Towards a syntactically motivated analysis of modifiers in German

Ines Rehbein & Hagen Hirschmann

KONVENS 2014

October 8, 2014
Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)

Relative frequencies of modifying POS in the TIGER corpus
Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)

Closed classes (e.g. *nicht* for PTKNEG – negation), relatively infrequent, relatively homogeneous syntax per class

Relative frequencies of modifying POS in the TIGER corpus
Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)

Closed classes (e.g. nicht for PTKNEG – negation), relatively infrequent, relatively homogeneous syntax per class

Prenominal adjectives, fixed syntactic position, easy to parse

Relative frequencies of modifying POS in the TIGER corpus
Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)

Closed classes (e.g. *nicht* for PTKNEG – negation), relatively infrequent, relatively homogeneous syntax per class.

- **PTKVZ**: very heterogeneous, open, residual class, hard to parse
- **PTKNEG**: prenominal adjectives, fixed syntactic position, easy to parse
- **PTKU**:
- **PTKA**:
- **PTKANT**:

Relative frequencies of modifying POS in the TIGER corpus
The problem with ADV

Manual parse for a clause with four consecutive 'ADV': TIGER07, s17263
(“In this case, more than 30 legal proceedings are still waiting for Aksoy.”)
The problem with ADV

- Syntactic underspecification (heterogeneity) of many single word modifiers in parser input data
- Parsing difficulties: No clues for attachment and grammatical function from POS

Manual parse for a clause with four consecutive 'ADV': TIGER07, s17263
(“In this case, more than 30 legal proceedings are still waiting for Aksoy.”)

“In this case, more than 30 legal proceedings are still waiting for Aksoy.”
Resulting research question

- Does a syntactically motivated extension of the STTS category ADV help to improve parsing accuracy?
Redefining ADV and ADJD

- ADV-ADJD distinction according to STTS guidelines (Schiller et al. 1999)
  - [...] vielleicht/ADV wäre es ihm ähnlich ergangen (...) “Perhaps he would have experienced something similar” (TIGER07, s9814)
  - [...] wahrscheinlich/ADJD wird er nicht einmal gebilligt (...) “Probably, he will not even be approved” (TIGER07, s17581)
Redefining ADV and ADJD

• ADV-ADJD distinction according to STTS guidelines (Schiller et al. 1999)
  
  • (...)

  • **vielleicht/ADV** wäre es ihm ähnlich ergangen (...) “*Perhaps he would have experienced something similar*” (TIGER07, s9814)

  • (...)

  • **wahrscheinlich/ADJD** wird er nicht einmal gebilligt (...) “*Probably, he will not even be approved*” (TIGER07, s17581)

• Syntactic definition:

  • ADJD: modifiers of nouns
    (criterion: complement to copula verb)

  • ADV: modifiers of verbs or clauses
    (criterion: all other clause constituents)
New categories: MODP & PTK...

- Class: modal particle
- Criterion: Sentence modifier with topological restrictions
- Test: no pre-field position

- Class: particle
- Criterion: Modifier within a clause constituency
- Test: pre-field position within clause constituency
New categories: PTK...

- **PTKFO: Nur Peter gewinnt (Only Peter wins)**
  - Class: Focus particle
  - Criterion: specification of set of alternatives
  - Test: naming alternatives

- **PTKINT: Sehr oft geschieht das (It happens very often)**
  - Class: Intensifier
  - Criterion: graduation or quantification of head
  - Test: naming equivalent gradual/intensifying expression

- **PTKLEX: Immer noch regnet es (It’s still raining)**
  - Class: part of non-compositional multi word expression
  - Criterion: lexical meaning is not equivalent to meaning in phrase
  - Test: comparing meaning in different contexts
Annotation Experiment

Data

- Developing the guidelines and training the annotators
  - 1000 sentences randomly selected from TIGER (Brants et al. 2004)
  - manually reassign labels to all tokens tagged as either ADJD, ADV, VAPP or VVPP

