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 treat 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**
1. Identifying non-restrictive relative clauses
   - Syntactic clues: rel or spec tags
   - Lexical clues: relative pronoun
   - Punctuation: clause within “,”

2. Simplifying:
   - sentence splitting

\[ W \ S, \ Pron \ Yrel, \ Z. \rightarrow W \ S \ Z. \ S \ Yrel. \]

**Example:**

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

More than 20 people have been bitten by gold piranhas. \textcolor{red}{Gold} piranhas live in the waters of the Sanchuri dam.
Example: Passive voice

1. Identifying passive voice
   - Syntactic clues: tag $fA_{pass}$ in parser output

2. Simplifying:
   - Transformation to active voice

   \[
   \text{Agent Pred Subj. } \rightarrow \text{Subj Pmod Agent.}
   \]

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

\[
\text{Main} \ Dm \ Sub. \rightarrow \ Sub. \ Dm' \ Main. 
\]

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

\[
\text{Predicate Object Subj. } \rightarrow \text{Subj Predicate Object.}
\]

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

<table>
<thead>
<tr>
<th>Simplification Operations</th>
<th># Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-simplification</td>
<td>2638</td>
</tr>
<tr>
<td>Subject-verb-object ordering</td>
<td>44</td>
</tr>
<tr>
<td>Transformation to active voice</td>
<td>154</td>
</tr>
<tr>
<td>Inversion of clause ordering</td>
<td>265</td>
</tr>
<tr>
<td>Splitting sentences</td>
<td>1103</td>
</tr>
</tbody>
</table>
Evaluation

1. Phenomenon Identification

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

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