

Automatic English Text Simplification for Statutes

Akshaya Muralidharan, 27.09.2021, Master Thesis final presentation

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de

Motivation

Research questions

Corpus creation

Characteristics of accessible legal text

State-of-the-art simplification techniques

Takeaways



Low written language comprehension (76% of the non-native-English speakers, 37% of the native-English-speakers) [1]



Inability to comprehend legal language



Simplified text improves readability of laws [2]



Makes legal information more accessible



Makes downstream tasks such as parsing and summarization more accessible

Motivation

Research questions

- Characteristics of simple legal text
- Text simplification methods

Corpus creation

Characteristics of accessible legal text

State-of-the-art simplification techniques

Takeaways



What are the features of simple legal text?

How can we find and automatically align easy legal text with laws in their original form to create a parallel aligned corpus?

How can we use this corpus to understand the features of easy and hard to read legal text?



How do the state-of-the-art simplification models perform on legal data?



Corpus creation

Existing text simplification corpora
Possible sources of simple legal text
Text alignment techniques



Readability metrics

Grade-based metrics
Direct measurements
Feature based metrics
Count based statistics



State of the art text simplification techniques

Motivation

Research questions

Corpus creation

- Sources of simple and standard legal text
- Text alignment methods
- Collected data

Characteristics of accessible legal text

State-of-the-art simplification techniques

Takeaways



Find

Find easily readable legal text and standard legal counterpart



Scrape and store

Scrape and store text in appropriate data structure



Align

Align easily readable text with standard legal text



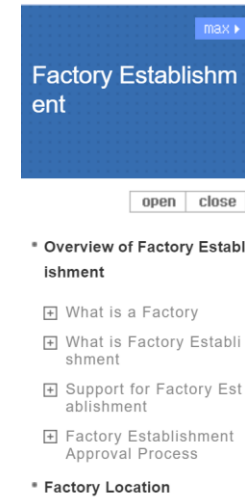
Evaluate

Manually evaluate corpus by examining random samples

Sources of simple and standard legal text

EasyLaw (South Korea)

- Government regulated
- “Simple and Easy Law Info” service
- Laws translated into English and made simpler
- Laws divided into Topics and Subtopics



The screenshot shows the 'Factory Establishment' page on the EasyLaw website. It features a blue header with the title and a 'max' button. Below the header is a navigation menu with links: 'Overview of Factory Establishment', 'What is a Factory', 'What is Factory Establishment', 'Support for Factory Establishment', 'Factory Establishment Approval Process', and 'Factory Location'. The 'Overview of Factory Establishment' link is selected.

Business Type and Scale of Factory

This information is drafted as of August 15, 2020.

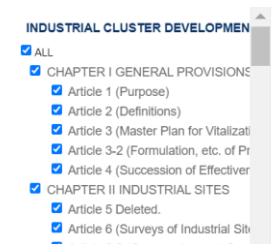
SAVE PRINT BOOKMARK SEND EMAIL

Scope of Regulation in Type and Scale of Factory

- The business type of the factory subject to approval is limited to the manufacturing business (Article 2.-1 of the [Industrial Cluster Development and Factory Establishment Act] and Article 2.(1) of the enforcement decree of the [Industrial Cluster Development and Factory Establishment Act]).
 - ※ The term "manufacturing business" refers to the industrial activity that transforms raw materials to new products carrying different properties through physical and chemical reactions. It follows the classification of manufacturing business(category:C) under the [Korea Standard Industrial Classification] (the notification of Statistics Korea).
- The factory establishment shall comply with standard factory construction area of factory building and others under the [Industrial Cluster Development and Factory Establishment Act] . The scope of regulation of factory establishment differs depending upon the factory construction area and the total floor area of the factory building.

E-Law South Korea

- Government regulated
- Laws translated into English
- Acts in original form



The screenshot shows the 'INDUSTRIAL CLUSTER DEVELOPMENT AND FACTORY ESTABLISHMENT ACT' page on the E-Law South Korea website. It features a sidebar with a tree view of the act's contents, including 'CHAPTER I GENERAL PROVISIONS' and 'CHAPTER II INDUSTRIAL SITES'. The 'ALL' checkbox is selected.

INDUSTRIAL CLUSTER DEVELOPMENT AND FACTORY ESTABLISHMENT ACT

Gov. Body : Trade, Industry and Energy

Search by Pages

Article 2 (Definitions)

The definitions of terms used in this Act shall be as follows: <Amended by Act No. 10252, Apr. 12, 2010>

1. The term "factory" means a place of business prescribed by Presidential Decree for conducting manufacturing business prescribed by Presidential Decree, equipped with buildings or structures, manufacturing facilities such as machinery or equipment which constitutes manufacturing processes, and incidental facilities (hereinafter referred to as "manufacturing facilities, etc.");

```
{
  "Meaning of International Marriage": [
    {
      "What is an International Marriage?": [
        {
          "Definition of International Marriage": [
            "- An international marriage generally refers to \"a marriage between a man and a wo"
          ]
        },
        {
          "Civil Marriage": [
            "- As the Republic of Korea acknowledges civil marriages, the rights and obligations"
          ]
        }
      ]
    }
  ]
}
```

Source: easylaw.go.kr

Scraped using BeautifulSoup

Translated styles into a nested json

Data contains 717 articles

```
{
  "TRIAL OF SMALL CLAIMS ACT": [
    "Act No. 7427, Mar. 31, 2005",
    "Act No. 2547, Feb. 24, 1973\nAmended by Act No. 2821, Dec. 31, 1975\nAct No. 3246, Jan. 4, 1980\nAct No. 4205, Jan. 1
2001\nAct No. 6630, Jan. 26, 2002\nAct No. 7427, Mar. 31, 2005",
    {
      "Article 1 (Purpose)": [
        "The purpose of this Act is to prescribe special cases of the Civil Procedure Act for the purpose of dealing with s
courts and their branch courts."
      ]
    },
    {
      "Article 2 (Scope of Application, etc.)": [
        "(1) This Act shall apply to civil cases provided by the Supreme Court Regulations (hereinafter referred to as the
courts and their branch courts. <Amended by Act No. 2821, Dec. 31, 1975; Act No. 3246, Jan. 4, 1980>",
        "(2) The cases referred to in paragraph (1), except as otherwise prescribed in this Act, shall be applied by the pr
      ]
    },
    {
      "Article 3 (Final Appeal and Reappeal)": [
        "A final appeal or reappeal against a judgment to the Supreme Court, the decision or order on small claims rendered
instance, shall be made only in a case falling under any of the following subparagraphs:",
        "1. Where the decision on the constitutionality of Acts, orders, regulations or dispositions and the decision on th
unfair; and",
        "2. Where the decision is in contravention of the judicial precedents of the Supreme Court."
      ]
    },
    {
      "Article 4 (Institution of Action by Oral Statement)": [
        "(1) An action may be instituted by oral statements.",
        "(2) If any person institutes an action by oral statements, it shall be made in the presence of an administrative o
clerk of a court (hereinafter referred to as the \"junior administrative officer, etc. of a court\"). <Amended by Act No. 6
        "(3) In the case of paragraph (2), the junior administrative officer, etc. of a court shall draw up a protocol for
<Amended by Act No. 6410, Jan. 29, 2001>"
      ]
    },
    {
      "Article 5 (Institution of Action by Voluntary Appearance)": [
```

