Parser evaluation across text types

Yannick Versley
Seminar für Sprachwissenschaft
Universität Tübingen
E-mail: versley@sfs.uni-tuebingen.de

Workshop on Treebanks and Linguistic Theories,
December 2005

Outline

Parsing German

- (Beil et al., 2002; Schulte im Walde, 2003)
  Unsupervised training of a hand-written PCFG
- (Dubey and Keller, 2003)
  Lexicalized parsing for German isn't better than unlexicalized parsing, but keeping some horizontal context helps

Parsing German

- (Beil et al., 2002; Schulte im Walde, 2003)
  Unsupervised training of a hand-written PCFG
- (Dubey and Keller, 2003)
  Lexicalized parsing for German isn't better than unlexicalized parsing, but keeping some horizontal context helps
- (Foth et al., 2004)
  Weighted Constraint Dependency Grammar
  Parsing as optimization with declarative rules

Parsing German

- (Beil et al., 2002; Schulte im Walde, 2003)
  Unsupervised training of a hand-written PCFG
- (Dubey and Keller, 2003)
  Lexicalized parsing for German isn't better than unlexicalized parsing, but keeping some horizontal context helps
- (Foth et al., 2004)
  Weighted Constraint Dependency Grammar
  Parsing as optimization with declarative rules
- (Müller, 2004)
  Extended finite-state approach

Parsing German

- (Beil et al., 2002; Schulte im Walde, 2003)
  Unsupervised training of a hand-written PCFG
- (Dubey and Keller, 2003)
  Lexicalized parsing for German isn't better than unlexicalized parsing, but keeping some horizontal context helps
- (Foth et al., 2004)
  Weighted Constraint Dependency Grammar
  Parsing as optimization with declarative rules
- (Müller, 2004)
  Extended finite-state approach
- (Schiehlen, 2004), (Dubey, 2005)
  Accurate unlexicalized PCFGs

Evaluations

	Dubey	Schiehlen	Foth	Müller	This
TüBa-D/Z (Dep)			*		*
TüBa-D/Z (GF)			*	+	*
Negra (Dep)		+	+		*
Negra (CS)	+	+			
CDG Corpus (Dep)			+		*

- Different measures
  (Parseval vs. Dependencies vs. Grammatical Functions)
- Different text types
  TüBa/Negra: newspaper text
  CDG Corpus: newswire text, trivial/serious literature, law text

Statistical parsing of German

- Freer word order and richer morphology than English
- More nonprojectivity
- Lexicalization helps for other highly configurational languages such as French (Arun and Keller, 2005), but only very little for German (Dubey and Keller, 2003)

[1] = all sentences of the first release

Statistical parsing of German

- Freer word order and richer morphology than English
- More nonprojectivity
- Lexicalization helps for other highly configurational languages such as French (Arun and Keller, 2005), but only very little for German (Dubey and Keller, 2003)
- As an example for a state of the art statistical parser, we developed an unlexicalized PCFG parser using BitPar (Schmid, 2004) and TüBa-D/Z

[1] = all sentences of the first release

Statistical parsing of German

- Freer word order and richer morphology than English
- More nonprojectivity
- Lexicalization helps for other highly configurational languages such as French (Arun and Keller, 2005), but only very little for German (Dubey and Keller, 2003)
- As an example for a state of the art statistical parser, we developed an unlexicalized PCFG parser using BitPar (Schmid, 2004) and TüBa-D/Z
- Contrary to Negra, TüBa-D/Z (2nd release) contains morphology information for 15 260 sentences[1].

[1] = all sentences of the first release

From a treebank to a PCFG

- Different goals for treebanking and PCFG construction:
  - annotating linguistic information in a convenient way vs.
  - independence assumptions of a PCFG

