Hybrid Morphosyntactic Disambiguation for Parsing Swiss Law Texts in German

University of Zurich[UZH]

Kyoko Sugisaki

Background

- We develop a tool for **automatically detecting violations of domain-specific style rules** in drafts of **legislative texts written in German**.
- To detect the violations of **syntax-related style rules**, we develop a **domain-specific shallow functional parser** for Swiss law texts.
- One key problem for **parsing German texts** is **morphosyntactic disambiguation**.

Challenges

Morphosyntactic disambiguation is challenging especially for laws texts written in German:

- **German** exhibits a relatively high degree of **morphological syncretism**, has a relatively **free word order** and can contain **discontinuous verbal complexes**.
- **Legislative language** is **syntactically complex** due to the frequency of **embedded clauses**, **coordination structures**, **light verb constructions**, **appositions** and **bare nouns**.

Approach

We propose a **hybrid architecture for morphosyntactic disambiguation**:

- Linguistically motivated **hard constraints** are modeled in a **rule-based** component.
- Data-driven **soft constraints** are modeled in a **statistical** component.

Tasks

Our system performs the following tasks:

1. **Morphological analysis** (Gertwol and TreeTagger)
2. **Morphosyntactic disambiguation of Verbs**: by applying linguistically motivated **hard constraints based on the topological field model**.
3. **Morphosyntactic ambiguity reduction of nouns**: by applying linguistically motivated **hard constraints** (cf. the right side of the poster).
4. **Grammatical function recognition**: by applying **soft constraints**.

Disambiguation by Hard Constraints

In Sugisaki and Höfler (submitted), we explore how much linguistically motivated **hard constraints** can contribute to the **morphosyntactic disambiguation of nouns**. To this aim, we have implemented a three-step procedure:

- **Step 1**: Local phrase-level feature unification (e.g. simple NPs)
- **Step 2**: Upper phrase-level feature unification (e.g. coordinated NPs, participle phrases, PPs)
- **Step 3**: Clause-level feature unification (e.g. subject-verb agreement, argument structures, diathesis, topological fields)

In Table 1, this three-step ambiguity reduction is illustrated for the nouns in the following example sentence:

(1) Sie berücksichtigt dabei den der Tierhalterin oder dem Tierhalter entstehenden Aufwand und das Wohlergehen der Tiere.
('In doing so, it [the agency] takes into account the expenses arising for the animal owners and the welfare of the animals.')

Results

We tested our system on 239 sentences (4,789 tokens) that were randomly selected from the Swiss Legislative Corpus (Höfler and Piotrowski 2011) and found that the rate of morphosyntactically ambiguous nouns could be reduced from 91.12% to 32.31% (cf. Table 2).

	1 analysis per token	more than 1 analysis per token
Input	148 (8.87%)	1,520 (91.12%)
Step 1	387 (23.20%)	1,281 (76.79%)
Step 2	917 (54.97%)	751 (45.02%)
Step 3	1,129 (67.68%)	539 (32.31%)

Table 2: Number of unambiguous and ambiguous analyses after each step

Our results indicate that a major part (67.68%) of the morphosyntactic ambiguity of nouns can be resolved by applying linguistically motivated hard constraints before data-driven **soft constraints** (e.g. **word order**, **animacy**, **definiteness** and **information structures**).

References

Höfler, S., & Piotrowski, M. (2011). Building corpora for the philological study of Swiss legal texts. Journal for Language Technology and Computational Linguistics (JLCL), 26(2):77–89.

Sugisaki, K, & Höfler, S. (submitted). Incremental Morphosyntactic Disambiguation of Nouns in German for Parsing Law Texts.

Example of Three Step Ambiguity Reduction

	Sie	*Tierhalterin*	*Tierhalter*	*Aufwand*	*Wohlgehen*	*Tiere*
Input: Gertwol	NOM AKK	NOM AKK DAT GEN	NOM AKK DAT GEN	NOM AKK DAT	NOM AKK DAT	NOM AKK DAT GEN
Step1: Local phrase-level feature unification	NOM AKK	DAT GEN	DAT	NOM AKK DAT	NOM AKK	DAT GEN
Step2: Upper phrase-level feature unification	NOM AKK	DAT	DAT	AKK	AKK	DAT GEN
Step3: Clause-level feature unification	NOM	DAT	DAT	AKK	AKK	DAT GEN

Table 1: The case features of pronouns and nouns of sentence (1) are incrementally disambiguated by being unified with those in various contexts.

Contact

Address
Institute of Computational Linguistics
Binzmühlestrasse 14, CH-8050 Zürich

This project is funded under
Swiss National Science
Foundation grant no. 134701.

Email
Kyoko Sugisaki
sugisaki@cl.uzh.ch