

SIMPLIFICA: an authoring system targeting simplified texts in Brazilian Portuguese

Carolina Scarton, Matheus de Oliveira, Arnaldo Candido Jr.,
Caroline Gasperin, and Sandra Maria Aluísio

Department of Computer Sciences, University of São Paulo
Av. Trabalhador São-Carlense, 400. 13560-970 - São Carlos/SP, Brazil
{carolina@grad.,matheusol@grad.,arnaldoc,cgasperin@sandra}@icmc.usp.br
<http://www.icmc.usp.br>

Abstract. SIMPLIFICA is an authoring tool for producing simplified texts in Portuguese. It provides functionalities for lexical and syntactic simplification and for readability assessment. This tool is the first of its kind for Portuguese; it brings innovative aspects for simplification tools in general, since the authoring process is guided by readability assessment based on the levels of literacy of the Brazilian population.

Key words: text simplification and readability assessment

1 Introduction

In Brazil, since 2001, the National Indicator of Functional Literacy (INAF) index has been computed annually to measure the levels of literacy of the Brazilian population. These literacy levels are defined as: (1) Illiterate; (2) Rudimentary Literate; (3) Basic Literate; and (4) Advanced/Fully Literate.

We present in this paper the current version of an authoring tool named SIMPLIFICA. It helps authors to create simple texts targeted at poor literate readers. It extends the previous version presented in [1] with two new modules: lexical simplification and the assessment of the level of complexity of the input texts. The study is part of the PorSimples project (Simplification of Portuguese Text for Digital Inclusion and Accessibility) [2].

2 SIMPLIFICA authoring tool

SIMPLIFICA is a web-based WYSIWYG editor, based on TinyMCE web editor. The user inputs a text in the editor and customizes the simplification settings, where he/she can choose: (i) strong simplification, where all the complex syntactic phenomena are treated for each sentence, or customized simplification, where the user chooses one or more syntactic simplification phenomena to be treated for each sentence, and (ii) one or more thesauri to be used in the syntactic and lexical simplification processes. Then the user activates the readability assessment module to predict the complexity level of a text. This module maps

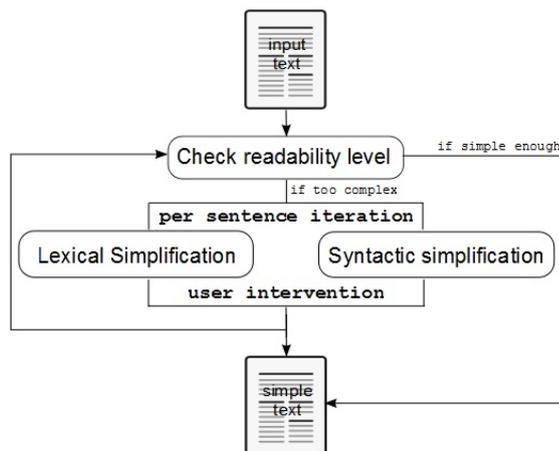


Fig. 1. Steps of the authoring process.

the text to one of the three levels of literacy defined by INAF: rudimentary, basic or advanced using Coh-Metrix-Port metrics [3]. According to the resulting readability level the user can trigger the lexical and/or syntactic simplifications modules, revise the automatic simplification and restart the cycle by checking the readability level of the current version of the text. Figure 1 summarizes how the three modules are integrated.

In the PROPOR demonstration section we will run all steps of the authoring process - readability assessment, lexical simplification and syntactic simplification - in order to demonstrate the use of the tool in producing a text for basic and rudimentary readers of Portuguese. We outline a script of our demonstration at http://nilc.icmc.usp.br/~cgasperin/PorSimples/demo/demo_script.htm

References

1. Candido Jr, A., Maziero, E., Gasperin, C., Pardo, T., Specia, L., Aluísio, S. M.: Supporting the Adaptation of Texts for Poor Literacy Readers: a Text Simplification Editor for Brazilian Portuguese. In the Proceedings of the NAACL HLT Workshop on Innovative Use of NLP for Building Educational Applications, pages 34-42, Boulder, Colorado, June 2009.
2. Aluísio, S. M., Specia, L., Pardo, T., Maziero, E., Fortes, R.: Towards Brazilian Portuguese Automatic Text Simplification Systems. In the Proceedings of The Eight ACM Symposium on Document Engineering (DocEng 2008), 240-248, São Paulo, Brasil.
3. Scarton, C. E., Almeida, D. M., Aluísio, S.M.: Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. In: the Proceedings of STIL 2009 1 CD-ROM v1. ISSN 2175-6201.