Tutorial on Universal Dependencies

Making use of UD in NLP and linguistics

Joakim Nivre\textsuperscript{1}  Daniel Zeman\textsuperscript{2}  Filip Ginter\textsuperscript{3}  Francis M. Tyers\textsuperscript{4,5}

\textsuperscript{1}Department of Linguistics and Philology, Uppsala University, Sweden

\textsuperscript{2}Institute of Formal and Applied Linguistics, Charles University in Prague, Czech Republic

\textsuperscript{3}Department of Information Technology, University of Turku, Finland

\textsuperscript{4}Giela ja kultuvra instituhtta, UiT Norgga árktalaš universitehta, Tromsø, Norway

\textsuperscript{5}Arvutiteaduse instituut, Tartu Ülikool, Estonia
Parsing

- Monolingual parsing
  - Benchmarking for many different languages
  - Off-the-shelf models:
    - UDPipe
    - SyntaxNet

- Cross-lingual parsing
  - Usually motivated by a low-resource scenario
  - Three main approaches
    - Annotation projection (Hwa et al., 2002)
    - Model transfer (Zeman and Resnik, 2008)
    - Treebank translation (Tiedemann et al., 2014)

- Universal parsing
  - A single model for all languages (Ammar et al., 2016)
### Cross-Lingual Parsing

<table>
<thead>
<tr>
<th>Target Test Language</th>
<th>da</th>
<th>de</th>
<th>el</th>
<th>en</th>
<th>es</th>
<th>it</th>
<th>nl</th>
<th>pt</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>da</td>
<td>79.2</td>
<td>45.2</td>
<td>44.0</td>
<td>45.9</td>
<td>45.0</td>
<td>48.6</td>
<td>46.1</td>
<td>48.1</td>
<td>47.8</td>
</tr>
<tr>
<td>de</td>
<td>34.3</td>
<td>83.9</td>
<td>53.2</td>
<td>47.2</td>
<td>45.8</td>
<td>53.4</td>
<td>55.8</td>
<td>55.5</td>
<td>46.2</td>
</tr>
<tr>
<td>el</td>
<td>33.3</td>
<td>52.5</td>
<td>77.5</td>
<td>63.9</td>
<td>41.6</td>
<td>59.3</td>
<td>57.3</td>
<td>58.6</td>
<td>47.5</td>
</tr>
<tr>
<td>en</td>
<td>34.4</td>
<td>37.9</td>
<td>45.7</td>
<td>82.5</td>
<td>28.5</td>
<td>38.6</td>
<td>43.7</td>
<td>42.3</td>
<td>43.7</td>
</tr>
<tr>
<td>es</td>
<td>38.1</td>
<td>49.4</td>
<td>57.3</td>
<td>53.3</td>
<td>79.7</td>
<td>68.4</td>
<td>51.2</td>
<td>66.7</td>
<td>41.4</td>
</tr>
<tr>
<td>it</td>
<td>44.8</td>
<td>56.7</td>
<td>66.8</td>
<td>57.7</td>
<td>64.7</td>
<td>79.3</td>
<td>57.6</td>
<td>69.1</td>
<td>50.9</td>
</tr>
<tr>
<td>nl</td>
<td>38.7</td>
<td>43.7</td>
<td>62.1</td>
<td>60.8</td>
<td>40.9</td>
<td>50.4</td>
<td>73.6</td>
<td>58.5</td>
<td>44.2</td>
</tr>
<tr>
<td>pt</td>
<td>42.5</td>
<td>52.0</td>
<td>66.6</td>
<td>69.2</td>
<td>68.5</td>
<td>74.7</td>
<td>67.1</td>
<td>84.6</td>
<td>52.1</td>
</tr>
<tr>
<td>sv</td>
<td>44.5</td>
<td>57.0</td>
<td>57.8</td>
<td>58.3</td>
<td>46.3</td>
<td>53.4</td>
<td>54.5</td>
<td>66.8</td>
<td>84.8</td>
</tr>
</tbody>
</table>

