

PORSIMPLES: SIMPLIFICATION OF PORTUGUESE TEXT FOR DIGITAL INCLUSION AND ACCESSIBILITY

Sandra Maria Aluísio

São Carlos Institute of Mathematics and Computer Sciences / University of São Paulo (USP)

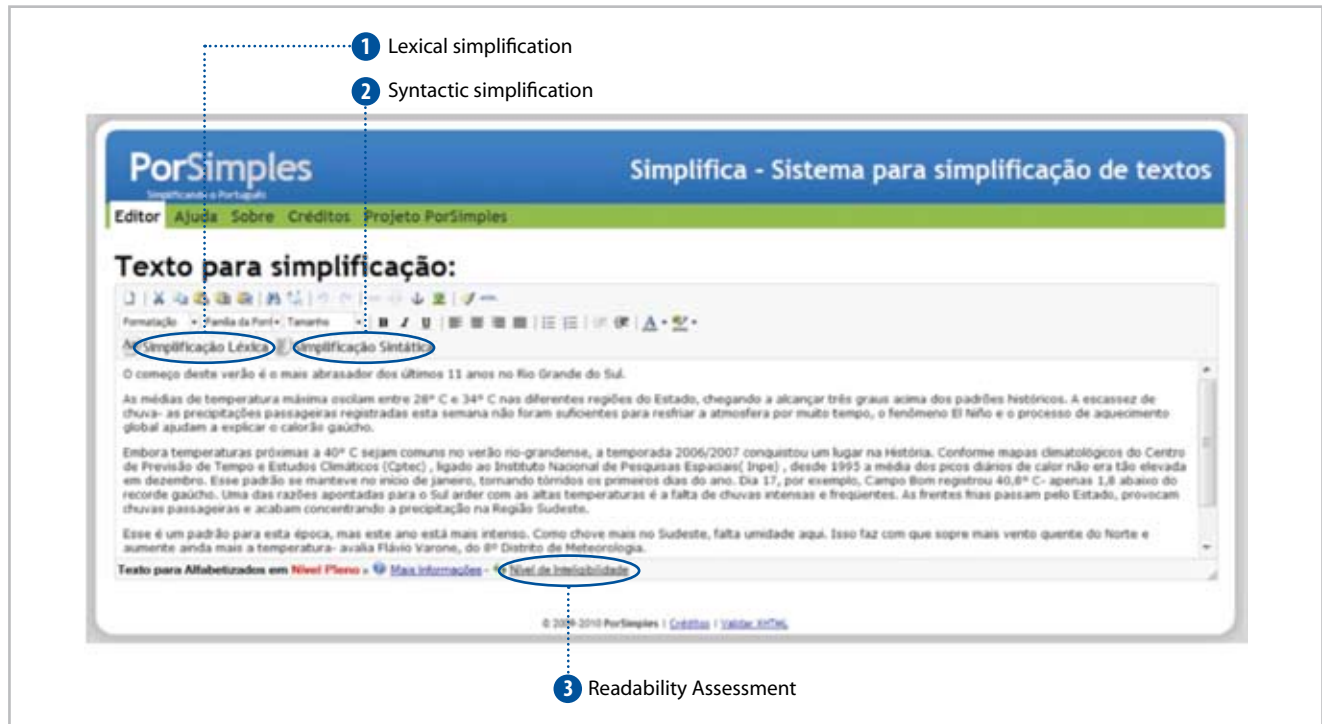


Figure 1. Main page of Simplifica

The main goal of PorSimples was to develop Natural Language Processing (NLP) technologies related to Text Adaptation (TA) to promote digital inclusion and accessibility for people with low levels of literacy. There are two general different approaches for TA: Text Simplification and Text Elaboration. The first can be defined as any task that reduces the lexical or syntactic complexity of a text, while trying to preserve meaning and information, and can be subdivided into Lexical and Syntactic Simplification, Automatic Summarization, and other techniques. Text Elaboration aims at clarifying and explaining information and making connections explicit in a text, for example, providing definitions or synonyms for words known to only a few speakers of a language. The technologies developed in PorSimples are available by means of three systems aimed to distinct users: 1)

an authoring system, called Simplifica, to help authors to produce simplified texts targeting people with low literacy levels (*figure 1*); 2) an assistive technology system, called Facilita, which explores the tasks of summarization and simplification to allow poor literate people to read Web content (*figure 2*), and 3) a web content adaptation tool, named Educational Facilita, for assisting low-literacy readers to perform detailed reading. It exhibits questions that clarify the semantic relations linking verbs to their arguments, highlights the associations amongst the main ideas of the texts and the named entities, and perform lexical elaboration. Currently, Educational Facilita only explores the NLP tasks of lexical elaboration and named entity labeling.

Website: <http://caravelas.icmc.usp.br/wiki>



MAIN RESULTS

Watanabe WM, Cândido A, Amancio MA, Oliveira M, Pardo TAS, Fortes RPM, Aluísio SM. 2010. Adapting Web content for low-literacy readers by using lexical elaboration and named entities labeling. *New Review of Hypermedia and Multimedia*. **16**: 303-327.

