Enhanced Universal Dependencies

Daniel Zeman

April 27, 2023
• 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.
Kate went to Florida and Jane (went) to Europe
Kate went to Florida and Jane (went) to Europe.
Kate went to Florida and Jane (went) to Europe
UD V2 Basic Dependencies: The *orphan* Relation

Kate went to Florida and Jane (went) to Europe

Diagram:
- Kate (nsubj)
- went (cc)
- to (obl)
- Florida (case)
- and
- Jane (nsubj)
- (went) (cc)
- to (obl)
- Europe (case)

Dependencies:
- Kate -> went
- went -> to
- to -> Florida
- and
- Jane -> (went)
- (went) -> to
- to -> Europe
- conj
- orphan
Kate goes to Florida and Jane (goes) to Europe.
Kate goes to Florida.
Kate goes to Florida.

Katka jede na Floridu.
Kate goes to Florida.

Katka jede na Floridu.

凯特住在佛罗里达。
Kate goes to Florida.

Katka jede na Floridu.

Case=Acc

Katriina menee Floridaan.

Case=Ill
Enhanced UD: Shared Dependent of Coordination

- Enhanced Universal Dependencies

Diagram:

- sweet
- apples
- and
- oranges
- amod
- conj
- cc

Question mark indicating uncertainty or a question about the dependency relationships.
Jane eats apples and oranges
Enhanced UD: Coordination

Jane eats sweet apples and oranges
Kate wants to go to Florida
Kate made me go to Florida.
“And where to look for shampoo that works?”
How to Recognize Relative Clauses

Dorazil he-arrived at města, kde bydlí.

he-arrived at city, where he-lives.
How to Recognize Relative Clauses

Odpověděl na otázku, kde bydlí.

he-answered to question, where he-lives.
Kate wants to go to Florida and Jane wants to go to Europe.
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
• Part of Stanford CoreNLP (Java)
• Rules from basic to enhanced UD
• Part of Stanford CoreNLP (Java)
• Rules from basic to enhanced UD

• **Gapping:** embeddings for similarity of arguments
• 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
• 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
Analytical layer of Prague-style treebanks: *shared dependents of coordination* are known!

Karel píše a Lucie prodává knihy.

**Conversion from non-UD Data: Extra Information?**

Karel writes and Lucie sells books.