Revista de Sistemas de Informação da FSMA
n. 18 (2017) pp. 36-42

SISTEMAS DE INFORMAÇÃO
http://www.fsma.edu.br/si/sistemas.html

TRILHA PRINCIPAL

Pylinguistics: an open source library for readability assessment of texts written in Portuguese.

Suya Castilhos, *BSc in Computer Science, UFRGS*,
Vinicius Woloszyn, *PhD Student in Computer Science, PPGC-UFRGS*,
Dante Barone, *Full Professor, PPGC-UFRGS*,
Leandro Krug Wives, *Associate Professor, PPGC-UFRGS*

***Abstract*—Readability assessment is an important task in automatic text simplification that aims identify the text complexity by computing a set of metrics. In this paper, we present the development and assessment of an open source library called *Pylinguistics* to readability assessment of texts written in Portuguese. Additionally, to illustrate the possibilities of our tool, this work also presents an empirical analysis of readability of Brazilian scientific news dissemination.**

***Index Terms*—text simplification, readability, e-accessibility**

I. Introduction

Automatic text simplification is a Natural Language Processing (NLP) task that reduces text syntactic and lexical complexity while preserving, in essence, the original content. The simplified version of the text becomes easier to read and to understand than the original one. This process can be considered a digital inclusion initiative that promotes information access to people with cognitive disabilities (e.g. aphasia and dyslexia) or hearing-impaired people who communicate with each other using sign languages like LIBRAS (Brazilian Sign Language), since the structural differences between LIBRAS and Portuguese make it difficult to understand complex texts [1]. Additionally, text simplification can help people of poor literacy improve their reading skills, including children learning to read different genres of text, second language learners, adults being alphabetized and students undertaking Distance Education, in which text intelligibility is of major importance [2].

The comprehension difficulty of a text does not depend only on the linguistic aspects, but also on the reader's reading skills. Some people might easily understand different kinds of text, from scientific papers to intricate novels, while others may find it difficult to read newspaper reports, being necessary an adaptation of the text to their particular personal characteristics.

Considering the text aspects, not all texts are of the same genre, and they differ in degrees of complexity. For instance, the Scientific Journalistic genre should exhibit some typical properties, such as relative abstractness, technicality, and informational density while the Journalistic genre destined to general public text should present a higher incidence of nouns and verbs that would decrease the comprehension difficulty of a document [3].

Related to the automatic simplification problem is the issue of measuring textual readability with the goal of developing metrics that can associate a readability score to texts. Unfortunately, to the best of our knowledge, the current readability tools are private or are only available via limited web interfaces that do not allow automatic processing of large amounts of text [1]. This noticeable lack of *public*, functional and of easy use (in terms of automatization and integration) linguistic tools makes the study of automatic text simplification difficult.

Our research is mainly concerned with the development of an *open source* library[2] for natural language analysis for texts in Portuguese to serve as a measuring instrument and basis to new studies on the readability area. Additionally, we illustrate a use case of the proposed library, this work also provides an empirical analysis on readability aspects of Brazilian scientific journalism and a comparison with general public journalism, providing an insight on textual characteristics that could make the scientific work more accessible to the general public.

The rest of this paper is structured as follows. In Section II presents some works found in the literature that are related to the context of our research and which comprises the state of the art in terms of Automatic Text Assesment. Section III presents the set of metrics chosen and a description of how they were implemented. In Section IV we present our approach. Section V presents a performance comparison with another similar tool as a mean to validate our tool. Chapter VI contextualizes a real world problem and illustrates the use of our tool by creating a model for automatically distinguishing journalistic genres. Lastly, on Chapter **??**, we conclude discussing lessons learned in the process as well as a reflection on the contributions and

[1] In fact, there is a very interesting and relevant Brazilian project known as Coh-Metrix-Dementia (http://www.nilc.icmc.usp.br/coh-metrix-port/), which also implements some metrics found in the literature, but until the current date it appears not to be totally functional.

[2] https://github.com/vwoloszyn/pylinguistics

possibilities of further work.

II. Related work

Many works have addressed the problem of quantifying text readability. Early works, focused on the Flesch-Kincaid and Flesch Reading-Ease tests [4], which rely on word frequency and word and sentence length to quantify textual complexity [5].

