

Studies on text complexity

Maria José Bocorny Finatto, PhD (mariafinatto@gmail.com)

Federal University of Rio Grande do Sul (UFRGS - Programa de Pós-graduação em Letras - Lexicography, Terminology, Translation: Textual Relations)

Bianca Franco Pasqualini, MA (bianca.pasqualini@gmail.com)

Federal University of Rio Grande do Sul (UFRGS - Programa de Pós-graduação em Letras - Lexicography, Terminology, Translation: Textual Relations)

Aix-Marseille University (LIF-TALEP - Laboratoire d'Informatique Fondamentale - Traitement Automatique du Langage Ecrit et Parlé)

Aline Evers, MA (aline.evers@gmail.com)

Federal University of Rio Grande do Sul (UFRGS - Programa de Pós-graduação em Letras - Lexicography, Terminology, Translation: Textual Relations)

Université Paris-Est Marne-la-Vallée (UPEM-LIGM - Laboratoire d'Informatique Gaspard-Monge)

We are going to present a synthesis of a few studies in development at Federal University of Rio Grande do Sul (UFRGS) on the topics of text complexity and automated essay evaluation for Brazilian Portuguese. By addressing current possibilities of automating tasks of simplification and evaluation of texts according to complexity level, we intend to demonstrate successful linguistic descriptions and natural language processing techniques. Also, we are going to address some of the problems and limitations regarding the automation of such tasks, as well as difficulties and obstacles we have been facing in our ongoing studies.

- **First Talk: Studies on text complexity, vocabulary of popular newspapers and text simplification for adults with low literacy**

Bianca Franco Pasqualini

In the first part of this talk, we'll present PorPopular Project, which is a systematic collection of texts from Brazilian popular newspapers whose intended audience is composed primarily by people with low literacy levels and low income. The main goal of the project is profiling the lexicon and language used in these texts as a possible way to model text simplicity in Brazilian Portuguese in terms of vocabulary and syntax. In the second part, we'll show the results of three studies: (1) text complexity of compared literary texts in Brazilian Portuguese and English, (2) text complexity and automated text simplification for adults with poor reading skills, currently ongoing, and (3) analysis and comparison of two *corpora* of simple texts for readers with low literacy: "La Traversée" (contemporary fiction novels in French) and "É Só o Começo" (adaptations of classic novels in Brazilian Portuguese), also ongoing.

- **Second Talk: Automated essay evaluation in Brazilian Portuguese**

Aline Evers

We'll show a few procedures we have been testing and some of the results we have so far on automated essay evaluation of Brazilian Portuguese texts: (1) previous experience on automating the task of text evaluation applied to Brazilian Portuguese essays produced by students of Portuguese as a second language, (2) manual annotation process of a *corpus* of essays written by high school students using BRAT, (3) construction of graphs using UNITEX to describe, extract and select features of these essays that may be used to automatically distinguish well written texts from texts with poor writing quality, and (4) possibilities of using available resources in Brazilian Portuguese and other languages to identify and quantify orthographic and grammar mistakes and map their distributions in order to help with the process of automatically evaluating essays in Brazilian Portuguese.

PorPopular:

<<http://www.ufrgs.br/textecc/porlexbras/porpopular/>>

PorLexBras:

<<http://www.ufrgs.br/textecc/porlexbras/escrivendo/>>