



**University of
Zurich** ^{UZH}

Centre for Legislative Studies

Assessing the Quality of Legislation by means of an Automatically Annotated Corpus

Transius Symposium on Corpus Analysis in Legal Research and Legal
Translation Studies

Stefan Höfler

Contents

1. Motivation

How can we assess the quality of legislation?

2. Requirements

What corpora are needed for the task?

3. Challenges

How can such corpora be built?

Is the Quality of Legislation Plummeting?

(Headline of an article in the NZZ in 2013)



Well-known complaints

- “There is **too much** legislation!”
- “Too many legislative texts are **too hard to read and understand!**”
- “The **quality of legislation** is plummeting!”

Based on **opinion, anecdote, ...**

Questions

- Are these claims **true**?
- Can they be **verified empirically**?

Aspects of “Good” Legislation

The question about what constitutes “good” legislation has **two aspects**:

1. Substance

“Good” legislation is fair, necessary, effective, efficient, ...

2. Form

“Good” legislation is easily accessible, clear, precise, ...

Question

→ How can we gain empirical data on the **formal quality** of legislation –
i.e. the formal quality of **legislative texts**?

Drafting Guidelines

Some aspects of the formal quality of legislation have been described in **drafting guidelines**, e.g. through rules on:

- text structure
- sentence structure
- word use
- special drafting techniques
(legal definitions, cross references, ...)

} “drafting style”

Questions

- How closely do current legislative texts **adhere** to these rules?
- Do certain areas of legislation **differ** in style from others?
- Has there been a **shift** in style over time?

These are questions best answered with a **corpus!**

Contents

1. Motivation

How can we assess the quality of legislation?

2. Requirements

What corpora are needed for the task?

3. Challenges

How can such corpora be built?

Example: “Weight” of Articles

Claim

“Articles and paragraphs have become **longer** and thus **less accessible.**”

Rule

An article should not contain more than **three paragraphs.**

A paragraph should not contain more than **one sentence.**



Eugen Huber
(1849–1923)
Author of the
Swiss Civil Code

Example: Federal Supreme Court Act

2. Kapitel: Allgemeine Verfahrensbestimmungen

1. Abschnitt: Zuständigkeit

Art. 29 Prüfung

¹ Das Bundesgericht prüft seine Zuständigkeit von Amtes wegen.

² Bestehen Zweifel, ob das Bundesgericht oder eine andere Behörde zuständig ist, so führt das Gericht mit dieser Behörde einen Meinungs austausch.

Art. 30 Unzuständigkeit

¹ Erachtet sich das Bundesgericht als nicht zuständig, so tritt es auf die Sache nicht ein.

² Hat sich in einem Meinungs austausch die Zuständigkeit einer anderen Behörde ergeben oder erscheint die Zuständigkeit einer anderen Bundesbehörde als wahrscheinlich, so überweist das Bundesgericht die Sache der betreffenden Behörde.

Art. 31 Vorfragen

Ist das Bundesgericht in der Hauptsache zuständig, so befindet es auch über die Vorfragen.

Example: Federal Supreme Court Act

Art. 42 Rechtsschriften

¹ Rechtsschriften sind in einer Amtssprache abzufassen und haben die Begehren, deren Begründung mit Angabe der Beweismittel und die Unterschrift zu enthalten.

² In der Begründung ist in gedrängter Form darzulegen, inwiefern der angefochtene Akt Recht verletzt. Ist eine Beschwerde nur unter der Voraussetzung zulässig, dass sich eine Rechtsfrage von grundsätzlicher Bedeutung stellt oder ein besonders bedeutender Fall nach Artikel 84 vorliegt, so ist auszuführen, warum die jeweilige Voraussetzung erfüllt ist.

³ Die Urkunden, auf die sich die Partei als Beweismittel beruft, sind beizulegen, soweit die Partei sie in Händen hat; richtet sich die Rechtsschrift gegen einen Entscheid, so ist auch dieser beizulegen.

⁴ Bei elektronischer Zustellung muss das Dokument, das die Rechtsschrift und die Beilagen enthält, von der Partei oder ihrem Vertreter beziehungsweise ihrer Vertreterin mit einer anerkannten elektronischen Signatur versehen werden. Das Bundesgericht bestimmt in einem Reglement, in welchem Format die elektronische Zustellung erfolgen kann.

⁵ Fehlen die Unterschrift der Partei oder ihrer Vertretung, deren Vollmacht oder die vorgeschriebenen Beilagen oder ist die Vertretung nicht zugelassen, so wird eine angemessene Frist zur Behebung des Mangels angesetzt mit der Androhung, dass die Rechtsschrift sonst unbeachtet bleibt.

⁶ Unleserliche, ungebührliche, unverständliche, übermässig weitschweifige oder nicht in einer Amtssprache verfasste Rechtsschriften können in gleicher Weise zur Änderung zurückgewiesen werden.

