Treebanks of German – Annotation Schemes and NLP Applications

Erhard Hinrichs

{eh@sfs.uni-tuebingen.de

SfS-CL, University of Tübingen, Germany

Seminar für Sprachwissenschaft
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Talk Outline

- Clause Structure of German
- Negra/TIGER and TüBa-D/Z treebanks for German
- Using Treebanks for NLP tasks:
  - Statistical Parsing of German: Is it really that difficult to parse German?
  - Computational Anaphora Resolution for German

Clause Types

(1) a. V2 clause:

Peter **wird** das Buch gelesen haben.
Peter will the book read have

'Peter will have read the book.'

b. V1 clause:

Wird Peter das Buch gelesen haben?
Will Peter the book have read

'Will Peter have read the book?'

c. VL clause:

dass Peter das Buch gelesen haben **wird.**
that Peter the book read have will

'... that Peter will have read the book.'

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Flexible Phrase Ordering

(2) a. Der Mann hat gestern den Roman gelesen.

The man has yesterday the novel read

'The man read the novel yesterday.'

b. Gestern hat der Mann den Roman gelesen

c. Den Roman hat der Mann gestern gelesen

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Discontinuous Constituents

(3) Der Mann hat gestern den Roman gelesen, den ihm
The man has yesterday the novel read which him
Peter empfahl.
Peter recommended

'Yesterday the man read the novel which Peter recommended to him.'

(4) Peter soll dem Mann empfohlen haben, den Roman zu
Peter is to the man recommended have the novel to
lesen.
read

'Peter is said to have recommended to the man to read the novel.'

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Topological Fields

(5) a. $[_{VF}$ $[_{NP}$ Peter $]$ $]$ $[_{LK}$ wird $]$ $[_{MF}$ $[_{NP}$ das Buch $]$ $]$
$[_{RK}$ $[_{VC}$ gelesen haben. $]$ $]$

b. $[_{LK}$ Wird $]$ $[_{MF}$ $[_{NP}$ Peter $]$ $[_{NP}$ das Buch $]$ $]$
$[_{RK}$ $[_{VC}$ gelesen haben? $]$ $]$

c. $[_{LK}$ $[_{CF}$ dass $]$ $]$ $[_{MF}$ $[_{NP}$ Peter $]$ $[_{NP}$ das Buch $]$ $]$
$[_{RK}$ $[_{VC}$ gelesen haben wird. $]$ $]$

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

The TüBa-D/Z treebank

- Linguistically annotated German newspaper corpus
- Based on data taken from the daily issues of ’die tageszeitung’ (taz)
- Manual annotation supported by annotate tool
- release 3.0: appr. 27 000 sentences (470 000 words).