Another work on the topic, the *Lexile Framework* [6] provides ways to measure student's ability and text difficulty on the same scale, called Lexile. The *Lexile measure* is a quantitative representation of the reader's ability or the text difficulty. The *Lexile scale* is a developmental scale for reading, ranging from 200L, in the case of beginner readers, to above 1700L, for advanced texts. When the reader and the text are appropriately matched, a reader can enjoy a comprehension rate of about 75%.

Nevertheless, these properties reveal superficial readability aspects due to their incapacity in capturing other linguistic elements that would aid the reader to connect mentally ideas on the text [7].

More recently, however, most works focus on different linguistic layers, such as lexical, syntactic, discursive, and conceptual representation to provide a measure of complexity. One example is *Coh-Metrix* [8], [9], which is a system for computing the cohesion and coherence of English texts. All texts must have some formal aspects that establish a relationship with each sentence, such as cohesion and coherence that give meaning to the text. Textual cohesion is the relationship, i.e., the connection, between the words, phrases, or sentences of text. Coherence is related to understanding, i.e., the interpretation of what is said or written. *Coh-Metrix 2.0* provides a set of 60 metrics to constituents, word frequency, connectives, logic operators, pronouns, hyperonyms, and ambiguity.

Coh-Metrix-Port [10] is an adaptation of the original *Coh-Metrix* ported to the Brazilian Portuguese. It implements 48 of the original metrics. The performance of the metrics is intrinsically related to the performance of the part-of-speech (POS) tagging module used. By default, it uses *PALAVRAS* [11], which is a parser for Portuguese. *PALAVRAS* achieves 99% in terms of morphosyntax (word class and flexion), and 97% in terms of syntax [12].

Considering the studies above, the contribution of this work lies in readability assessment area, with the development of an open source set of metrics in the mold of *Coh-Metrix-Port*, to support the readability assessment of texts written in Portuguese.

III. Metrics

One important issue in text simplification research is the aspects that make a text more or less readable for a target user group. For instance, the PSET project [13], addressed text simplification for people with aphasia, while the PorSimples project [1] looked into simplification for people with poor literacy rate. Finally, Finatto et al. [3] showed the readability aspects of journalistic texts by comparing newspapers geared to two different target audiences. The development of the metrics used to quantify complexity aspects of text in our work was guided by previous works on readability [3], [6], [8]–[10]. Currently, there are 38 metrics already operative, as follow.

A. Descriptive

Descriptive metrics provide basic information about the text, such as the number of sentences and words. They help checking the output to make sure that the information makes sense, as well as being important basic values for computing more complex metrics. In this work we used 16 descriptive metrics.

- **Word, sentence and syllable count.** The total number of words, sentences and syllables in a text.
- **Words per sentence.** Some statistical data of the size of sentences in words. We calculate the mean, the median, four percentiles (25, 50, 75 and 90) and also the percentage of sentences above 30 words long.
  The average (mean) sentence length is a classical feature in readability, the work of François [14] explored the use of the 90th percentile sentence and inspired the computing of the other percentiles and the median. Daoust [15] also inspired the developing of calculating the percentage of sentences above size 30, with the use of such metric in his work for the French language.
- **Syllables per word.** Statistical data of the length of words in syllables. We compute the mean, median and four percentiles (25, 50, 75 and 90). Flesch (1948) found that the mean number of syllables per word had a correlation of .66 with comprehension difficulty. In fact, mean word length in syllables combined with mean sentence length are classical readability metrics that make the core principles of the Flesch-Kindcaid and the Lexile readability scores, remaining some of the most important metrics in measuring textual readability.
  The median and percentiles were added by us inspired by the data calculated for sentence size, with the goal of providing extra metrics and possible insights in text analysis.

B. Word information

Metrics of word information are based on the concept that to each word is assigned a syntactic part-of-speech category, these categories are separated in content words (adjectives, nouns, verbs and adverbs) and function words (determiners, adpositions[3], pronouns and conjunctions).

Some words can perform multiple syntactic roles. For example, the word "andar" can be a noun ("Seu andar era rápido") or a verb ("Vou andar hoje"). Through the use of the parser of our choice, *NLPNET* [16], which achieves around 97% of accuracy when compared to other state-of-the-art taggers for the Portuguese language [17]. A single

[3]adpositions is a cover term for prepositions and postpositions

part-of-speech category is attributed to each word of the text based on its syntactic context. In this work we used 7 descriptive metrics.