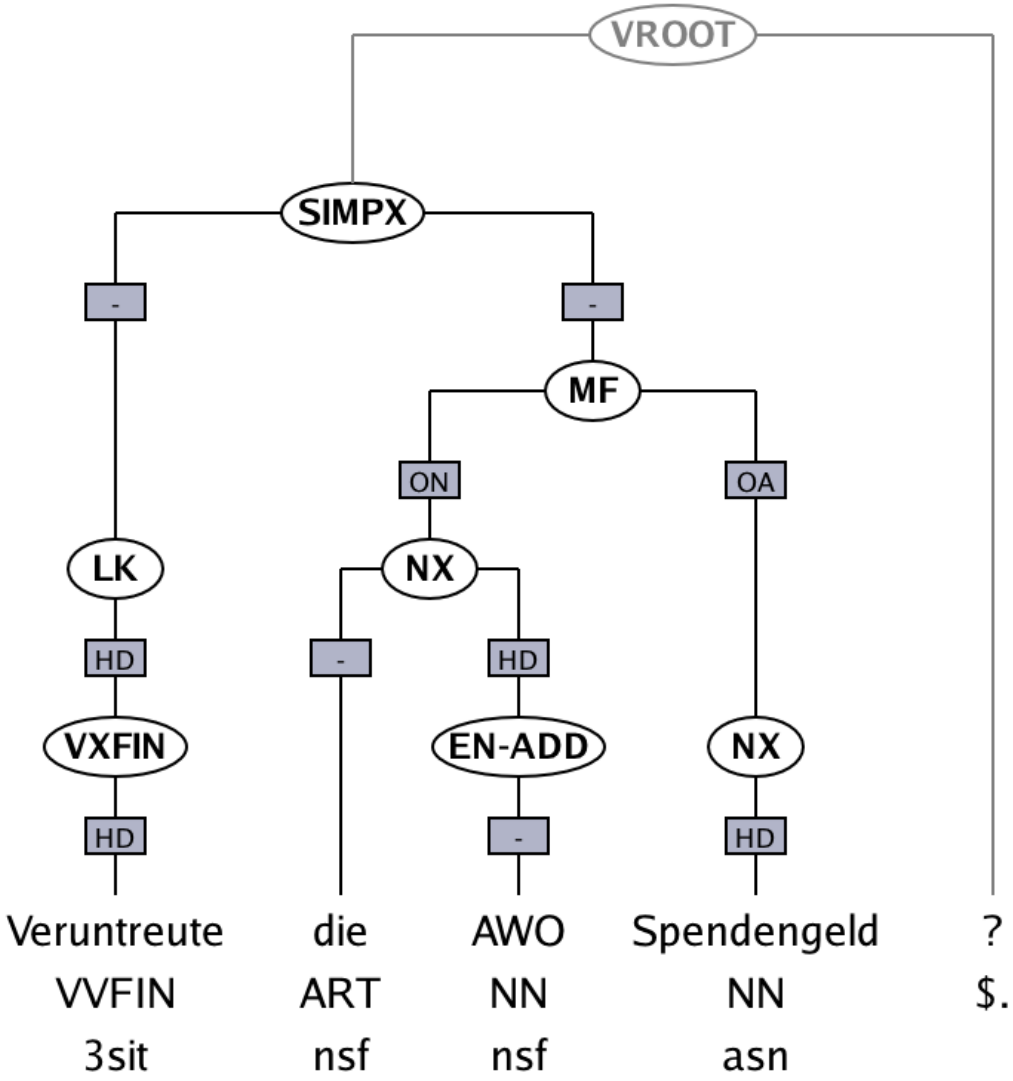
From a treebank to a PCFG

- Different goals for treebanking and PCFG construction:
  - annotating linguistic information in a convenient way vs.
  - independence assumptions of a PCFG
- Enriching node labels by adding information that would be lost to the PCFG
  - Incorporate morphological information
  - Subcategorize clauses by their form
  - Annotate arguments of verbs in the topological fields

