Conclusion and Outlook

Joakim Nivre

Uppsala University
Department of Linguistics and Philology
joakim.nivre@lingfil.uu.se
1. Synthesis
2. Ensemble Systems
3. Domain Adaptation
4. Cross-Language Variation
5. Unsupervised Parsing
- Feature-rich discriminative factored models
- Global learning (conditional or discriminative)
- Dynamic programming or beam search decoding
Parser outputs (for sentence $x$): $y_1, \ldots, y_m$

Arc scores: $\text{Score}(i, l, j, x) = |\{y_k : (i, l, j, x) \in y_k\}|$

Maximum spanning tree parsing
<table>
<thead>
<tr>
<th></th>
<th>WSJ</th>
<th>Brown</th>
<th>Genia</th>
<th>SWBD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>89.7</td>
<td>84.1</td>
<td>76.2</td>
<td>76.7</td>
</tr>
<tr>
<td>Language</td>
<td>Arabic</td>
<td>Basque</td>
<td>Catalan</td>
<td>Chinese</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>76.5</td>
<td>76.9</td>
<td>88.7</td>
<td>84.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th>English</th>
<th>Greek</th>
<th>Hungarian</th>
<th>Italian</th>
<th>Turkish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>89.6</td>
<td>76.3</td>
<td>80.3</td>
<td>84.4</td>
<td>79.8</td>
</tr>
</tbody>
</table>
Universal Dependencies (http://universaldependencies.github.io/docs/)

- Google universal part-of-speech tags
- Interset morphological features
- Stanford universal dependencies

- Guidelines, version 1, October 2014
- First set of treebanks (10 languages), January 2015
\[ P(x|y) = \prod_{1 \leq i < j \leq n} P(x_i, \ldots, x_j|y_{ij})P(x_{i-1}, x_{j+1}|y_{ij}) \]
\[ P(T(h)) = \prod_{d \in l,r} \left[ \prod_{a \in D(h,d)} P_!(-!|h, d, ?)P_v(a|h, d)P(T(a)) \right] P_!(|!|h, d, ?) \]