Readability Assessment for Text Simplification

Sandra Aluisio[1], Lucia Specia[2], Caroline Gasperin[1] and Carolina Scarton[1]

[1]Center of Computational Linguistics (NILC)
University of São Paulo
São Carlos - SP, Brazil
{sandra,cgasperin}@icmc.usp.br,
carol.scarton@gmail.com

[2]Research Group in Computational Linguistics
University of Wolverhampton
Wolverhampton, UK
L.Specia@wlv.ac.uk

Abstract

We describe a readability assessment approach to support the process of text simplification for poor literacy readers. Given an input text, the goal is to predict its readability level, which corresponds to the literacy level that is expected from the target reader: rudimentary, basic or advanced. We complement features traditionally used for readability assessment with a number of new features, and experiment with alternative ways to model this problem using machine learning methods, namely classification, regression and ranking. The best resulting model is embedded in an authoring tool for Text Simplification.

1 Introduction

In Brazil, the National Indicator of Functional Literacy (INAF) index has been computed annually since 2001 to measure the levels of literacy of the Brazilian population. The 2009 report presented a worrying scenario: 7% of the individuals are illiterate; 21% are literate at the rudimentary level; 47% are literate at the basic level; only 25% are literate at the advanced level (INAF, 2009). These literacy levels are defined as:

(1) Illiterate: individuals who cannot perform simple tasks such as reading words and phrases;

(2) Rudimentary: individuals who can find explicit information in short and familiar texts (such as an advertisement or a short letter);

(3) Basic: individuals who are functionally literate, i.e., they can read and understand texts of average length, and find information even when it is necessary to make some inference; and

(4) Advanced: fully literate individuals, who can read longer texts, relating their parts, comparing and interpreting information, distinguish fact from opinion, make inferences and synthesize.

In order to promote digital inclusion and accessibility for people with low levels of literacy, particularly to documents available on the web, it is important to provide text in a simple and easy-to-read way. This is a requirement of the Web Content Accessibility Guidelines 2.0's principle of comprehensibility and accessibility of Web content[1]. It states that for texts which demand reading skills more advanced than that of individuals with lower secondary education, one should offer an alternative version of the same content suitable for those individuals. While readability formulas for English have a long history – 200 formulas have been reported from 1920 to 1980s (Dubay, 2004) – the only tool available for Portuguese is an adaptation of the *Flesch Reading Ease* index. It evaluates the complexity of texts in a 4-level scale corresponding to grade levels (Martins et al., 1996).

In the PorSimples project (Aluísio et al., 2008) we develop text adaptation methods (via text simplification and elaboration approaches) to improve the comprehensibility of texts published on government websites or by renowned news agencies, which are expected to be relevant to a large audience with various literacy levels. The project provides automatic simplification tools to aid (1) poorly literate readers to understand online content – a browser plug-in for automatically simplifying websites – and (2) authors producing texts for this audience – an authoring tool for guiding the creation of simplified versions of texts.

This paper focuses on a readability assessment approach to assist the simplification process in the authoring tool, SIMPLIFICA. The current version of SIMPLIFICA offers simplification operations addressing a number of lexical and syntactic phenomena to make the text more readable. The au-

[1] http://www.w3.org/TR/WCAG20/

Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, pages 1–9,
Los Angeles, California, June 2010.

thor has the freedom to choose when and whether to apply the available simplification operations, a decision based on the level of complexity of the current text and on the target reader.

A method for automatically identifying such level of complexity is therefore of great value. With our readability assessment tool, the author is able to automatically check the complexity/readability level of the original text, as well as modified versions of such text produced as he/she applies simplification operations offered by SIMPLIFICA, until the text reaches the expected level, adequate for the target reader.

In this paper we present such readability assessment tool, developed as part of the PorSimples project, and discuss its application within the authoring tool. Different from previous work, the tool does not model text difficulty according to linear grade levels (e.g., Heilman et al., 2008), but instead maps the text into the three levels of literacy defined by INAF: rudimentary, basic or advanced. Moreover, it uses a more comprehensive set of features, different learning techniques and targets a new language and application, as we discuss in Section 4. More specifically, we address the following research questions:

1. Given some training material, is it possible to detect the complexity level of Portuguese texts, which corresponds to the different literacy levels defined by INAF?
2. What is the best way to model this problem and which features are relevant?