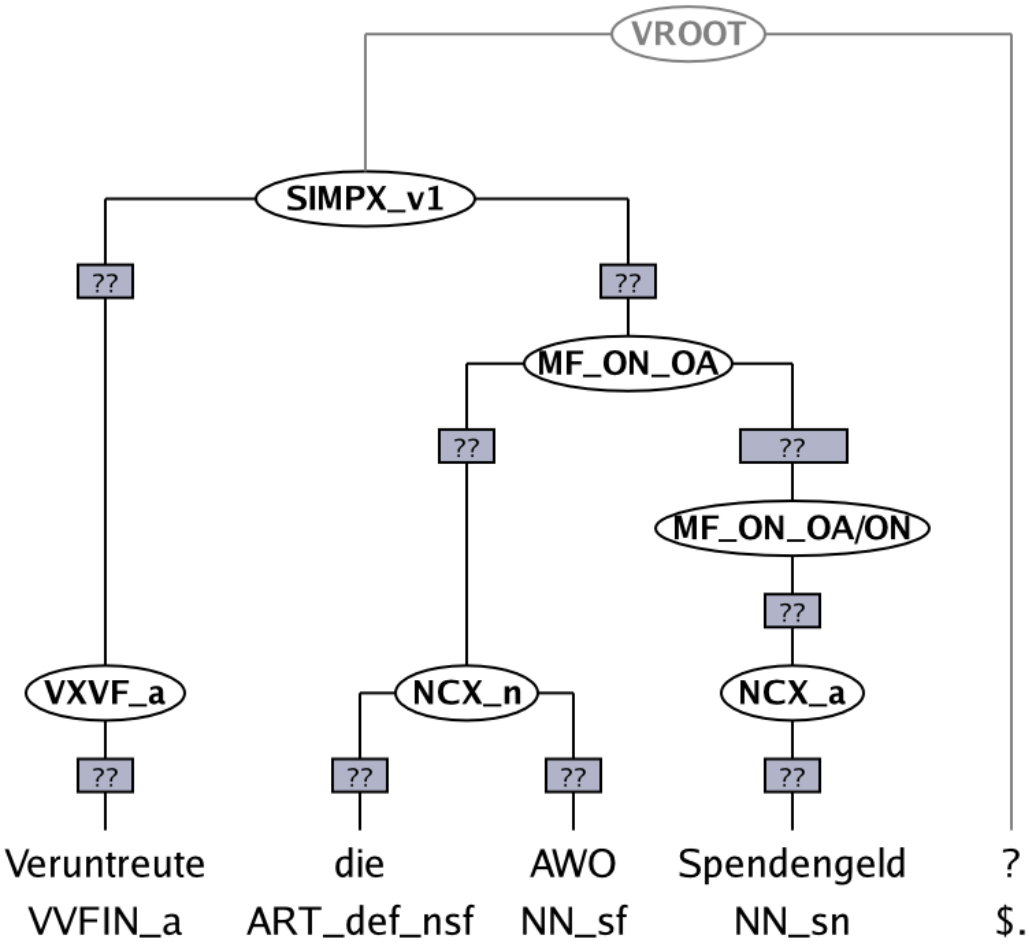
From a treebank to a PCFG

- Different goals for treebanking and PCFG construction:
  - annotating linguistic information in a convenient way vs.
  - independence assumptions of a PCFG
- Enriching node labels by adding information that would be lost to the PCFG
  - Incorporate morphological information
  - Subcategorize clauses by their form
  - Annotate arguments of verbs in the topological fields
- Limiting yourself to the information that can be estimated from the treebank
  - Underspecified morphology
  - Disregarding rare verb complements in subcategorization
  - Better unknown word classification