Source: [Korean Law Translation Center - Statutes of the Republic of ...](https://elaw.klri.re.kr)

<https://elaw.klri.re.kr>

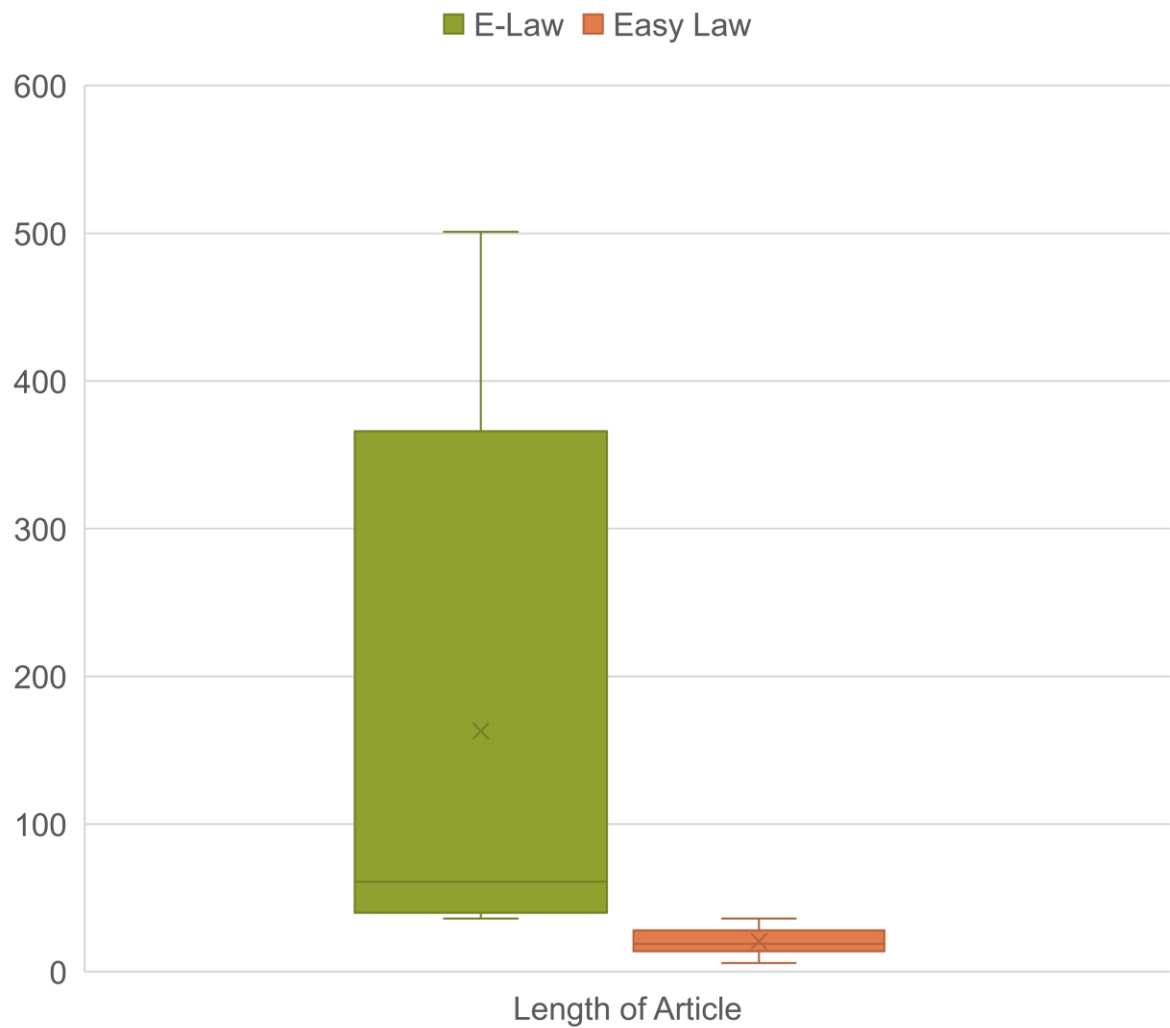
Scraped using BeautifulSoup

Translated styles into a nested json

Data contains 2183 acts

Some preliminary statistics

Article size

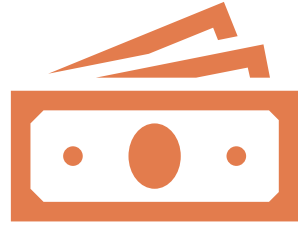


Number of words per sentence



Text Alignment Algorithm

Clean	Clean the text by removing special characters
Extract	Extract references to Act, Article, Section, Subsection from EasyLaw text using regular expressions
Build	Build a search index of Acts to be looked up
Compute	Compute sentence embeddings for all Act names
Find Min	For each Act, find embedding with min cosine distance
Match Act	If distance < threshold, match
Match Article	If match found, match article, section, subsection with # lookup