Gasperin C, Maziero E, Aluísio, SM. 2010. Challenging choices for text simplification. In: *Proceedings of PROPOR 2010*, 40-50.

Antônio Branco, Aldebaro Klautau, Renata Vieira, Vera Lúcia Strube de Lima (Eds.): Computational Processing of the Portuguese Language, 9th International Conference, PROPOR 2010, Porto Alegre, RS, Brazil. April 27-30, 2010. *Proceedings. Spring, 2010*. **6001**: 40-50.

Aluísio SM, Gasperin C. 2010. Fostering digital inclusion and accessibility: the PorSimples project for simplification of portuguese texts. *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*. New York: ACL, 2010. **1**: 46-53.

Aluísio SM, Specia L, Gasperin C, Scarton CE. Readability assessment for text simplification. In: NAACL 5th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-2010), 2010, Los Angeles. *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*. New York: ACL, 2010. **1**: 1-9.

Scarton CE, Gasperin C, Aluísio SM. Revisiting the readability assessment of texts in portuguese. In: IBERAMIA 2010, 2010, Bahia Blanca. *Lecture Notes in Computer Science. Heidelberg: Springer, 2010*. **6433**: 306-315.

Watanabe WM, Candido Jr. A, Uzêda V, Fortes RPM, Pardo TAS, Aluísio SM. Facilita: reading assistance for low-literacy readers. In: *Proceedings of ACM SIGDOC 2009 - ACM International Conference on Design of Communication*, 2009, Bloomington, In. **1**: 29-36.

Scarton CE, Aluísio SM. 2010. Análise da inteligibilidade de textos via ferramentas de processamento de língua natural adaptando as métricas do Coh-Metrix para o português. *Linguamática* (Online journal about natural language processing of Iberic languages). **2(1)**: 45-61.

Candido Jr. A, Maziero E, Gasperin C, Pardo T, Specia L, Aluísio S. 2009. Supporting the adaptation of texts for poor literacy readers: a text simplification editor for brazilian portuguese. In: *Proceedings of the NAACL HLT Workshop on Innovative Use of NLP for Building Educational Applications*, 34-42.

Gasperin C, Specia L, Pereira T, Aluísio S. Learning when to simplify sentences for natural text simplification. In: *CSBC*

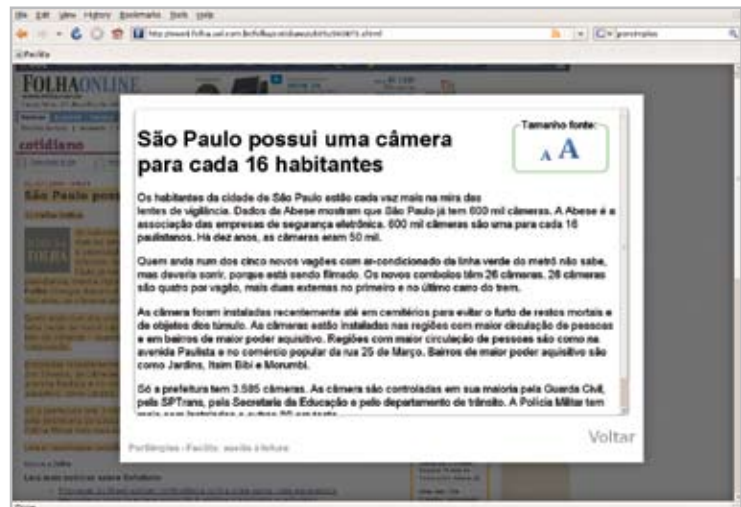


Figure 2. Facilita Demo in action: a summarized and simplified text (shown in front) of a text selected by the user from an online Brazilian newspaper (shown in back)

2009 - XXIX Congresso da Sociedade Brasileira de Computação, 2009, Bento Gonçalves, RS. ENIA 2009 – Encontro Nacional de Inteligência Artificial, 2009. 809-818.

PORSIMPLES IN NUMBERS

- Started in November 2007 and finished in April 2010
- Team:
 - 6 researchers/students supported by MSR-FAPESP
 - 11 other students joint the project
- Publications:
 - 28 papers (conferences and journal)
 - 6 Demos/Posters (short papers)
 - 12 Technical Reports
 - 1 submitted paper
- Research Collaborations: 13 senior researchers from
 - Psycholinguistics
 - Statistics
 - Natural Language Processing
 - Human-Computer Interaction
- Products:
 - 3 main systems
 - 6 types of text adaptation methods
 - 4 data resources
 - 3 supporting tools

Sandra Maria Aluísio

Instituto de Ciências Matemáticas e de Computação
de São Carlos – Universidade de São Paulo (USP)
Av. Trabalhador São-carlense, 400 – Centro
Caixa Postal 668
CEP 13560-970 – São Carlos, SP – Brasil

sandra@icmc.usp.br
55.16.33739663