

Comparando Avaliações de Inteligibilidade Textual entre Originais e Traduções de Textos Literários

Bianca Franco Pasqualini¹, Carolina Evaristo Scarton², Maria José B. Finatto³

¹ Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (PPG-Let/UFRGS), Porto Alegre-RS, Brasil

² Núcleo Interinstitucional de Linguística Computacional (NILC) - Instituto de Ciências Matemáticas e de Computação (ICMC) – Universidade de São Paulo (USP)
Caixa Postal 668-13.560-970 – São Carlos – SP – Brasil

³ Instituto de Letras, Departamento de Linguística, Filologia e Teoria Literária – Universidade Federal do Rio Grande do Sul, Programa de Pós-Doutorado da Universidade de São Paulo (USP-ICMC), CNPq – Porto Alegre-RS, Brasil

bianca.pasqualini@gmail.com, carol.scarton@gmail.com,
mfinatto@terra.com.br

Abstract. *In this paper, textual intelligibility is evaluated in a set of selected source texts and translations in Portuguese (PT) and English (EN) using the Coh-Metrix and Coh-Metrix-Port tools. Both PT to EN and EN to PT directions of translation are analyzed, making notice that the tools used are independent, without reference to the texts under scrutiny being source texts or translations. The results show, at least regarding the Flesch Index, that the translations to PT are more complex than those to EN. Because some metrics are affected by specificities of each language and by the direction of translation, a refinement to the metrics of both tools is indicated in order to properly compare translations to source texts.*

Resumo. *Compara-se a inteligibilidade textual entre originais e traduções em um conjunto de contos em português e inglês utilizando-se as ferramentas Coh-Metrix e Coh-Metrix-Port. São consideradas as direções tradutórias português-inglês e a inversa, salientando-se que não estão previstas traduções e originais no uso dessas ferramentas, que são independentes. Aparentemente, pelo menos no índice Flesch, as traduções para o português tornam-se mais complexas do que as traduções para o inglês. Conclui-se que é preciso aperfeiçoar as métricas, criando sincronização entre o português e o inglês com a finalidade de examinar traduções, pois algumas métricas mostram-se afetadas por especificidades de cada língua e também pela direção tradutória.*

1. Introdução

No Brasil, no âmbito dos Estudos da Linguagem, o tema da complexidade ou da inteligibilidade textual (IT) tem sido tratado sob diferentes perspectivas, especialmente sob a ótica da Leitura [Perini, 1988, 2005; Kleiman 1987-89-93-97; Leffa, 1996; Fulgêncio e Liberato 2004, entre outros]. De outro lado, no cenário da Psicologia internacional, desde muito tempo, buscam-se fórmulas ou modelos – sempre muito

discutidos e duramente criticados – para prever elementos textuais associados à dificuldade de compreensão de um texto escrito, de modo que pudessem ser gerados textos de acesso facilitado para leitores de diferentes perfis. Essa busca também tem sido bastante intensa em Psicolinguística e Processamento da Linguagem Natural (PLN), integrando estudos de Inteligência Artificial [Dubay, 2004].

Entre os Estudos de Tradução, a escola funcionalista [p. ex., Reiss e Vermeer, 1984] propõe que a tradução deve ter uma finalidade, apresentando coesão interna para a comunidade a que se destina e mantendo coerência com o texto-fonte. Nessa perspectiva, no que diz respeito à avaliação da qualidade da tradução, um fator preponderante são as regras e necessidades da comunidade leitora, e não as da língua de partida, e a função do texto traduzido dentro dessa comunidade. A noção de função, nessa ótica, está ligada principalmente ao perfil linguístico e cultural da comunidade à qual a tradução se destina. No entanto, os tradutores – como outros profissionais do texto, incluindo revisores – dispõem de poucos recursos que os auxiliem na tarefa de caracterizar linguisticamente os perfis das comunidades a que se dirigem, sabendo quais formulações textuais lhes seriam mais ou menos inteligíveis, restringindo-se ao que lhes ofereçam dicionários, gramáticas e sua intuição linguística. Carecem, portanto, de ferramentas que os auxiliem a analisar, como um todo, o texto original e a tradução feita. Nesse sentido, tratar de sistemas que contemplem medidas de inteligibilidade e aplicá-los a traduções e originais, ainda que haja uma série de limitações a superar, poderia representar importante contribuição para ajudar a ponderar-se sobre maior ou menor acessibilidade de determinados grupos de leitores brasileiros a esses textos.

