Experiments in Parsing German

Sisay Fissaha, Daniel Olejnik, Ralf Kornberger, Karin Müller, Detlef Prescher

Universität des Saarlandes, IAI & Computational Linguistics Department
Universiteit van Amsterdam, ILLC, Language and Inference Technology Group
Outline

• Motivation

• Symbolic versus Probabilistic Parsing
  ▶ Symbolic Component
  ▶ Probabilistic Component

• Experiments

• Evaluation

• Discussion
Motivation

Develop a **broad-coverage** probabilistic parser for German which recognizes constituents and **grammatical functions**.

- **Method**
  - Extract a symbolic context-free grammar from the NEGRA treebank.
  - Train probabilistic grammar versions using a treebank training method (similar to Charniak:1996), and some grammar transformation techniques like parent encoding (similar to Johnson:1998).
  - Evaluate the performance of the grammar on a test corpus.
Background Information

- Symbolic Parsing:
  - parse a given sentence and assign all possible structures (syntactic trees).

- Probabilistic Parsing:
  - use a symbolic parsing component to get all possible syntactic structures of a sentence
  - disambiguate the different analyses by using rule probabilities, i.e. choose the most probable analysis.
Symbolic Component

Example:
“Der Verein sucht noch kreative Erwachsene mit viel Elan.”
(The association still searches for creative adults full of verve.)

Possible structures are:

▶ [Der Verein [sucht noch [ kreative Erwachsene [ mit Elan ]]]],
  \([NP \text{ kreative Erwachsene } [PP \text{ mit Elan }]]\)
▶ [Der Verein [sucht noch [kreative Erwachsene] [mit Elan]]].
  \([VP \text{ sucht noch } [NP \ldots ] [PP \text{ mit Elan }]]\)

Problem: ambiguity
Probabilistic Component

- Resource: Negra treebank, a newspaper corpus (20,000), manually annotated with syntactic structure.
- Read the rules from the trees, collect the rules and their frequencies.
- Compute rule probabilities of each rule $p(r)$.

$$p(r) = \frac{|r|}{\sum_{r' \in \{r'|\lambda(r') = \lambda(r)\}} |r'|}$$

- Several grammar variants comprising constituents with/without grammatical functions, and (partial) parent encoding have been investigated.
Computing the probabilities

(1) 0.0031 \( S \rightarrow \) NP-SB VVFIN-HD NP-OA PP-MNR
(2) 0.0109 \( S \rightarrow \) NP-SB VVFIN-HD NP-OA
(3) 0.0059 NP-OA \( \rightarrow \) ADJA-NK NN-NK PP-MNR
(4) 0.2396 NP-SB \( \rightarrow \) ART-NK NN-NK
(5) 0.0518 NP-OA \( \rightarrow \) ADJA-NK NN-NK
(6) 0.0355 PP-MNR \( \rightarrow \) APPR-NK ADJA-NK NN-NK

- Part-of-speech tags are assigned to each sentence by a probabilistic tagger, i.e. no rules

\[ \text{POS} \rightarrow \text{word} \]

are in the grammar.
Parsing a sentence

- **Example:** *The association searches for creative adults full of verve.*
  
  Der Verein sucht kreative Erwachsene mit viel Elan.

  ART    NN    VVFIN   ADJA    NN    APPR    ADJA    NN

- **Compute probability of the two readings:**

  (1)  \((S \rightarrow NP-SB \text{ VVFIN-HD NP-OA PP-MNR})\)  0.0031 *
      (NP-SB \rightarrow ART-NK NN-NK)  0.2396 *
      (NP-OA \rightarrow ADJA-NK NN-NK)  0.0518 *
      (PP-MNR \rightarrow APPR-NK ADJA-NK NN-NK)  0.0355

      \[= 1.36e^{-06}\]

  (2)  \((S \rightarrow NP-SB \text{ VVFIN-HD NP-OA})\)  0.0109 *
      (NP-SB \rightarrow ART-NK NN-NK)  0.2396 *
      (NP-OA \rightarrow ADJA-NK NN-NK PP-MNR)  0.0059 *
      (PP-MNR \rightarrow APPR-NK ADJA-NK NN-NK)  0.0355

      \[= 5.47e^{-07}\]

- **Choose the most probable one.**
Experiments

- Split the corpus into 18,000 sentences for training, 2000 sentences for testing.

- Training corpus is divided into pieces of 1000 to investigate the influence of the size of the training data.

- Training
  - with different grammar variants

- Evaluate on the test corpus using standard evaluation measures (PARSEVAL).
Experiment 1:

Grammar contains constituent structure & grammatical functions. The PCFG can predict both after training.

e.g. a noun phrase consisting of an article and a noun occur more often as subjects than as direct or indirect objects:

NP-SB → ART-NK NN-NK (3,498)
NP-OA → ART-NK NN-NK (1,385)
NP-DA → ART-NK NN-NK (338)
Results of Experiment 1

The graph shows the results of Experiment 1 for PARSEVAL Constituent Structure and Grammatical Functions. The x-axis represents corpus size, ranging from 2000 to 18000, while the y-axis measures PARSEVAL Constituent Structure and Grammatical Functions.