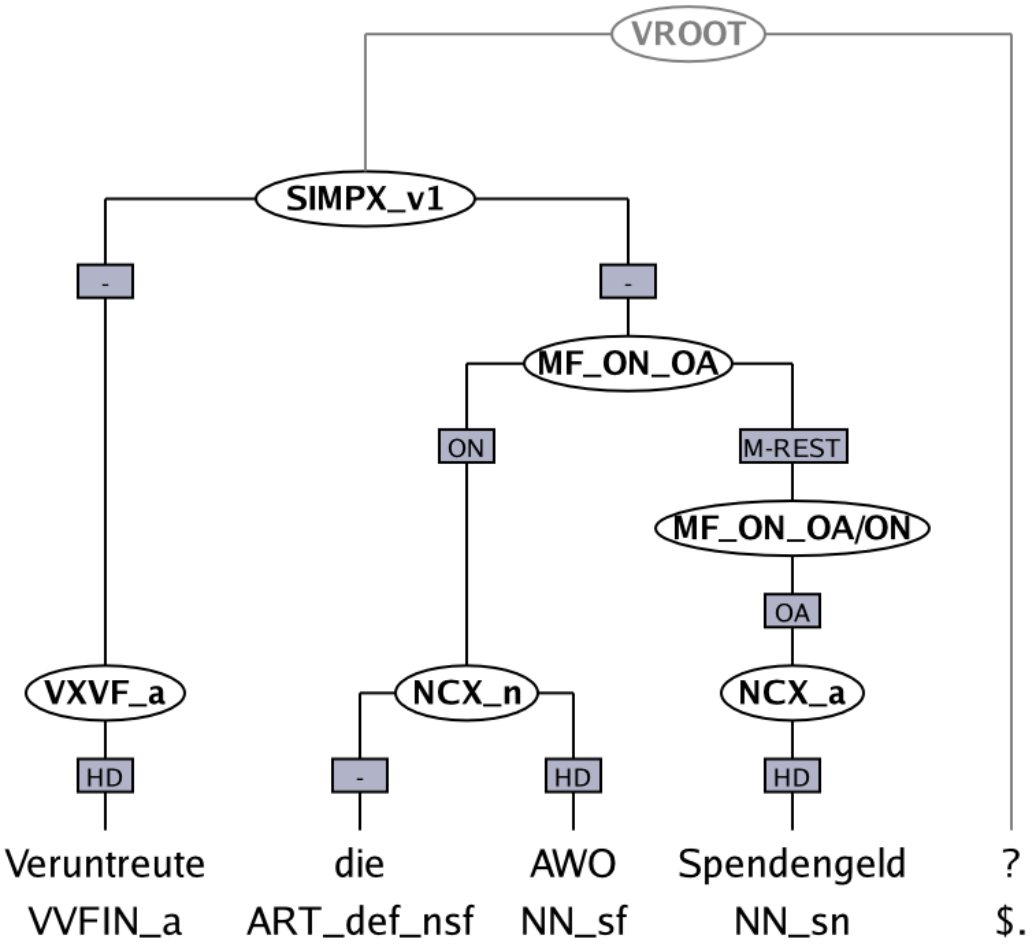
An Example: Treebank



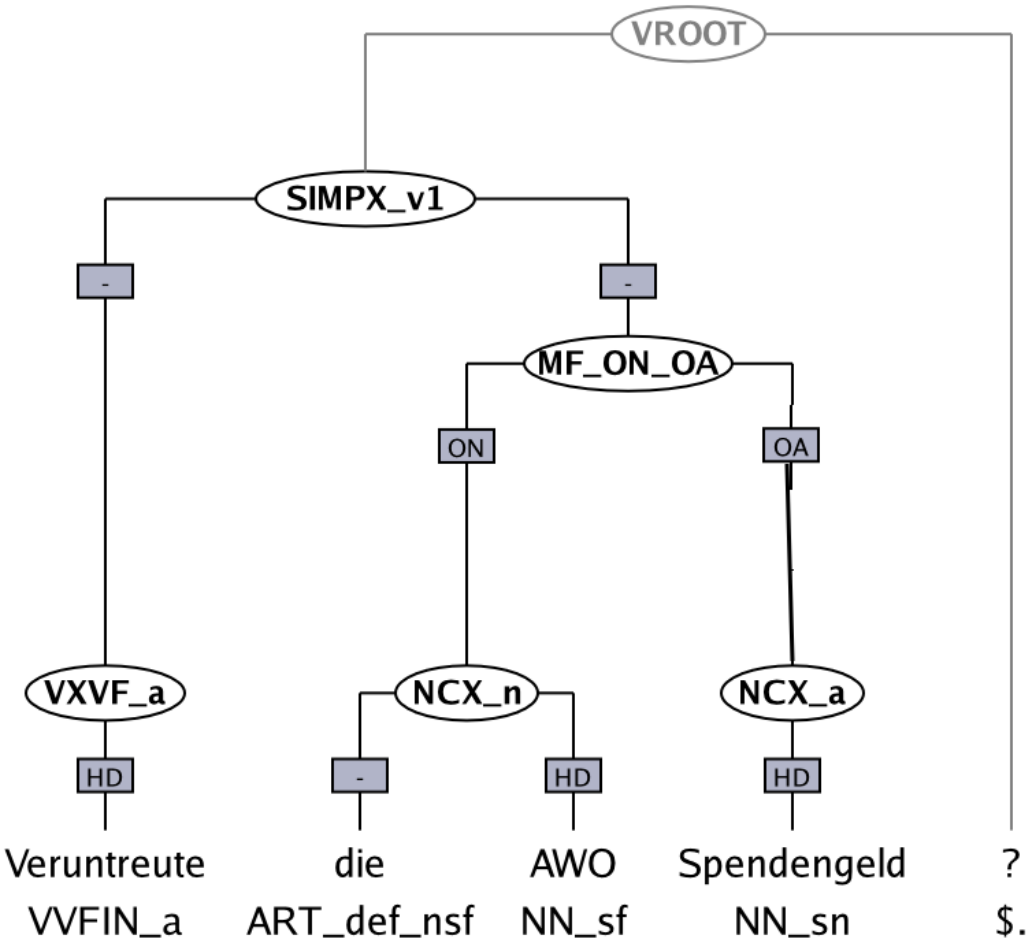
An Example: PCFG



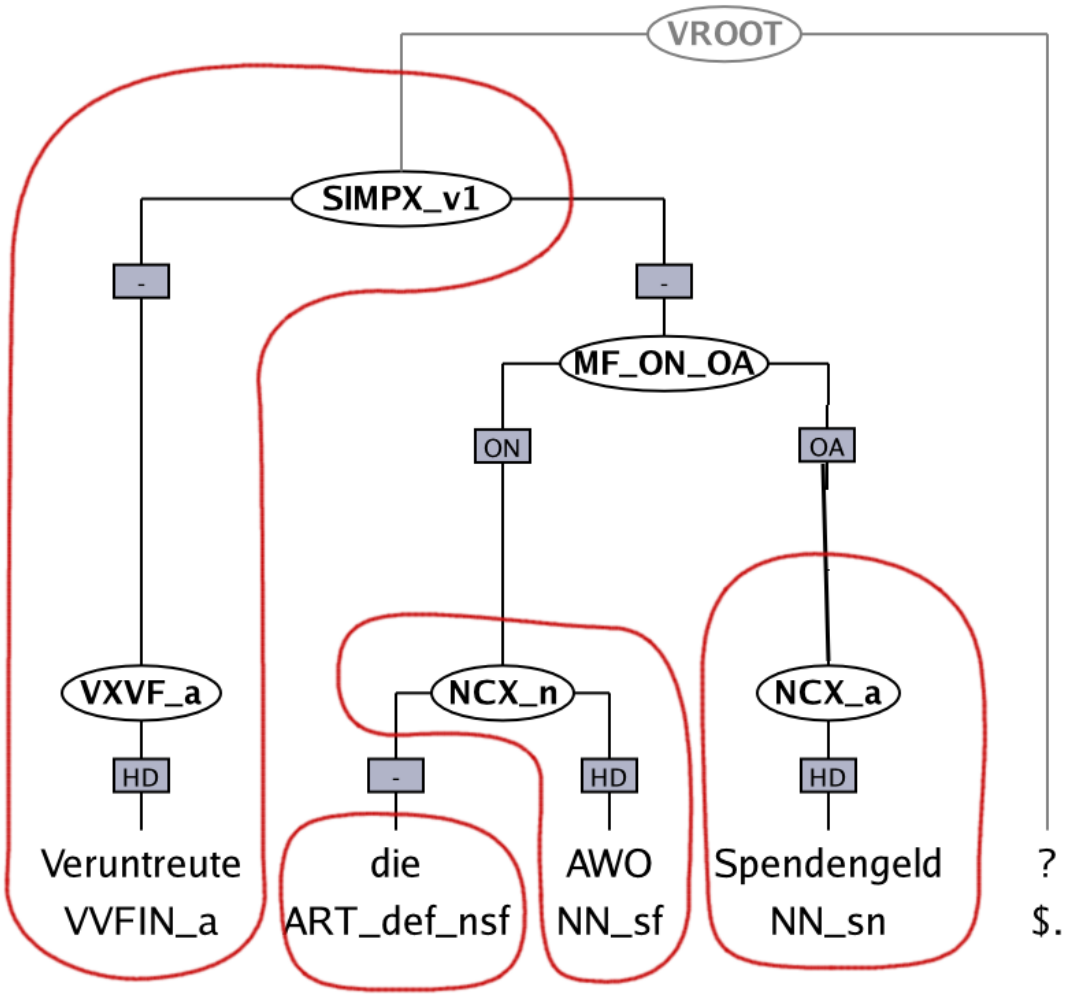
An Example: Relabeling



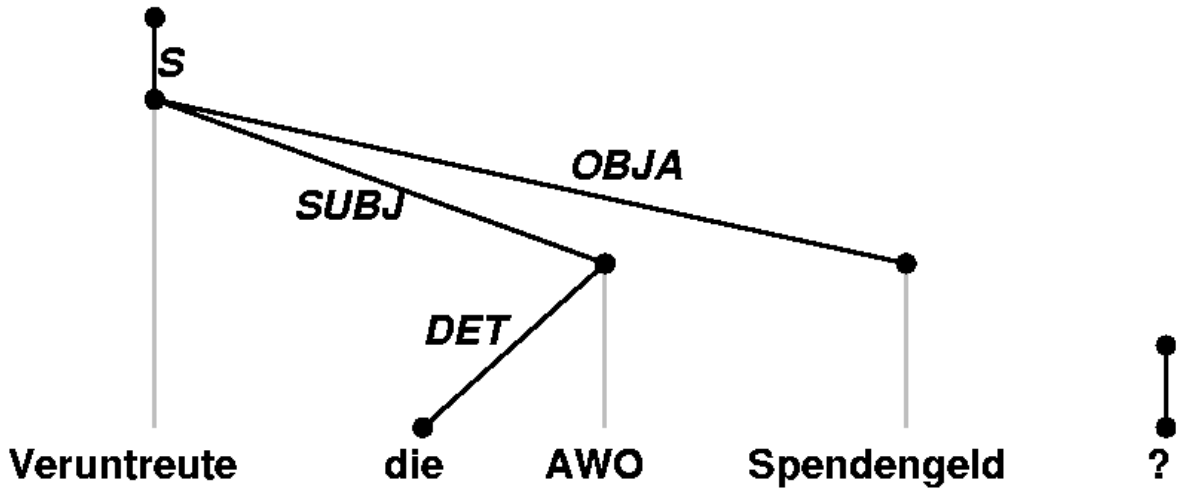
An Example: Undo Markovization



An Example: Head projection



An Example: Dependencies



Results (quantitative)

- Baseline for the unmodified treebank grammar:
  80.0% unlabeled F-measure against manually converted trees
  78.0% on Negra
- TüBa-D/Z:
  85.4% unlabeled F-measure against manually converted trees
  87.2% against automatically converted trees
- Negra:
  84.1% on development set (sent. 1-3000),
  83.6% on unseen test set (sent. 18 603-19 602)
- Grammatical functions on TüBa-D/Z (F-measure):
  86.9% (SUBJ), 73.5% (OBJA), 53.8% (OBJD)

Comparison to previous work

- For the grammatical function task, Foth's WCDG parser performs slightly better than Müller's parser, which in turn visibly outperforms the PCFG parser
- On the Negra test set, our PCFG parser performs better than Schiehlen's (83.6% vs. 81.7% unlabeled F-measure)
  Foth's WCDG parser outperforms both (89.0% unlabeled F-measure)