We experiment with nominal, ordinal and interval-based modeling techniques and exploit a number of the cognitively motivated features proposed by Coh-Metrix 2.0 (Graesser et al., 2004) and adapted to Portuguese (called Coh-Metrix-PORT), along with a set of new features, including syntactic features to capture simplification operations and n-gram language model features.

In the remainder of this paper, we first provide some background information on the need for a readability assessment tool within our text simplification system (Section 2) and discuss prior work on readability assessment (Section 3), to then present our features and modeling techniques (Section 4) and the experiments performed to answer our research questions (Section 5).

2. Text Simplification in PorSimples

Text Simplification (TS) aims to maximize reading comprehension of written texts through their simplification. Simplification usually involves substituting complex by simpler words and breaking down and changing the syntax of complex, long sentences (Max, 2006; Siddharthan, 2003).

To meet the needs of people with different levels of literacy, in the PorSimples project we propose two types of simplification: *natural* and *strong*. The first type results in texts adequate for people with a basic literacy level and the second, rudimentary level. The difference between these two is the degree of application of simplification operations to complex sentences. In strong simplification, operations are applied to all complex syntactic phenomena present in the text in order to make it as simple as possible, while in natural simplification these operations are applied selectively, only when the resulting text remains “natural”. One example of original text (a), along with its natural (b) and strong (c) manual simplifications, is given in Table 1.

(a)	The cinema theaters around the world were showing a production by director Joe Dante in which a shoal of piranhas escaped from a military laboratory and attacked participants of an aquatic show. (...) More than 20 people were bitten by palometas (Serrasalmus spilopleura, a species of piranhas) that live in the waters of the Sanchuri dam.
(b)	The cinema theaters around the world were showing a production by director Joe Dante. In the production a shoal of piranhas escaped from a military laboratory and attacked participants of an aquatic show. (…) More than 20 people were bitten by palometas that live in the waters of the Sanchuri dam. Palometas are Serrasalmus spilopleura, a species of piranhas.
(c)	The cinema theaters around the world were showing a movie by director Joe Dante. In the movie a shoal of piranhas escaped from a military laboratory. The shoal of piranhas attacked participants of an aquatic show. (...). Palometas have bitten more than 20 people. Palometas live in the waters of the Sanchuri dam. Palometas are Serrasalmus spilopleura, a species of piranhas.

Table 1: Example of original and simplified texts

The association between these two types of simplification and the literacy levels was identified by means of a corpus study. We have manually built a corpus of simplified texts at both natural and strong levels and analyzed their linguistic structures according to the description of the two literacy levels. We verified that strong simplified sentences are more adequate for rudimentary level readers, and natural ones for basic level readers. This claim is supported by several studies which relate capabilities and performance of the working memory with reading levels (Siddharthan, 2003; McNamara et al., 2002).

2.1 The Rule-based Simplification System

The association between simplification operations and the syntactic phenomena they address is implemented within a rule-based syntactic simplification system (Candido Jr. et al., 2009). This system is able to identify complex syntactic phenomena in a sentence and perform the appropriate operations to simplify each phenomenon.

The simplification rules follow a manual for syntactic simplification in Portuguese also developed in PorSimples. They cover syntactic constructions such as apposition, relative clauses, coordination and subordination, which had already been addressed by previous work on text simplification (Siddharthan, 2003). Additionally, they address the transformation of sentences from passive into active voice, normalization of sentences into the Subject-Verb-Object order, and simplification of adverbial phrases. The simplification operations available are: sentence splitting, changing particular discourse markers by simpler ones, transforming passive into active voice, inverting the order of clauses, converting to subject-verb-object order, relocating long adverbial phrases.

2.2 The SIMPLIFICA Tool

The rule-based simplification system is part of SIMPLIFICA, an authoring tool for writers to adapt original texts into simplified texts. Within SIMPLIFICA, the author plays an active role in generating natural or strong simplified texts by accepting or rejecting the simplifications offered by the system on a sentence basis and post-editing them if necessary.

Despite the ability to make such choices at the sentence level, it is not straightforward for the author to judge the complexity level of the *text as whole* in order to decide whether it is ready for a certain audience. This is the main motivation for the development of a readability assessment tool.

The readability assessment tool automatically detects the level of complexity of a text at any moment of the authoring process, and therefore guides the author towards producing the adequate simplification level according to the type of reader. It classifies a text in one of three levels: rudimentary, basic or advanced.