- **Incidences of part-of-speech elements.** We calculate the incidence of word categories (adjectives, nouns, verbs, adverbs, pronouns) per 1000 words in the text, and also the incidence of content and function words per 1000 words. These metrics are reflective of elements of a text that are likely to support a reader's construction of a coherent situation model [8].

C. Diversity

Metrics of diversity provide information on the variety of words in the text, a high word diversity means many unique words need to be decoded and integrated with the discourse context, which should make comprehension more difficult. In contrast, if some words are being repeated often in a text, it tends to increase cohesion, and thus, make for a more readable text.

Word diversity is a common metric on the measuring of readability, with works as early as Lorge (1948), who found a correlation between difficulties of passages and the mean frequency of the words in such passages. Both diversity metrics implemented in this work were inspired by Coh-Metrix diversity metrics.

- **Lexical diversity.** This metric measures how varied is the total vocabulary of a text. It's defined by the ratio of unique words that appear in the text in comparison to the total number of words in the text.
- **Content diversity.** This metric is similar to Lexical diversity, but it only takes in consideration content words (adjectives, nouns, verbs and adverbs). It's defined by the ratio of unique content words that appear in the text in comparison to the total number of content words in the text.

D. Connectives

The traditional unit for analyzing grammatical complexity has been the sentence, defined by a staring capital letter and ending on a punctuation mark. However, Coleman (1962) found evidence that independent clauses might be a more valid unit of analysis. Since sentences with connectives such as "Ele acordou e ele foi a aula." (He woke up and he went to school) may actually be treated as if it were two separate syntactic units. The conjunction "e" (and) serves roughly the same function as a punctuation mark. Hence, although the presence of connectives creates longer sentences, their use might decrease the difficulty of understanding of a text by creating cohesive links between ideas and providing clues about text organization [8] (Cain Nash, 2011; Crismore, Markkanen, Steffensen, 1993; Longo, 1994; Sanders Noordman, 2000; van de Kopple, 1985). In this work we computed 11 connectives metrics.

- **Incidence of connectives.** The connectives are divided into five general classes (Halliday Hasan, 1976; Louwerse, 2001), additive, logic, temporal, causal and negative. *Pylinguistics* calculates the incidence of the total number of connectives, as well as the incidence of each separate category. For this, we created a dictionary of connectives in the Portuguese language based on the correspondent connective categories existing in English, as well as some additions based on material found on the web.
- **Logic operators.** Within the logical connectives, there are some specific logical particles "e" (and), "ou" (or), "se" (if), and "não", "nem", "nenhum", etc... (not, neither, none, etc...). Measuring such logical particles and how they relate with cohesion and readability in a text was explored by Coh-Metrix-Port [10], and inspired 5 metrics of *Pylinguistics*, one for each group and one metric for the sum of all groups.

E. Readability

There are many traditional methods for assessing text difficulty. Klare stated that more than 40 readability formulas have been developed over the years (Klare, 1974-1975). The most common, however, are the Flesch-Kincaid formulas. Since the Flesch reading ease score has a validated Portuguese adaptation [5], it was our metric of choice for the readability measure in the Portuguese version of *Pylinguistics*.

- **Flesch reading ease, Portuguese version.** The Flesch reading ease adaptation for Brazilian Portuguese developed by Martins [5], basically consists on a shift of 42 points on the result of the original Flesch reading ease formula to compensate for the fact that words in Portuguese typically have a bigger number of syllables. It is as follows

  $$FleschPT = 248.835 - (1.015 \times ASL) - (84.6 \times ASW)$$

  Being ASL the average sentence length in words, and ASW the average number of syllables per word.

F. Ratios

In contrast to incidence metrics, which compute the number of a specific word type in a span of 1000 words. Ratios are a more relative measure, comparing the incidence of a certain class of units to the incidence of another class of units. Bormuth (1966) demonstrated that the ratios of some part-of-speech elements of the text can be a good predictor of the text's style and genre. At the current state of development, *Pylinguistics* computes a single ratio metric.

- **Ratio of pronouns on prepositions.** Inspired on the model of François [14] for readability on the French language, we implemented a metric of pronouns on prepositions as a measure of syntactic complexity of sentences. This specific ratio was found to be related to readability of texts in the French language, what indicates that it could be an interesting metric to compute for Portuguese.