EasyLaw

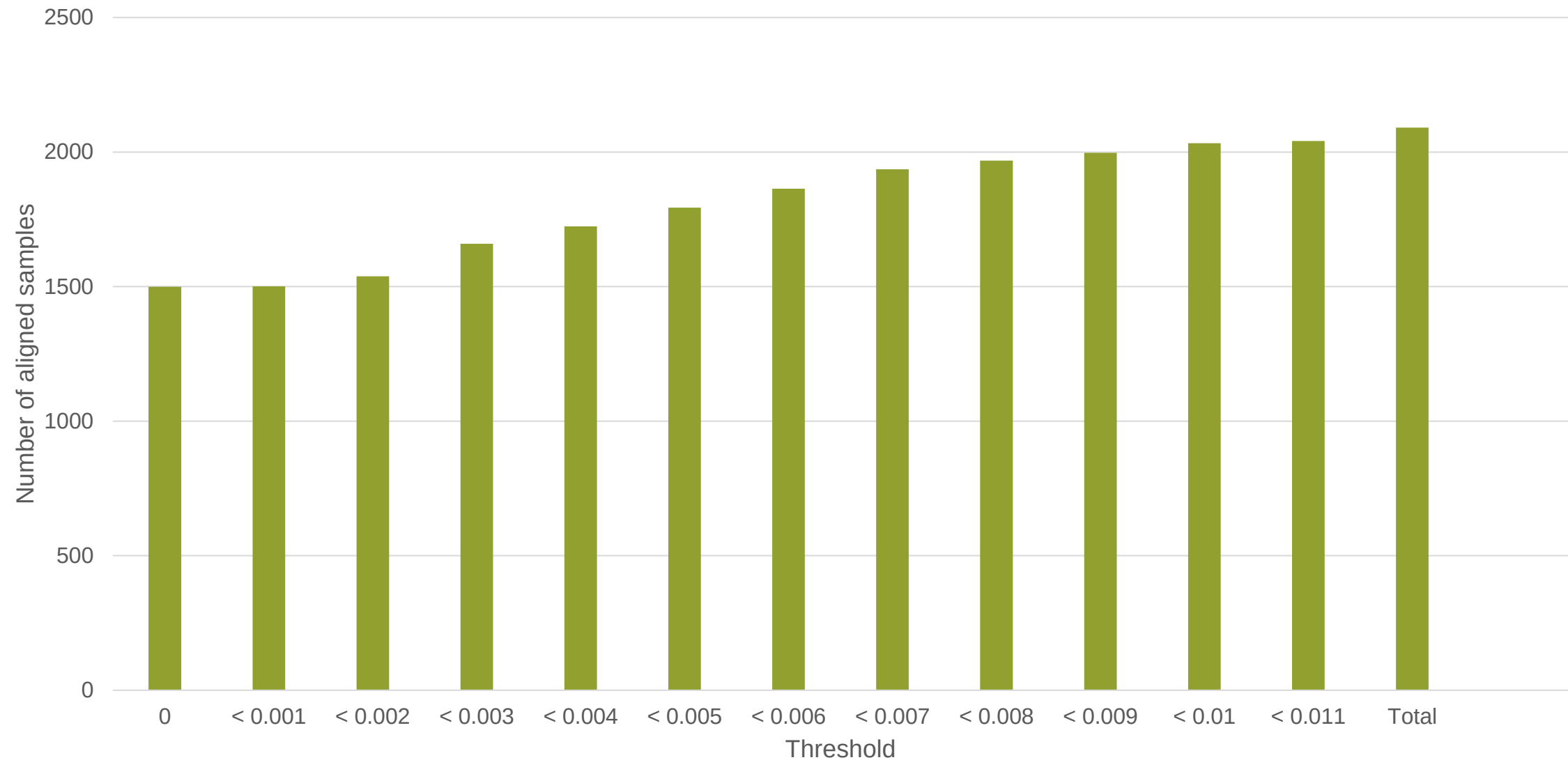
Remuneration (construction payment) must be made at a time that is mutually agreed upon, but where no time has been specified, custom shall prevail. In the absence of any established custom in regards to remuneration period, it has to be paid after the completion of the work that was agreed. For clear progress, remuneration period should be entered in the contract.



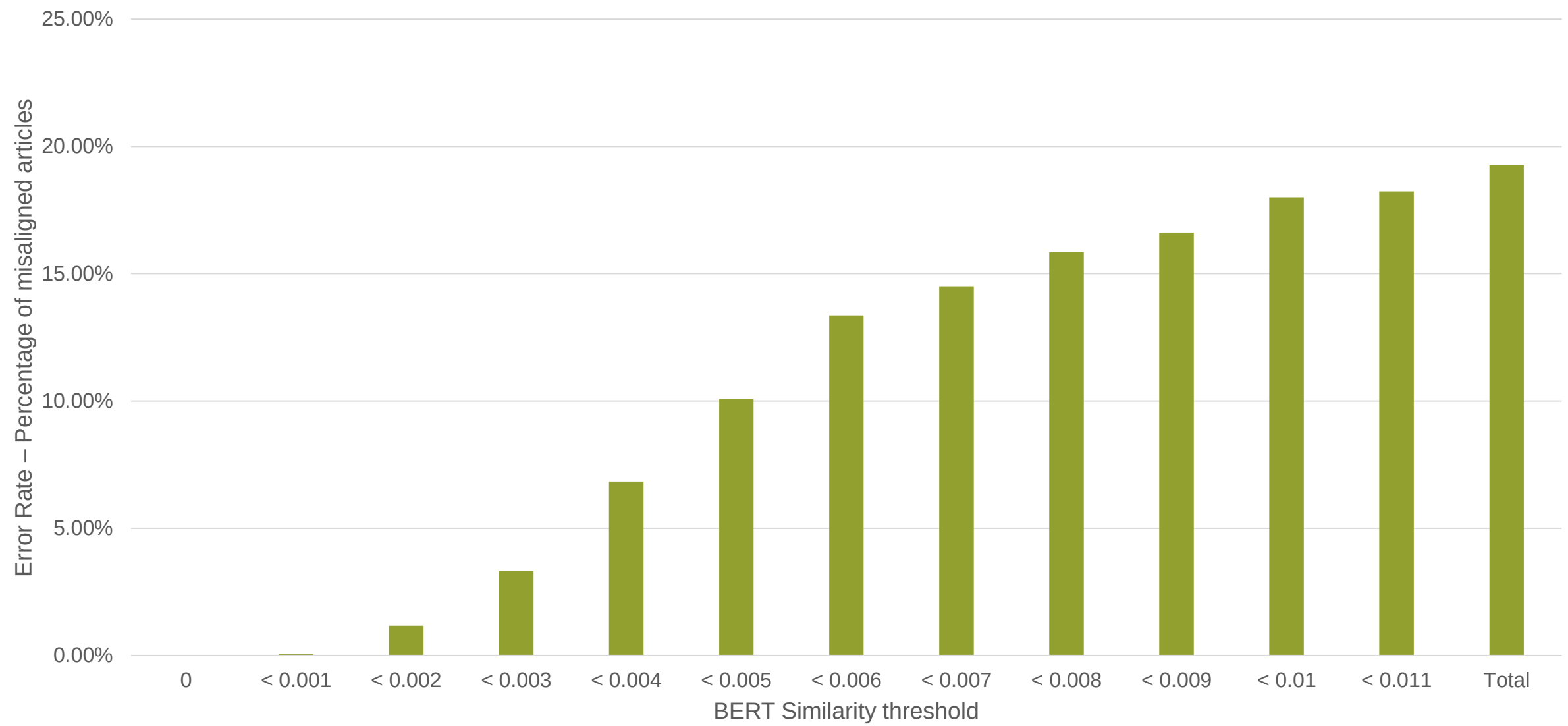
E-Law

- (1) The remuneration shall be made simultaneously with the delivery of the finished object of the work: Provided, That if the delivery of the object of the work is not required, remuneration shall be paid without delay after the work has been performed.
- (2) The provisions of Article 656 (2) shall apply mutatis mutandis to the remuneration mentioned in paragraph (1).