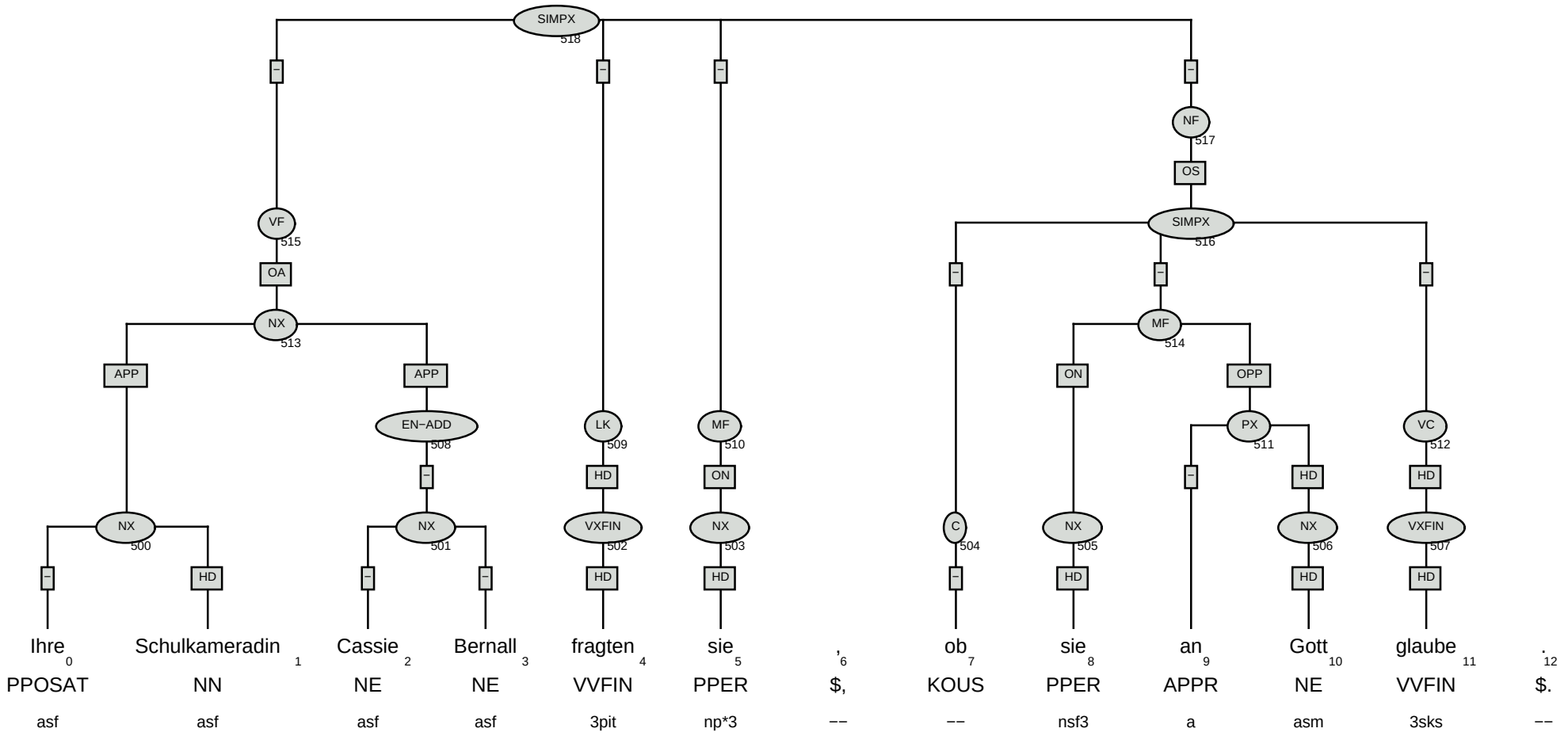
Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

TüBa-D/Z: Levels of Annotation

- inflectional morphology
- syntactic constituency
- grammatical functions
- (complex) named entities
- anaphora and coreference relations

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Example



They asked their fellow student Cassie Bernall whether she believes in God.

