

O USO DE SOFTWARE PARA MEDIR A COMPLEXIDADE DO TEXTO

Adriana Riess Karnal
(PUC-RS, UNISINOS)
Vera Wanmacher Pereira
(PUC-RS)

RESUMO: Este artigo apresenta a pesquisa sobre leiturabilidade/inteligibilidade e o uso de software disponível on-line para medir a complexidade dos textos. Esta é uma área efervescente em língua inglesa e que agora começa a ganhar espaço em língua portuguesa. Descreve-se o software chamado ATOS, o mais utilizado nas escolas americanas, e a adaptação do Coh-Metrix para o Português brasileiro.

PALAVRAS-CHAVE: readability, medidores de leitura, ensino de leitura.

ABSTRACT: This article presents the research about readability/intelligibility and the use of software available online to measure text complexity. This is an important area in the English language and it is now growing in Portuguese. We describe a software named ATOS, the most used in American schools, and the adaptation of Coh-Metrix to Brazilian Portuguese.

KEYWORDS: readability, reading measurers, teaching reading.

Introdução

Uma das questões discutidas na escola em relação à leitura é a escolha adequada dos livros de acordo com a série escolar. Esta questão está ligada à compreensão leitora que os professores esperam em cada uma das séries. Nesse sentido, os professores acabam escolhendo os livros de acordo com sua experiência ou *expertise*. No entanto, já existe uma pesquisa na área de leiturabilidade (ou inteligibilidade) que visa desenvolver *softwares*

que medem a complexidade dos textos e podem indicar para qual série escolar o livro é indicado (SCARTON; ALUÍSIO, 2010).

Este artigo tem o objetivo de apresentar a pesquisa sobre leiturabilidade e compreensão leitora e as ferramentas disponíveis *on-line* que podem ser acessadas para medir a complexidade dos textos. Embora esta seja uma área mais desenvolvida nos Estados Unidos e a maioria dos programas estejam em inglês, ela agora começa a ser desenvolvida em língua portuguesa. Em especial, professores de línguas estrangeiras são beneficiados com os “medidores” *on-line*. É possível que em poucos anos estas ferramentas estejam disponíveis para a sala de aula regular no Brasil.

O artigo está organizado em quatro seções. A primeira delas descreve os conceitos de leiturabilidade ou inteligibilidade e compreensão leitora (PERFETTI,1999, 2007), bem como a legibilidade. A segunda e terceira seções abordam o estado da arte desta pesquisa nos Estados Unidos e como a língua portuguesa está incluída nos *softwares* já disponíveis .A última seção, por fim, apresenta tais *softwares* de acesso *on-line* (PERFETTI et al., 2012; SKENAZI, 2012) que podem ser utilizados pelos professores gratuitamente.

1. Conceitos importantes

Quando o objetivo é medir o nível de complexidade de um texto, um conceito importante é o de *leiturabilidade*, tradução literal do inglês para *readability*. No português o termo também adotado é *inteligibilidade*, segundo a pesquisadora Sandra Aluísio, da área de linguística computacional, que adaptou o programa

Coh-Metrix no país. Há mais de 50 anos, Dale e Chall (1948) utilizaram o termo *readability* para expressar o conceito de facilidade na leitura, aquilo que permite a leitura. Hoje, o conceito de inteligibilidade pode ser utilizado como a qualidade da linguagem que é compreensível.

Autores como Perfetti (1999, 2007) e Caldwell (2008) e nas línguas estrangeiras Koda (2005) e Grabe (2009) utilizam o termo compreensão leitora, por se tratar de uma ideia mais contemporânea segundo a qual não só o papel do texto é importante na leitura, mas sua interação com o leitor, incluindo o papel do próprio autor. A leiturabilidade envolve a função do leitor, sua competência na atividade, suas características, seus conhecimentos, seus objetivos e sua experiência. Os conhecimentos do leitor aqui envolvem tanto o nível linguístico, a sintaxe e o vocabulário, por exemplo, quanto seu conhecimento de mundo.

Outro termo que pode ser confundido nesta área é a *legibilidade*. Isto porque ele também determina a facilidade do texto, no entanto, é mais utilizado pela área de tipografia. Esta palavra lida com a estética, isto é, o tipo de fonte, a cor e contraste da letra com o fundo, até o espaçamento e margens.