*McDonald et al. (2011)*
## Cross-Lingual Parsing

<table>
<thead>
<tr>
<th>Target Test Language</th>
<th>da</th>
<th>de</th>
<th>el</th>
<th>en</th>
<th>es</th>
<th>it</th>
<th>nl</th>
<th>pt</th>
<th>sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>da</td>
<td><strong>79.2</strong></td>
<td>45.2</td>
<td>44.0</td>
<td>45.9</td>
<td>45.0</td>
<td>48.6</td>
<td>46.1</td>
<td>48.1</td>
<td>47.8</td>
</tr>
<tr>
<td>de</td>
<td>34.3</td>
<td><strong>83.9</strong></td>
<td>53.2</td>
<td>47.2</td>
<td>45.8</td>
<td>53.4</td>
<td>55.8</td>
<td>55.5</td>
<td>46.2</td>
</tr>
<tr>
<td>el</td>
<td>33.3</td>
<td>52.5</td>
<td><strong>77.5</strong></td>
<td>63.9</td>
<td>41.6</td>
<td>59.3</td>
<td>57.3</td>
<td>58.6</td>
<td>47.5</td>
</tr>
<tr>
<td>en</td>
<td>34.4</td>
<td>37.9</td>
<td>45.7</td>
<td><strong>82.5</strong></td>
<td>28.5</td>
<td>38.6</td>
<td>43.7</td>
<td>42.3</td>
<td>43.7</td>
</tr>
<tr>
<td>es</td>
<td>38.1</td>
<td>49.4</td>
<td>57.3</td>
<td>53.3</td>
<td><strong>79.7</strong></td>
<td>68.4</td>
<td>51.2</td>
<td>66.7</td>
<td>41.4</td>
</tr>
<tr>
<td>it</td>
<td>44.8</td>
<td>56.7</td>
<td>66.8</td>
<td>57.7</td>
<td>64.7</td>
<td><strong>79.3</strong></td>
<td>57.6</td>
<td>69.1</td>
<td>50.9</td>
</tr>
<tr>
<td>nl</td>
<td>38.7</td>
<td>43.7</td>
<td>62.1</td>
<td>60.8</td>
<td>40.9</td>
<td>50.4</td>
<td><strong>73.6</strong></td>
<td>58.5</td>
<td>44.2</td>
</tr>
<tr>
<td>pt</td>
<td>42.5</td>
<td>52.0</td>
<td>66.6</td>
<td>69.2</td>
<td>68.5</td>
<td>74.7</td>
<td>67.1</td>
<td><strong>84.6</strong></td>
<td>52.1</td>
</tr>
<tr>
<td>sv</td>
<td><strong>44.5</strong></td>
<td>57.0</td>
<td>57.8</td>
<td>58.3</td>
<td>46.3</td>
<td>53.4</td>
<td>54.5</td>
<td>66.8</td>
<td><strong>84.8</strong></td>
</tr>
</tbody>
</table>

**McDonald et al. (2011)**

<table>
<thead>
<tr>
<th>Source Training Language</th>
<th>Unlabeled Attachment Score (UAS)</th>
<th>Target Test Language</th>
<th>Labeled Attachment Score (LAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Germanic</td>
<td>Romance</td>
<td>Germanic</td>
</tr>
<tr>
<td></td>
<td>DE</td>
<td>EN</td>
<td>SV</td>
</tr>
<tr>
<td>DE</td>
<td>74.86</td>
<td>55.05</td>
<td>65.89</td>
</tr>
<tr>
<td>EN</td>
<td>58.50</td>
<td>83.33</td>
<td><strong>70.56</strong></td>
</tr>
<tr>
<td>SV</td>
<td><strong>61.25</strong></td>
<td><strong>61.20</strong></td>
<td>80.01</td>
</tr>
<tr>
<td>ES</td>
<td>55.39</td>
<td>58.56</td>
<td>66.84</td>
</tr>
<tr>
<td>FR</td>
<td>55.05</td>
<td>59.02</td>
<td>65.05</td>
</tr>
<tr>
<td>KO</td>
<td>33.04</td>
<td>32.20</td>
<td>27.62</td>
</tr>
</tbody>
</table>