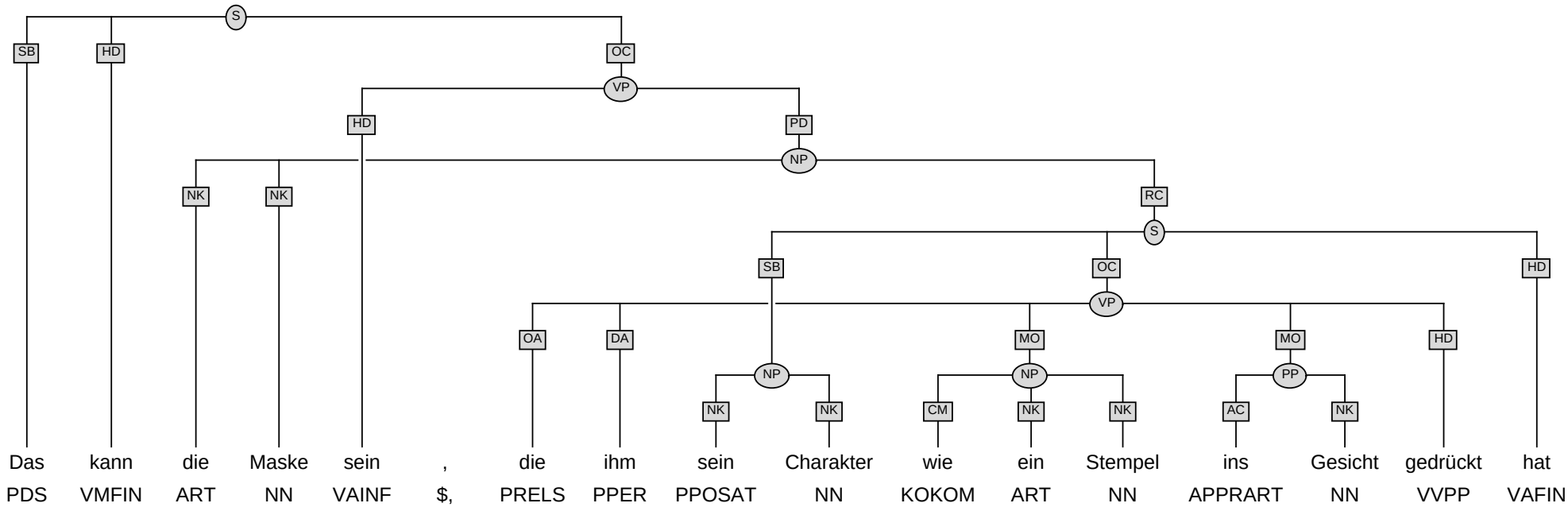
Negra and TüBa-D/Z

- both are based on newspaper texts: Frankfurter Rundschau, taz
- both use same POS tagset: STTS
- both annotate constituent structure and function-argument structure
- Negra/Tiger: 40 000 sentences; TüBa-D/Z: 27 000 sentences

- **different annotation decisions**

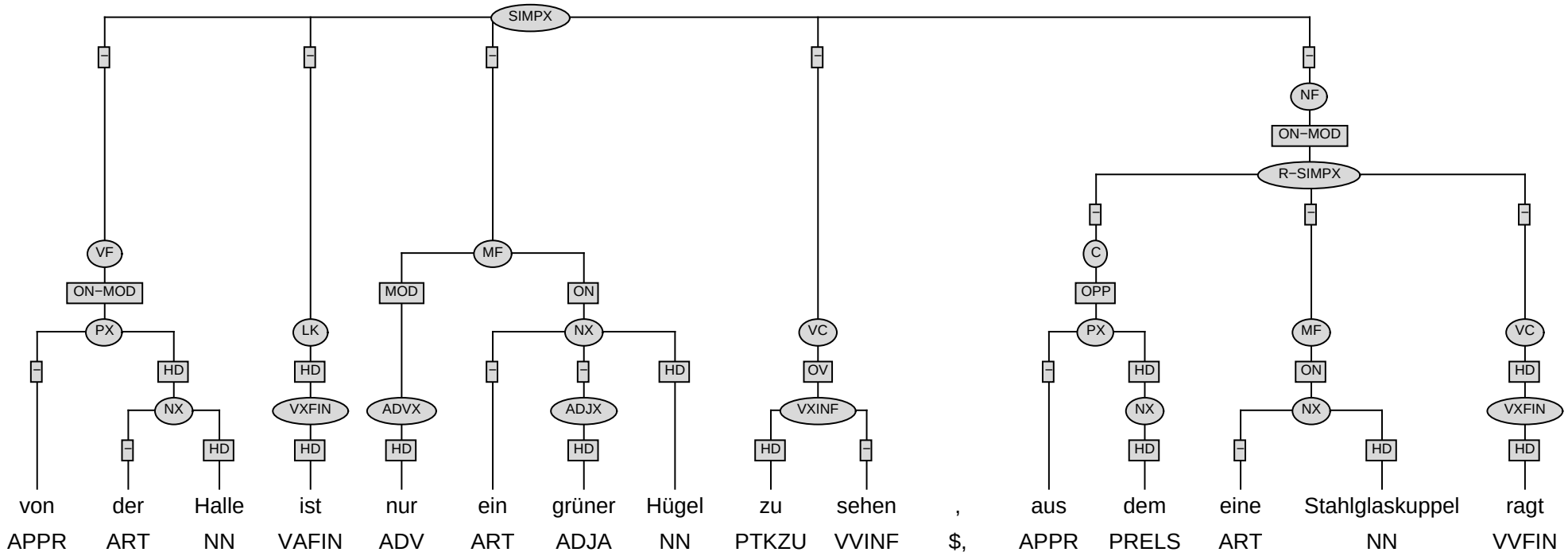
Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

A Tree from Negra



That could be the mask that his character pressed like a stamp on his face.