The graph illustrates the coverage, labeled f-measure, and exact-match rate against corpus size. The learning phase is indicated by the horizontal bar at the bottom of the graph.
Results of Experiment 1 - Alternative

[Graph showing the relationship between coverage and other metrics such as labeled f-measure and exact-match rate.]
Experiment 2: Influence of Grammatical Functions

- Compare two grammar versions
  - using constituents only
    \[ S \rightarrow NP \text{ VVFIN NP} \]
  - using grammatical functions (previous experiment)
    \[ S \rightarrow NP-SB \text{ VVFIN-HD NP-OA} \]
Results of Experiment 2

- PARSEVAL
- constituent structure only
- coverage
- labeled f-measure
- exact-match rate
- learning phase
- parsing of constituents
- parsing of constituents + functions

Experiments in Parsing German, 10 September 2003, TSD 2003
Copyright © Karin Müller
Experiment 3: Parent Encoding

Soften the independence assumption.
Lead to improvement of about 7% for English.

- full parent encoding, e.g., ART-NK•NP-OA
- partial parent encoding (constituent labels), ART-NK•NP
- partial parent encoding (grammatical functions), ART-NK•OA
Results of Experiment 3

- PARSEVAL
- constituent structure and grammatical functions
- coverage
- exact-match rate
- no parent encoding
- full parent encoding
- partial parent encoding (functions)
- partial parent encoding (constituents)

Experiments in Parsing German, 10 September 2003, TSD 2003
Copyright © Karin Müller
Discussion (i)

- (i) High coverage parser providing both grammatical functions and constituent structure (18000): coverage 97.8%, precision 53%, recall 58%.
- (ii) In comparison to the constituent structure parsers, a parser using additionally grammatical functions perform better on comparable level of coverage.
- (iii) Full parent encoding outperformed partial parent encoding with functions and constituents at same level of coverage.

<table>
<thead>
<tr>
<th>parser</th>
<th>no PE</th>
<th>partial PE (const)</th>
<th>partial PE (GF)</th>
<th>full PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>coverage</td>
<td>97.8%</td>
<td>91.1%</td>
<td>80.8%</td>
<td>70.2%</td>
</tr>
</tbody>
</table>
Discussion (ii)
Compared to English, the precision for German is significantly lower. Possible reasons:

- free-er word order, complements and adjuncts can be shifted
- NEGRA follows the dependency grammar tradition
  - flat syntactic representation, e.g., no S → NP VP rule), rather subject, objects are sisters.
  - coordination is expressed by a different category (e.g., CS → NP VVFIN, S → NP VVFIN).
- additional information: grammatical function labels.
Comparison to German

Results of Dubey & Keller 2003 (tagging with TnT):

<table>
<thead>
<tr>
<th></th>
<th>LP★</th>
<th>LR★</th>
<th>Cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>71%</td>
<td>67%</td>
<td>94%</td>
</tr>
<tr>
<td>Baseline + Grammatical F</td>
<td>70%</td>
<td>65%</td>
<td>79%</td>
</tr>
<tr>
<td>C &amp; R</td>
<td>68%</td>
<td>60%</td>
<td>94%</td>
</tr>
<tr>
<td>C &amp; R + pool</td>
<td>69%</td>
<td>61%</td>
<td>94%</td>
</tr>
<tr>
<td>C &amp; R + Grammatical F</td>
<td>68%</td>
<td>60%</td>
<td>79%</td>
</tr>
<tr>
<td>Collins</td>
<td>68%</td>
<td>66%</td>
<td>95%</td>
</tr>
</tbody>
</table>

★ on constituent structure only
Comparison to English

Penn Treebank: 30,000 sentences,
Negra Treebank: 20,000 sentences

<table>
<thead>
<tr>
<th>language</th>
<th>LP</th>
<th>LR</th>
<th>Cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>English constituents (Charniak 1996)</td>
<td>79%</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>German constituents</td>
<td>71.8%</td>
<td>67.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>constituents + GF</td>
<td>56.4%</td>
<td>61.8%</td>
<td>97.8%</td>
</tr>
</tbody>
</table>
Evaluation measures

PARSEVAL defines standard evaluation measures:

- **precision**: percentage of spans that also appear in the treebank
- **recall**: percentage of spans from the treebank that also appear in the most probable parse of the parser
- **f-score**: harmonic mean of precision and recall
- **coverage**: percentage of sentences which are parsed
- **exact match**: number of sentences which are totally correct parsed