



**Universität
Zürich** ^{UZH}

Institut für Computerlinguistik

Verbal Morphosyntactic Disambiguation through Topological Field Recognition in German-Language Law Texts

SFCM 2013

Kyoko Sugisaki and Stefan Höfler

Background and Motivation

Aim:

- To develop a German **style checker** for law texts

Task:

- For the reliable detection of the violations of **syntax-related style rules**, existing POS taggers have to be adopted to the domain.

Current situation:

- Need an annotated corpus for training
- Difficulty in modeling long-distance dependencies

Our approach:

- **Rule-based** approach for correcting POS tags using **morphosyntactic** information

Overview

- Introduction
- Morphosyntactic disambiguation for the recognition of clause boundaries
- Evaluation
- Conclusion

Ambiguity and Clause Boundaries

Verb or Adjective ?

Stellt die Zollverwaltung Unregelmässigkeiten fest, so verweigert sie den Abschluss des Transitverfahrens und hält die Sicherheit zurück, bis die mit bedingter Zahlungspflicht veranlagten Einfuhrzollabgaben bezahlt sind.

Preposition or subordinating conjunction?

'If the customs administration recognises irregularities, it refuses the completion of the transit procedure and retains the security until the import customs fees rated with conditioned duty of payment have been paid.'

Er kann die Anforderungen festlegen an die Aus- und Weiterbildung der Tierhalterinnen und Tierhalter sowie der Personen, die Tiere ausbilden.

definite article or relative pronoun?

'It may define the requirements for the training and continuing education of animal keepers and persons who train animals.'

Ambiguity and Long Distance Dependency

Verb or Adjective ?

Stellt die Zollverwaltung Unregelmässigkeiten fest, so verweigert sie den Abschluss des Transitverfahrens und hält die Sicherheit zurück, bis die mit bedingter Zahlungspflicht veranlagten Einfuhrzollabgaben bezahlt sind.

Preposition or subordinating conjunction?

Er kann die Anforderungen festlegen an die Aus- und Weiterbildung der Tierhalterinnen und Tierhalter sowie der Personen, die Tiere ausbilden.

definite article or relative pronoun?

Clause Boundaries and Topological Fields

Stellt die Zollverwaltung Unregelmässigkeiten fest, so verweigert sie den Abschluss des Transitverfahrens und hält die Sicherheit zurück, bis die mit bedingter Zahlungspflicht veranlagten Einfuhrzollabgaben bezahlt sind.

Er kann die Anforderungen festlegen an die Aus- und Weiterbildung der Tierhalterinnen und Tierhalter sowie der Personen, die Tiere ausbilden.

Left brackets of clauses

Right brackets of clauses

Morphosyntactic dependencies

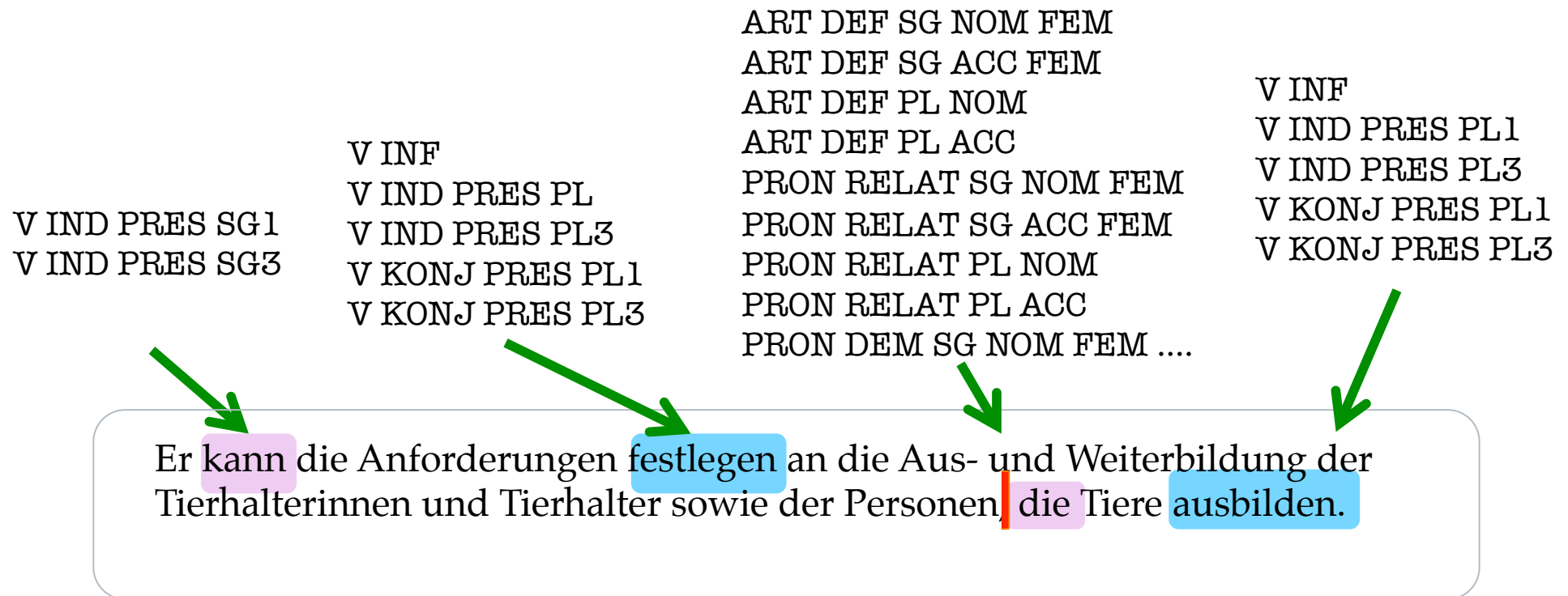
→ Rule-based modeling of morphosyntactic dependencies

