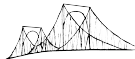


# Towards a syntactically motivated analysis of modifiers in German

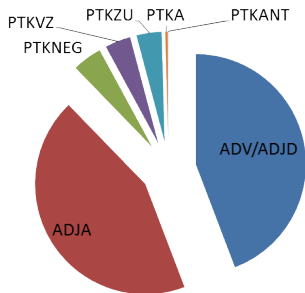
Ines Rehbein & Hagen Hirschmann

KONVENS 2014

October 8, 2014



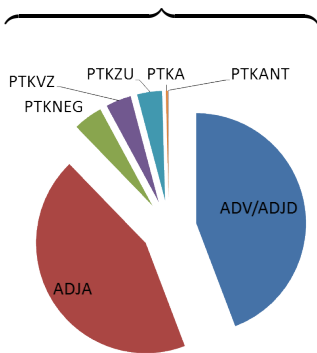
# Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)



Relative frequencies of modifying POS in the TIGER corpus

# Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)

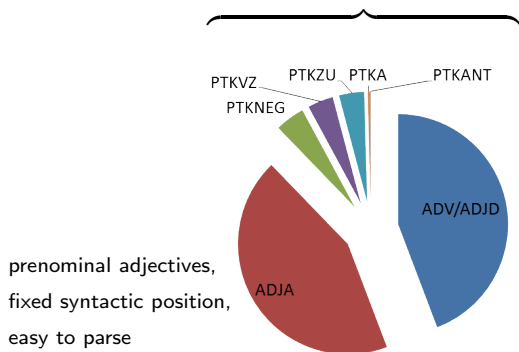
Closed classes (e.g. *nicht* for PTKNEG – negation), relatively infrequent, relatively homogeneous syntax per class



Relative frequencies of modifying POS in the TIGER corpus

# Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)

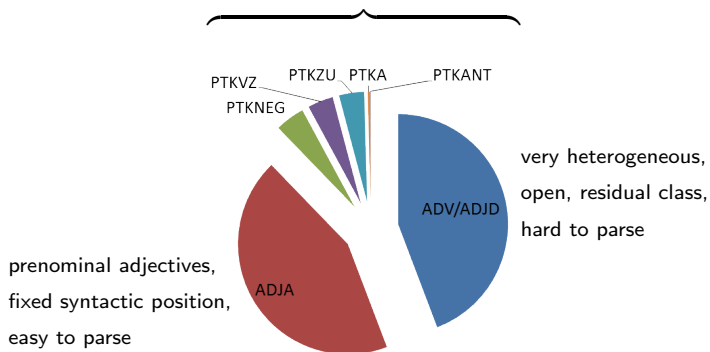
Closed classes (e.g. *nicht* for PTKNEG – negation), relatively infrequent, relatively homogeneous syntax per class



Relative frequencies of modifying POS in the TIGER corpus

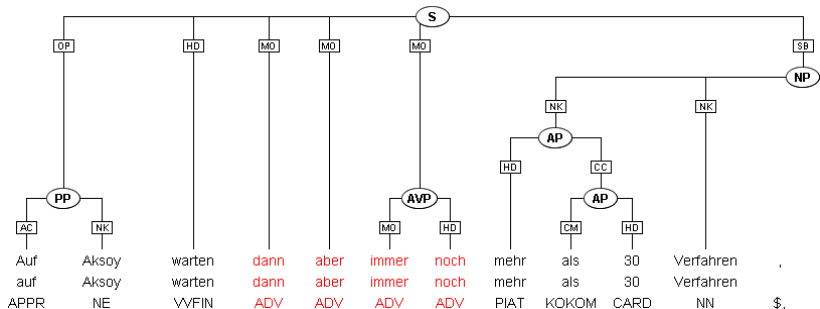
# Modifying parts of speech (POS) in the Stuttgart-Tübingen Tagset (STTS)

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



Relative frequencies of modifying POS in the TIGER corpus

# 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.")



# Resulting research question

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



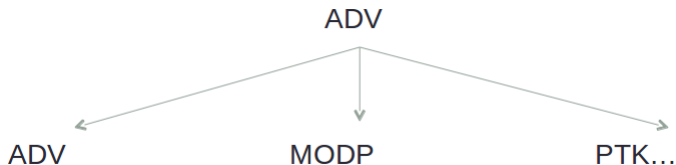
## Redefining ADV and ADJD

- 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

POS	# STTS	# new	# agr.	Fleiss' $\kappa$
VAPP	21	21	21	1.000
VVPP	173	172	172	0.989
ADJD	191	74	63	0.891
ADV	445	378	343	0.800
PTKFO	-	80	67	0.797
PTKINT	-	63	49	0.788
PTKLEX	-	33	17	0.594
MODP	-	12	6	0.515
<b>total</b>	<b>830</b>	<b>833</b>	<b>88.3%</b>	<b>0.838</b>