Figure 1 shows the interface of SIMPLIFICA, where the complexity level of the current text as given by the readability assessment tool is shown at the bottom, in red (in this case, "*Nível Pleno*", which corresponds to *advanced*). To update the readability assessment of a text the author can choose "*Nível de Inteligibilidade*" (r*eadability level*) at any moment.

The text shown in Figure 1 is composed of 13 sentences, 218 words. The lexical simplification module (not shown in the Figure 1) finds 10 candidate words for simplification in this text, and the syntactic simplification module selects 10 sentences to be simplified (highlighted in gray).

When the author selects a highlighted sentence, he/she is presented with all possible simplifications proposed by the rule-based system for this sentence. Figure 2 shows the options for the first sentence in Figure 1. The first two options cover non-finite clause and adverbial adjuncts, respectively, while the third option covers both phenomena in one single step. The original sentence is also given as an option.

It is possible that certain suggestions of automatic simplifications result in ungrammatical or inadequate sentences (mainly due to parsing errors). The author can choose not to use such suggestions as well as manually edit the original or automatically simplified versions. The impact of the author's choice on the *overall* readability level of the text is not always clear to the author. The goal of the readability assessment function is to provide such information.

Simplified texts are usually longer than the original ones, due to sentence splittings and repetition of information to connect such sentences. We acknowledge that low literacy readers prefer short texts, but in this tool the shortening of the text is a responsibility of the author. Our focus is on the linguistic structure of the texts; the length of the text actually is a feature considered by our readability assessment system.

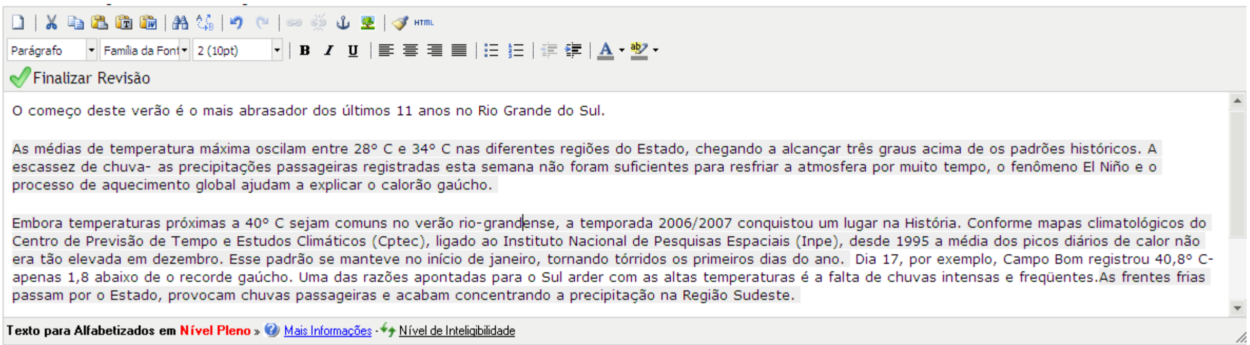

Figure 1: SIMPLIFICA interface

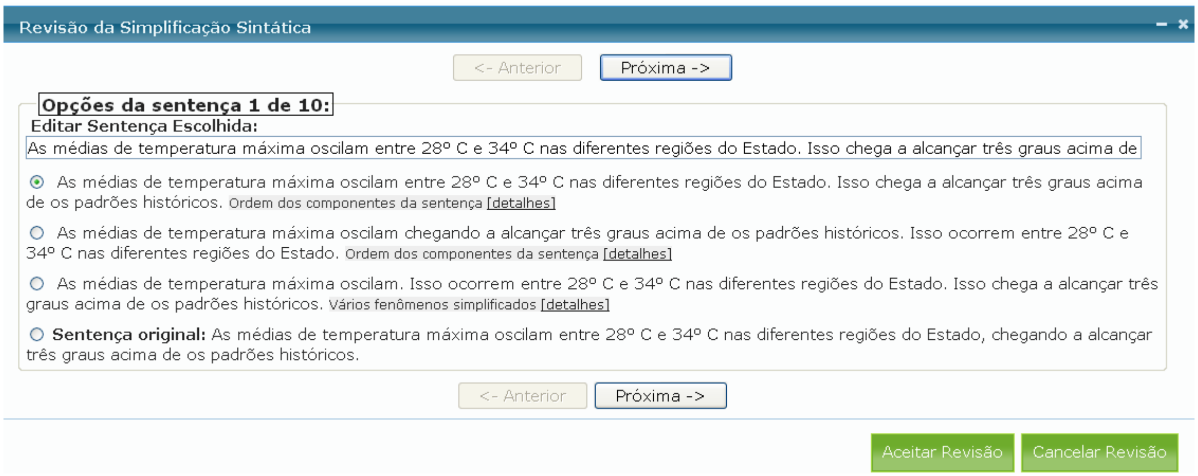

Figure 2. Simplification options available for the first sentence of the text presented in Figure 1