IV. Pylinguistics

The language chosen for the development of *Pylinguistics* was *Python*. *Python* is a high-level general purpose programming language, which supports many programming paradigms, such as object oriented, procedural and functional programming. It is a popular language that is employed in a wide variety of contexts. Particularly in the area of Natural Language Processing, *Python* is a common choice due to many factors:

- **Dynamically typed** The type definition of the data-structures emerge as we code.
- **Comprehensive standard library** Rich built-in support of data-structures.
- **Large amount of open-source code available for use** Many open-source text processing libraries that are easy to integrate and provide many resources for classification, tokenization, stemming, tagging and parsing for both English and Portuguese. Such as Google's Natural Language Toolkit, *NLTK*.
- **High level language** Expressive and succinct, the expressions are intuitive and make the final code easy to read.
- **Scripting language** *Python* code can run in any environment in which there is a *Python* interpreter, making it significantly more portable and easier to use than other programming languages.

All these characteristics make *Python* an ideal choice to easily prototype, test and develop code in the area of *NLP*. Among its few disadvantages there is the fact that it has a slow speed of execution, which can be a problem for some applications, but is a completely manageable issue in the context of text analysis.

A. Natural Language Parsing

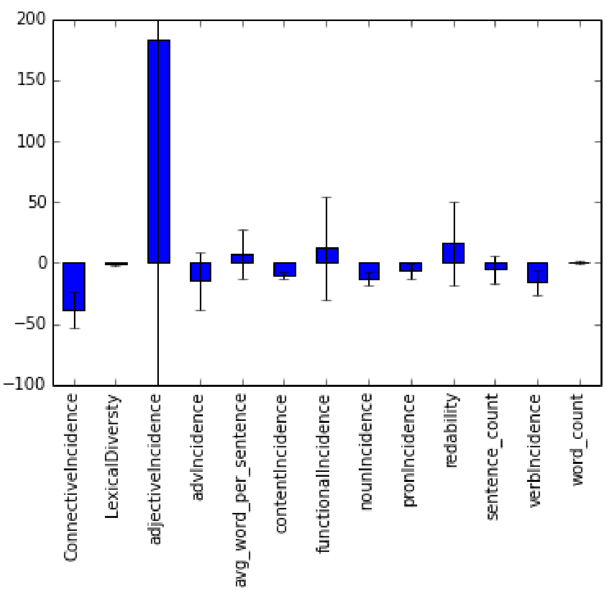
A Natural Language Parser is a program which breaks down a string of natural language text into small part-of-speech components relative to the form, function, and syntactic relationship of each part. For the Portuguese version of *Pylinguistics*, we have decided on *nlpnet* [17], which is a Python library for Natural Language Processing tasks based on neural networks and specially tailored for working with the Portuguese language. *nlpnet* contains a state-of-the-art parser, performing 97.33% token accuracy on part-of-speech tagging in Portuguese.

V. Evaluation

In order to provide an assessment of our tool, we performed a comparative evaluation with the similar tool *Coh-Metrix-Port* [10], which provides a web interface where texts could be individually uploaded and the metrics that result from this text analysis could be downloaded. However, not all of the features implemented by us in *Pylinguistics* were also available in *Coh-Metrix-Port*, therefore it was just possible to compare a subset of 13 out of our 38 features, those 13 metrics being implemented in both tools. For the comparison, we randomly selected a set of 20 articles from the corpus of Fapesp, a magazine in Brazilian Portuguese focused on scientific production.

Figure 1 represents the results of the comparison for each metric, with the bar being the average difference in results, positive if greater, negative if smaller than the results of the tool *Coh-Metrix-Port*, and the lines represent the standard deviation of the respective values. The results show a similar performance for nearly all metrics, all of them with a difference smaller than 50%, with the exception of the Adjective Incidence metric, in which our tool was classifying a much larger number of words as adjectives than our comparison tool.

Fig. 1. Box plot of the comparison of metrics with *Coh-Metrix-Port*, bars represent mean difference in the result of *Pylinguistics* compared with the results of *Coh-Metrix-Port*, lines represent the respective standard deviation



In order to better understand the disparity in terms of adjective incidence, we performed an exhaustive manual count of all adjectives on the 20 texts of our sample. Afterwards, we calculated the difference from the expected results to the results yield by each tool, as well as the mean and standard deviation of such differences, both shown in Table I. This analysis lead to the conclusion that *Pylinguistics* adjective labeling comes closer to the number expected from the manual count than *Coh-Metrix-Port*, as well as having a smaller variation.