**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  $\rightarrow$  PTKFO

```
[cat="NP"] >@I [pos="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

## Setup

- 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

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

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

# Parsing Experiments

## Results

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

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

# Summary

## Contribution

- Extension to the STTS → 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

- Brants, S., Dipper, S., Eisenberg, P., Hansen, S., König, E., Lezius, W., Rohrer, C., Smith, G., and Uszkoreit, H. (2004). TIGER: Linguistic Interpretation of a German Corpus. *Journal of Language and Computation*, 2004 (2), 597-620.
- Brill, E. (1992). A simple rule-based part of speech tagger. *3rd conference on Applied natural language processing (ANLC'92)*, Trento, Italy.
- Dickinson, M. and Meurers, D. W. (2003). Detecting Errors in Part-of-Speech Annotation. *10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03)*, Budapest, Hungary.
- Dickinson, M. (2006). An Investigation into Improving Part-of-Speech Tagging. *Proceedings of the Third Midwest Computational Linguistics Colloquium (MCLC-06)*, Urbana-Champaign, IL.
- Dligach, D. and Palmer, M. (2011). Reducing the Need for Double Annotation. *Proceedings of the 5th Linguistic Annotation Workshop (LAW V '11)*, Portland, Oregon.
- Eskin, E. (2000). Automatic Corpus Correction with Anomaly Detection. In *1st Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Seattle, Washington.
- Květoň, P. and Oliva, K. (2002). (Semi-)Automatic Detection of Errors in PoS-Tagged Corpora, *19th International Conference on Computational Linguistics (COLING-02)*.
- Loftsson, H. (2009). Correcting a POS-Tagged Corpus Using Three Complementary Methods. *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, Athens, Greece.

## References (2)

- Manning, C. D. (2011). Part-of-speech Tagging from 97% to 100%: Is It Time for Some Linguistics?. *Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing'11)*, Tokyo, Japan.
- Rehbein, I., Schalowski, S. and Wiese, H. (2014). The KiezDeutsch Korpus (KiDKo) Release 1.0. *The 9th International Conference on Language Resources and Evaluation (LREC-14)*, Reykjavik, Iceland.
- Rocio, V. Silva, J. and Lopes, G. (2007). Detection of Strange and Wrong Automatic Part-of-speech Tagging. *Proceedings of the Artificial Intelligence 13th Portuguese Conference on Progress in Artificial Intelligence (EPIA'07)*, Guimarães, Portugal.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *International Conference on New Methods in Language Processing*, Manchester, UK.
- Schiller, A., Teufel, S., Stöckert, C. and Thielen, C. (1999). *Guidelines für das Tagging deutscher Textkorpora mit STTS*, Universität Stuttgart, Universität Tübingen.  
<http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.
- Toutanova, K. and Manning, C. D. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. *Proceedings of the conference on Empirical methods in natural language processing and very large corpora (EMNLP '00)*, Hong Kong.
- van Halteren, H. (2000). The Detection of Inconsistency in Manually Tagged Text. *Proceedings of the COLING-2000 Workshop on Linguistically Interpreted Corpora*, Centre Universitaire, Luxembourg.

# Backup slides

# Error Analysis

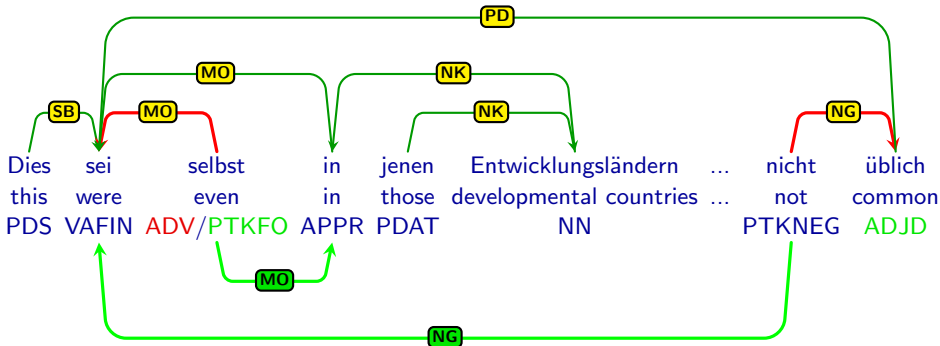


Figure: Parser output tree for orig. (red) and new tags (green)

# Annotation Experiment

## Confusion Matrix

	ADJD	ADV	PFO	PINT	PLEX	MODP
ADJD	63	6	0	0	0	0
ADV	6	343	15	6	6	5
PFO	0	12	67	2	1	0
PINT	0	9	0	49	2	0
PLEX	0	9	0	1	17	0
MODP	0	5	0	0	1	6

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

# Annotation Experiment

## Confusion Matrix

	ADJD	ADV	PFO	PINT	PLEX	MODP
ADJD	63	6	0	0	0	0
ADV	6	343	15	6	6	5
PFO	0	12	67	2	1	0
PINT	0	9	0	49	2	0
PLEX	0	9	0	1	17	0
MODP	0	5	0	0	1	6

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

# Annotation Experiment

## Confusion Matrix

	ADJD	ADV	PFO	PINT	PLEX	MODP
ADJD	63	6	0	0	0	0
ADV	6	343	15	6	6	5
PFO	0	12	67	2	1	0
PINT	0	9	0	49	2	0
PLEX	0	9	0	1	17	0
MODP	0	5	0	0	1	6

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



# Annotation Experiment

## Confusion Matrix

	ADJD	ADV	PFO	PINT	PLEX	MODP
ADJD	63	6	0	0	0	0
ADV	6	343	15	6	6	5
PFO	0	12	67	2	1	0
PINT	0	9	0	49	2	0
PLEX	0	9	0	1	17	0
MODP	0	5	0	0	1	6

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

# Annotation Experiment

## Ambiguous Cases

### (1) ADV vs PTKFO

Hennemann hatte seinen Rückzug **bereits** im September  
Hennemann had his withdrawal already in September  
angeboten.  
offered.

“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”

# Annotation Experiment

## Ambiguous Cases

### (3) ADV vs PTKLEX

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

“Again and again, these become abused”