3. Readability Assessment

Recent work on readability assessment for the English language focus on: (i) the **feature set** used to capture the various aspects of readability, to evaluate the contribution of lexical, syntactic, semantic and discursive features; (ii) the **audience** of the texts the readability measurement is intended to; (iii) the **genre** effects on the calculation of text difficult; (iv) the type of **learning technique** which is more appropriate: those producing nominal, ordinal or interval scales of measurement, and (v) providing an **application** for the automatic assessment of reading difficulty.

Pitler and Nenkova (2008) propose a unified framework composed of vocabulary, syntactic, elements of lexical cohesion, entity coherence and discourse relations to measure text quality, which resembles the composition of rubrics in the area of essay scoring (Burstein et al., 2003).

The following studies address readability assessment for specific audiences: learners of English as second language (Schwarm and Ostendorf, 2005; Heilman et al., 2007), people with intellectual disabilities (Feng et al., 2009), and people with cognitive impairment caused by Alzheimer (Roark at al, 2007).

Sheehan et al. (2007) focus on models for literary and expository texts, given that traditional metrics like Flesch-Kincaid Level score tend to overpredict the difficulty of literary texts and underpredict the difficulty of expository texts.

Heilman et al. (2008) investigate an appropriate scale of measurement for reading difficulty – nominal, ordinal, or interval – by comparing the effectiveness of statistical models for each type of data. Petersen and Ostendorf (2009) use classification and regression techniques to predict a readability score.

Miltsakali and Troutt (2007; 2008) propose an automatic tool to evaluate reading difficulty of Web texts in real time, addressing teenagers and adults with low literacy levels. Using machine learning, Glöckner et al. (2006) present a tool for automatically rating the readability of German texts using several linguistic information sources and a global readability score similar to the *Flesch Reading Ease*.

4. A Tool for Readability Assessment

In this section we present our approach to readability assessment. It differs from previous work in the following aspects: (i) it uses a feature set with cognitively-motivated metrics and a number of additional features to provide a better explanation of the complexity of a text; (ii) it targets a new audience: people with different literacy levels; (iii) it investigates different statistical models for non-linear data scales: the levels of literacy defined by INAF, (iv) it focus on a new application: the use of readability assessment for text simplification systems; and (v) it is aimed at Portuguese.

4.1 Features for Assessing Readability

Our feature set (Table 2) consists of 3 groups of features. The first group contains cognitively-motivated features (features 1-42), derived from the Coh-Metrix-PORT tool (see Section 4.1.1). The second group contains features that reflect the incidence of particular syntactic constructions which we target in our text simplification system (features 43-49). The third group (the remaining features in Table 2) contains features derived from n-gram language models built considering unigrams, bigrams and trigrams probability and perplexity plus out-of-vocabulary rate scores. We later refer to a set of *basic features*, which consist of simple counts that do not require any linguistic tool or external resources to be computed. This set corresponds to features 1-3 and 9-11.

4.1.1 Coh-Metrix-Port

The Coh-Metrix tool was developed to compute features potentially relevant to the comprehension of English texts through a number of measures informed by linguistics, psychology and cognitive studies. The main aspects covered by the measures are cohesion and coherence (Graesser et al., 2004).

Coh-Metrix 2.0, the free version of the tool, contains 60 readability metrics. The Coh-Metrix-PORT tool (Scarton et al., 2009) computes similar metrics for texts in Brazilian Portuguese. The major challenge to create such tool is the lack of some of the necessary linguistic resources. The following metrics are currently available in the tool (we refer to Table 2 for details):