Since labeling words with their respective part-of-speech tag is a parser task, this difference is most likely due to the difference in terms of the parser used by both tools. For *Pylinguistics* we used the NLP parser, Coh-Metrix-Port 2.0 claims to use *PALAVRAS* [11]. Taking in consideration the *PALAVRAS* parser yields correctness rates of over 99% for part-of-speech tagging, we theorize that either Coh-Metrix-Port 2.0 uses a different parser or there is some bug concerning the adjective count in its code. Especially since other part-of-speech parser tasks such as number of nouns and verbs is well within the expected value.

TABLE I
DIFFERENCE IN NUMBER OF ADJECTIVES FOUND

Tool	Mean	Standard Deviation
Pylinguistics	**8.45%**	**8.04**
Coh-Metrix-Port	26.56%	19.59

The set of 13 metrics analyzed in this comparative assessment is rather diverse, comprising of some metrics relative to the performance of our parser, some related to counts implemented by us, and also the readability metric, which is the Portuguese adaptation of the Flesch reading ease formula. Since several of the non-tested metrics are relative to the parser's performance, and simple mathematical variations of them (mean, median and percentiles), and the tested metrics showed a satisfactory result in comparison with a state-of-the-art tool, we consider that the results of the comparative assessment were enough for validation of *Pylinguistics*.

VI. Case Study

TIn order to illustrate the use of our tool, we contextualize a real world problem, the complexity and intelligibility of the scientific journalism.

There are still few computational linguistic studies devoted to observe their textual constitution with particular emphasis on the characterization of stylistics elements of this textual genre. A thorough scientific journalism description can be extremely important for many of the core problems that computational linguists are concerned with. For example, parsing accuracy could be increased by taking genre into account, for instance, certain object-less constructions occur only in recipes in English. Similarly for POS-tagging, where the frequency of uses of *trend* as a verb in the *Journal of Commerce* is 35 times higher than in *Sociological Abstracts*. In information retrieval, genre classification could enable users to sort search results according to their immediate interests, for example scholarly articles about supercollider, novels about the French Revolution, and so forth.

A. Methodology

Crossley et al. [18] present a text readability analysis using the Coh-Metrix tool by comparing one complex and one simple corpus. Similarly, to illustrate one of the utilities of the proposed tool, we use *Pylinguistics* to compare scientific news dissemination with simple texts to highlight readability features that would make the scientific work more accessible to the general public.

TABLE II
DESCRIPTION OF THE CORPORA USED IN THIS STUDY

Corpus	Articles	Words	Words per article
a) FAPESP	3,866	6,266,831	1,621.01
b) FSP	3,808	1,330,335	349.3

The two corpora used in this study are geared towards different groups. Thus, they employ different vocabularies and textual structures that can be classified into different levels of complexity. In this study, we compared two corpora: *Pesquisa Fapesp*[4] a Brazilian specialized science magazine; and Folha de São Paulo[5] (FSP) a Brazilian newspaper aimed at the general public.

The FAPESP corpus is composed of articles obtained from the magazine *Pesquisa Fapesp* that has its primary focus on the national (Brazilian) scientific production. It is composed of 3880 articles from 237 editions collected along 19 years. The FSP is a set of 3808 journalistic articles published from 1994 to 1995 by Folha de São Paulo. Table II presents some descriptive statistics on the corpora used in this work.

TABLE III
READABILITY METRICS COMPUTED BY *Pylinguistics* WHERE $p < 0.05$

Metrics	FAPESP		FSP	
	mean	**std**	**mean**	**std**
1. Avg word per sentence	28.1	5.2	20.01	7.8
2. Syllable count	3638.2	2107.03	712.7	594.8
3. Avg syllables per word	2.2	0.09	2.1	0.1
4. Adjective incidence	78.2	12.9	64.02	22.9
5. Noun incidence	330.3	31.7	349.8	54.6
6. Verb incidence	111.7	19.9	122.2	32.9
7. Adv incidence	30.1	11.3	30.3	18.08
8. Pron incidence	47.9	13.6	43.2	23.6
9. Content incidence	549.3	20.1	565.3	37.3
10. Functional incidence	380.4	17.8	363.3	34.7
11. Lexical diversity	0.4	0.06	0.6	0.1
12. Content diversity	0.6	0.07	0.7	0.09
13. Connective incidence	49.2	9.1	43.9	17.5
14. Flesch Readability	27.24	13.97	49.99	16.38

B. Results

The results are structured in two parts. We present an analysis of the text features found in the scientific journalism genre and afterwards discuss our model for text classification.