A partir do momento em que a Linguística de Corpus (LC) torna-se mais conhecida no Brasil [Berber Sardinha, 2004] entre linguistas, passa-se a reconhecer o quanto instrumentos estatísticos e computacionais podem contribuir para a descrição da linguagem [Rajagopalan, 2008]. E, à medida que esses linguistas se aproximam do PLN [Dias-da-Silva, 1996], o potencial de contribuição mútua entre áreas diferentes fica ainda mais claro. Com acesso a pesquisas em PLN, descobrem que temas como, p.ex., *Readability Assessment* e métricas de IT já haviam sido tratados por pesquisadores de Ciência da Computação (CC) em parceria com psicólogos e psicolinguistas [Davidson, Green, 1988; DuBay, 2004].

Fruto da integração entre estudiosos de CC, PLN, LC, psicolinguistas e linguistas de diferentes formações, dispõe-se hoje também no Brasil de sistemas automáticos para simplificação textual e sumarização, incluindo aplicações para avaliar e classificar textos em função da sua inteligibilidade para leitores de diferentes graus de escolaridade, além de recursos gratuitos *on-line* para gerar versões facilitadas de textos complexos com opções de regulagem da simplificação para o léxico ou sintaxe. Um bom exemplo disso são os produtos do Projeto PorSimples (www.caravelas.icmc.usp.br). Entre eles, destaca-se aqui o sistema Coh-Matrix-Port [Scarton, Aluísio, 2010], o qual fornece, automaticamente, uma série de métricas indicativas de IT. Esse sistema é uma versão em português para o sistema original em inglês [Graesser et al., 2004].

De outro lado, embora a LC brasileira tenha impulsionado também vários trabalhos sobre tradução com apoio computacional, seja tradução literária ou técnica [Tagnin, Vale, 2008], poucos reuniram PLN e IT em *corpora* e questões sobre perfis de tradução. No trabalho de Pasqualini, Evers e Finatto (2001), p. ex., relata-se

experimento inicial com uso dos sistemas Coh-Metrix para comparar textos-fonte em inglês e suas traduções para o português e vice-versa. A hipótese do trabalho era a de que textos literários traduzidos do inglês para o português brasileiro se tornariam mais complexos, enquanto textos literários originalmente escritos em português brasileiro ficariam menos complexos traduzidos para o inglês. Os *corpora*, entretanto, eram bastante pequenos.

O objetivo deste trabalho, expandindo o anteriormente citado, é examinar, contrastivamente, a inteligibilidade textual entre 28 contos de autores variados da literatura de língua inglesa e portuguesa do Brasil produzidos entre 1830 e 1940 e suas traduções (detalhes em Anexos), na direção inglês-português e português-inglês. Para esta análise, foram utilizados os sistemas Coh-Metrix em inglês e português, com preferência para textos curtos, sem ultrapassar 30 mil caracteres cada (dada a limitação de processamento das ferramentas). As métricas estudadas dos dois sistemas foram as que, a princípio, fossem linguística e gramaticalmente equivalentes entre as duas línguas. Foi realizado o teste estatístico *t-Student*, para cada métrica e entre as traduções português-inglês e inglês-português, para avaliar se a diferença entre as médias dos resultados obtidos eram significativamente diferentes com confiança de 95% ($p\text{-value} < 0,05$). Uma maior heterogeneidade de autores e de tradutores dos textos visa apreender, para estudos futuros, indícios de prováveis efeitos do gênero literário conto, estilos de autores e tradutores e da direção tradutória sobre elementos da IT. Também são indicados pontos de melhoria para métricas do português diretamente adaptadas do inglês.

A importância deste trabalho reside não só na descrição de índices lexicais, sintáticos e semânticos dos textos das línguas abordadas separadamente, mas também no levantamento de índices contrastivos entre o português e o inglês em tradução. A análise dos resultados, sob o ponto de vista linguístico, é útil para qualificar as ferramentas usadas. Além disso, a análise contrastiva das métricas é um primeiro passo para uma possível automatização de índices de IT em traduções, ou seja, para a criação de uma ferramenta de processamento da inteligibilidade de textos traduzidos do inglês para o português.