1. Readability metric: feature 12.
2. Words and textual information:
   - Basic counts: features 1 to 11.

1	Number of words
2	Number of sentences
3	Number of paragraphs
4	Number of verbs
5	Number of nouns
6	Number of adjectives
7	Number of adverbs
8	Number of pronouns
9	Average number of words per sentence
10	Average number of sentences per paragraph
11	Average number of syllables per word
12	*Flesch* index for Portuguese
13	Incidence of content words
14	Incidence of functional words
15	Raw Frequency of content words
16	Minimal frequency of content words
17	Average number of verb hypernyms
18	Incidence of NPs
19	Number of NP modifiers
20	Number of words before the main verb
21	Number of high level constituents
22	Number of personal pronouns
23	Type-token ratio
24	Pronoun-NP ratio
25	Number of “e” (and)
26	Number of “ou” (or)
27	Number of “se” (if)
28	Number of negations
29	Number of logic operators
30	Number of connectives
31	Number of positive additive connectives
32	Number of negative additive connectives
33	Number of positive temporal connectives
34	Number of negative temporal connectives
35	Number of positive causal connectives
36	Number of negative causal connectives
37	Number of positive logic connectives
38	Number of negative logic connectives
39	Verb ambiguity ratio
40	Noun ambiguity ratio
41	Adverb ambiguity ratio
42	Adjective ambiguity ratio
43	Incidence of clauses
44	Incidence of adverbial phrases
45	Incidence of apposition
46	Incidence of passive voice
47	Incidence of relative clauses
48	Incidence of coordination
49	Incidence of subordination
50	Out-of-vocabulary words
51	LM probability of unigrams
52	LM perplexity of unigrams
53	LM perplexity of unigrams, without line break
54	LM probability of bigrams
55	LM perplexity of bigrams
56	LM perplexity of bigrams, without line break
57	LM probability of trigrams
58	LM perplexity of trigrams
59	LM perplexity of trigrams, without line break

Table 2. Feature set

- Frequencies: features 15 to 16.
- Hypernymy: feature 17.

3. Syntactic information:
   - Constituents: features 18 to 20.
   - Pronouns: feature 22
   - Types and Tokens: features 23 to 24.
   - Connectives: features 30 to 38.
4. Logical operators: features 25 to 29.

The following resources for Portuguese were used: the MXPOST POS tagger (Ratnaparkhi, 1996), a word frequency list compiled from a 700 million-token corpus[2], a tool to identify reduced noun phrases (Oliveira et al., 2006), a list of connectives classified as positives/negatives and according to cohesion type (causal, temporal, additive or logical), a list of logical operators and WordNet.Br (Dias-da-Silva et al., 2008).

In this paper we include seven new metrics to Coh-Metrix-PORT: features 13, 14, 21, and 39 to 42. We used TEP[3] (Dias-da-Silva et al., 2003) to obtain the number of senses of words (and thus their ambiguity level), and the *Palavras* parser (Bick, 2000) to identify the higher level constituents. The remaining metrics were computed based on the POS tags.

According to a report on the performance of each Coh-Metrix-PORT metric (Scarton et al., 2009), no individual feature provides sufficient indication to measure text complexity, and therefore the need to exploit their combination, and also to combine them with the other types of features described in this section.

4.1.2 Language-model Features

Language model features were derived from a large corpus composed of a sample of the Brazilian newspaper *Folha de São Paulo* containing issues from 12 months taken at random from 1994 to 2005. The corpus contains 96,868 texts and 26,425,483 tokens. SRILM (Stolcke, 2002), a standard language modelling toolkit, was used to produce the language model features.

4.2 Learning Techniques

Given that the boundaries of literacy level classes are one of the subjects of our study, we exploit three different types of models in order to check which of them can better distinguish among the three literacy levels. We therefore experiment with three types of machine learning algorithms: a standard classifier, an ordinal (ranking) classifier and a regressor. Each algorithm assumes different relations among the groups: the classifier assumes no relation, the ordinal classifier assumes that the groups are ordered, and the regressor assumes that the groups are continuous.

As classifier we use the Support Vector Machines (SVM) implementation in the Weka[4] toolkit (SMO). As ordinal classifier we use a meta classifier in Weka which takes SMO as the base classification algorithm and performs pairwise classifications (OrdinalClassClassifier). For regression we use the SVM regression implementation in Weka (SMO-reg). We use the linear versions of the algorithms for classification, ordinal classification and regression, and also experiment with a radial basis function (RBF) kernel for regression.

5. Experiments

5.1 Corpora

In order to train (and test) the different machine learning algorithms to automatically identify the readability level of the texts we make use of manually simplified corpora created in the PorSimples project. Seven corpora covering our three literacy levels (advanced, basic and rudimentary) and two different genres were compiled. The first corpus is composed of general news articles from the Brazilian newspaper *Zero Hora* (*ZH original*). These articles were manually simplified by a linguist, expert in text simplification, according to the two levels of simplification: natural (*ZH natural*) and strong (*ZH strong*). The remaining corpora are composed of popular science articles from different sources: (a) the *Caderno Ciência* section of the Brazilian newspaper *Folha de São Paulo*, a mainstream newspaper in Brazil (*CC original*) and a manually simplified version of this corpus using the natural (*CC natural*) and strong (*CC strong*) levels; and (b) advanced level texts from a popular science magazine called *Ciência Hoje* (*CH*). Table 3 shows a few statistics about these seven corpora.