1) Feature Analysis: To illustrate the use of our tool we computed the textual features of the cited corpora using the 38 metrics provided by *Pylinguistics*. With the mean and standard deviation of the features, we used Student's t-test to find which metrics have the most different values between the two genres. Table III shows the mean and standard deviation (std) of the 14 metrics that performed best as predictors of genre ($p < 0.05$). Among these metrics, we chose *Average of Word per Sentence* and *Average of syllables per word* to be discussed in this Section.

- *Average syllables per word.* The presence of long words in a text can increase its complexity and hinder user comprehension [10], [19]. Among the analysed corpora, FAPESP is the most complex having 2.2 syllables per word in average, while FSP has 2.1. The FAPESP complexity can be reduced by performing a lexical simplification [20]. This process aims to replace long terms with terms of lesser complexity without loss of meaning, making the text easier to understand.

[4] http://revistapesquisa.fapesp.br/
[5] http://www.folha.com.br/

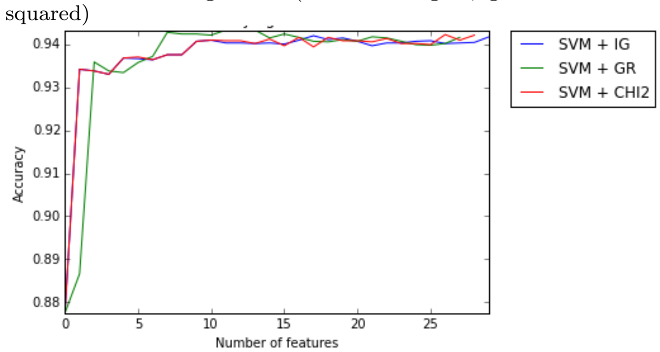
- *Average words per sentence.* Ideally, each sentence should only contain one idea. Easy texts include fairly short sentences, avoiding subordinate clauses whenever possible [19]. However, FAPESP presents a high average of words per sentence (28.1) that makes the user comprehension more difficult than texts from Folha de São Paulo (20.01). The FAPESP corpus complexity can be reduced making the text more accessible to the general public through a sentence reduction process [19]. This task consists of splitting sentences into two or more pieces, or even removing elements unnecessary to the understanding of the idea, to reduce the complexity.

For example, the long sentence "*Com o objetivo de agilizar e tornar mais eficiente o processo de importação de itens concedidos pela FAPESP*" (With the goal of speeding and making more efficient the process of importing items conceded by FAPESP), found in an article of the FAPESP magazine can be replaced with negligible loss of information by "*Para agilizar e tornar mais eficiente a importação de itens dados pela FAPESP*" (To speed and make more eficient the importing of items given by FAPESP), which has less syllables per word and less words per sentence, scoring better in readability and being presumptively more accessible to the general public. Further simplification could still be made, but at a cost of losing information and specificity.

2) Model Analysis: Once the predictive power of each metric is calculated, the last step is to select the most informative subset of features and combine them into a model for text classification. Several ways of combining the best possible subset of predictors are possible. We decided on using a support vector machine (SVM) since previous works have shown that it yields better results for similar models [14].

We then assessed the predictive power of our features based on 3 feature selection algorithms commonly used for text categorization: Information Gain, Gain Ratio and Chi-square [21]. Figure 2 shows the performance of the SVM when varying the number of features selected for each method. It shows that with only 2 features we can already predict the genre with over 93% accuracy. Additionaly by 7 metrics it already reaches the best possible result (97%). Finding a small subset of predictors is important to avoid over-fitting.

Fig. 2. Performance of the SVM varying the number of features selected for each algorithm (information gain, gain ratio and chi squared)



VII. Conclusion and future work

Previous works already have reported the use of linguistic metrics to provide comparison and understanding of the adequacy of text to a target audience [4], [5], [8]–[10]. However, existing tools are private or just available via limited web interfaces that do not allow the processing of large amounts of text [6], [8], [9]. Thus, *Pylinguistics* provides a set of open-source metrics that can be employed in a large volume of data. This opens up possibilities to a wide range of different analysis and visualization of linguistic aspects, it can help teachers pick texts that better suit their students, as well as helping people with reading handicaps enjoy a more comprehensive understanding of texts.