→ Constraint Grammar

Recognition of Topological fields

Incremental 2-step morphosyntactic disambiguation

- Preprocessing: Morphological Analyzer (Gertwol)
- **Step1: Left brackets of Clauses**
- **Step2: Right brackets of Clauses**



Recognition of Topological fields

Incremental 2-step morphosyntactic disambiguation

- **Step1: Left brackets of Clauses**
- **Step2: Right brackets of Clauses**

; ART DEF SG NOM FEM
 ; ART DEF SG ACC FEM
 ; ART DEF PL NOM
 ; ART DEF PL ACC
 PRON RELAT SG NOM FEM
 PRON RELAT SG ACC FEM
 PRON RELAT PL NOM
 PRON RELAT PL ACC
 ; PRON DEM SG NOM FEM

; V INF
 V IND PRES PL
 V IND PRES PL3
 V KONJ PRES PL1
 V KONJ PRES PL3

Comma before

Period after

REL.PRON → Finite verb

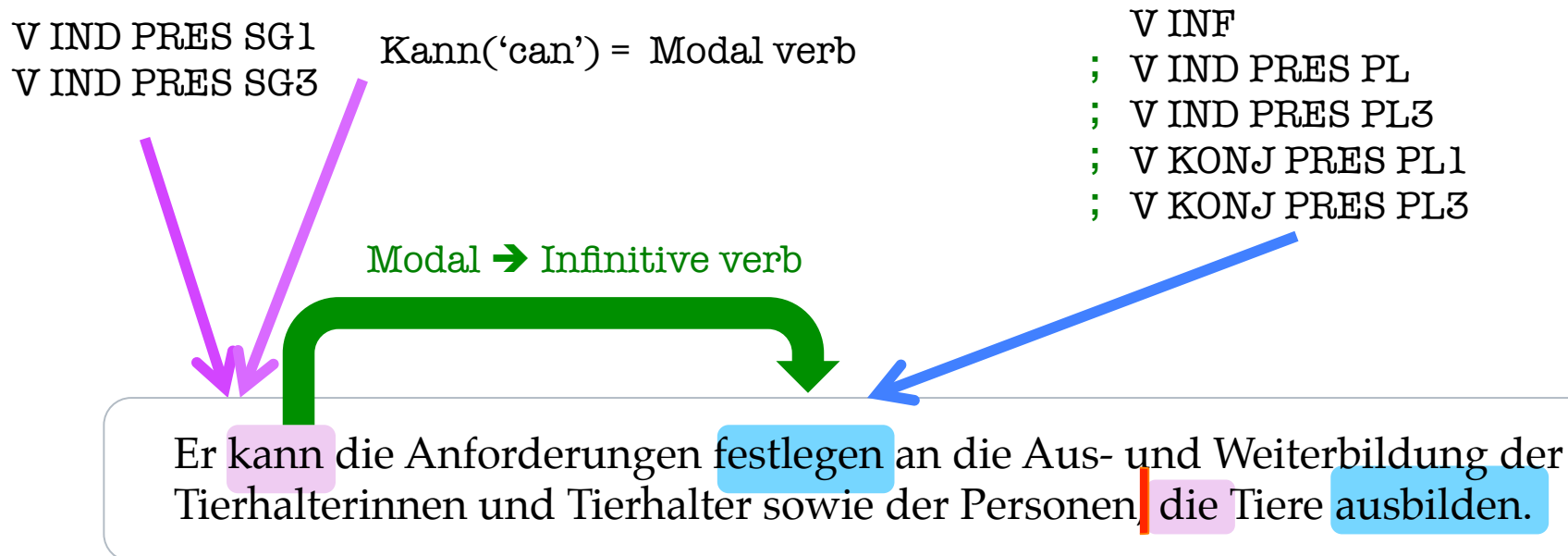
Er kann die Anforderungen festlegen an die Aus- und Weiterbildung der Tierhalterinnen und Tierhalter sowie der Personen, die Tiere ausbilden.

```
SELECT (PRON RELAT) IF (-1 comma) AND (1 * finite_verb LINK 1 punctuation)
ADD (&LB) IF (OC (PRON RELAT))
SELECT (finite_verb) IF (-1 * rc_pron BARRIER Left_right_brackets)
ADD (&RB) IF (OC(INF)) AND (-1C* (PRON RELAT) BARRIER Left_right_brackets )
```


