

Automatic simplification of legal texts in Portuguese using machine learning

Alexandre ALVES¹, Péricles MIRANDA²,
Rafael MELLO^{1,2}, André NASCIMENTO^{1,2}

¹Cesar School, Recife, Brazil

²UFRPE, Recife, Brazil





Context

Legal documents in Brazilian judicial proceedings, such as judgments, appeals, documents, orders, decisions, sentences among others, use elaborate **vocabulary** in the Portuguese language and heavy **legal jargon**.

Documents are **difficult to read and understand** for the average citizen

The large volume of documents demands automatic and scalable solutions



Automatic Text Simplification (ATS)

Main objective is to produce a **simpler version** of the original text through **lexical** and **syntactic transformations**, without altering the original meaning of the text, and consequently **improve communication** and easy understanding of the texts

Automatic Text Simplification (ATS)



original
the quintessential birds

automatic
the typical birds

decoration
the quintessential birds
↓
the typical birds

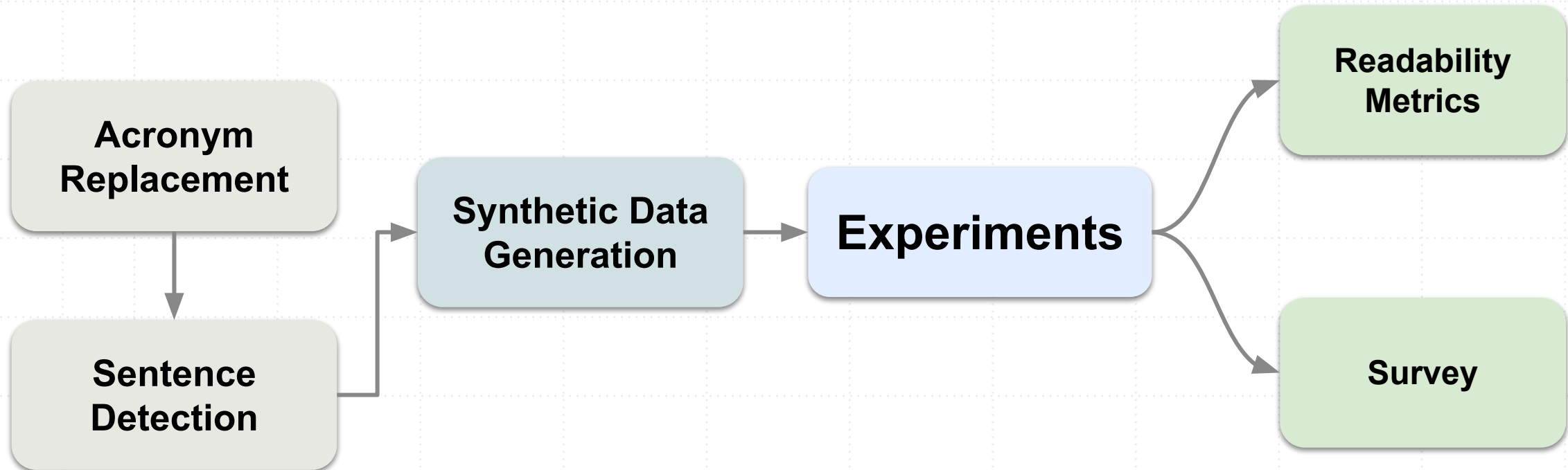
pop-up
typical
the quintessential birds



Research Questions

- **RQ1:** *How to increase the average readability score of simplified texts, in relation to the originals, on a set of summaries, so that they are simpler to read and understand?*
- **RQ2:** *How to generate a parallel corpus in Portuguese and in the legal environment to be used in the training of models that use supervised learning?*

Methodology



Flesch Reading Ease (FRE)

The **FRE** gives a text a score between 1 and 100, with 100 being the highest readability score

Adaptation for the Portuguese Language

$$FRE_{PT} = 248,835 - (1,015 \times ASL) - (84.6 \times ASW)$$

ASL: Average Sentence Length

ASW: Average Number of Syllables per Word

FRE Score	Descriptive Categories	Estimated Grade
90 - 100	Very Easy	5th grade
80 - 90	Easy	6th grade
70 - 80	Fairly Easy	7th grade
60 - 70	Standard	8th - 9th grade
50 - 60	Fairly Difficult	10th - 12th grade
30 - 40	Difficult	College
0 - 30	Very Difficult	College Graduate

Datasets

- Data was extracted from two major Brazilian Federal Courts (TRF5 and STF)
 - STF corpus was obtained from the RullingBR dataset¹
- The datasets comprise 200 summarized documents (100 for each court), addressing topics from different legal domains
- On average, TRF5 summaries have 424 tokens (words), while STF summaries have an average of 191 tokens.

¹ D. Feijo, Summarizing Legal Rulings (2021)

Datasets

A synthetic (parallel) corpus was generated by the MUSS(PT) method based on the decisions of the STF

A random sample of 10,574 summaries from the RulingBR dataset

The synthetic corpus was used for training of the supervised learning methods

(i.e., Transformer and NMT + Attention):

- **NMT + Attention:** parallel corpus with 132,879 pairs of sentences (original and simplified) (2h).
- **Transformer:** a subsample (10%) of the parallel corpus with 13,288 pairs of sentences (120h).

Test Datasets:

- 100 decisions from the TRF5 (Regional Federal Court of the 5th Region)
- 100 decisions from the STF (Federal Supreme Court)



Methods

- ATS based on **Transformer** architecture and attention mechanisms
- ATS based on **Multilingual Unsupervised Sentence Simplification by Mining Paraphrases (MUSS)**
 - English Model (PT→ EN Translation)
 - Portuguese Model
- ATS based on **Neural Machine Translation (NMT)** with attention mechanisms

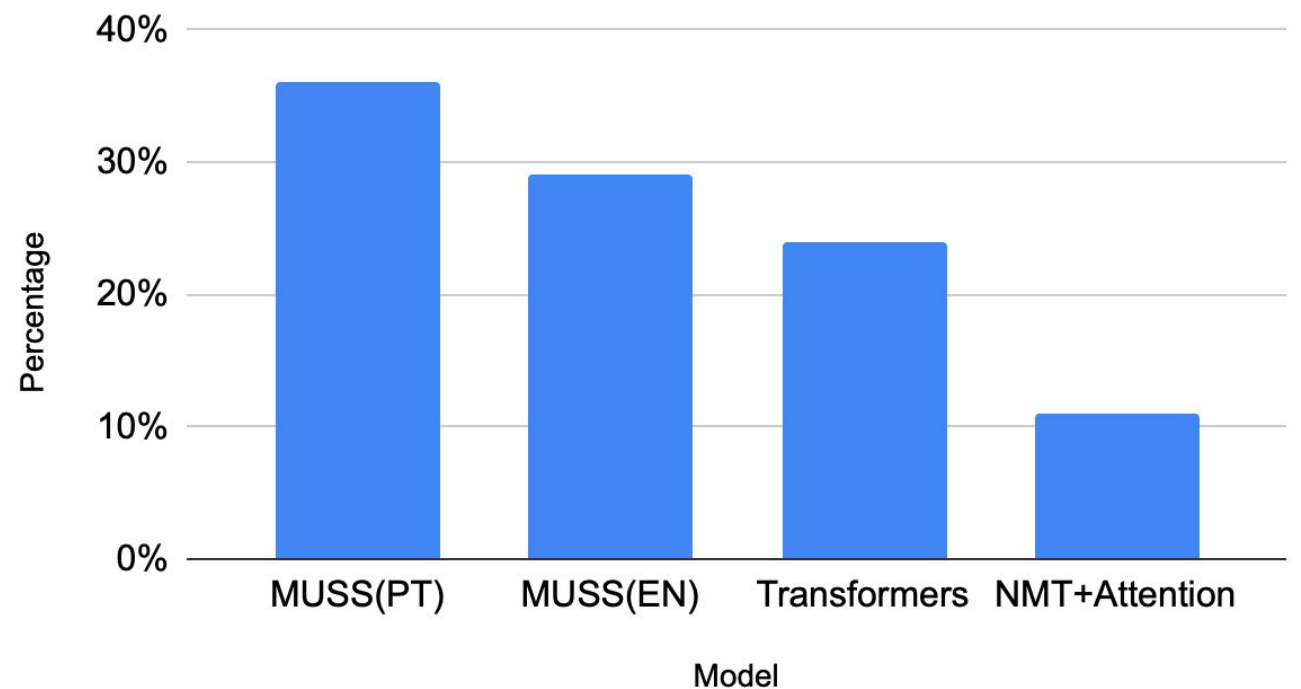
Results

Dataset	Method	Number of Sentences	ASL	ASW	Readability
TRF5	Original	28,92 ±8,43	17,52 ±3,45	2,11 ±0,10	54,86 ±9,66
	MUSS(EN)	27,44 ±18,26	19,02 ±5,70	2,00 ±0,10	62,82 ±8,90
	MUSS(PT)	29,20 ±8,39	15,37 ±3,08	2,11 ±0,12	56,64 ±10,28
	TRANSFORMERS	30,19 ±8,27	13,16 ±2,75	2,05 ±0,12	64,71 ±11,13
	NMT+ATTENTION	26,61 ±7,82	12,93 ±4,07	2,57 ±0,12	19,85 ±10,24
STF	Original	15,41 ±9,56	11,25 ±6,78	2,29 ±0,20	45,09 ±18,79
	MUSS(EN)	13,22 ±8,32	12,63 ±6,18	2,10 ±0,19	60,02 ±17,21
	MUSS(PT)	15,32 ±9,15	10,28 ±5,49	2,30 ±0,22	45,23 ±19,82
	TRANSFORMERS	15,63 ±9,14	10,30 ±4,40	2,29 ±0,23	46,12 ±20,41
	NMT+ATTENTION	14,66 ±8,70	10,09 ±4,55	2,74 ±0,22	7,96 ±18,41

Questionnaire

- A survey with **16 people** was made to assess the quality of the simplifications.
- Overall, **69% of the respondents** believe that the models **were able to simplify the texts**

Best Simplification according to respondents



Example

Original Sentence

3. Entendeu o Juízo a quo, em síntese, não haver falar em suspensão da presente execução fiscal em razão do deferimento da recuperação judicial da empresa devedora, ao contrário do que ocorre em relação a outras demandas judiciais, por força da previsão do art. 6º, § 7º, da Lei nº 11.101/05.

MUSS (EN)

3. O tribunal de primeira instância não considerou que esta execução fiscal tenha sido suspensa com base na definição de recuperação judicial da empresa devedora, nos termos do artigo 6.o, n.o 7, do ato n.o 11.101/05, relativamente a outros processos judiciais.

Example

Original Sentence

3. Entendeu o Juízo a quo, em síntese, não haver falar em suspensão da presente execução fiscal em razão do deferimento da recuperação judicial da empresa devedora, ao contrário do que ocorre em relação a outras demandas judiciais, por força da previsão do art. 6º, § 7º, da Lei nº 11.101/05.

MUSS (PT)

3. Entendeu o Juízo a quo, em síntese, não haver falar em suspensão da presente execução fiscal por causa do deferimento da recuperação judicial da empresa devedora, ao contrário do que ocorre em relação a outras demandas judiciais.

Example

Original Sentence

3. Entendeu o Juízo a quo, em síntese, não haver falar em suspensão da presente execução fiscal em razão do deferimento da recuperação judicial da empresa devedora, ao contrário do que ocorre em relação a outras demandas judiciais, por força da previsão do art. 6º, § 7º, da Lei nº 11.101/05.

Transformers

3. Entendendo - se em síntese, não haver necessidade de suspensão da execução fiscal em razão do deferimento da recuperação judicial da empresa Recorrente, ao argumento de que a previsão do art. 20, § 1o, da Lei no 11. 101 / 05 foi feita por meio do arto.

Example

Original Sentence

3. Entendeu o Juízo a quo, em síntese, não haver falar em suspensão da presente execução fiscal em razão do deferimento da recuperação judicial da empresa devedora, ao contrário do que ocorre em relação a outras demandas judiciais, por força da previsão do art. 6º, § 7º, da Lei nº 11.101/05.

Transformers

. o juizo a quo, em sintese, nao haver falar em suspensao da presente execucao fiscal da empresa devedora da empresa devedora do transito em razao da pratica determinada pela empresa devedora da empresa devedora, ao contrario do que ocorre em relacao a outras demandas judiciais, por forza da previsao do art., o, da lei no..



Conclusion

The experiments demonstrated that ATS with pre-trained models can **improve readability** of legal documents, achieving more **compact sentences** with **shorter words**.

As future work, we believe that LLM's are a promising approach to the Legal ATS task

Automatic simplification of legal texts in Portuguese using machine learning

Alexandre ALVES¹, Péricles MIRANDA²,
Rafael MELLO^{1,2}, André NASCIMENTO^{1,2}

¹Cesar School, Recife, Brazil

²UFRPE, Recife, Brazil