A Tree from TüBa-D/Z



From the hall, only a green hill can be seen, over which looms a steel and glass cupola.

Main Differences in Annotation

- phrase structure
  - Negra: extremely flat
  - TüBa-D/Z: premodification flat, postmodification high
- clause structure
  - Negra: VP $\Rightarrow$ crossing branches
    no unary nodes
  - TüBa-D/Z: topological fields
- long-distance relationships
  - Negra: crossing branches
  - TüBa-D/Z: pure tree structure + special labels

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

State of the Art in Stat. Parsing

	prec.	recall	F-score
English			
Collins (1999)	88.7	88.6	88.6
Charniak (2000)	89.8	89.6	89.7
Klein & Manning (2003)	86.9	85.7	86.3
German			
Dubey & Keller (2003)	73.9	74.2	74.0
Dubey & Keller lex.	67.9	66.1	67.0
Schiehlen (2005)			68.4

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Research Questions

comparison of 2 treebanks, Negra und TüBa-D/Z:

1. different lexicalization: Stanford parser
2. test different parsers: comparison of treebanks with LoPar and Stanford Parser; with and without markovization
3. structure vs. content: comparison of treebanks concerning grammatical functions (Stanford Parser)

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Experimental Setup

- preprocessing:
  - both: insert virtual root that covers single trees of a sentence
  - TüBa-D/Z: attach parentheses to surrounding tree
  - NEGRA: resolve crossing branches: attach non-head daughters higher
- 10 fold cross validation
- labeled / unlabeled precision and recall

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Results: Lexicalization

		precision	recall	F-score
Negra				
Stanford PCFG	unlab.	71.24	72.68	71.95
	labeled	66.26	67.59	66.92
Stanford lex.	unlab.	71.31	73.12	72.20
	labeled	66.30	67.99	67.13
TüBa-D/Z				
Stanford PCFG	unlab.	93.07	89.41	91.20
	labeled	88.25	84.78	86.48
Stanford lex.	unlab.	91.60	91.21	91.36
	labeled	89.12	88.65	88.88

Results: Lexicalization

		precision	recall	F-score
Negra				
Stanford PCFG	unlab.	71.24	72.68	71.95
	labeled	66.26	67.59	66.92
Stanford lex.	unlab.	71.31	73.12	72.20
	labeled	66.30	67.99	67.13
TüBa-D/Z				
Stanford PCFG	unlab.	93.07	89.41	91.20
	labeled	88.25	84.78	86.48
Stanford lex.	unlab.	91.60	91.21	91.36
	labeled	89.12	88.65	88.88

Results: Lexicalization

		precision	recall	F-score
Negra				
Stanford PCFG	unlab.	71.24	72.68	71.95
	labeled	66.26	67.59	66.92
Stanford lex.	unlab.	71.31	73.12	72.20
	labeled	66.30	67.99	67.13
TüBa-D/Z				
Stanford PCFG	unlab.	93.07	89.41	91.20
	labeled	88.25	84.78	86.48
Stanford lex.	unlab.	91.60	91.21	91.36
	labeled	89.12	88.65	88.88

Results: Lexicalization

- slightly improved results for lexicalization
  - different data split from Dubey and Keller?
  - Stanford Parser miracle cure?

- TüBa-D/Z ca. 20 points better than Negra
  - Stanford Parser miracle cure?
  - deeper structures easier to parse?

