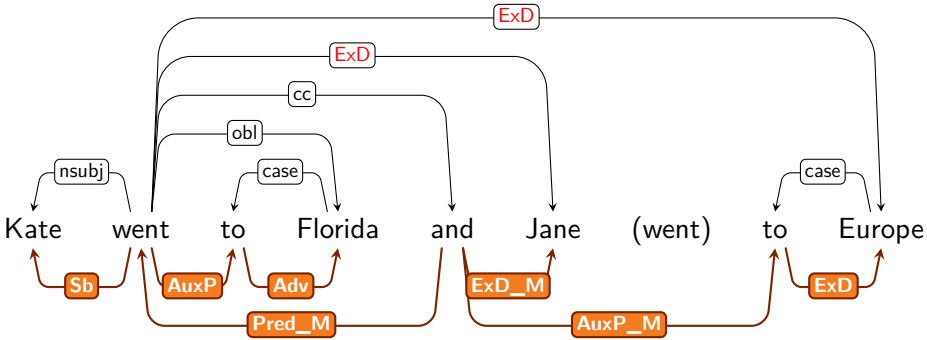


- Some treebanks would use an empty node to represent the second *went*.
- UD enhanced representation allows empty nodes!
- But the basic representation sticks with the overt words.

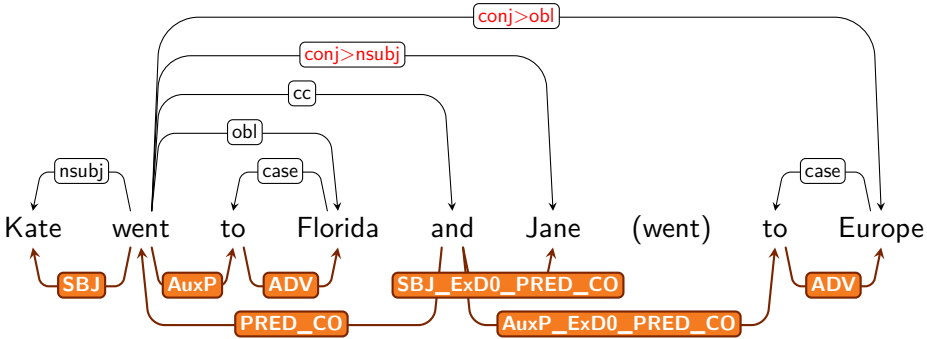
UD V1: The `remnant` Relation



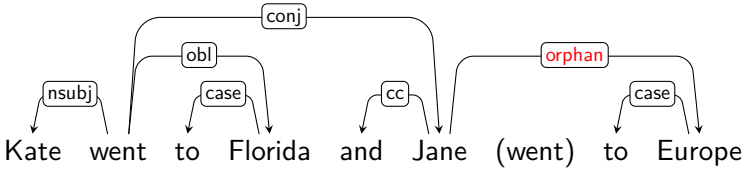
PDT: The ExD Relation



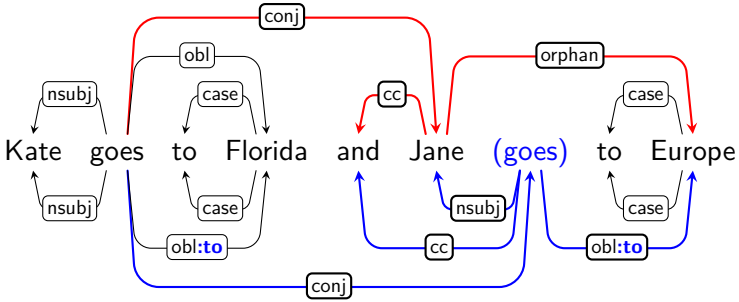
Perseus Treebanks: Chained Relations



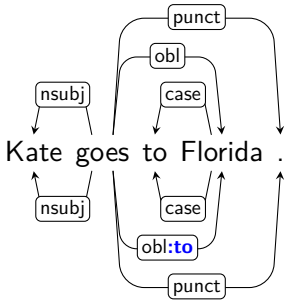
UD V2 Basic Dependencies: The `orphan` Relation