Number of samples vs BERT similarity threshold



Number of misaligned articles vs similarity threshold



Motivation

Research questions

Corpus creation

Characteristics of accessible legal text

- Grade based readability metrics
- Structural metrics
- Grammatical characteristics

State-of-the-art simplification techniques

Takeaways

Derive characteristics of easy text from corpus



Count statistics



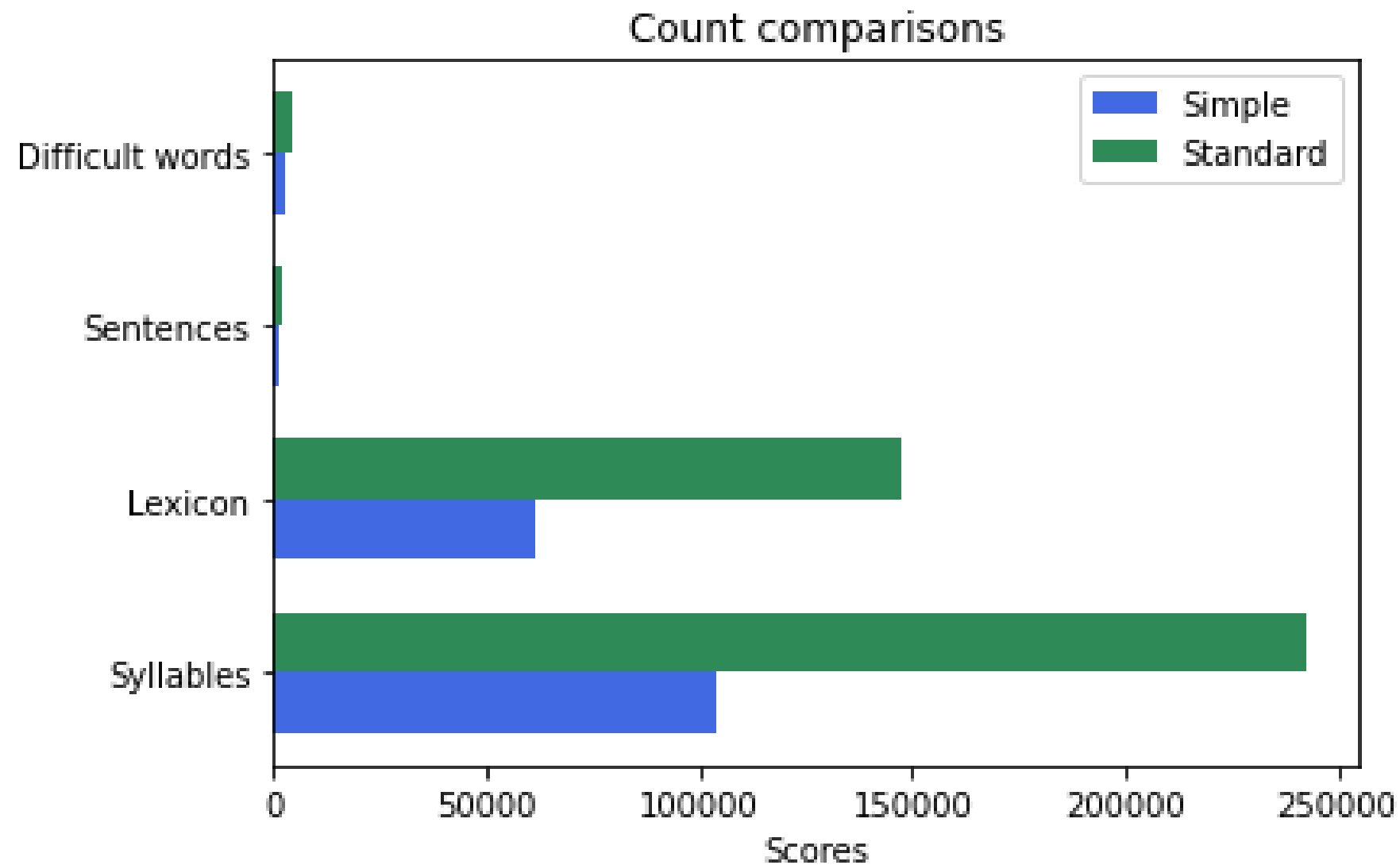
Grade based
readability scores



Structuring
statistics



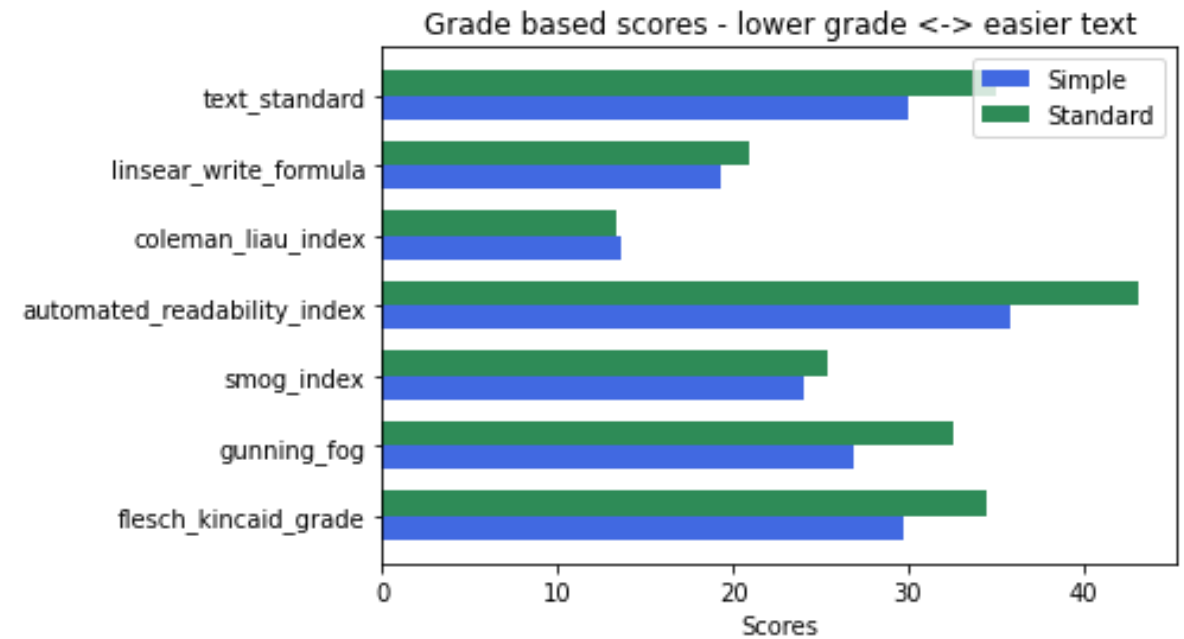
Grammar
frequencies



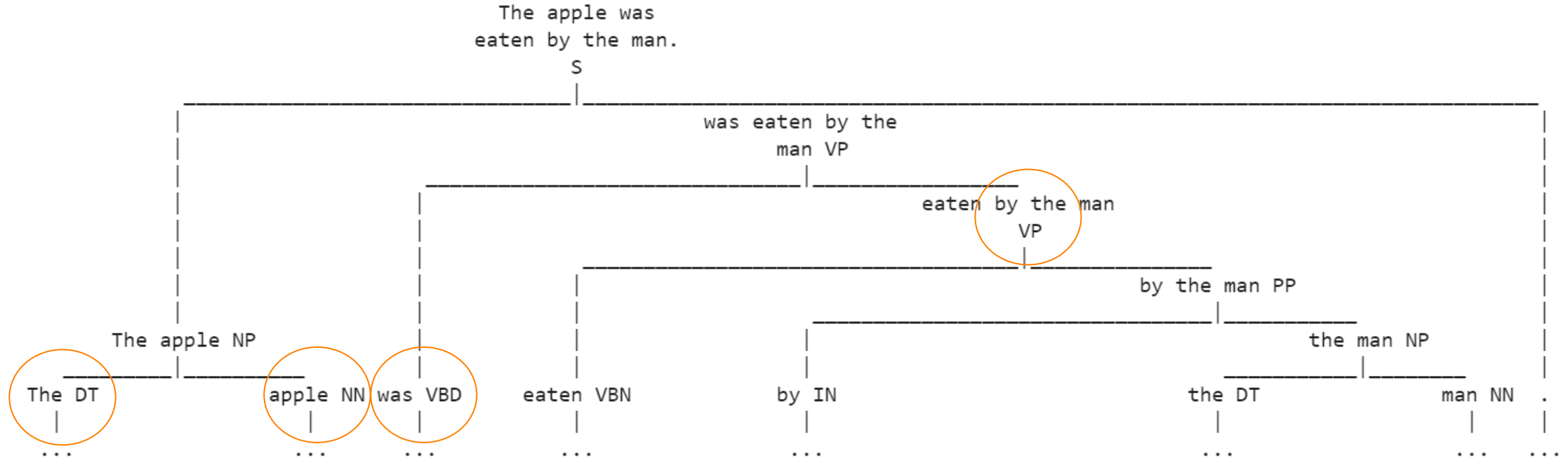
Grade based readability statistics

Readability formulae depend on

- Average sentence length
- average word length in syllables
- number of one-syllable words per 100 words
- average sentence length in words
- words with more than two syllables
- words with more than three syllables



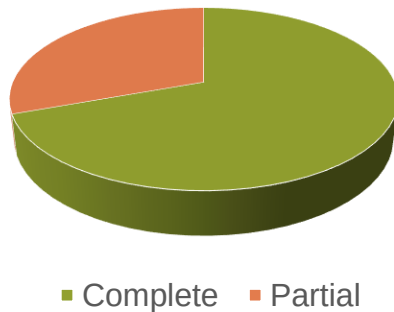
- Compare grammatical structures
- Study [7] indicates that 3rd level parse tree good indicator of sentence structure



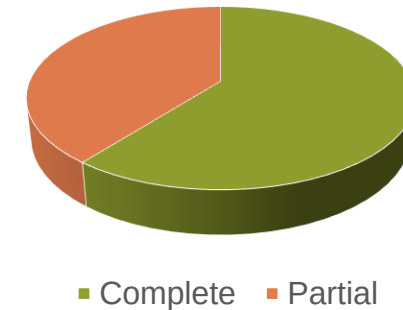
Grammar stats - Frequency

	Elaw Rank	EasyLaw Rank	Example
S [NP [NP PP] VP [MD VP]]	1	1	An individual under influence may not drive.
NP [NP [DT NN] SBAR [WHNP S]]	2	-	An individual who is under influence
S [NP [NP SBAR] VP [MD VP]]		2	An individual who is under influence may not drive.

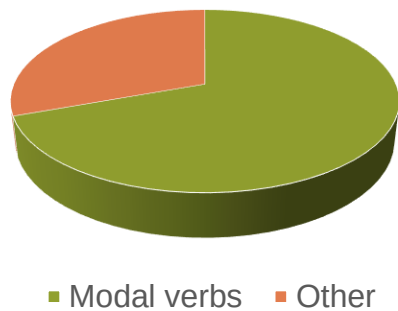