⁷ Rechtsschriften, die auf querulatorischer oder rechtsmissbräuchlicher Prozessführung beruhen, sind unzulässig.

Example: Federal Supreme Court Act

2. Kapitel: Allgemeine Verfahrensbestimmungen

1. Abschnitt: Zuständigkeit

Art. 29 Prüfung

¹ Das Bundesgericht prüft seine Zuständigkeit von Amtes wegen.

² Bestehen Zweifel, ob das Bundesgericht oder eine andere Behörde zuständig ist, so führt das Gericht mit dieser Behörde einen Meinungs austausch.

Art. 30 Unzuständigkeit

¹ Erachtet sich das Bundesgericht als nicht zuständig, so tritt es auf die Sache nicht ein.

² Hat sich in einem Meinungs austausch die Zuständigkeit einer anderen Behörde ergeben oder erscheint die Zuständigkeit einer anderen Bundesbehörde als wahrscheinlich, so überweist das Bundesgericht die Sache der betreffenden Behörde.

Art. 31 Vorfragen

Ist das Bundesgericht in der Hauptsache zuständig, so befindet es auch über die Vorfragen.

Art. 42 Rechtsschriften

¹ Rechtsschriften sind in einer Amtssprache abzufassen und haben die Begehren, deren Begründung mit Angabe der Beweismittel und die Unterschrift zu enthalten.

² In der Begründung ist in gedrängter Form darzulegen, inwiefern der angefochtene Akt Recht verletzt. Ist eine Beschwerde nur unter der Voraussetzung zulässig, dass sich eine Rechtsfrage von grundsätzlicher Bedeutung stellt oder ein besonders bedeutender Fall nach Artikel 84 vorliegt, so ist auszuführen, warum die jeweilige Voraussetzung erfüllt ist.

³ Die Urkunden, auf die sich die Partei als Beweismittel beruft, sind beizulegen, soweit die Partei sie in Händen hat; richtet sich die Rechtsschrift gegen einen Entscheid, so ist auch dieser beizulegen.

⁴ Bei elektronischer Zustellung muss das Dokument, das die Rechtsschrift und die Beilagen enthält, von der Partei oder ihrem Vertreter beziehungsweise ihrer Vertreterin mit einer anerkannten elektronischen Signatur versehen werden. Das Bundesgericht bestimmt in einem Reglement, in welchem Format die elektronische Zustellung erfolgen kann.

⁵ Fehlen die Unterschrift der Partei oder ihrer Vertretung, deren Vollmacht oder die vorgeschriebenen Beilagen oder ist die Vertretung nicht zugelassen, so wird eine angemessene Frist zur Behebung des Mangels angesetzt mit der Androhung, dass die Rechtsschrift sonst unbeachtet bleibt.

⁶ Unleserliche, ungebührliche, unverständliche, übermässig weitschweifige oder nicht in einer Amtssprache verfasste Rechtsschriften können in gleicher Weise zur Änderung zurückgewiesen werden.

⁷ Rechtsschriften, die auf querulatorischer oder rechtsmissbräuchlicher Prozessführung beruhen, sind unzulässig.

Example: “Weight” of Articles

Claim

“Articles and paragraphs have become **longer** and thus **less accessible**.”

Rule

An article should not contain more than **three paragraphs**.

A paragraph should not contain more than **one sentence**.

Requirement

Text segmentation:

articles, paragraphs, sentences

Verification

Count the number of paragraphs per article and sentences per paragraph.



Eugen Huber
(1849–1923)
Author of the
Swiss Civil Code

Example: Legal Definitions

Claim

“The number of **legal definitions** in Swiss legislative texts has increased under the influence of EU legislation.”

Example

Clearing is deemed to be the permanent or temporary misuse of forest soil.

Requirement

Content type recognition:

Legal definitions

Verification

Count the number of legal definitions per text.

Example: Cross References

Claim

“The **use of text-internal cross references** has increased.”

Rule

Avoid cross references, as they impair **readability**.

Requirement

Content type recognition:

Cross references

Verification

Count the number of internal cross references per text.



Eugen Huber
(1849–1923)
Author of the
Swiss Civil Code

Example: Syntactic Complexity

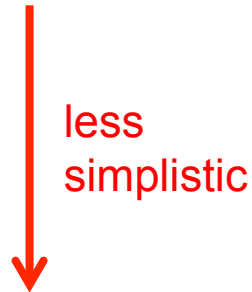
Claim

“The **sentences** in legislative texts are becoming ever more **complex**.”

Problem

How can we “**measure**” syntactic complexity?

- sentence length
- noun/verb ratio
- number of embedded clauses
- ...



Requirement

Linguistic analysis:

Part-of-speech tagging, (chunk) parsing, ...

Contents

1. Motivation

How can we assess the quality of legislation?