5.2 Feature Analysis

As a simple way to check the contribution of different features to our three literacy levels, we com-

[2] http://www2.lael.pucsp.br/corpora/bp/index.htm
[3] http://www.nilc.icmc.usp.br/tep2/index.htm
[4] http://www.cs.waikato.ac.nz/ml/weka/

Corpus	Doc	Sent	Words	Avg. words per text (std. deviation)	Avg. words p. sentence
ZH original	104	2184	46190	444.1 (133.7)	21.1
ZH natural	104	3234	47296	454.7 (134.2)	14.6
ZH strong	104	3668	47938	460.9 (137.5)	13.0
CC original	50	882	20263	405.2 (175.6)	22.9
CC natural	50	975	19603	392.0 (176.0)	20.1
CC strong	50	1454	20518	410.3 (169.6)	14.1
CH	130	3624	95866	737.4 (226.1)	26.4

Table 3. Corpus statistics

puted the (absolute) Pearson correlation between our features and the expected literacy level for the two sets of corpora that contain versions of the three classes of interest (original, natural and strong). Table 4 lists the most highly correlated features.

	Feature	**Corr.**
1	Words per sentence	**0.693**
2	Incidence of apposition	0.688
3	Incidence of clauses	0.614
4	Flesch index	0.580
5	Words before main verb	0.516
6	Sentences per paragraph	0.509
7	Incidence of relative clauses	0.417
8	Syllables per word	0.414
9	Number of positive additive connectives	0.397
10	Number of negative causal connectives	0.388

Table 4: Correlation between features and literacy levels

Among the top features are mostly basic and syntactic features representing the number of appositive and relative clauses and clauses in general, and also features from Coh-Metrix-PORT. This shows that traditional cognitively-motivated features can be complemented with more superficial features for readability assessment.

5.3 Predicting Complexity Levels

As previously discussed, the goal is to predict the complexity level of a text as original, naturally or strongly simplified, which correspond to the three literacy levels of INAF: rudimentary, basic and advanced level.

Tables 5-7 show the results of our experiments using 10-fold cross-validation and standard classification (Table 5), ordinal classification (Table 6) and regression (Table 7), in terms of F-measure (F), Pearson correlation with true score (Corr.) and mean absolute error (MAE). Results using our complete feature set (All) and different subsets of it are shown so that we can analyze the performance of each group of features. We also experiment with the Flesch index on its own as a feature.

Features	Class	F	Corr.	MAE
All	original	**0.913**	**0.84**	**0.276**
	natural	**0.483**		
	strong	**0.732**		
Language Model	original	0.669	0.25	0.381
	natural	0.025		
	strong	0.221		
Basic	original	0.846	0.76	0.302
	natural	0.149		
	strong	0.707		
Syntactic	original	0.891	0.82	0.285
	natural	0.32		
	strong	0.74		
Coh-Metrix-PORT	original	0.873	0.79	0.290
	natural	0.381		
	strong	0.712		
Flesch	original	0.751	0.52	0.348
	natural	0.152		
	strong	0.546		

Table 5: Standard Classification

Features	Class	F	Corr.	MAE
All	original	**0.904**	**0.83**	**0.163**
	natural	0.484		
	strong	**0.731**		
Language Model	original	0.634	0.49	0.344
	natural	**0.497**		
	strong	0.05		
Basic	original	0.83	0.73	0.231
	natural	0.334		
	strong	0.637		
Syntactic	original	0.891	0.81	0.180
	natural	0.382		
	strong	0.714		
Coh-Metrix-PORT	original	0.878	0.8	0.183
	natural	0.432		
	strong	0.709		
Flesch	original	0.746	0.56	0.310
	natural	0.489		
	strong	0		

Table 6: Ordinal classification

The results of the standard and ordinal classification are comparable in terms of F-measure and correlation, but the mean absolute error is lower for the ordinal classification. This indicates that ordinal classification is more adequate to handle our classes, similarly to the results found in (Heilman et al., 2008). Results also show that distinguishing between natural and strong simplifications is a harder problem than distinguishing between these and original texts. This was expected, since these two levels of simplification share many features. However, the average performance achieved is considered satisfactory.