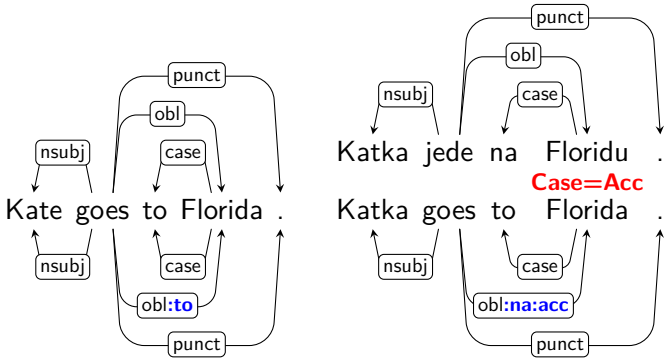
UD Enhanced Dependencies: Gapping and Stripping



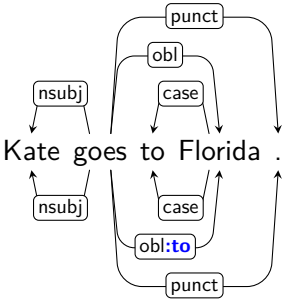
Enhanced UD: Case Information in Dependency Label

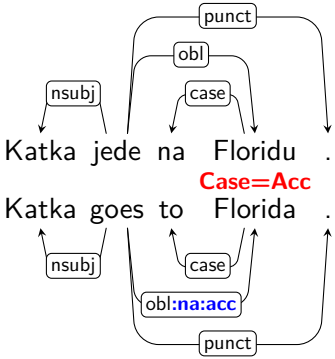


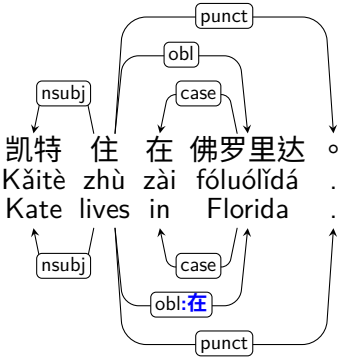
Enhanced UD: Case Information in Dependency Label



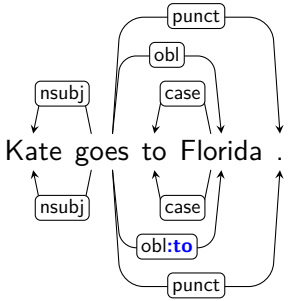
Enhanced UD: Case Information in Dependency Label

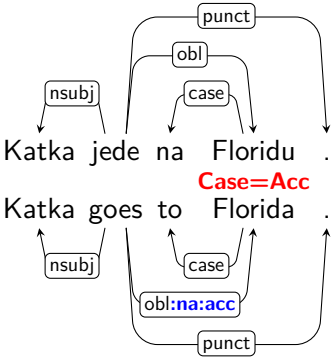


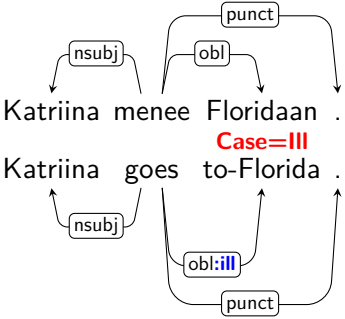




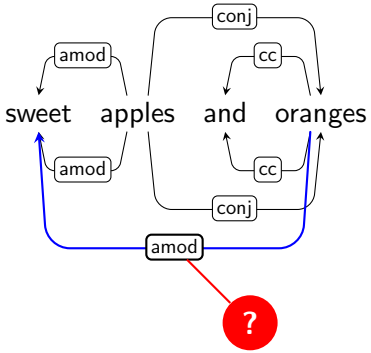
Enhanced UD: Case Information in Dependency Label



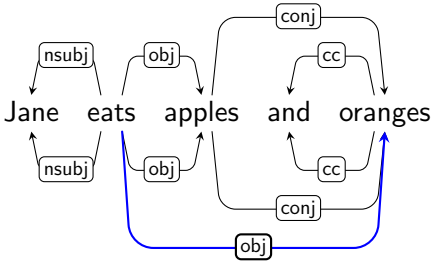




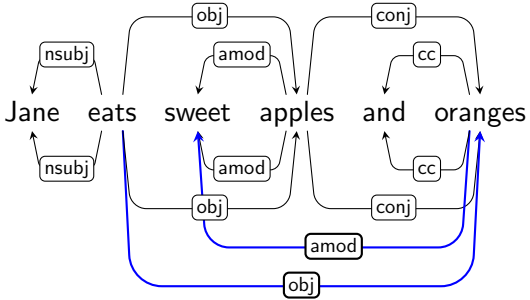
Enhanced UD: Shared Dependent of Coordination