2. Requirements

What corpora are needed for the task?

3. Challenges

How can such corpora be built?


The Swiss Legislation Corpus (SLC)



Texts

The SLC contains all Swiss federal legislation currently in force:

- the **federal constitution** and all **cantonal constitutions**
- all **federal acts** and all **ordinances** of federal authorities.

~ 1900 texts,
~ 165,000 sentences
~ 4.7 M tokens, 
~ 123,000 types

Languages

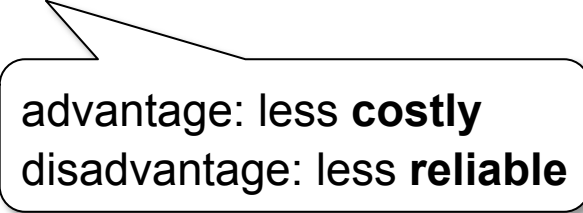
- Each text is available in all three official languages:
German, French, Italian.
- Some of the texts are also available in **Romansh** and in **English**.

Diachronicity

- Texts date from 1875 to 2016 (*inter-textual diachronicity*).
- Texts have been modified for up to 122 years (*intra-textual diachronicity*).

Construction

The texts in the SLC have been **annotated (by machine)** with various kinds of **structural and linguistic information**:



advantage: less **costly**
disadvantage: less **reliable**

Pipeline

1. Tokenisation

Download from the web and conversion into XML

2. Text segmentation

Identifying sections, articles, paragraphs, sentences, enumeration items, ...

3. Linguistic analysis

Lemmatisation, part-of-speech tagging, morphological analysis, syntactic analysis (chunk parsing)

4. Content type recognition

Identifying legal definitions, cross references, transitional provisions, ...

Challenges

1. Text segmentation

- **domain-specific** text segments: articles, paragraphs, enumerations, ...
- **variety** of formatting standards
- **erroneous** formatting

→ *domain-specific analysis (line-based pattern-matching with look-around)*

2. Part-of-speech tagging

- **unknown words**: archaisms, technical terms, ...
- **unknown abbreviations**: government units, other laws, ...
- **unknown ordinal numbers**: *paragraph 2^{bis}, articles 4a–6, ...*

→ *off-the-shelf tools augmented with domain-specific items*

Challenges

3. Syntactic analysis

- **constructions not frequent in other domains:**
co-ordinations of prepositional phrases, subject clauses, ...
- **domain-specific sentence structures:**
enumeration sentences (with parenthetical sentences)

→ *domain-specific pre-processing, machine learning + heuristics*

4. Content type recognition

- **content types are mostly semantic concepts:**
legal definitions, transitional provisions, ...

→ *domain-specific heuristics based on text patterns, position, keywords, ...*

Conclusions

Corpora of legislative texts can be used to **study legislative drafting style**, but there are challenges:

1. Rule concretisation

Abstract drafting rules need to be broken down into **concrete queries**.

2. Corpus annotation

Corpora need to be enriched with **structural and linguistic annotations**.

3. Tool adaptation

NLP tools need to be **adapted to the domain** of legislative texts.

References

- Höfler, Stefan; Sugisaki, Kyoko (2014). Constructing and exploiting an automatically annotated resource of legislative texts. In: Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, 26 May 2014 – 31 May 2014, 175–180.
- Uhlmann, Felix (2014). Qualität der Gesetzgebung – Wünsche an die Empirie. In: Griffel, Alain (ed.), Vom Wert einer guten Gesetzgebung. Bern: Stämpfli, 171–181.
- Sugisaki, Kyoko; Höfler, Stefan (2013). Verbal morphosyntactic disambiguation through topological field recognition in German-language law texts. In: Mahlow, Cerstin; Piotrowski, Michael (eds.), Systems and Frameworks for Computational Morphology. Berlin Heidelberg: Springer, 136–147.
- Sugisaki, Kyoko; Höfler, Stefan (2013). Incremental morphosyntactic disambiguation of nouns in German-language law texts. In: ESSLLI-13 Workshop on Extrinsic Parse Improvement (EPI), Düsseldorf, 12 August 2013 – 16 August 2013.
- Bratschi, Rebekka (2012). Unterscheiden sich Behördentexte von Alltagstexten? Eine quantitative Untersuchung. Diplomarbeit, Institut für mathematische Statistik und Versicherungslehre der Universität Bern, Bern.
- Höfler, Stefan; Sugisaki, Kyoko (2012). From drafting guideline to error detection: Automating style checking for legislative texts. In: EACL 2012 Workshop on Computational Linguistics and Writing, Avignon, France, 23 April 2012 – 23 April 2012, 9–18.
- Höfler, Stefan; Piotrowski, Michael (2011). Building corpora for the philological study of Swiss legal texts. *Journal for Language Technology and Computational Linguistics*, 26(2):7–89.