Concerning the regression model (Table 7), the RBF kernel reaches the best correlation scores among all models. However, its mean error rates are above the ones found for classification. A linear SVM (not shown here) achieves very poor results across all metrics.

Features	Corr.	MAE
All	**0.8502**	**0.3478**
Language Model	0.6245	0.5448
Basic	0.7266	0.4538
Syntactic	0.8063	0.3878
Coh-Metrix-PORT	0.8051	0.3895
Flesch	0.5772	0.5492

Table 7: Regression with RBF kernel

With respect to the different feature sets, we can observe that the combination of all features consistently yields better results according to all metrics across all our models. The performances obtained with the subsets of features vary considerably from model to model, which shows that the combination of features is more robust across different learning techniques. Considering each feature set independently, the syntactic features, followed by Coh-Metrix-PORT, achieve the best correlation scores, while the language model features performed the poorest.

These results show that it is possible to predict with satisfactory accuracy the readability level of texts according to our three classes of interest: original, naturally simplified and strongly simplified texts. Given such results we embedded the classification model (Table 5) as a tool for readability assessment into our text simplification authoring system. The linear classification is our simplest model, has achieved the highest F-measure and its correlation scores are comparable to those of the other models.

6. Conclusions

We have experimented with different machine learning algorithms and features in order to verify whether it was possible to automatically distinguish among the three readability levels: original texts aimed at advanced readers, naturally simplified texts aimed at people with basic literacy level, and strongly simplified texts aimed at people with rudimentary literacy level. All algorithms achieved satisfactory performance with the combination of all features and we embedded the simplest model into our authoring tool.

As future work, we plan to investigate the contribution of deeper cognitive features to this problem, more specifically, semantic, co-reference and mental model dimensions metrics. Having this capacity for readability assessment is useful not only to inform authors preparing simplified material about the complexity of the current material, but also to guide automatic simplification systems to produce simplifications with the adequate level of complexity according to the target user.

The authoring tool, as well as its text simplification and readability assessment systems, can be used not only for improving text accessibility, but also for educational purposes: the author can prepare texts that are adequate according to the level of the reader and it will also allow them to improve their reading skills.

References

Sandra M. Aluísio, Lucia Specia, Thiago A. S. Pardo, Erick G. Maziero, Renata P. M. Fortes (2008). Towards Brazilian Portuguese Automatic Text Simplification Systems. In the *Proceedings of the 8th ACM Symposium on Document Engineering*, pp. 240-248.

Eckhard Bick (2000). *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework.* PhD Thesis. University of Århus, Denmark.

Jill Burstein, Martin Chodorow and Claudia Leacock (2003). CriterionSM Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. In the *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence*, Acapulco, Mexico.

Arnaldo Candido Jr., Erick Maziero, Caroline Gasperin, Thiago A. S. Pardo, Lucia Specia, and Sandra M. Aluisio (2009). Supporting the Adaptation of Texts for Poor Literacy Readers: a Text Simplification Editor for Brazilian Portuguese**.** In *NAACL-HLT Workshop on Innovative Use of NLP for Building Educational Applications*, pages 34–42, Boulder'.

Helena de M. Caseli, Tiago de F. Pereira, Lúcia Specia, Thiago A. S. Pardo, Caroline Gasperin and Sandra Maria Aluísio (2009). Building a Brazilian Portuguese Parallel Corpus of Original and Simplified Texts. In the *Proceedings of CICLing*.

Max Coltheart (1981). The MRC psycholinguistic database. In *Quartely Jounal of Experimental Psychology*, 33A, pages 497-505.

Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer e Richard Harshman (1990). Indexing By Latent Semantic Analysis. In *Journal of the American Society For Information Science*, V. 41, pages 391-407.

Bento C. Dias-da-Silva and Helio R. Moraes (2003). A construção de um thesaurus eletrônico para o português do Brasil. In *ALFA- Revista de Lingüística*, V.

47, N. 2, pages 101-115.

Bento C Dias-da-Silva, Ariani Di Felippo and Maria das Graças V. Nunes (2008). The automatic mapping of Princeton WordNet lexical conceptual relations onto the Brazilian Portuguese WordNet database. In *Proceedings of the 6th LREC*, Marrakech, Morocco.