Experiment with Different Parsers

- LoPar and Stanford
- Stanford with and without markovization
- Markov parameters: vertical = 1; horizontal = 2
- no lexicalization, only constituent structure

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Results: Different Parsers

	lab. prec.	lab. rec.	lab. F-score
Negra			
LoPar	65.86	67.41	66.62
Stanford	66.26	67.59	66.92
Stanford + markov	69.96	69.95	69.95
TüBa-D/Z			
LoPar	87.39	83.57	85.44
Stanford	88.25	84.78	86.48
Stanford + markov	89.86	88.51	89.18

Results: Different Parsers

	lab. prec.	lab. rec.	lab. F-score
Negra			
LoPar	65.86	67.41	66.62
Stanford	66.26	67.59	66.92
Stanford + markov	69.96	69.95	69.95
TüBa-D/Z			
LoPar	87.39	83.57	85.44
Stanford	88.25	84.78	86.48
Stanford + markov	89.86	88.51	89.18

Results: Different Parsers

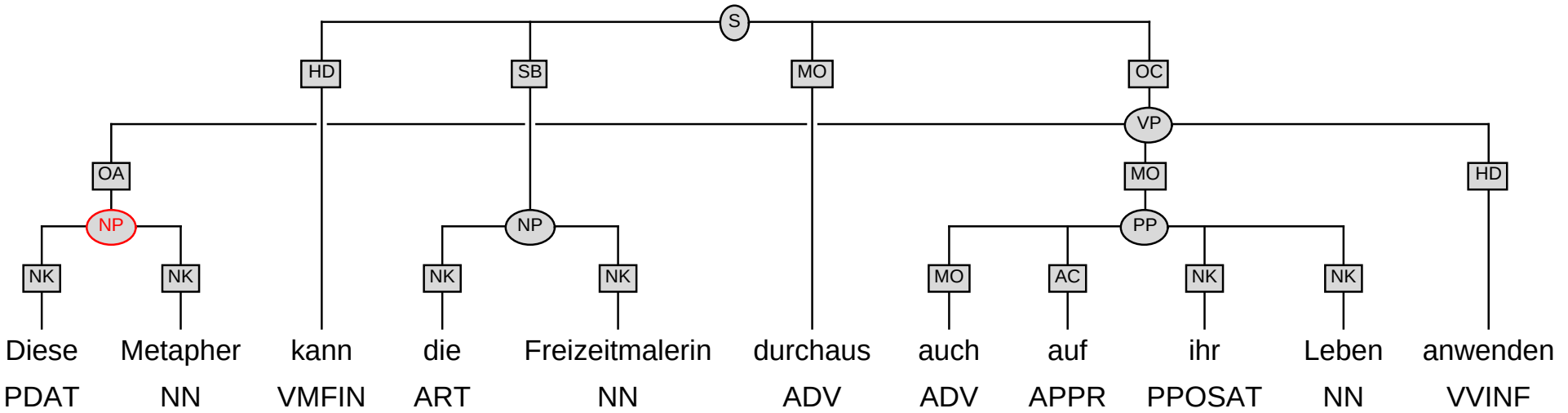
	lab. prec.	lab. rec.	lab. F-score
Negra			
LoPar	65.86	67.41	66.62
Stanford	66.26	67.59	66.92
Stanford + markov	69.96	69.95	69.95
TüBa-D/Z			
LoPar	87.39	83.57	85.44
Stanford	88.25	84.78	86.48
Stanford + markov	89.86	88.51	89.18
Stanford lex.	89.12	88.65	88.88

Results: Different Parsers

- difference between treebanks constant
  - deeper structures better than many daughters
  - good structure by topological fields
  - conversion from crossing branches to CFG tree leads to inconsistencies
- markovization better than lexicalization
- Stanford Parser PCFG better than LoPar

