Universal Dependencies
A Framework for Morphosyntactic Annotation

Joakim Nivre
Uppsala University
Department of Linguistics and Philology

Based on collaborative work with Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher Manning, Ryan McDonald, Natalia Silveira, Slav Petrov, Sampo Pyysalo, Sebastian Schuster, Reut Tsarfaty, Francis Tyers, Daniel Zeman and many others
Why?

Cross-linguistically consistent morphosyntactic annotation

Facilitate multilingual research in NLP and linguistics

• Meaningful linguistic analysis across languages
• Syntactic parsing in cross-lingual settings
• NLP systems for multiple languages
• Facilitate resource-building for new languages

Complement – not replace – language-specific schemes
Focus on grammatical relations between (content) words
Who?

Open community effort – a big tent

UD v2.2: 71 languages, 122 treebanks, 266 contributors

Come join us at http://universaldependencies.org
En kattjager och mossa det
A cat chases rats and mice
Toutefois, les filles adorent les desserts.
toutefois, la fille adorer les dessert.
Universal Dependencies

http://universaldependencies.org

Part-of-speech tags
Universal Dependencies

http://universaldependencies.org

Syntactic relations

Part-of-speech tags

Morphological features

Toutefois, les filles adorent les desserts.

ADV PUNCT DET NOUN VERB DET NOUN PUNCT

Definite=Def Gender=Fem Number=Plur Number=Plur Person=3 Tense=Pres

Definite=Def Gender=Masc Number=Plur Number=Plur

Enkattjagar och mossa dhett.

Google

En katjager rottor och mus.

ADV PUNCT DET NOUN VERB DET NOUN PUNCT

Definite=Def Gender=Fem Number=Plur Number=Plur Person=3 Tense=Pres

Definite=Def Gender=Masc Number=Plur Number=Plur

A cat chases rats and mice.

ADV PUNCT DET NOUN VERB DET NOUN PUNCT

Definite=Def Gender=Fem Number=Plur Number=Plur Person=3 Tense=Pres

Definite=Def Gender=Masc Number=Plur Number=Plur

Toutefois, les filles adorent les desserts.

ADV PUNCT DET NOUN VERB DET NOUN PUNCT

Definite=Def Gender=Fem Number=Plur Number=Plur Person=3 Tense=Pres

Definite=Def Gender=Masc Number=Plur Number=Plur
Morphology

Le chat chasse les chiens .
Morphology

• Lemma representing the semantic content of the word
Morphology

- Lemma representing the semantic content of the word
- Part-of-speech tag representing its grammatical class
Morphology

- **Lemma** representing the semantic content of the word
- **Part-of-speech tag** representing its grammatical class

<table>
<thead>
<tr>
<th>Open</th>
<th>Closed</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>ADP</td>
<td>PUNCT</td>
</tr>
<tr>
<td>ADV</td>
<td>AUX</td>
<td>SYM</td>
</tr>
<tr>
<td>INTJ</td>
<td>CCONJ</td>
<td>X</td>
</tr>
<tr>
<td>NOUN</td>
<td>DET</td>
<td></td>
</tr>
<tr>
<td>PROP</td>
<td>NUM</td>
<td></td>
</tr>
<tr>
<td>VERB</td>
<td>PART</td>
<td></td>
</tr>
<tr>
<td>PRON</td>
<td>SCONJ</td>
<td></td>
</tr>
</tbody>
</table>
Morphology

- Lemma representing the semantic content of the word
- Part-of-speech tag representing its grammatical class
- Features representing lexical and grammatical properties of the lemma or the particular word form
Morphology

- Lemma representing the semantic content of the word
- Part-of-speech tag representing its grammatical class
- Features representing lexical and grammatical properties of the lemma or the particular word form
The cat could have chased all the dogs down the street.
• Content words are linked by grammatical relations
- Content words are linked by grammatical relations
- Function words attach to the content word they modify
• Content words are linked by grammatical relations
• Function words attach to the content word they modify
• Punctuation attach to head of phrase or clause
The dog was chased by the cat.

Hunden jagades av katten.
The dog was chased by the cat.
The dog was chased by the cat.

Hunden jagades av katten.
The dog was chased by the cat.

Hunden jagades av katten.
Syntactic Relations
Syntactic Relations

Taxonomy of 37 universal syntactic relations

- Three types of structures: nominals, clauses, modifiers
- Core arguments vs. obliques (not arguments vs. adjuncts)
- Language-specific subtypes
Syntactic Relations

Taxonomy of 37 universal syntactic relations
  • Three types of structures: nominals, clauses, modifiers
  • Core arguments vs. obliques (not arguments vs. adjuncts)
  • Language-specific subtypes

Basic and enhanced representations
  • Basic representation forms a (possibly non-projective) tree
  • Additional relations in the enhanced representation
## Syntactic Relations

<table>
<thead>
<tr>
<th>Core Predicate Dep</th>
<th>Nominal</th>
<th>Clause</th>
<th>Modifier Word</th>
<th>Function Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsubj</td>
<td>csubj</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>obj</td>
<td>ccomp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iobj</td>
<td>xcomp</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-Core Predicate Dep</th>
<th>Nominal Dep</th>
<th>Coordination</th>
<th>MWE</th>
<th>Loose</th>
<th>Special</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>obl</td>
<td>nmod</td>
<td>conj</td>
<td>fixed</td>
<td>parataxis</td>
<td>orphan</td>
<td>punct</td>
</tr>
<tr>
<td>vocative</td>
<td>appos</td>
<td>cc</td>
<td>flat</td>
<td>list</td>
<td>goeswith</td>
<td>root</td>
</tr>
<tr>
<td>expl dislocated</td>
<td>nummod</td>
<td>compound</td>
<td></td>
<td></td>
<td>reparandum</td>
<td>dep</td>
</tr>
</tbody>
</table>
Papers


Papers


Papers


Papers


Papers


Research Problems

Annotation
- Produce guidelines and/or annotation for language X
- Study the annotation of construction Y across languages

Parsing
- Develop and/or evaluate a parser for language X
- Study cross-lingual transfer learning and/or annotation projection

Typology
- Study word order patterns in language X
- Compare the realisation of construction Y across languages
Questions?