Summary

We develop a tool for automatically detecting violations of domain-specific controlled language rules in drafts of legislative texts. The project’s most important innovative contribution is the enhancement of the method of error modelling, which was developed for controlled language checking in technical writing, to meet the requirements of legislative editing – a domain largely out of reach for state-of-the-art controlled language checkers. We focus on German-language legislative drafting in Switzerland.

Approach

The key method we apply is that of error modelling:

- **Individual violations** of controlled language rules are anticipated.
- The draft texts are then searched for specific typographical and linguistic features that indicate the presence of these violations.

Tasks

1. **Automatic preprocessing** of the draft text: tokenisation, text segmentation, part-of-speech tagging, morphological analysis, parsing
2. **Automatic error detection** in the preprocessed text: searching for features that indicate a violation of a controlled language rule
3. **Generation of user-feedback**: inserting a comment into the original Word document explaining which controlled language rule the highlighted passage potentially violates (cf. Fig. 1)

Challenges

Controlled language checking is more challenging for laws than for technical documents:

- **Legislative language** is relatively complex and idiosyncratic. The pre-processing tools have to be adapted to the domain.
- **Controlled language rules for legislative texts** are often relatively abstract and highly domain-specific (cf. example to the right). Extensive domain-specific linguistic error modelling is required in order to be able to detect violations of such rules automatically.

System Architecture

![Architecture of the controlled language checker.](image)

Figure 2: Architecture of the controlled language checker.

Example: Output of the Checker

**Example: Only one Proposition per Sentence**

Legislative drafting guidelines contain rules controlling linguistic phenomena both at the sentence level and at the text level (cf. Höfler 2012). One prominent rule states that sentences should not contain more than one proposition. To make such an abstract rule accessible to controlled language checking, domain-specific error modelling is required: In Höfler (2011), we investigate key phrases and syntactic constructions that can serve as linguistic indicators for the detection of sentences that contain more than one proposition. Examples of such constructions are sentence coordination, relative clauses introduced by the adverb *wobei* (‘whereby’), and prepositional phrases beginning with *vorbehältlich* (‘subject to’) or with *mit Ausnahme von* (‘with the exception of’).

References


Contact

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

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

**Email**
Stefan Höfler (project leader)
hoefler@cl.uzh.ch