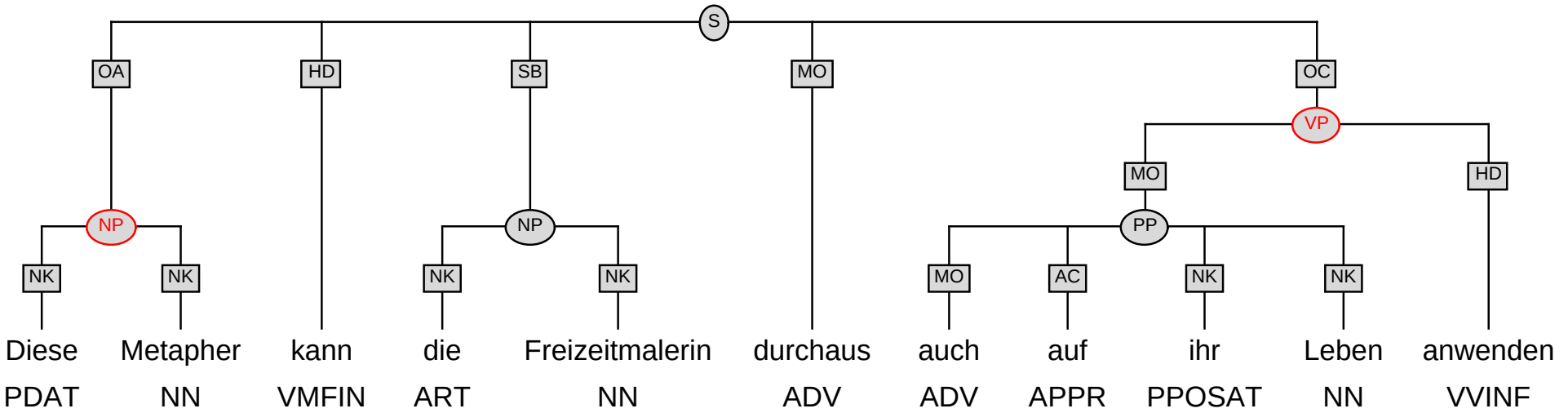
Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Crossing Branches

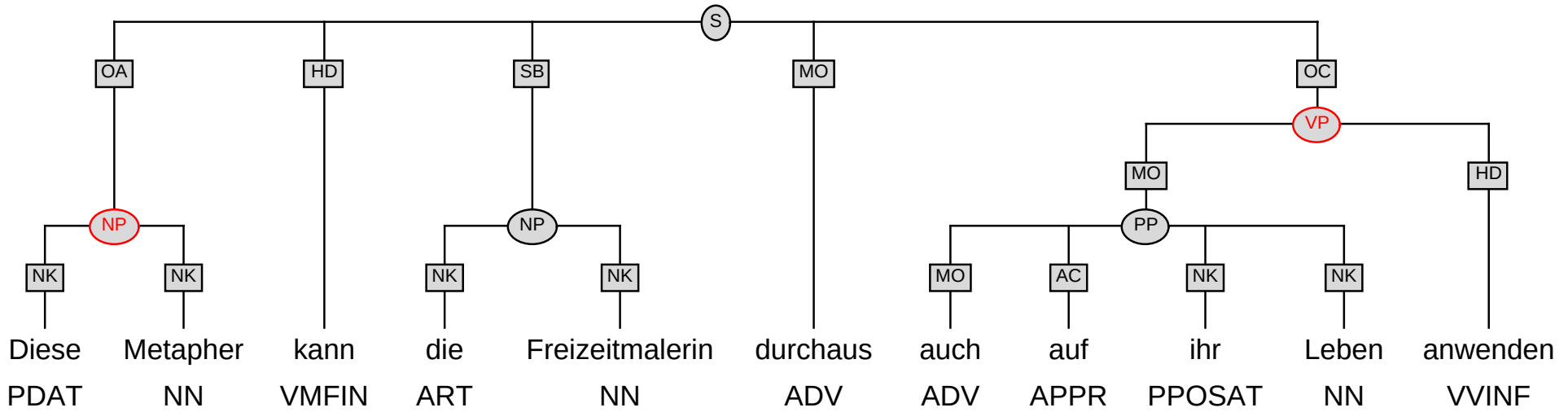


The amateur painter can by all means apply this metaphor also to her life.