- Test set for inter-annotator agreement
  - 500 sentences from TIGER (sentences 9,501-10,000)
# Annotation Experiment

## Inter-annotator Agreement

<table>
<thead>
<tr>
<th>POS</th>
<th># STTS</th>
<th># new</th>
<th># agr.</th>
<th>Fleiss’ κ</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAPP</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>1.000</td>
</tr>
<tr>
<td>VVPP</td>
<td>173</td>
<td>172</td>
<td>172</td>
<td>0.989</td>
</tr>
<tr>
<td>ADJD</td>
<td>191</td>
<td>74</td>
<td>63</td>
<td>0.891</td>
</tr>
<tr>
<td>ADV</td>
<td>445</td>
<td>378</td>
<td>343</td>
<td>0.800</td>
</tr>
<tr>
<td>PTKFO</td>
<td>-</td>
<td>80</td>
<td>67</td>
<td>0.797</td>
</tr>
<tr>
<td>PTKINT</td>
<td>-</td>
<td>63</td>
<td>49</td>
<td>0.788</td>
</tr>
<tr>
<td>PTKLEX</td>
<td>-</td>
<td>33</td>
<td>17</td>
<td>0.594</td>
</tr>
<tr>
<td>MODP</td>
<td>-</td>
<td>12</td>
<td>6</td>
<td>0.515</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>830</strong></td>
<td><strong>833</strong></td>
<td><strong>88.3%</strong></td>
<td><strong>0.838</strong></td>
</tr>
</tbody>
</table>

**Table:** Distribution (STTS, new) and agreement (percentage agreement and Fleiss’ $\kappa$) for the different tags
Outline

Expanding the STTS – The Tagset

Annotation Experiment

Parsing Experiments
Related Work

Refine POS tagset to improve tagging accuracy

- MacKinlay and Baldwin (2005)
  - experimented with more fine-grained tagsets
  - refined tagsets did not improve tagging accuracy → data sparseness?

- Dickinson (2006)
  - re-define POS for ambiguous words: add complex tags which reflect ambiguity
  - yields slight improvements on test set, but less robust to errors than original tagger
Related Work

Refine POS tagset to improve tagging accuracy

- MacKinlay and Baldwin (2005)
  - experimented with more fine-grained tagsets
  - refined tagsets did not improve tagging accuracy
    → data sparseness?

- Dickinson (2006)
  - re-define POS for ambiguous words:
    add complex tags which reflect ambiguity
  - yields slight improvements on test set,
    but less robust to errors than original tagger

Hypothesis:

- Syntactically motivated POS distinctions can improve parsing accuracy
Related Work (2)

Impact of POS tagsets on parsing

- Kübler & Maier (2014), Maier et al. (2014) compare the influence of different POS tagsets on constituency parsing
  
  1. universal POS tagset (Petrov et al., 2006) (12 tags)
  2. STTS (54 tags)
  3. fine-grained morphological tagset (>700 tags)

  → slightly lower results for coarse-grained tags
  → morphological tags seem too sparse
Related Work (3)

• Plank et al (2014)
  • incorporate annotator disagreements into the loss function of the tagger
    • improves tagging results as well as the accuracy of a chunker
      → information on ambiguous words can improve parsing

• Difference to Plank et al (2014):
  • they incorporate the ambiguity in the tagging model
  • we reduce the ambiguity in the data by refining the tagset
Parsing Experiments

Data Expansion

1. Define patterns
2. Apply to the first 5000 sentences in TIGER
3. Relabel with new tags

Example: ADV → PTKFO

$[\text{cat} = "\text{NP}"] > @! [\text{pos} = "\text{ADV}"]$

& lemma = ("allein" | "auch" | ... | "zwar")

Overall: 49 patterns, coverage: 90.9%

• Manual clean-up:
  • assign tags to the remaining tokens
  • check for potential errors
Parsing Experiments