Recognition of Topological fields

Incremental 2-step morphosyntactic disambiguation

- **Step1: Left brackets of Clauses**
- **Step2: Right brackets of Clauses**



```
ADD (&LB) IF (0 finite_verb) AND (0 modal_verb) AND (1 *(INF))
SELECT infinite_verb IF (-1 * modal LINK 0 finite_verb)
ADD(&RB) IF (-1 * (&LB)) (0 infinite_verb)
```

Evaluation: Topological field recognition

- Test data: 100 sentences (2370 tokens) from the Swiss Legislation Corpus.

Recall	Precision	F1
95.33	99.65	97.44

- Error analysis
 - Relative pronouns \Leftrightarrow definite article
 - Mistakes from Gertwol

POS correction by recognizing topological fields

- Test data: 100 sentences(2370 tokens) from the Swiss Legislation Corpus.
- Right and left bracket elements (verbs, subordinating conjunction, etc.)

RECALL	Analysed correctly	Analysed incorrectly	Total: Annotation Outputs
TreeTagger	281 Tokens (89.77%)	32 Tokens (10.22%)	313 Tokens (100%)
Our system	308 Tokens (98.40%)	5 Tokens (1.59%)	313 Tokens (100%)

PRECISION	Analysed correctly	Analysed incorrectly	Total: System Outputs
TreeTagger	281 Tokens (98.25%)	5 Tokens (1.74%)	286 Tokens (100%)
Our system	308 Tokens (99.67%)	1 Tokens (0.32%)	309 Tokens (100%)

- ➔ Reducing POS-tagging mistakes from 10.22 to 1.59%
- ➔ Improvement: the tags of right bracket elements

Discussion & Summary

- The system presented is capable of reducing the rate of POS-tagging mistakes from 10.2% to 1.6%.
- The evaluation shows that this reduction is mostly gained through checking the compatibility of morphosyntactic features within the long-distance syntactic relationships of discontinuous verbal elements.
- The present study shows that in law texts, the average distance between the left and right bracket of clauses is relatively large (9.5 tokens), and that in this domain, a wide context window is therefore necessary for the morphosyntactic disambiguation of verbs.



Thank you

Acknowledgement

We thank

The Swiss National Foundation, Switzerland

Prof. Dr. Michael Hess, Institute of Computational Linguistics, University of Zurich

Prof. Dr. Felix Uhlmann, Institute of Law, University of Zurich

Dr. Rebekka Bratschi, Swiss Federal Chancellery

for their support of our project.

Our project “*Automated Detection of Style Guide Violations in Legislative Drafts*”:

http://www.cl.uzh.ch/research/maschinellestilpruefung/gesetzestextanalyse_en.html