Resolved Structure



Resolved Structure



S → NP VMFIN NP ADV VP

Experiment with Gramm. Functions

- Stanford Parser unlexicalized, with markovization
- constituent labels + grammatical functions: NP-SB, NX-ON, etc.
- evaluation of all grammatical functions
- evaluation of single, selected functions: arguments: subject, accusative object, dative object

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Results: Gramm. Functions

		lab. prec.	lab. rec.	lab. F-score
Negra	no GF	69.96	69.95	69.95
	all GF	47.20	56.43	51.41
	subject	52.50	58.02	55.12
	acc. object	35.14	36.30	35.71
	dat. object	8.38	3.58	5.00
TüBa	no GF	89.86	88.51	89.18
	all GF	75.73	74.93	75.33
	subject	66.82	75.93	71.08
	acc. object	43.84	47.31	45.50
	dat. object	24.46	9.96	14.07

Results: Gramm. Functions

		lab. prec.	lab. rec.	lab. F-score
Negra	no GF	69.96	69.95	69.95
	all GF	47.20	56.43	51.41
	subject	52.50	58.02	55.12
	acc. object	35.14	36.30	35.71
	dat. object	8.38	3.58	5.00
TüBa	no GF	89.86	88.51	89.18
	all GF	75.73	74.93	75.33
	subject	66.82	75.93	71.08
	acc. object	43.84	47.31	45.50
	dat. object	24.46	9.96	14.07

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Results: Gramm. Functions

		lab. prec.	lab. rec.	lab. F-score
Negra	no GF	69.96	69.95	69.95
	all GF	47.20	56.43	51.41
	subject	52.50	58.02	55.12
	acc. object	35.14	36.30	35.71
	dat. object	8.38	3.58	5.00
TüBa	no GF	89.86	88.51	89.18
	all GF	75.73	74.93	75.33
	subject	66.82	75.93	71.08
	acc. object	43.84	47.31	45.50
	dat. object	24.46	9.96	14.07

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Computational Anaphora Resolution

- German:

  Die Mutter gab dem Mädchen eine Tomate. Es mochte sie nicht.

- English:

  The mother gave the girl a tomato. It did not like her.

High-quality MT is hard

- German:

  Die Mutter gab dem Mädchen [neuter] eine Tomate [fem]. Es [neuter] mochte sie [fem] nicht.

High-quality MT is hard

- German:

  Die Mutter gab dem Mädchen [neuter] eine Tomate [fem]. Es [neuter] mochte sie [fem] nicht.

- English:

  The mother gave the girl a tomato. She did not like it.

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Markables

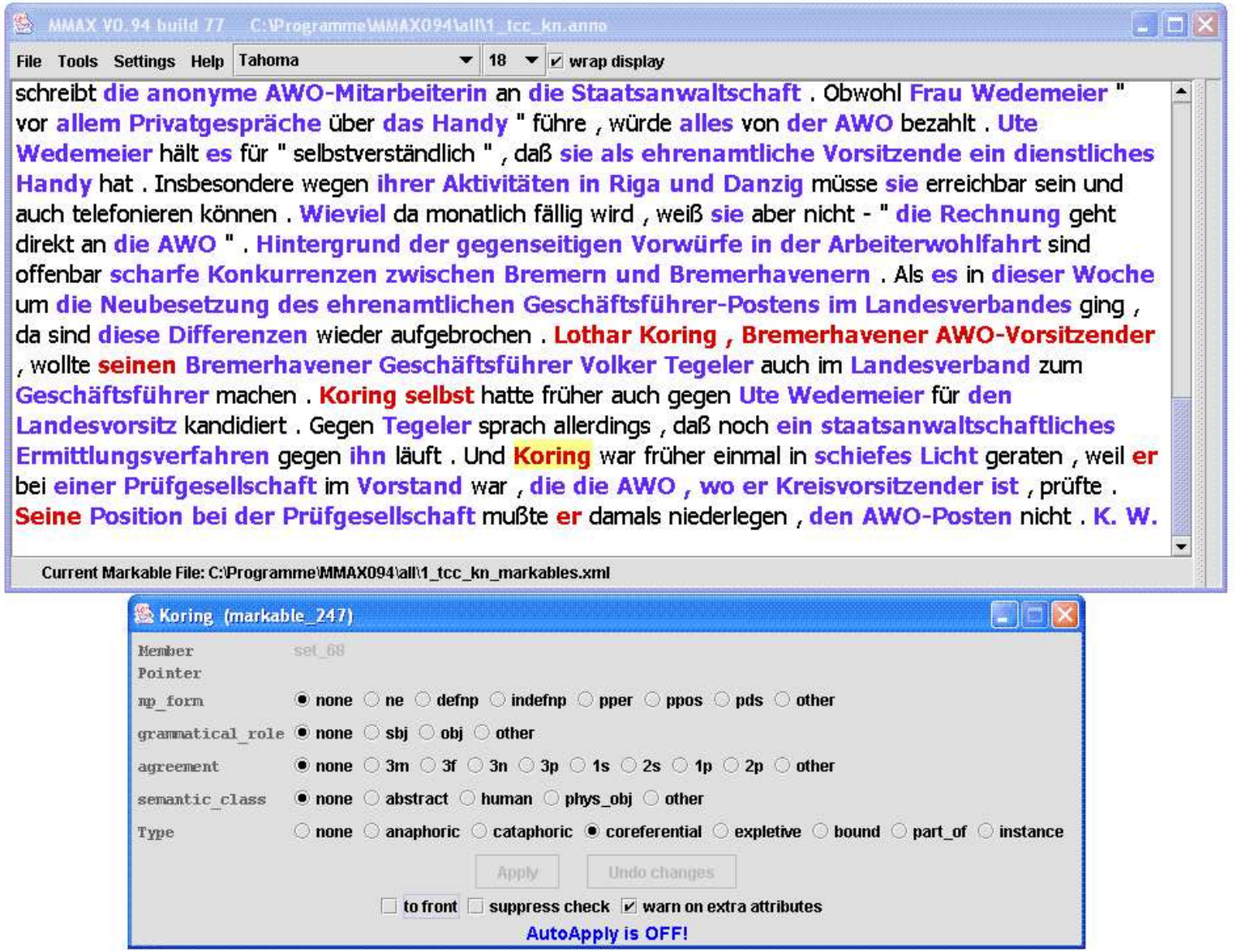
- definite NPs
- personal, possessive, relative, reflexive and reciprocal pronouns
- demonstrative and indefinite pronouns
- automatically extracted from the treebank