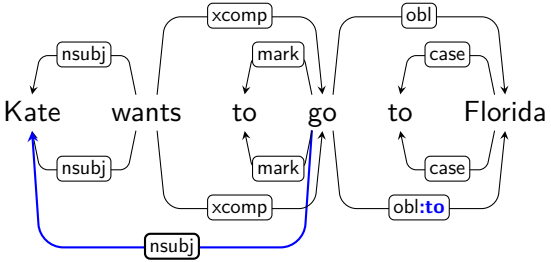
Enhanced UD: Parent of Coordination



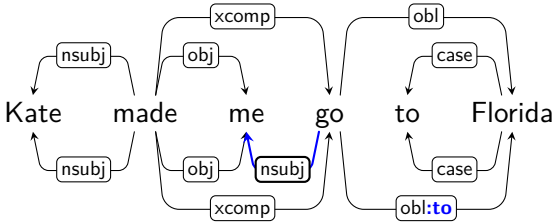
Enhanced UD: Coordination



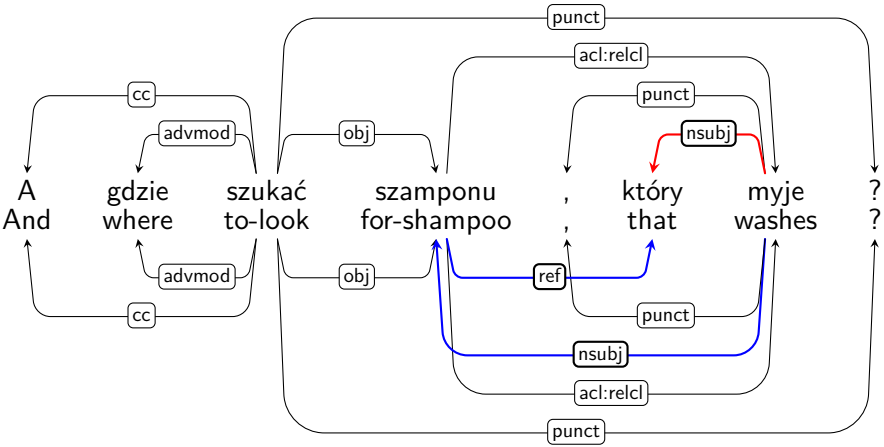
Enhanced UD: External Subject of Controlled Predicate