O estudo em leiturabilidade é importante, pois um dos objetivos é simplificar os textos para serem entendidos por um número maior de leitores. Este é o caso dos disléxicos ou semi-letrados, e nas línguas estrangeiras é feita uma adaptação do texto e sua classificação por níveis, como em básico, intermediário e avançado.

2. A pesquisa nos Estados Unidos

Nos Estados Unidos, a leitura e a possibilidade de ser medida é investigada há muitos anos e muitas fórmulas surgiram para calcular um valor aproximado do nível de dificuldade. Os cálculos são feitos através de características superficiais do texto, como o número de palavras em sentenças e o número de letras ou sílabas por palavra. Alguns “medidores” são mais complexos e podem medir a qualidade do vocabulário, como é o caso do *Person Reading Maturity Metric*. Na quarta seção deste trabalho, estes softwares serão apresentados.

Um dos objetivos principais de se utilizar softwares para medir a dificuldade do texto decorre da implantação do *Common Core State Standards* na educação americana (PERFETTI et al, 2012). Estes padrões de núcleo comum são uma tentativa de padronizar os currículos escolares, fazendo com que a comunidade tenha conhecimento prévio dos conteúdos a serem aprendidos em cada série (<http://www.corestandards.org>). Assim, o currículo é similar em todas as escolas públicas dos Estados Unidos. Para os incentivadores desta padronização, esta é uma forma de garantir a qualidade na educação desde o início até a universidade.

Na área de ensino de línguas, e especificamente na literatura, a ideia de usar um software deste tipo é que ele irá adequar o nível de complexidade do livro a uma série escolar. Por exemplo, no site da *Renaissance Learning* se encontra o programa *ARBookFinder*, que é um buscador de livros. O usuário informa o título do livro que está interessado e no menu a esquerda ele indica o nível e outras obras relacionadas àquele nível. A figura 1 *Book Finder* abaixo é um exemplo do livro *Harry Potter*.

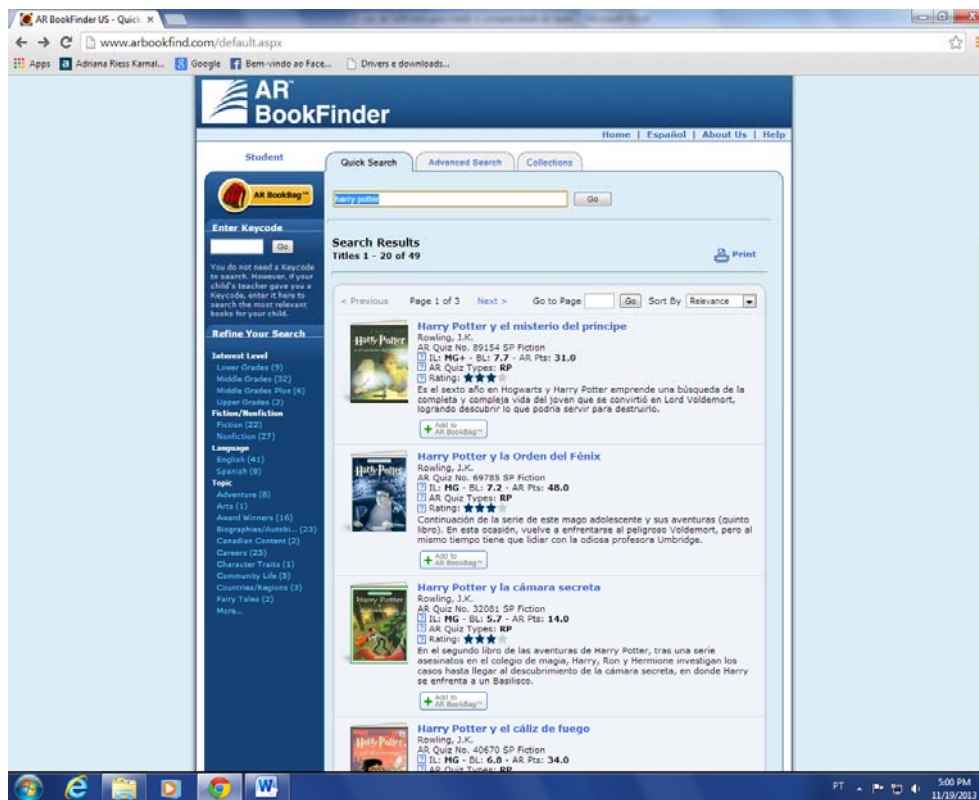


Figura 1: AR Book Finder.

