Multiword expressions and lexical complexity in Portuguese and French

Bianca Pasqualini

Advisor: Maria José B. Finatto

PhD study – Long term goal:

To develop a tool focused mainly on metrics related to lexical, syntactical and semantical complexity (considering a text as a unit of analysis).
Ideally, the tool will indicate whether a text is suitable or not for a reader according to their literacy level (rudimentary, basic or advanced).

So far, by the end of my second year, I have done the following:

Study of theoretical aspects concerning linguistic complexity.

Experimental study (accepted to the XII Encontro de Linguística de Corpus – nov-2014)

Collected and prepared a parallel corpus composed of Brazilian Portuguese literary texts and their simplified version (made by linguists).

Multiword expressions and lexical complexity in Portuguese and French

Bianca Pasqualini

Advisor: Maria José B. Finatto

Study plan in Marseille – draft

Studies of lexical complexity prediction with Nuria Gala

FIRST SEMESTER

Expansion of the corpus:
Inclusion of school textbooks (in Brazilian Portuguese)
Inclusion of texts from popular Brazilian newspapers

SECOND SEMESTER

Focus on lexical aspects:
Distribution of multiword expressions frequencies among all the corpora (literary texts; textbooks; newspapers)
Creation of lexical metrics especific to analysis or prediction of text complexity, according to multiword expressions results.
Paper reporting the research.
Build a lexicon, a resource for portuguese, applying the metrics of multiword expressions complexity – depending on the corpus, and if applicable.

Boca de Rua

Experiment in popular journalism:
newspaper written by homeless people

Corpus: 2005-2014
4 editions/year