Enhanced UD: External Subject in Object-Control Construction

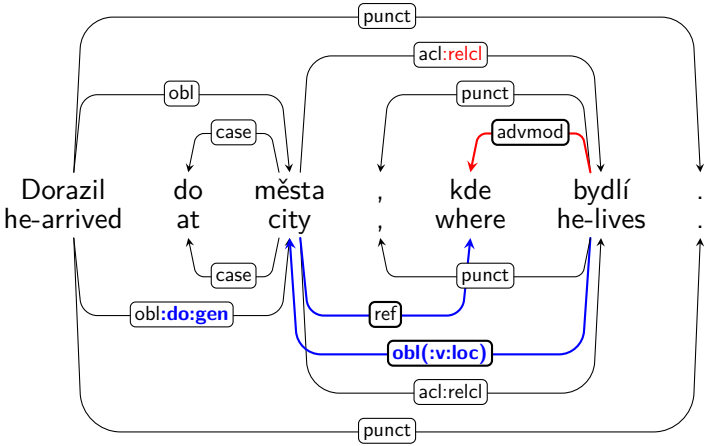


Enhanced UD: Relative Clauses

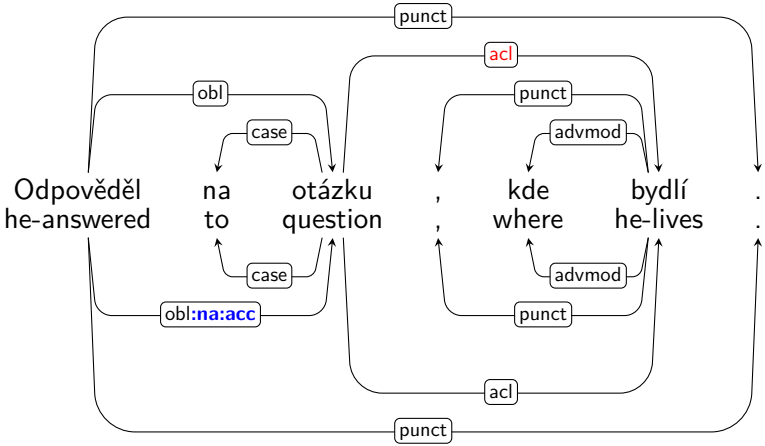


"And where to look for shampoo that works?"

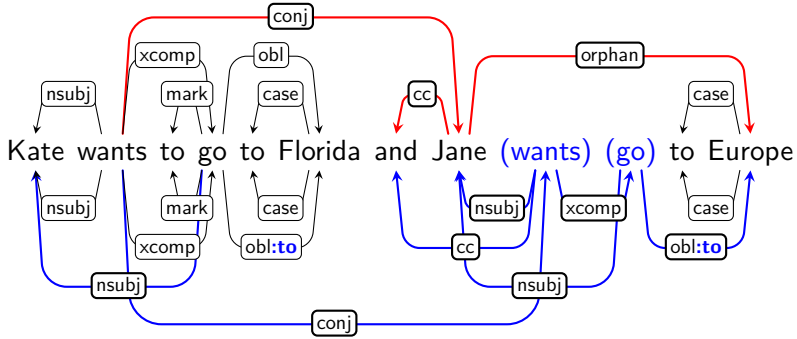
How to Recognize Relative Clauses



How to Recognize Relative Clauses



Enhanced UD: Gapping + Control Verb Construction



Enhanced UD: Six Enhancements

- Null nodes for gapping (27 treebanks in UD 2.11)
  - But in 4 of them the null nodes are actually not used (only) for gapping
- Coordination: Common parent (27 treebanks)
- Coordination: Shared dependents (26 treebanks)
- External subjects of controlled predicates (23 treebanks)
- Cyclic dependencies to/from relative clauses (25 treebanks)
- Case-enhanced dependency labels (26 treebanks)

- All 6 types: 18 treebanks, 11 languages
- At least 1 type: 33 treebanks, 21 languages
- Only basic UD: 210 treebanks

Stanford Enhancer

- Part of Stanford CoreNLP (Java)
- Rules from basic to enhanced UD

Stanford Enhancer

- Part of Stanford CoreNLP (Java)
- Rules from basic to enhanced UD

- Gapping: embeddings for similarity of arguments

Stanford Enhancer

- Part of Stanford CoreNLP (Java)
- Rules from basic to enhanced UD

- Gapping: embeddings for similarity of arguments
- Coordination:
  - Parent propagation: deterministic
  - Shared dependents: heuristics (human desirable!)

Stanford Enhancer

- Part of Stanford CoreNLP (Java)
- Rules from basic to enhanced UD

- Gapping: embeddings for similarity of arguments
- Coordination:
  - Parent propagation: deterministic
  - Shared dependents: heuristics (human desirable!)
- External subjects: heuristics (subject vs. object control)

Stanford Enhancer

- Part of Stanford CoreNLP (Java)
- Rules from basic to enhanced UD

- Gapping: embeddings for similarity of arguments
- Coordination:
  - Parent propagation: deterministic
  - Shared dependents: heuristics (human desirable!)
- External subjects: heuristics (subject vs. object control)
- Relative clauses: need `acl:relcl` and list of relative pronouns

Stanford Enhancer

- Part of Stanford CoreNLP (Java)
- Rules from basic to enhanced UD

- Gapping: embeddings for similarity of arguments
- Coordination:
  - Parent propagation: deterministic
  - Shared dependents: heuristics (human desirable!)
- External subjects: heuristics (subject vs. object control)
- Relative clauses: need `acl:relcl` and list of relative pronouns
- Case-enhanced labels: deterministic

Conversion from non-UD Data: Extra Information?

- Analytical layer of Prague-style treebanks: shared dependents of coordination are known!