EasyLaw Sentences



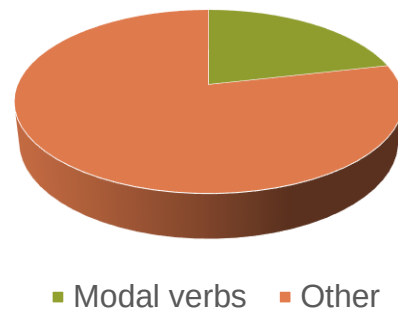
E-Law Sentences



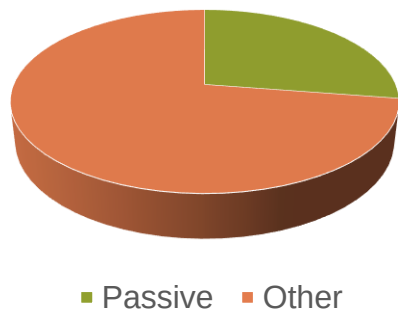
EasyLaw Sentences



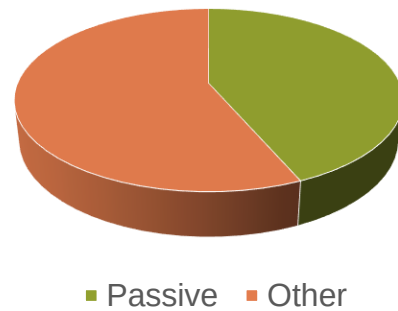
E-Law Sentences



EasyLaw Sentences



E-Law Sentences



	References	Total lines	Percentance
E-Law	96539	800k	12%
EasyLaw	339	15k	2%

The driver of any motor vehicle or tram or the rider of any horse who has violated Article 5, 13 (1) through (3) (in cases falling under Article 13 (3), excluding a person who drives a motor vehicle intentionally violating it on an expressway, an exclusive road for motor vehicles, or a road with median strip), or (5), 14 (2), (3), or (5), 15 (3) (including where the provisions are applied mutatis mutandis in Article 61 (2)), 15-2 (3), 16 (2), 17 (3), 18, 19 (1), (3), or (4), 21 (1), (3), or (4), 24, 25 through 28, 32, 33, 34-3, 37 (excluding paragraph (1) 2), 38 (1), 39 (1), (3), (4), and (5), 48 (1), 49 (excluding a person who drives a motor vehicle or tram in violation of paragraph (1) 1 or 3 of the same Article and a person who drives a motor vehicle with a device installed on which to interfere with functions of traffic regulation equipment, of the offenses referred to in subparagraph 4 of the same paragraph), 50 (5) through (7), 51, 53 (1) or (2) (excluding a driver who fails to make children or infants fasten their seat belts), 62, or 73 (2) (excluding subparagraph 1 of the same paragraph)