William H. DuBay (2004). *The principles of readability*. Costa Mesa, CA: Impact Information: http://www.impact-information.com/impactinfo/readability02.pdf

Christiane Fellbaum (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Lijun Feng, Noémie Elhadad and Matt Huenerfauth (2009). Cognitively Motivated Features for Readability Assessment. *In the Proceedings of EACL 2009*, pages 229-237.

Ingo Glöckner, Sven Hartrumpf, Hermann Helbig, Johannes Leveling and Rainer Osswald (2006b). An architecture for rating and controlling text readability. In *Proceedings of KONVENS 2006*, pages 32-35. Konstanz, Germany.

Arthur C. Graesser, Danielle S. McNamara, Max M. Louwerse and Zhiqiang Cai (2004). Coh-Metrix: Analysis of text on cohesion and language. In *Behavioral Research Methods, Instruments, and Computers*, V. 36, pages 193-202.

Ronald K. Hambleton, H. Swaminathan and H. Jane Rogers (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Press.

Michael Heilman, Kevyn Collins-Thompson, Jamie Callan and Max Eskenazi (2007). Combining lexical and grammatical features to improve readability measures for first and second language texts. In the *Proceedings of NAACL HLT 2007*, pages 460-467.

Michael Heilman, Kevyn Collins-Thompson and Maxine Eskenazi (2008). An Analysis of Statistical Models and Features for Reading Difficulty Prediction. *In Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications*, pages 71-79.

INAF (2009). Instituto P. Montenegro and Ação Educativa. *INAF Brasil - Indicador de Alfabetismo Funcional - 2009*. Available online at http://www. ibope.com.br/ipm/relatorios/relatorio_inaf_2009.pdf

Teresa B. F. Martins, Claudete M. Ghiraldelo, Maria das Graças V. Nunes e Osvaldo N. de Oliveira Jr. (1996). *Readability formulas applied to textbooks in brazilian portuguese*. ICMC Technical Report, N. 28, 11p.

Aurélien Max (2006). Writing for Language-impaired Readers. In *Proceedings of CICLing*, pages 567-570.

Danielle McNamara, Max Louwerse, and Art Graesser, 2002. Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Grant proposal. http://cohmetrix.memphis.edu/cohmetrixpr/publications.html

Eleni Miltsakaki and Audrey Troutt (2007). Read-X: Automatic Evaluation of Reading Difficulty of Web Text. In the *Proceedings of E-Learn 2007*, Quebec, Canada.

Eleni Miltsakaki and Audrey Troutt (2008). Real Time Web Text Classification and Analysis of Reading Difficulty. In the *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications*, Columbus, OH.

Cláudia Oliveira, Maria C. Freitas, Violeta Quental, Cícero N. dos Santos, Renato P. L. and Lucas Souza (2006). A Set of NP-extraction rules for Portuguese: defining and learning. In *7th Workshop on Computational Processing of Written and Spoken Portuguese*, Itatiaia, Brazil.

Sarah E. Petersen and Mari Ostendorf (2009). *A machine learning approach to reading level assessment*. Computer Speech and Language 23, 89-106.

Emily Pitler and Ani Nenkova (2008). Revisiting readability: A unified framework for predicting text quality. In *Proceedings of EMNLP, 2008*.

Adwait Ratnaparkhi (1996). A Maximum Entropy Part-of-Speech Tagger. In *Proceedings of the First Empirical Methods in Natural Language Processing* Conference, pages133-142.

Brian Roark, Margaret Mitchell and Kristy Hollingshead (2007). Syntactic complexity measures for detecting mild cognitive impairment. In the *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, Prague, Czech Republic.

Caroline E. Scarton, Daniel M. Almeida, Sandra M. Aluísio (2009). Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. In *Proceedings of STIL-2009*, São Carlos, Brazil.

Sarah E. Schwarm and Mari Ostendorf (2005). Reading Level Assessment Using Support Vector Machines and Statistical Language Models. In the *Proceedings of the 43rd Annual Meeting of the ACL*, pp 523–530.

Kathleen M. Sheehan, Irene Kostin and Yoko Futagi (2007). Reading Level Assessment for Literary and Expository Texts. In D. S. McNamara and J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society*, page 1853. Austin, TX: Cognitive Science Society.

Advaith Siddharthan (2003). *Syntactic Simplification and Text Cohesion*. PhD Thesis. University of Cambridge.

Andreas Stolcke. *SRILM -- an extensible language modeling toolkit*. In Proceedings of the International Conference on Spoken Language Processing, 2002.