A existência de uma ferramenta dessa natureza, que auxilie a avaliar os textos em sua complexidade, faz-se ainda mais importante à medida que iniciativas recentes do Ministério da Educação do Brasil (MEC) visam popularizar o acesso a clássicos da literatura nacional e internacional para *neoleitores* (conforme o MEC: “Jovens com mais de 15 anos e adultos que participam do programa Brasil Alfabetizado em todo o país e nas escolas públicas com turmas de educação de jovens e adultos [EJA].” <http://portal.mec.gov.br>) por meio de versões mais facilitadas dos textos para um primeiro contato de leitores iniciantes, em projetos como o “Leitura para Todos” e “É Só o Começo”. Essas políticas públicas brasileiras baseiam-se em dados de pesquisas sobre alfabetismo e leitura, como as realizadas pelo Indicador de Alfabetismo Funcional (INAF), uma parceria entre a organização Ação Educativa e o Instituto Paulo Montenegro, e a pesquisa “Retratos da Leitura no Brasil” (RLB), do Instituto Pró-Livro, em parceria com a Imprensa Oficial do Estado de São Paulo.

Segundo Ribeiro (2010), a matriz de referência usada pelo INAF para medição do alfabetismo abarca quatro habilidades funcionais: localização, integração, elaboração

e avaliação, as quais servem como parâmetro para a descrição de níveis de proficiência. A autora salienta que “tarefas mais difíceis envolvem (...) palavras pouco usuais, palavras polissílabas e com padrões silábicos complexos”. Já a pesquisa RLB mostra que, do universo de 77,1 milhões de entrevistados que se declaram não leitores (45% do total da população estudada), 16% afirmaram que não costumam ler porque leem muito devagar ou não compreendem a maior parte do que leem (7%) [Amorim, 2008]. Além disso, dados recentes da Fundação Getúlio Vargas apontam para um crescimento vertiginoso da classe C – classe que, de acordo com a pesquisa RLB, é a maior consumidora de livros no Brasil (47% do total de leitores que compram livros) [Gasparin, 2011].

Pensando nesses dados é que a seleção de originais e traduções deste estudo centrou-se em autores conhecidos no universo dos clássicos da literatura estrangeira em língua inglesa, tendo como base para comparação autores também clássicos da literatura nacional. Nesse particular, os contos selecionados para exame de IT nas ferramentas Coh-Metrix oferecem um *continuum* histórico importante da literatura nacional e estrangeira que vai de 1839 até 1940, contemplando apenas traduções produzidas no século XX. Até o momento, não se tem registro de que as adaptações facilitadas de algumas obras literárias já distribuídas pelos programas mencionados tenham sido produzidas com aproveitamento de índices de IT: na verdade, os textos têm sido adaptados por linguistas e tradutores a partir de critérios subjetivos tais como “frases muito longas, com muitas orações; encaixe de orações mediante processos sofisticados, como o que ocorre quando se usa ‘cujo’, etc.” [Fischer, 2010].

2. Materiais e métodos

2.1. Corpus reunido

O *corpus* compõe dois blocos. No Bloco 1, originais em inglês (EN) e respectivas traduções para o português do Brasil (PT), há 14 contos dos seguintes autores: Edgar Allan Poe (10); Nathaniel Hawthorne (01), O. Henry (01), Virginia Woolf (01) e James Joyce (1). No Bloco 2, originais em português e respectivas traduções para o inglês, também há 14 contos: Machado de Assis (06), Coelho Neto (02), Humberto de Campos (03) e Lima Barreto (03). No Bloco 1, os textos originais têm uma média aproximada de 1.800 palavras (*tokens*) cada; no Bloco 2, 1.600. A seleção de autores para a amostra foi guiada por uma homogeneidade relativa, centrada no gênero conto curto, salvaguardadas todas as diferenças, visto que são autores que marcam diferentes mudanças da literatura do século XIX para o século XX. As traduções utilizadas foram todas produzidas a partir de 1980, e o conjunto de tradutores inclui experientes e novatos.