Motivation

Research questions

Corpus creation

Characteristics of accessible legal text

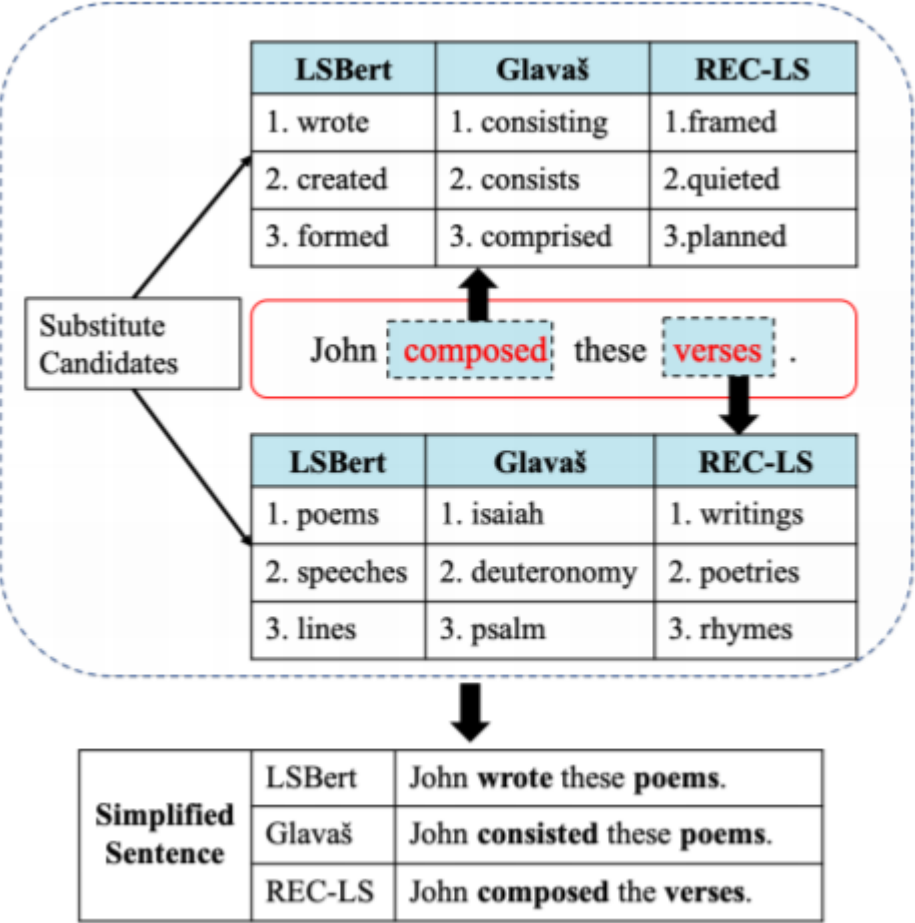
State-of-the-art simplification techniques

- LS-BERT
- MUSS + ACCESS
- Comparison with corpus

Takeaways

Evaluate text simplification models using corpus

Lexical simplification model LS-BERT [5]



Sentence simplification model MUSS + ACCESS [6]

Source	<NbChars_0.3> <LevSim_0.4> <i>He settled in London , devoting himself chiefly to practical teaching .</i>
Target	<i>He teaches in London .</i>

E-Law A minor shall obtain the consent of both parents in order to marry. If one parent is unable to exercise the right of consent, the minor shall obtain the consent of the other parent, and if neither parent is able to exercise the right of consent, the minor shall obtain the consent of his or her guardian of minor.

EasyLaw In case a minor or an adult ward marries, he/she shall go through the following procedures for consent. A minor (under the age of 19) shall obtain the consent of both parents in order to marry. If one parent is unable to exercise the right of consent, the minor shall obtain the consent of the other parent, and if neither parent is able to exercise the right of consent, the minor shall obtain the consent of his/her guardian of minor .

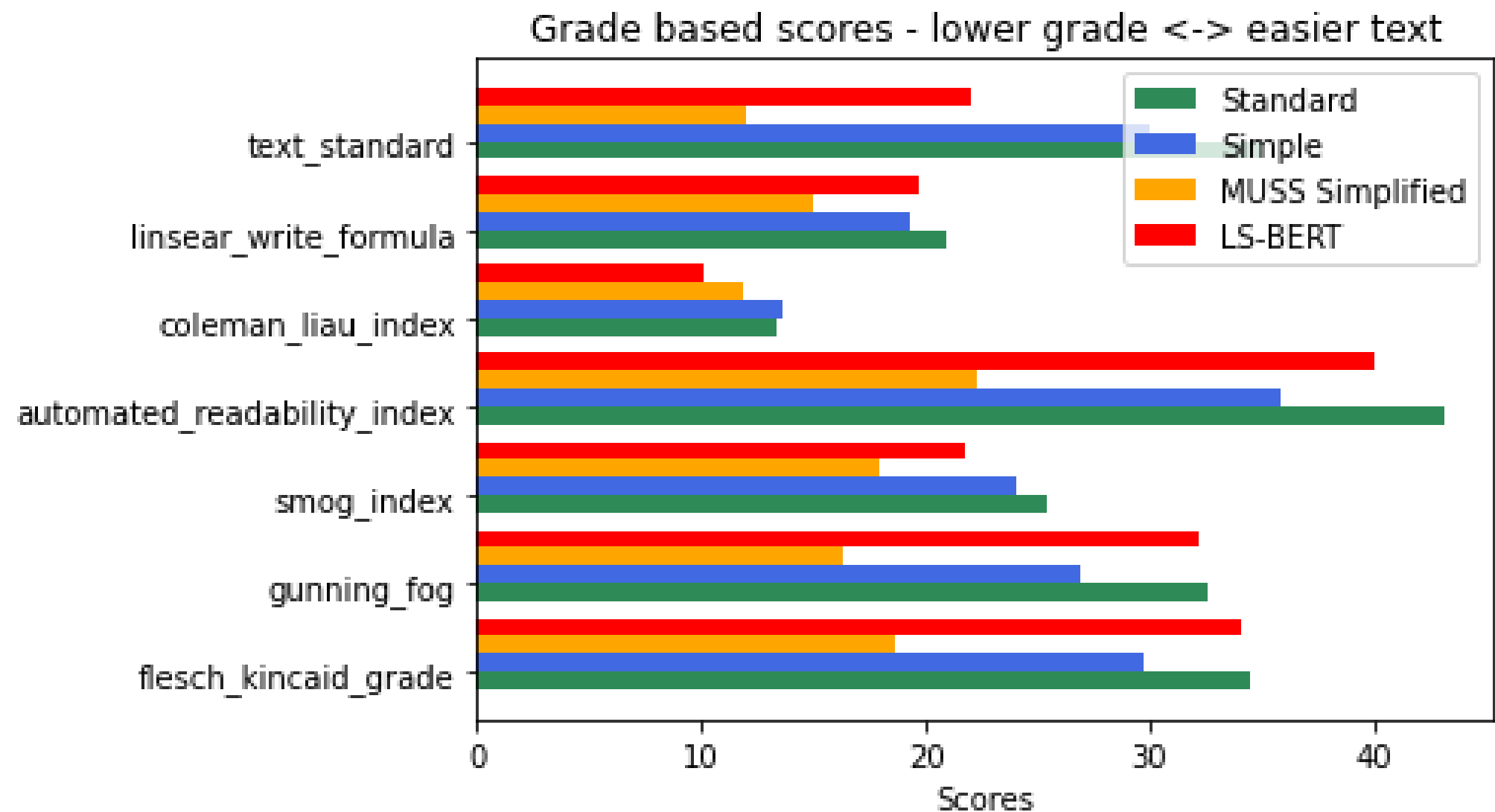
MUSS + ACCESS A minor **must get** the consent of both parents in order to marry. If one parent is **not able to give consent**, the minor **must get** the consent of the other parent. If neither parent is able to **give consent**, the minor **must get** the consent of the minor's guardian.

