Basic Issues in Syntactic Parsing

Joakim Nivre

Uppsala University
Department of Linguistics and Philology
joakim.nivre@lingfil.uu.se
Syntactic Parsing

Structured prediction: $\mathcal{X} \rightarrow \mathcal{Y}$

1. Input ($\mathcal{X}$)
2. Output ($\mathcal{Y}$)
3. Mapping ($\mathcal{X} \rightarrow \mathcal{Y}$)
4. Models
5. Evaluation
Sentence $x = x_1, \ldots, x_n \in \mathcal{X}$
ev (the) house
evler (the) houses
evleriniz your houses
evlerinizden from your houses
evlerinizdendi (he/she/it) was from your houses
<table>
<thead>
<tr>
<th>Token</th>
<th>Lemma</th>
<th>Tag</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je</td>
<td>být</td>
<td>B</td>
<td>Num=S</td>
</tr>
<tr>
<td>tento</td>
<td>tento</td>
<td>D</td>
<td>Gen=Y</td>
</tr>
<tr>
<td>reklamní</td>
<td>reklamní</td>
<td>A</td>
<td>Gen=l</td>
</tr>
<tr>
<td>slogan</td>
<td>slogan</td>
<td>N</td>
<td>Gen=l</td>
</tr>
<tr>
<td>pravdivý</td>
<td>pravdivý</td>
<td>A</td>
<td>Gen=l</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td></td>
<td>_</td>
</tr>
</tbody>
</table>
Syntactic representation $y \in \mathcal{Y}$
Economic news had little effect on financial markets.
Economic news had little effect on financial markets.
Parsing as optimization: $y^* = \arg\max_y f_M(x, y)$
$M = (\text{GEN, EVAL})$

\[ \text{GEN}(x) \subseteq \mathcal{Y} \]

\[ \text{EVAL}(x, y) \in \mathbb{R} \]

\[ y^* = \arg\max_{y \in \text{GEN}(x)} \text{EVAL}(x, y) \]
Test set $T = \{(x^1, y^1), \ldots, (x^m, y^m)\}$