2.2. Métricas do sistema Coh-Metrix para o inglês (EN) e o português (PT)

A ferramenta Coh-Metrix [Graesser et al., 2004] utiliza 60 índices psicolinguísticos que avaliam a inteligibilidade de um texto (em inglês) nos níveis léxico, sintático e semântico. A ferramenta Coh-Metrix-Port [Scarton e Aluísio, 2010] adapta o sistema para o português, contando com 48 métricas. Entretanto, nem todas as métricas de ambos os recursos podem ser comparadas, pois há métricas próprias de cada recurso e métricas incompatíveis devido aos recursos utilizados (*wordnet*, por exemplo). As 31 métricas que puderam ser diretamente comparadas e, portanto, puderam ser utilizadas neste trabalho, são:

- Contagens básicas: número de palavras, número de sentenças, número de parágrafos, sentenças por parágrafos, palavras por sentenças e sílabas por palavras.
- O índice Flesch.
- Constituintes: incidência de sintagmas nominais, modificadores por sintagmas nominais e palavras antes de verbos principais.
- Conectivos: incidência de todos os conectivos, incidência de conectivos aditivos positivos, incidência de conectivos aditivos negativos, incidência de conectivos temporais positivos, incidência de conectivos temporais negativos, incidência de conectivos causais positivos, incidência de conectivos causais negativos, incidência de conectivos lógicos positivos e incidência de conectivos lógicos negativos.
- Operadores lógicos: incidência de operadores lógicos e número de negações.
- Pronomes, tipos e *tokens*: incidência de pronomes pessoais, pronomes por sintagmas nominais e relação tipo/*token*.
- Correferências: sobreposição do argumento em sentenças adjacentes, sobreposição de argumento, sobreposição do radical de palavras em sentenças adjacentes, sobreposição do radical de palavras, sobreposição de palavras de conteúdo em sentenças adjacentes.
- Anáforas: referência anafórica em sentenças adjacentes e referência anafórica.

3. Resultados e análises

Na Tabela 1, estão os resultados obtidos comparando-se textos originais em PT com suas traduções para o EN e textos originais em EN com suas traduções para o PT.

Métricas lexicais - Sílabas por palavras e Índice Flesch – linhas 1 e 2 da Tabela 1: O Flesch engloba a métrica “sílabas por palavras”. Houve diminuição considerável nessas métricas do PT ao EN: de 62,7 (coluna 1) para 74,6 (coluna 3), lembrando que quanto maior, mais fácil o texto. Nas traduções EN-PT, ocorre o contrário: parte-se em EN de 64,37 (coluna 5); a média dos índices Flesch das traduções para o português é 48,20 (coluna 7). Nesse quesito, quando traduzidos para português, temos textos que exigem mais esforço de compreensão.

Métricas sintáticas – **a)** Conectivos (temporais positivos; causais positivos; lógicos positivos; lógicos negativos) e negações – da linha 3 à linha 7 na Tabela 1: conectivos compõem a tessitura do texto, sendo, em tese, facilitadores da leitura. Assim, os resultados apontam para um maior índice de coesão em PT e, por conseguinte, maior legibilidade. **b)** Sintagmas nominais, modificadores por sintagmas, pronomes por sintagmas e pronomes pessoais – da linha 8 à linha 11 da Tabela 1: em PT, em função da desinência verbal, não é necessário usar pronomes pessoais com tanta frequência. De outro lado, não é surpresa o número elevado de pronomes pessoais em EN, visto que precisam ser expressos em função da gramática dessa língua. Essa relação também fica aparente no índice maior de pronomes por sintagmas em EN, pois essa métrica contempla também os pronomes pessoais. No entanto, o número de modificadores por sintagma – métrica que não envolve pronomes, mas adjetivos, advérbios e artigos – é inferior em PT nas duas direções, o que indica legibilidade maior neste quesito em PT.

Métricas semânticas – a) Referência anafórica/referência anafórica adjacente – linhas 12 e 13 da Tabela 1: os resultados apontam que o número maior de sintagmas nominais encontrados nos textos em EN, tanto de partida quanto de chegada, indicam uso maior de substantivos e menor de pronomes. **b)** Sobreposição de argumentos/sobreposição de argumentos (adjacente) – linhas 14 e 15 da Tabela 1: as duas métricas aumentam nas traduções PT-EN (colunas 1 e 3) e diminuem nas traduções EN-PT (colunas 5 e 7). Em função de a sobreposição de argumentos poder gerar dificuldade de leitura, pois pode induzir erro, os textos em EN, tanto os textos-fonte quanto os traduzidos, poderiam ser considerados mais complexos. **c)** Sobreposição de palavras de conteúdo (adjacente), sobreposição de radical de palavras e sobreposição de radical de palavras (adjacente) – da linha 16 à linha 18 da Tabela 1: essas métricas dizem respeito ao fluxo do texto e à manutenção de tópicos. Todos os textos em EN, tanto traduzidos como originais, têm índices menores nessas três métricas, o que indica menor repetição de palavras e de radicais e uma possível maior dificuldade de leitura.