LSBERT A **child** will **receive** the **permission** of both parents in order to **meet** . If one parent is unable to **use** the right of permission , the child must **get** the permission of the other parent , and if neither parent is able to use the right of permission , the child will get the permission of his or her **father of major** .

Neither model can simplify

Any of the following persons shall be punished by imprisonment with labor for not more than one year or by a fine not exceeding ten million won:

- 1. A person who enters the Republic of Korea without undergoing an entry inspection, in violation of Article 6 (1)*
- 2. A person who violates any of the conditions imposed on conditional entry permission under Article 13 (2)*
- 3. A person who lands without obtaining emergency landing permission under Article 15 (1), landing permission in distress under Article 16 (1), or temporary landing permission for refugees under Article 16-2 (1)*
- 4. A person who violates any of the terms or conditions of permission under Article 15 (2), 16 (2) or 16-2 (2)*



Motivation

Research questions

Corpus creation

Characteristics of accessible legal text

State-of-the-art simplification techniques

Takeaways



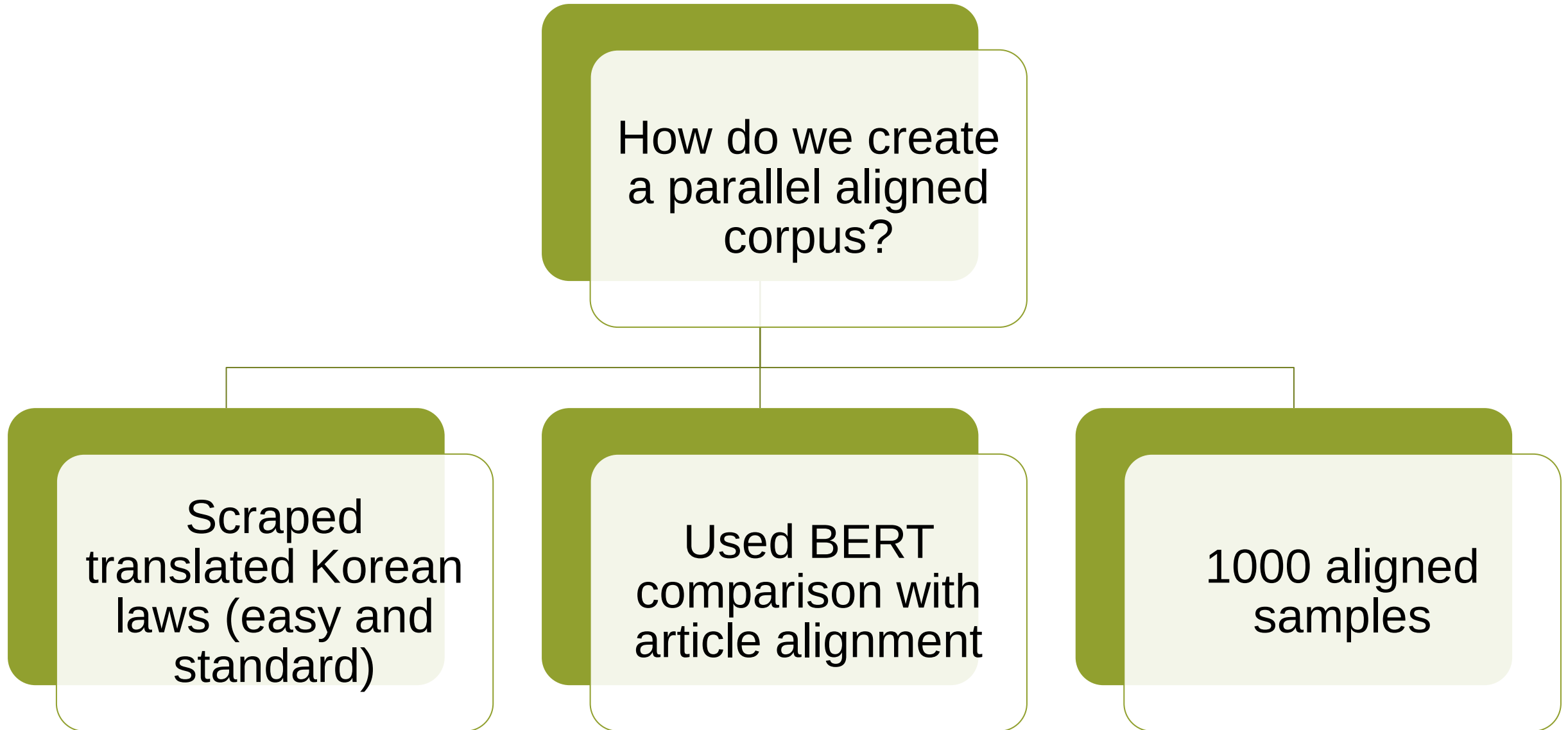
How can we find and automatically align easy legal text with laws in their original form to create a parallel aligned corpus?