Set up

- Two data-driven, language-independent dependency parsers:
  - Malt parser (Nivre et al., 2007)
  - MATE parser (Bohnet, 2010)

- Trained on the expanded training set (CoNLL)
  1. with original STTS tags
  2. with new tags

- Evaluation: 10-fold crossvalidation
### Parsing Experiments

#### Results

<table>
<thead>
<tr>
<th>fold</th>
<th>Malt</th>
<th></th>
<th>MATE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>orig</td>
<td>new</td>
<td>orig</td>
<td>new</td>
</tr>
<tr>
<td>1</td>
<td>84.0</td>
<td>84.3</td>
<td>85.4</td>
<td>86.3</td>
</tr>
<tr>
<td>2</td>
<td>84.2</td>
<td>84.7</td>
<td>87.1</td>
<td>87.6</td>
</tr>
<tr>
<td>3</td>
<td>89.0</td>
<td>89.3</td>
<td>91.7</td>
<td>91.7</td>
</tr>
<tr>
<td>4</td>
<td>85.3</td>
<td>85.9</td>
<td>88.5</td>
<td>89.1</td>
</tr>
<tr>
<td>5</td>
<td>89.0</td>
<td>88.9</td>
<td>91.2</td>
<td>91.5</td>
</tr>
<tr>
<td>6</td>
<td>86.0</td>
<td>85.5</td>
<td>88.0</td>
<td>88.4</td>
</tr>
<tr>
<td>7</td>
<td>86.0</td>
<td>86.2</td>
<td>88.7</td>
<td>89.2</td>
</tr>
<tr>
<td>8</td>
<td>89.1</td>
<td>89.2</td>
<td>91.6</td>
<td>91.9</td>
</tr>
<tr>
<td>9</td>
<td>89.7</td>
<td>89.8</td>
<td>92.0</td>
<td>92.1</td>
</tr>
<tr>
<td>10</td>
<td>85.0</td>
<td>85.9</td>
<td>87.4</td>
<td>88.1</td>
</tr>
<tr>
<td><strong>avg.</strong></td>
<td><strong>86.7</strong></td>
<td><strong>87.0</strong></td>
<td><strong>89.2</strong></td>
<td><strong>89.6</strong></td>
</tr>
</tbody>
</table>

**Table:** Parsing results (Malt and MATE parsers, LAS) for original and new tags
### Parsing Experiments

#### Results

<table>
<thead>
<tr>
<th>fold</th>
<th>Malt</th>
<th></th>
<th>MATE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>orig</td>
<td>new</td>
<td>orig</td>
<td>new</td>
</tr>
<tr>
<td>1</td>
<td>84.0</td>
<td>84.3</td>
<td>85.4</td>
<td>86.3</td>
</tr>
<tr>
<td>2</td>
<td>84.2</td>
<td>84.7</td>
<td>87.1</td>
<td>87.6</td>
</tr>
<tr>
<td>3</td>
<td>89.0</td>
<td>89.3</td>
<td>91.7</td>
<td>91.7</td>
</tr>
<tr>
<td>4</td>
<td>85.3</td>
<td>85.9</td>
<td>88.5</td>
<td>89.1</td>
</tr>
<tr>
<td>5</td>
<td>89.0</td>
<td>88.9</td>
<td>91.2</td>
<td>91.5</td>
</tr>
<tr>
<td>6</td>
<td>86.0</td>
<td>85.5</td>
<td>88.0</td>
<td>88.4</td>
</tr>
<tr>
<td>7</td>
<td>86.0</td>
<td>86.2</td>
<td>88.7</td>
<td>89.2</td>
</tr>
<tr>
<td>8</td>
<td>89.1</td>
<td>89.2</td>
<td>91.6</td>
<td>91.9</td>
</tr>
<tr>
<td>9</td>
<td>89.7</td>
<td>89.8</td>
<td>92.0</td>
<td>92.1</td>
</tr>
<tr>
<td>10</td>
<td>85.0</td>
<td>85.9</td>
<td>87.4</td>
<td>88.1</td>
</tr>
<tr>
<td></td>
<td><strong>avg.</strong></td>
<td><strong>86.7</strong></td>
<td><strong>87.0</strong></td>
<td><strong>89.2</strong></td>
</tr>
</tbody>
</table>