Nesse sentido, o professor tem a garantia de que o livro que indica para seus alunos lerem está adequado ao nível. O usuário ainda pode refinar sua busca por língua (também é oferecido em espanhol) e assunto. Todos os livros relacionados ao assunto que estão no mercado são encontrados pelo buscador e classificados por nível.

3. A pesquisa em língua portuguesa

A pesquisa sobre os medidores de complexidade de texto no Brasil é bastante recente. Até o período no qual este artigo foi escrito, pelo menos *on-line*, tem-se conhecimento do projeto de adaptação de métricas da ferramenta *Coh-Metrix* para o português do Brasil (*Coh-Metrix-Port*), coordenado pela professora Sandra Alúcio, da USP- São Carlos. O *Coh-Metrix-Port* é uma ferramenta desenvolvida com a finalidade de capturar a coesão e a dificuldade de um texto em vários níveis (léxico, sintático, discursivo e conceitual.)

O projeto inicial do *Coh-Metrix* para o inglês é da Universidade de Memphis. Na versão adaptada para o português do Brasil, foi incluído o corpus do Banco do Português compilado por Tony Sardinha, da PUC-SP. Nesse *corpus*, encontra-se uma lista de palavras com suas respectivas frequências. Além disso, foram incorporados outros recursos para as métricas, tais como o *Wordnet.br* com as relações de hiperonímia e o *TeP 2.0-Thesaurus* eletrônico do Português do Brasil. Alúcio (2010) diz que este *Thesaurus* oferece um conjunto completo de dados com 20.000 entradas distribuídas em 6.000 verbos, 2.000 substantivos e 12.000 adjetivos. Este recurso foi necessário para identificar o grau de ambiguidade das palavras.

Outra ferramenta que analisa a complexidade do texto em Português é o REAP (*REAders-specific Practice*) *Readability Tool*. O REAP é uma plataforma de ensino de línguas desenvolvido pela Universidade Carnegie Mellon, que faz uso de documentos da internet para o aprendizado do vocabulário da L2 e da compreensão leitora. Esta ferramenta adaptou um analisador para textos em

português europeu. No entanto, o uso ainda é restrito para alunos dessa universidade e não está disponível *on-line*.

4. Software on-line

Vários programas que medem a complexidade dos textos foram avaliados por Perfetti e seus colaboradores (2012) e considerados adequados ao *Common Core State Standard*, são eles: *Lexile®*, *ATOS*, *Degrees of Reading Power®: DRP® Analyzer*, *REAP (REAders-specific Practice) Readability Tool*, *SourceRater*, *Pearson Reading Maturity Metric* e *Coh-Metrix Text Easability Assessor*. Nem todos estão disponíveis on-line, no entanto, o ATOS, o mais utilizado nas escolas americanas, oferece acesso gratuito.

O ATOS foi lançado no ano de 2000 pela empresa americana de *software* educacional *Renaissance Learning*. Ele inclui duas fórmulas, uma para medir textos e outra para os livros, conforme o exemplo do *Harry Potter*, acima. Os parâmetros de análise da complexidade são o número de palavras por frase, média de palavras por nível, caractere por palavra e tamanho da palavra. No caso de livros, o número de páginas. Para que as palavras estejam ligadas ao seu grau de complexidade, o ATOS tem incorporada uma listagem de palavras previamente classificadas por níveis, chamada de *Graded Vocabulary List*, que as escolas americanas já utilizam. A figura 2, *Submissão*, abaixo, apresenta como um texto é submetido ao analisador. Para ter acesso, sugere-se a visita ao *site* <http://www.renlearn.com/atos/>

Submit a file by clicking the "Browse" button to select your file or type in the text you want analyzed, then click "Submit".

Language: English Spanish

Option 1:

Type In Your Text to Analyze

Suggested minimum of 150 characters, limit is 25,000 characters.