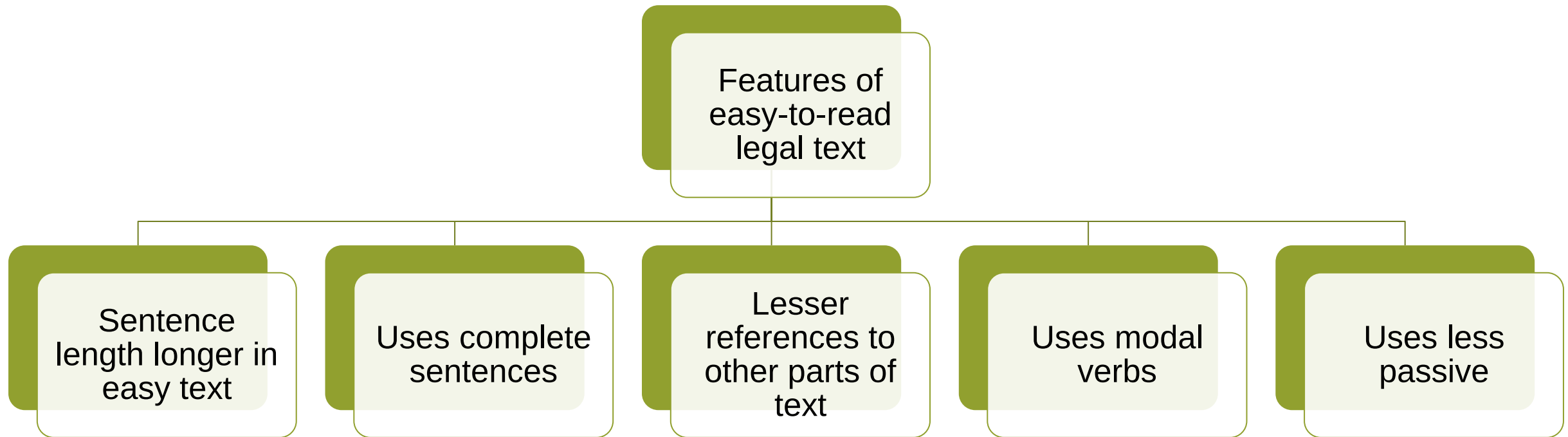


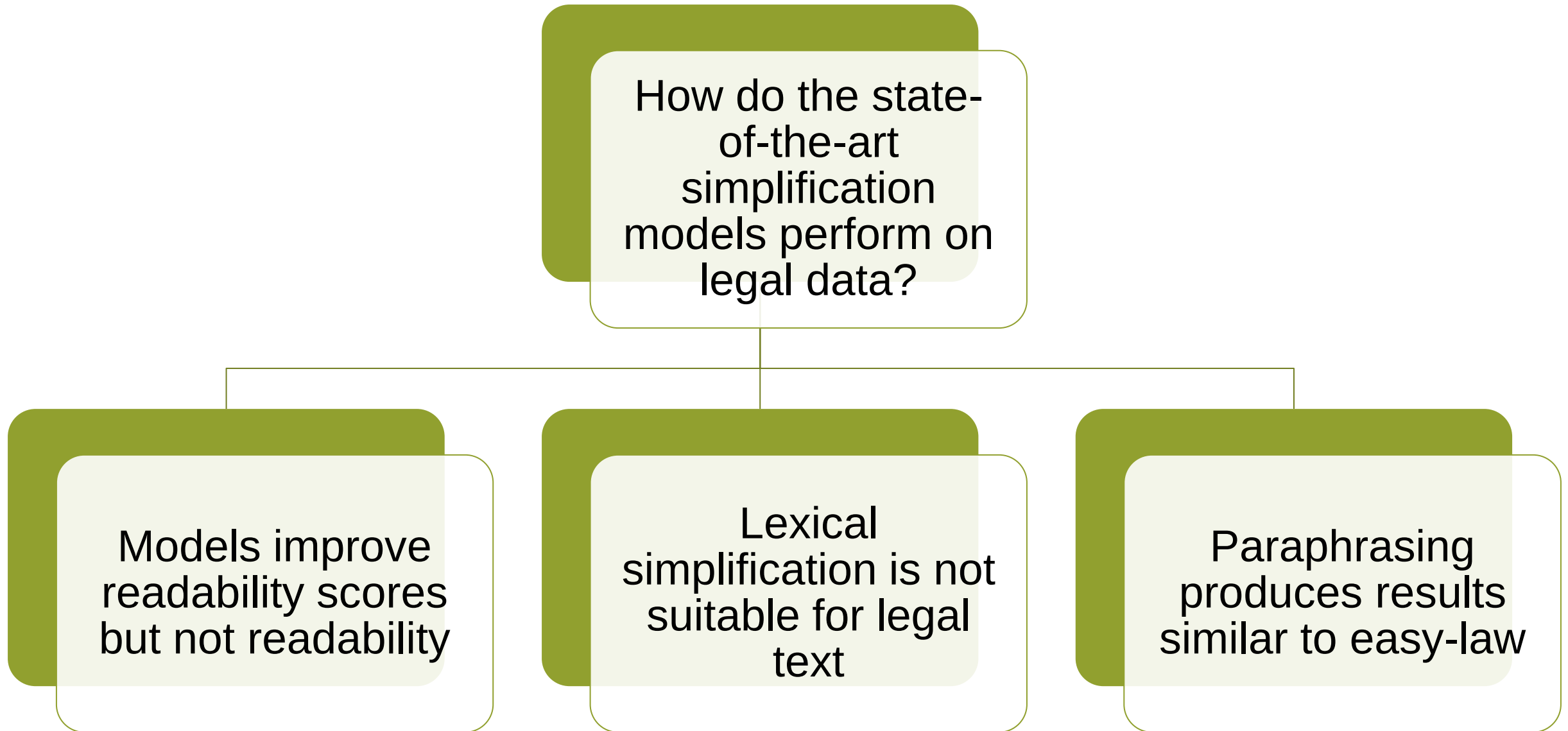
How can we use this corpus to characterize the features of easy and hard to read legal text?



How do the state-of-the-art simplification models perform on legal data?







Expand the aligned dataset

- Use legal-bert to perform alignment
- Extract more articles by expanding coverage of regex articles

Extract more characteristics of easy legal text

- Measure amount of information conveyed per sentence

Evaluation

- User study to quantitatively assess the collected dataset



Akshaya Muralidharan

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289.17132
Fax +49.89.289.17136

akshaya.muralidharan@in.tum.de
www.matthes.in.tum.de



-
- [1] Jin, Y., & Kling, J. (2009). Overcoming the language barrier: The literacy of non-native-English-speaking adults. US Department of Labor, Employment and Training Administration.
-
- [2] Iram Rubab, Mamona Yasmin Khan, and Tahira Asgher. Transformation of legal texts into simplified accounts to make the justice accessible.
-
- [3] Štajner, S., Franco-Salvador, M., Rosso, P., & Ponzetto, S. P. (2018, May). Cats: A tool for customized alignment of text simplification corpora. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).
-
- [4] Jiang, C., Maddela, M., Lan, W., Zhong, Y., & Xu, W. (2020). Neural CRF Model for Sentence Alignment in Text Simplification.
-
- [5] Qiang, Jipeng et al. “LSBert: A Simple Framework for Lexical Simplification.”
-
- [6] L. Martin, B. Sagot, E. De la Clergerie, A. Bordes, Controllable Sentence Simplification
-
- [7] Kauchak D, Leroy G, Hogue A. Measuring Text Difficulty Using Parse-Tree Frequency. *J Assoc Inf Sci Technol*.