We also highlight the future development of additional metrics, such as ambiguity, foreign terms, and infrequent terms. The use of these metrics can provide a better understanding of different readability aspects of the text. We also intend on developing an application program interface (API) to make the tool easily available to the general public, which we believe will open several new possibilities of use.

References

[1] S. M. Aluísio, L. Specia, T. A. Pardo, E. G. Maziero, and R. P. Fortes, "Towards Brazilian Portuguese automatic text simplification systems," in *ACM symposium on Document engineering*, 2008, pp. 240–248.

[2] J. Hyönä and R. K. Olson, "Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 21, no. 6, p. 1430, 1995.

[3] M. J. B. Finatto, C. E. Scarton, A. Rocha, and S. Aluísio, "Características do jornalismo popular: Avaliação da inteligibilidade e auxílio à descrição do gênero," in *Brazilian Symposium in Information and Human Language Technology*, 2011.

[4] R. Flesch, "A new readability yardstick.," *Journal of applied psychology*, vol. 32, no. 3, p. 221, 1948.

[5] T. B. Martins, C. M. Ghiraldelo, M. d. G. V. Nunes, and O. N. de Oliveira Junior, *Readability formulas applied to textbooks in Brazilian Portuguese.* Icmsc-Usp, 1996.

[6] C. Lennon and H. Burdick, "The lexile framework as an approach for reading measurement and success," *Electronic publication on www. lexile. com*, 2004.

[7] W. H. Dubay, *The principles of readability a brief introduction to readability research*, 2004.

[8] A. C. Graesser, D. S. McNamara, M. M. Louwerse, and Z. Cai, "Coh-metrix: Analysis of text on cohesion and language," *Behavior research methods, instruments, & computers*, vol. 36, no. 2, pp. 193–202, 2004.

[9] S. A. Crossley and D. S. McNamara, "Assessing l2 reading texts at the intermediate level: An approximate replication of crossley, louwerse, mccarthy & mcnamara (2007)," *Language Teaching*, vol. 41, no. 03, pp. 409–429, 2008.

[10] C. E. Scarton and S. M. Aluísio, "Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: Adaptando as métricas do coh-metrix para o português," *Linguamática*, vol. 2, no. 1, pp. 45–61, 2010.

[11] E. Bick, *The parsing system" Palavras": Automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus Universitetsforlag, 2000.

[12] ——, "Gramática constritiva na análise automática de sintaxe portuguesa," *A Língua Portuguesa no Computador. Campinas: Mercado de Letras, São Paulo: FAPESP. ISBN*, pp. 85–7591, 2005.

[13] J. Carroll, G. Minnen, Y. Canning, S. Devlin, and J. Tait, "Practical simplification of english newspaper text to assist aphasic readers," in *AAAI Workshop on Integrating Artif Intel and Assistive Tech*, 1998, pp. 7–10.

[14] T. François and C. Fairon, "An ai readability formula for french as a foreign language," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, 2012, pp. 466–477.

[15] F. Daoust, L. Laroche, and L. Ouellet, "Sato-calibrage: Présentation d'un outil d'assistance au choix et à la rédaction de textes pour l'enseignement," *Revue québécoise de linguistique*, vol. 25, no. 1, pp. 205–234, 1996.

[16] E. R. Fonseca and S. M. Aluísio, "A deep architecture for non-projective dependency parsing," in *Proceedings of NAACL-HLT*, 2015, pp. 56–61.

[17] E. R. Fonseca and J. L. G. Rosa, "Mac-morpho revisited: Towards robust part-of-speech tagging," in *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, 2013, pp. 98–107.

[18] S. A. Crossley, M. M. Louwerse, P. M. McCarthy, and D. S. McNamara, "A linguistic analysis of simplified and authentic texts," *The Modern Language Journal*, vol. 91, no. 1, pp. 15–30, 2007.

[19] S. Bott and H. Saggion, "Text simplification resources for spanish," *Language Resources and Evaluation*, vol. 48, no. 1, pp. 93–120, 2014.

[20] B. Drndarevic and H. Saggion, "Reducing text complexity through automatic lexical simplification: An empirical study for spanish," *Procesamiento del lenguaje natural*, vol. 49, pp. 13–20, 2012.

[21] F. Debole and F. Sebastiani, "Supervised term weighting for automated text categorization," in *Text mining and its applications*, Springer, 2004, pp. 81–97.