PP Attachment: where do we stand?
Daniël de Kok, Jianqiang Ma, Corina Dima, Erhard Hinrichs

The problem
Prepositional phrase (PP) attachment is a well-known structural ambiguity in natural language parsing. PP attachment is particularly difficult because resolving ambiguities requires semantic cues.

Research questions
- There are several proposed techniques to improve PP attachment disambiguation in the literature. What is the influence of these different techniques?
- Do association scores learnt from large, automatically annotated corpora still improve PP-attachment disambiguation in a neural network/embedding setup?
- What problems are remaining for a state-of-the-art disambiguator?

PP attachment with multiple attachment sites
Goetsch bezog sich auf einen Vorschlag der Opposition von voriger Woche .
Es gibt viel Beifall für Margret Mönig-Raane .

• PP attachment is traditionally treated as a binary classification task, where the attachment site is the preceding noun or the main verb.
• However: in reality multiple attachment sites at potentially large distances (3.15 on average in the German TüBa-D/Z treebank).

Association scores
- Normalized PMI [Bouma, 2009]:

\[
\text{N-PMI}(x, y) = \left( \log \frac{p(x, y)}{p(x)p(y)} \right) - \log p(x, y)
\]

- Specific interaction information [Van de Cruys, 2011]:

\[
\text{SI}_1(x, y, z) = \log \frac{p(x, y)p(y, z)p(x, z)}{p(x)p(y)p(z)p(x, y, z)}
\]

- Specific correlation [Van de Cruys, 2011]:

\[
\text{SI}_2(x, y, z) = \log \frac{p(x, y, z)}{p(x)p(y)p(z)}
\]

Normalized PMI for the triples:
- ⟨CANDIDATE,OBJECT⟩
- ⟨CANDIDATE,PREPOSITION⟩
- ⟨CANDIDATE,PREPOSITION+OBJECT⟩

Specific interaction information and total correlation for the triple:
- ⟨CANDIDATE,PREPOSITION,OBJECT⟩

Topological fields
Topological fields are commonly used to describe the regularities in German word order. The distributions of syntactic relations vary significantly across topological fields, which can inform attachment choices, including PP attachment.

\[
\begin{array}{llllll}
\text{VF} & \text{LK} & \text{MF} & \text{RK} & \text{NF} \\
\text{MC:} & \text{Gestern} & \text{hat} & \text{er} & \text{häufiger} & \text{angerufen} & \text{als heute} \\
\text{MC:} & \text{Er} & \text{ruft} & \text{häufig} & \text{an} & \text{up} \\
\text{SC:} & \text{der} & \text{noch häufiger} & \text{anruft} & \text{als er} & \text{more often} & \text{calls than him}
\end{array}
\]

Neural network scoring model
We use a feed-forward neural network to score each candidate attachment site, given the preposition (P), (preposition) object (Obj), and attachment site (S):

- Input layer: word embeddings, tag embeddings, and topological fields of P, Obj, and S. Normalized PMI, specific interaction information and total correlation scores as described above.
- Hidden layer: 100 neurons, with ReLU activations
- Output layer: The logistic function, trained to estimate the attachment probability.

The scoring model computes the probability of each attachment site and chooses the site with the highest probability.

Results
The association scores were computed using text from the German newspaper taz (1986 to 2009, 28.8 million sentences, 393.7 million tokens) after annotation with a state-of-the-art dependency parser. The PP disambiguator is trained using the dataset of [de Kok et al., 2017].

Qualitative analysis
We hand-classified a random sample of 100 errors made by the PP attacher.

<table>
<thead>
<tr>
<th>Type</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect</td>
<td>44</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>36</td>
</tr>
<tr>
<td>Discourse</td>
<td>13</td>
</tr>
<tr>
<td>Other</td>
<td>7</td>
</tr>
</tbody>
</table>

This analysis shows that more than half of the remaining errors are irrelevant or the attachments could not be solved at the sentence level.