Deep Lexical Grammars

- Deep grammars provide a full analysis, more semantic information than shallower tools
- Tendency to emphasise precision over recall can cause poor coverage
- HPSG grammars, parsing tools from the DELPH-IN initiative
- Our aim: take a “snapshot” of the grammar, examine potential for expansion
Analysis of a "Broad-Coverage" Grammar

- "Beauty and the Beast" (2004):
  - Use the ERG to parse 20K sentences from the BNC
  - Analyse sources of parse failures

- "Evaluating and Extending" (Today):
  - Use GG to parse 612K sentences from Frankfurter Rundschau
  - Evaluate errors over 1K sentences
  - Use lexical type prediction to increase coverage
Corpus Analysis of the Grammar

- Ran a large grammar out-of-the-box on a very different corpus
- Lexical span: ERG - 32%; GG - 28%
- Sentences with correct reading attested: ERG - 83%; GG - 85%

<table>
<thead>
<tr>
<th></th>
<th>No span</th>
<th>Span, no parse</th>
<th>≥ 1 parse</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERG</td>
<td>68%</td>
<td>14%</td>
<td>18%</td>
</tr>
<tr>
<td>GG</td>
<td>72%</td>
<td>16%</td>
<td>12%</td>
</tr>
</tbody>
</table>
Lexical Gaps for GG

Lexical gaps:

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexical entries</td>
<td>33%</td>
</tr>
<tr>
<td>proper nouns</td>
<td>22%</td>
</tr>
<tr>
<td>noun compounds</td>
<td>30%</td>
</tr>
<tr>
<td>tokenisation</td>
<td>12%</td>
</tr>
<tr>
<td>garbage strings</td>
<td>2%</td>
</tr>
</tbody>
</table>
Lexical entry:
Aufgrund des ruhigeren Geschäftsverlaufs rechnet Maier für 1992 mit einem “leicht rückläufigen” Ergebnis.

Noun compound:
Das Türelement läßt sich hinter die Verkleidung schieben und wird damit unsichtbar.

Sophisticated tokenisation could account for proper nouns, noun compounds, tokenisation errors.
Parsing Errors for GG:

<table>
<thead>
<tr>
<th>Error type</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>constructional gap</td>
<td>39%</td>
</tr>
<tr>
<td>lexical item gap</td>
<td>47%</td>
</tr>
<tr>
<td>multi–word expression</td>
<td>7%</td>
</tr>
<tr>
<td>spelling</td>
<td>4%</td>
</tr>
<tr>
<td>fragment</td>
<td>3%</td>
</tr>
</tbody>
</table>
Parsing Errors for GG

Constructional gap:
*BREMEN, 4. Februar.*

Lexical item gap:
*Beginn ist um 19 Uhr in der Stadthalle.*

Multi-word expression:
*Der Opfer dieser Verbrechen der Nationalsozialisten gedachte die Stadt Bad Homburg gestern abend.*

Similar distribution observed in Beauty and the Beast.
Baldwin (2005): use a range of morphological, syntactic, semantic features for predicting lexical type class of unknown token/type

e.g. *Katze* in *Die Katze ist schwarz.* is one of count-noun-le, mass-noun-le, count-noun-mass-unit-le, deverbal-noun-le...
Lexical Acquisition

- Feature set from Zhang and Kordoni (2006): prefixes/suffixes, 2 tokens of context, 2 types of context
- Token-wise prediction on the GG treebank (MaxEnt, cross-validation)
- Limit evaluation to “unknown words” (type-wise)
- Accuracy approaches 60%
Lexicon Extension

- Using the MaxEnt model from the treebank, predict lexical types for unknown tokens within Frankfurter Rundschau
- Intrinsic evaluation not possible
- Thresholding MaxEnt at 10% likelihood, add 1130 lexemes to the lexicon
- Further 9% coverage, 83% of these had at least one parse
Change of 12% of parsed sentences at 85% precision to about 20% at 84% precision. This means getting more “easy” sentences. Scope for improving the grammar, parsing strategy (shallow methods to improve deep parsing).