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Referential Annotation

- relations based on the MATE inventory
- categories used in annotation:
  - coreferential
  - anaphoric
  - cataphoric
  - bound
  - part-of
  - instance
  - expletive

Annotation in MMAX (EML Heidelberg)



Memory-based Anaphora Resolution

Hybrid architecture:

- Rule-based morphological pre-filter
  - Substantial reduction of candidate search space
  - Implemented using the Xerox XIP Tools
- Memory-based resolution module
  - Uses the Tilburg Memory-Based Learner (TiMBL)
  - Anaphora resolution is encoded as a binary encoding problem

Features

Feature	Comment
Pronoun type	personal possessive reflex-ive
Position	anaphoric cataphoric
Syntactic parallelism	parallel different n/a
Distance in sentences	loc 0/1/2/3
Distance in words	1...n
Features contributing to discourse history	ON OA OD OPP APP FOPP X-MOD PRED KONJ HD -- -

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

Experiments

3 Experiments with different encodings of discourse history during training:

- Experiment 1: Number of times members of the coreference chain occur in each of the 12 syntactic roles
- Experiment 2: Distance of the pronoun to the closest member in the coreference chain realized in each of the 12 syntactic roles
- Experiment 0: Only syntactic features of the closest antecedent
  $\rightarrow$ No information about discourse history (baseline)

Backoff when no antecedent found: Choose closest subject as the antecedent

Results

	Precision	Recall	F-Measure
Exp. 0	78.8%	63.7%	70.4%
Exp. 1	83.8%	66.8%	74.3%
Exp. 2	84.2%	66.4%	74.2%
Exp. 1.subj	**79.1%**	**75.1%**	**77%**
Exp. 2.subj	78.2%	74.1%	76.1%
RAP	*76.6%*	*76.5%*	*76.6%*

- Discourse history matters
- Competitive performance of machine-learning approaches given the usage of informative features

Summary

- Treebanks provide useful information for anaphora resolution
- Rule-based approach and memory-based system achieve roughly equal results on the same data
- Machine-learning approaches can successfully simulate functionality of hand-crafted systems, when given proper features

Future Research

- Incorporating lexical semantics (predicate-argument frequencies, GermaNet)
- Investigating alternative machine learning paradigms
- Using parsing input

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen