



Supporting the Adaptation of Texts for Poor Literacy Readers

Arnaldo Candido Jr., Erick Maziero,
Caroline Gasperin,
Thiago A. S. Pardo, Lucia Specia,
Sandra M. Aluisio

Universidade de São Paulo, Brazil



Overview

- **PorSimples project**
 - **Text simplification technology**
 - Brazilian Portuguese
 - Mainly news texts
 - Digital Inclusion and Accessibility
 - **Editor for creating simplified texts**
 - Resources:
 - **Corpus of simplified texts**



Motivation

- INAF - National Indicator of Functional Literacy:
 - **Rudimentary level:** 68% of 30.6 million Brazilians between 15 - 64 years-old who studied up to 4 years
 - Only able to find explicit information in short texts
 - **Basic level:** 75% of the 31.1 million who studied up to 8 years
 - Can process slightly longer texts and make simple inferences



Text simplification

- **Lexical simplification**

- Replacing uncommon words by simpler and frequent ones

- **Syntactic simplification**

- Getting rid of complex structures:
 - Apposition
 - Passive voice
 - Non-canonical ordering
 - Coordination, Subordination.

Previous work



- Siddharthan (2003)
 - Analysis; Transformation; Regeneration
 - Shallow parsing
- Inui et al. (2003)
 - Generate several paraphrases for each sentence
 - Train a classifier to select the simpler ones
 - It carries errors on the paraphrase generation
- PSET project (1998)
 - Full parsing; passive voice
- Chandrasekar et al. (1997)
 - Supertags; tree transformation
 - Preprocessing step for a parser

Our Text Simplification System

- Rule-based approach
- 2 steps:
 1. Identification of syntactic constructions
 - We treats **19** syntactic phenomena
 2. Simplification
 - We propose **6** simplification operations
- Resources
 - Syntactic parser for Portuguese, PALAVRAS (Bick, 2000)
 - List of discourse markers
 - Superficial information



Manual for Text Simplification

- *Plain Language* guidelines
 - Short sentences, active voice, up to 3 levels of subordination, etc.
- Previous work for English
 - Mainly Siddharthan's work
- Corpus study
 - simple accounts
- Portuguese grammar

Syntactic phenomena

1. **Passive voice**
2. Apposition
3. Relative clauses
 1. Restrictive
 2. Non-restrictive
4. Coordinate clauses
 1. Asyndetic
 2. Additive
 3. Adversative
 4. Correlated
 5. Result
 6. Reason
5. Subordinate clauses
 1. Reason
 2. Comparative
 3. Concessive
 4. Conditional
 5. Result
 6. Purpose
 7. Confirmative
 8. Time
 9. Proportional
6. **Non S-V-O order**
7. Adverbial phrases



Simplification operations

1. Sentence splitting
2. Transformation to active voice
3. Inverting order of clauses
 - Logical order of events
4. **S-V-O ordering**
5. Changing discourse markers
 - Replace uncommon markers by frequent ones
6. **(De)topicalization of adverbial phrases**

Example: Relative clauses

1. Identifying non-restrictive relative clauses

- Syntactic clues: **rel** or **spec** tags
- Lexical clues: relative pronoun
- Punctuation: clause within “,”

2. Simplifying:

- sentence splitting

W S, Pron Yrel, Z. → W S Z. S Yrel.

Example:

More than 20 people have been bitten by gold piranhas, which live in the waters of the Sanchuri dam.

More than 20 people have been bitten by gold piranhas. Gold piranhas live in the waters of the Sanchuri dam.

Example: Passive voice

1. Identifying passive voice
 - Syntactic clues: tag **fApass** in parser output
2. Simplifying:
 - Transformation to active voice

A_{gent} P_{red} S_{ubj} . → S_{ubj} P_{mod} A_{gent} .

**More than 20 people have been bitten by gold piranhas,
which live in the waters of the Sanchuri dam.**

**Gold piranhas, which live in the waters of the Sanchuri dam,
have bitten more than 20 people.**

Example: Subordination

1. Identifying concessive subordinate clauses
 - Discourse markers: **although** or **despite** ...
2. Simplifying:
 - Sentence splitting
 - Inversion of clause order
 - Change of discourse marker

Main Dm Sub. → Sub. Dm' Main.

The building hosting the Brazilian Consulate was also evacuated, although the diplomats have obtained permission to carry on working.

The diplomats have obtained permission to carry on working. But the building hosting the Brazilian Consulate was also evacuated.

Example: S-V-O

1. Identifying non-S-V-O order
 - Syntactic clues: P (Od **ou** Op) S
2. Simplifying:
 - S-V-O ordering

Predicate **O**bject **S**ubj. → **S**ubj **P**redicate **O**bject.

On the 9th of November of 1989, fell the wall that for almost three decades divided Germany.

On the 9th of November of 1989, the wall that for almost three decades divided Germany fell.”



Open issues

- (De)topicalization operation

- When and how

On the 9th of November of 1989, the wall that for almost three decades divided Germany fell.

The wall that for almost three decades divided Germany fell on the 9th of November of 1989.

- Other non-canonical orders

- Currently only treating V-O-S

- Broader list of discourse markers

Evaluation

- Corpus of manual simplifications
 - Created using an annotation tool that was able to record the simplification operations
 - 104 news articles
 - Operations are assigned to sentences

<i>Simplification Operations</i>	<i># Sentences</i>
Non-simplification	2638
Subject-verb-object ordering	44
Transformation to active voice	154
Inversion of clause ordering	265
Splitting sentences	1103

Evaluation

1. Phenomenon Identification

- Compare against the corpus of manual simplifications
 - Contains only operations assigned to sentences
 - Evaluation based on the annotation

<i>Operation</i>	<i>P</i>	<i>R</i>	<i>F</i>
Splitting sentences	64.07	82.63	72.17
Inversion of clause ordering	15.40	18.91	16.97
Transformation to active voice	44.29	44.00	44.14
Subject-verb-object ordering	1.12	4.65	1.81
ALL	51.64	65.19	57.62
Non-simplification	64.69	53.58	58.61

A decorative graphic consisting of six circles arranged in two rows. The top row has three circles: a solid light purple circle, an outlined light purple circle, and a solid light purple circle. The bottom row has three circles: a solid light purple circle, an outlined light purple circle, and a solid light purple circle.

Evaluation

2. Simplification itself

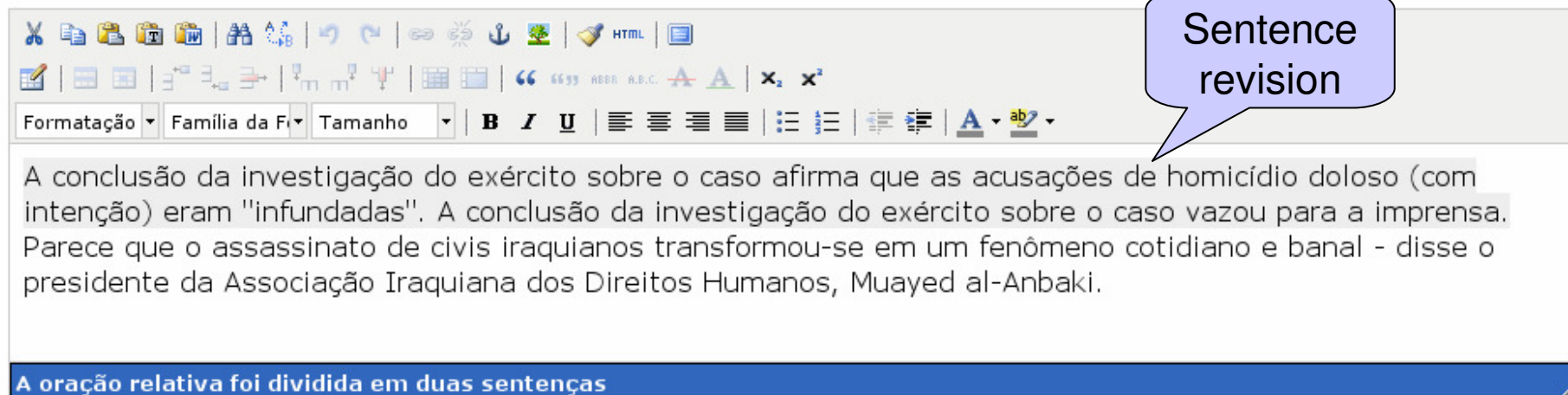
- Manual evaluation: future work



Our Editor

- Help writers to produce simplified texts
 - Teaching
 - Publications targeted to a poor literate audience
 - Government informative documents
- Uses our text simplification core
- The writer can revise the simplified text
 - The editor pass on to writers the decision about ambiguous cases

Our Editor



A screenshot of a web editor interface. The top part shows a toolbar with various icons for editing and formatting. Below the toolbar is a text area containing the following text: "A conclusão da investigação do exército sobre o caso afirma que as acusações de homicídio doloso (com intenção) eram "infundadas". A conclusão da investigação do exército sobre o caso vazou para a imprensa. Parece que o assassinato de civis iraquianos transformou-se em um fenômeno cotidiano e banal - disse o presidente da Associação Iraquiana dos Direitos Humanos, Muayed al-Anbaki." The text is highlighted in light gray. Below the text area is a blue status bar with the text "A oração relativa foi dividida em duas sentenças".

Sentence
revision

<- Anterior Próxima -> Finalizar revisão

Mais opções:

- A conclusão da investigação do exército sobre o caso afirma que as acusações de homicídio doloso (com intenção) eram "infundadas". A conclusão da investigação do exército sobre o caso vazou para a imprensa.
- A conclusão da investigação do exército sobre o caso afirma que as acusações de homicídio doloso (com intenção) eram "infundadas". O caso vazou para a imprensa.
- Sentença original:** A conclusão da investigação do exército sobre o caso, que vazou para a imprensa, afirma que as acusações de homicídio doloso (com intenção) eram "infundadas".

Alternative
simplifications



Comparison to other tools

- Text Adaptor (Burstein 2009)
 - Sentence highlighting, word substitution, text summarization and translation
 - Suggests that some sentences might be too complex
 - Using a shallow parser
 - No actual syntactic simplification
- Max (2006): Interactive editor
 - Within a word processor
 - Parser analyzes + rewriting rules
 - Ranks suggested simplifications by syntactic complexity
 - User chooses their preferred simplification (not autonomous)

Concluding remarks



- The potentialities of text simplification systems for education are evident:
 - Social inclusion
 - Facilitating and developing reading and writing skills
 - Teaching assistance
 - Creating text material to be used in classes
- Ours is the first text simplification system for Portuguese
 - Brazil has a clear need for such resource

Thanks!

- Acknowledgements

- FAPESP

- Microsoft Research

- Contact

- `cgasperin@icmc.usp.br`

- Search for “**PorSimples**” project