Enviar

Figura 2: Analisador de complexidade de textos ATOS

Após o texto ser submetido, o programa irá gerar o resultado. A tabela 1 a seguir apresenta o resultado de um texto analisado. O número 8.0 refere-se à complexidade desse texto.

ATOS Level:	8.0
Word Count:	40
Average Word Length:	5.8
Average Sentence Length:	13.3
Average Vocabulary Level:	4.3
Language selected:	English

Tabela 1: Resultado

Para saber para que série o texto submetido é indicado, deve-se olhar na tabela de classificação do ATOS. A tabela 2 a seguir descreve os números e a classificação por séries.

Instructional Use Only - Not Independent Reading	
CCSS Grade Bands	Recommended ATOS Level Ranges
2nd to 3rd grade	2.75 to 5.14
4th to 5th grade	4.97 to 7.03
6th to 8th grade	7.00 to 9.98
9th to 10th grade	9.67 to 12.01
11th grade to CCR	11.20 to 14.10

Tabela 2: Classificação da complexidade por série

Em geral, o número pode variar entre duas séries. Por exemplo, o texto de complexidade 8.0 está indicado para 6^a. até a 8^a. série. Já um texto cujo resultado é acima de 15 está indicado para o nível CCR (*College and Career Readiness*), que significa que o leitor está preparado para a leitura na sua carreira profissional, após a universidade.

5. Considerações finais

Este artigo apresentou as ferramentas computacionais utilizadas para medir a complexidade dos textos e a sua relação com cada série escolar. As métricas nas quais estas ferramentas estão fundamentadas ajudam a avaliar o texto e seu grau de dificuldade.

Para professores de inglês, em especial, utilizar estas ferramentas é uma vantagem, pois elas são capazes de apontar se o texto está adequado ao conhecimento de determinada turma. Do

mesmo modo, é possível verificar se o texto está de acordo com o nível básico, intermediário ou avançado. Muitas editoras já utilizam esse sistema, e no próprio livro está indicado o seu nível de dificuldade. No entanto, o professor pode ele mesmo fazer isso, ao lançar mão da ferramenta *on-line*.

Outros usos de um analisador de complexidade existem, por exemplo, discriminar para quem o texto está sendo dirigido. Nesse sentido, é possível avaliar se o texto é inteligível tanto para analfabetos funcionais, como para pessoas com dificuldades de leitura, tais como os disléxicos. De fato, esta é uma pesquisa em estágio inicial aqui no Brasil e carece de mais estudos que possam revelar todas as suas potencialidades.

Referências

CALDWELL, J. **Comprehension Assessment**. New York: Guilford Press, 2008

DALE , E. CHALL , J. S. Formula for predicting readability: instructions. **Educational Research Bulletin**, v. 27,p. 11-20, 37-54, 1948.

DUBAY, W. **The Principles of Readability A brief introduction to readability research**. 2004. Disponível em: <<http://www.impact-information.com/impactinfo/readability02.pdf>>
Acesso em: 15/12/2013

GRABE, W. **Reading in a second language: Moving from Theory to Practice**. Cambridge: Cambridge University Press, 2009

KODA, K. **Insights into second Language Reading: A crosslinguistics approach**. Cambridge: Cambridge University Press, 2005

M. ESKENAZI; DE LA ROSA, K. Impact of Word Sense Disambiguation on Ordering Dictionary Definitions in Vocabulary Learning Tutors. In: **Anais da 24th International Flairs Conference**. Flórida: 2013. Disponível em: <http://www.cs.cmu.edu/~max/mainpage_files/DelarosaEskenazi-FLAIRS-2011.pdf> Acesso em: 19/12/2013.

PERFETTI, C. et al. **Measures of Text complexity**. Relatório da Gates Foundation. 2012. Disponível em: <http://achievethecore.org/content/upload/nelson_perfetti_liben_measures_of_text_difficulty_research_eia.pdf> Acesso em: 19/12/2013

PIKULSKI, John J.; CHARD, David J. Fluency: Bridge between decoding and reading comprehension. **The Reading Teacher**, v. 58, n. 6, p. 510-519, 2005.

SCARTON, C.E.; ALUÍSIO, S.M. Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: Adaptando as métricas do Coh-Metrix para o Português. **Linguamática**, v. 2, n. 1, p. 45-61, 2010.