**McDonald et al. (2013)**
Cross-Lingual Parsing


- Three methods for cross-lingual dependency parsing
- The impact of not having gold part-of-speech tags
- Reveals weaknesses of delexicalized model transfer
- Parsing with multiple source and target languages
- Multilingual word embeddings and typological features
- Gain on "small" languages without loss on "big" languages

- Synthesizing treebanks for new (potential) natural languages
- Reorder N- and/or V-dependents in L₁ with model trained on L₂
Semantic Parsing

Reddy et al. (2016) Transforming Dependency Structures to Logical Forms for Semantic Parsing

Take syntactic dependencies

```
Pixar is the company which Disney acquired.
```

PROPN VERB DET NOUN DET PROPN VERB PUNCT

Deterministically infer logical form(s)

```
∃z. company(Pixar) ∧ acquired(z, Disney) ∧ arg₁(z, Disney) ∧ arg₂(z, Pixar)
```

Learn model to map logical form to KG to answer question
Semantic Parsing

Disney \text{ acquired } Pixar

\lambda z. \exists x y. \text{acquired}(z_e) \land \text{Pixar}(y_a) \land \text{Disney}(x_a) \land \text{arg}_1(z_e, x_a) \land \text{arg}_2(z_e, y_a)

- Cross-lingual UD-parsed corpora (using bilingual lexicon)
- Evaluated on word similarity and bilingual lexicon induction
- Outperforms cross-lingual/no syntax and monolingual/syntax
Futrell et al. (2015) Quantifying Word Order Freedom in Dependency Corpora

- Word order freedom = conditional entropy of order given tree
- Test hypotheses about case marking and word order freedom
Linguistic Typology

Relation Order Entropy of Subject and Object

Languages:
- Irish
- Persian
- French
- Arabic
- English
- Hebrew
- Danish
- Italian
- Estonian
- Swedish
- Portuguese
- Hindi
- Modern Greek
- Turkish
- Romanian
- Finnish
- Dutch
- Spanish
- Croatian
- German
- Catalan
- Japanese
- Bulgarian
- Bengali
- Ancient Greek
- Slovenian
- Czech
- Telugu
- Basque
- Slovak
- Russian
- Hungarian
- Latin
- Tamil

Entropy levels:
- none
- full
- dom

<table>
<thead>
<tr>
<th>Feature</th>
<th>Languages</th>
<th>Types</th>
<th>Tokens</th>
<th>Most common</th>
</tr>
</thead>
<tbody>
<tr>
<td>81A: Subject, Object, Verb (Dryer, 2013e)</td>
<td>342</td>
<td>85.4%</td>
<td>85.7%</td>
<td>SOV: 43.3%</td>
</tr>
<tr>
<td>82A: Subject, Verb (Dryer, 2013d)</td>
<td>376</td>
<td>89.4%</td>
<td>90.4%</td>
<td>SV: 79.8%</td>
</tr>
<tr>
<td>83A: Object, Verb (Dryer, 2013c)</td>
<td>387</td>
<td>96.4%</td>
<td>96.4%</td>
<td>VO: 54.8%</td>
</tr>
<tr>
<td>85A: Adposition, Noun Phrase (Dryer, 2013b)</td>
<td>329</td>
<td>94.8%</td>
<td>95.1%</td>
<td>Prep: 50.4%</td>
</tr>
<tr>
<td>87A: Adjective, Noun (Dryer, 2013a)</td>
<td>334</td>
<td>85.9%</td>
<td>88.0%</td>
<td>AdjN: 68.9%</td>
</tr>
</tbody>
</table>

- Word order study based on Bible translations (986 languages)
- Massively parallel alignment and UD annotation projection
- Evaluation against WALS for a subset of languages
Is UD really suitable for all it is used for?
Manning’s Law

The secret to understanding the design of UD is to realize that it is a very subtle compromise between approximately 6 things:

1. UD needs to be satisfactory for analysis of individual languages.
2. UD needs to be good for linguistic typology.
3. UD must be suitable for rapid, consistent annotation.
4. UD must be suitable for computer parsing with high accuracy.
5. UD must be easily comprehended and used by a non-linguist.
6. UD must provide good support for downstream NLP tasks.

It’s easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.
Questions?