**Table:** Parsing results (Malt and MATE parsers, LAS) for original and new tags
Summary

Contribution

• Extension to the STTS $\rightarrow$ more informative analysis of modification

Proof of concept

• A more detailed, syntactically motivated analysis of modification on the POS level can support data-driven syntactic parsing

Future Work

• Validate results on larger data set

• Show that the new tags be learned by a POS tagger (or parser) with sufficient accuracy to be useful
Thank You!
Questions?


References (2)


Backup slides
Error Analysis

Dies sei selbst in jenen Entwicklungsländern ... nicht üblich

PDS VAFIN ADV/PTKFO APPR PDAT NN PTKNEG ADJD

Figure: Parser output tree for orig. (red) and new tags (green)
### Annotation Experiment

#### Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>ADJD</th>
<th>ADV</th>
<th>PFO</th>
<th>PINT</th>
<th>PLEX</th>
<th>MODP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJD</td>
<td>63</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ADV</td>
<td>6</td>
<td>343</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>PFO</td>
<td>0</td>
<td>12</td>
<td>67</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PINT</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>49</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>PLEX</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>MODP</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

**Table:** Confusion matrix for adverbs (ADV), predicative adjectives (ADJD), focus-associated particles (PFO), intensifiers (PINT), lexicalised particles (PLEX) and modal particles (MODP)
### Table: Confusion matrix for adverbs (ADV), predicative adjectives (ADJD), focus-associated particles (PFO), intensifiers (PINT), lexicalised particles (PLEX) and modal particles (MODP)

<table>
<thead>
<tr>
<th></th>
<th>ADJD</th>
<th>ADV</th>
<th>PFO</th>
<th>PINT</th>
<th>PLEX</th>
<th>MODP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJD</td>
<td>63</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ADV</td>
<td>6</td>
<td>343</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>PFO</td>
<td>0</td>
<td>12</td>
<td>67</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PINT</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>49</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>PLEX</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>MODP</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>ADJD</td>
<td>ADV</td>
<td>PFO</td>
<td>PINT</td>
<td>PLEX</td>
<td>MODP</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>ADJD</td>
<td>63</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ADV</td>
<td>6</td>
<td>343</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>PFO</td>
<td>0</td>
<td>12</td>
<td>67</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PINT</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>49</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>PLEX</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>MODP</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Table: Confusion matrix for adverbs (ADV), predicative adjectives (ADJD), focus-associated particles (PFO), intensifiers (PINT), lexicalised particles (PLEX) and modal particles (MODP)
## Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>ADJD</th>
<th>ADV</th>
<th>PFO</th>
<th>PINT</th>
<th>PLEX</th>
<th>MODP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJD</td>
<td>63</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ADV</td>
<td>6</td>
<td>343</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>PFO</td>
<td>0</td>
<td>12</td>
<td>67</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PINT</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>49</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>PLEX</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>MODP</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

**Table:** Confusion matrix for adverbs (ADV), predicative adjectives (ADJD), focus-associated particles (PFO), intensifiers (PINT), lexicalised particles (PLEX) and modal particles (MODP)
(1) ADV vs PTKFO

Hennemann hatte seinen Rückzug **bereits** im September angeboten.

“Hennemann had already offered his withdrawal in September.”
Annotation Experiment
Ambiguous Cases

(2) ADV vs ADJD

Wer sich weigere, werde durch Drogen gefügig gemacht
Who himself refuses, is by drugs compliant made

“Who refuses is made compliant by drugs”
(3) ADV vs PTKLEX

Diese werden **immer wieder** missbraucht
These become always again abused

“Again and again, these become abused”