- For the "serious literature" texts, our PCFG parser is slightly better in terms of unlabeled F-measure (80.7% vs. 78.0%)
- For the law text, our PCFG parser performs much worse than Foth's WCDG parser (62.2% vs. 88.8%)

Conclusion

- With existing tools, it is possible to construct an unlexicalized PCFG parser from a treebank in a reasonable timeframe (about 6 months).
- Statistical parsing for German[2] is lacking not only in comparison with results for English, but also in comparison with manually constructed parsers for German
- Introducing morphological features is important, but easily leads to sparse data problems in a PCFG.
- Outlook: Reranking with global features

[2] This probably applies to other languages with freer word order

Thanks for listening

The End

Arun, A. and Keller, F. (2005).
Lexicalization in crosslinguistic probabilistic parsing: The case of french.
In *ACL'05*, pages 306–313.

Beil, F., Prescher, D., Schmid, H., and Schulte im Walde, S. (2002).
Evaluation of the Gramotron parser for German.
In *LREC'02*.

Dubey, A. (2005).
What to do when lexicalization fails: parsing German with suffix analysis and smoothing.
In *Proc. 43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005)*.

Dubey, A. and Keller, F. (2003).

Probabilistic parsing for german using sister-head dependencies.
In *ACL'2003*.

Foth, K., Daum, M., and Menzel, W. (2004).
A broad-coverage parser for German based on defeasible constraints.
In *KONVENS 2004, Beiträge zur 7. Konferenz zur Verarbeitung natürlicher Sprache*, pages 45–52, Wien.

Müller, F. H. (2004).
Annotating grammatical functions in german using finite-state cascades.
In *Proc. 20th Int. Conference on Computational Linguistics (COLING 2004)*.

Schiehlen, M. (2004).
Annotation strategies for probabilistic parsing in german.

In *Proc. 20th Int. Conference on Computational Linguistics (COLING-2004)*.

Schmid, H. (2004).

Efficient parsing of highly ambiguous context-free grammars with bit vectors.

In *Proc. 20th Int. Conference on Computational Linguistics (COLING 2004)*.

Schulte im Walde, S. (2003).

Experiments on the Automatic Induction of German Semantic Verb Classes.

PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.