Tabela 1: Métricas que possuem diferenças significativas em suas médias (p -value < 0.05)

	Português - Inglês				Inglês - Português			
	Português		Inglês		Inglês		Português	
	Média	Dsvp	Média	Dsvp	Média	Dsvp	Média	Dsvp
1) Sílabas por Palavras	2.64	0.11	1.40	0.08	2.85	0.14	1.44	0.09
2) Índice Flesch	62.67	5.48	74.60	7.19	48.20	10.29	64.37	11.61
3) Conectivos Temporais Positivos	14.88	2.35	11.03	2.65	14.06	3.86	10.29	2.57
4) Conectivos Causais Positivos	34.85	5.85	21.10	5.04	38.42	5.05	22.27	5.86
5) Conectivos Lógicos Positivos	28.39	5.22	18.63	4.39	31.54	4.44	20.93	6.92
6) Conectivos Lógicos Negativos	4.63	1.41	11.06	3.16	4.24	2.07	14.97	3.53
7) Negações	5.29	3.55	7.78	2.78	3.65	1.57	11.34	3.94
8) Sintagmas Nominais	248.88	14.39	298.74	12.44	234.35	23.83	278.53	19.36
9) Modificadores por Sintagmas	0.53	0.05	0.74	0.16	0.59	0.06	0.81	0.11
10) Pronomes por Sintagmas	0.23	0.07	0.37	0.08	0.21	0.05	0.35	0.11
11) Pronomes Pessoais	14.75	7.36	111.74	26.74	13.23	8.59	97.99	33.73
12) Referência Anafórica	0.44	0.22	0.24	0.08	0.49	0.22	0.27	0.11
13) Referência Anafórica Adjacente	0.31	0.16	0.40	0.12	0.34	0.17	0.49	0.18
14) Sobreposição de Argumentos	0.15	0.05	0.29	0.07	0.19	0.11	0.37	0.12
15) Sobreposição de Argumentos (Adjacentes)	0.20	0.09	0.36	0.10	0.25	0.13	0.44	0.14
16) Sobreposição de Palavras de Conteúdo (Adjacentes)	0.18	0.07	0.08	0.02	0.21	0.08	0.08	0.04
17) Sobreposição de Radical de Palavras	0.26	0.08	0.10	0.03	0.30	0.14	0.12	0.05
18) Sobreposição de Radical de Palavras (Adjacentes)	0.33	0.11	0.13	0.05	0.40	0.13	0.14	0.08

3.2. Resultados

Ainda que o EN comparativamente tenha mostrado muitas métricas com índices de complexidade mais altos, não se pode afirmar categoricamente que a tradução para o PT gere um texto mais fácil. Afinal, as métricas *sintagmas nominais* e *incidência de pronomes pessoais*, por exemplo, mostram particularidades específicas do EN que não correspondem *pari passu* ao PT, isto é, refletem a característica essencialmente lexicológica do EN em comparação ao PT, uma língua mais organizada pela gramática do que pelo léxico. Para resultados mais confiáveis, entretanto, é preciso mais do que a adaptação direta, ou seja, é preciso sincronizar as métricas entre as duas línguas a fim de

que possamos usá-las na comparação da IT entre textos originais e traduções, visto que são duas línguas com funcionamentos gramaticais bem diferentes. A sugestão é revisar métricas cujos resultados comparativos tenham mostrado as maiores diferenças significativas, como o caso dos pronomes pessoais, pois envolvem condições inerentes da gramática do EN. Além disso, revisar métricas que repercutem sobre outras, gerando um “efeito cascata”, pode qualificar o sistema para o PT.

Partindo desse princípio, seria possível incluir também métricas que ficaram de fora da análise empreendida aqui em função de basearem-se em bancos de dados distintos em PT e EN, como é o caso da métrica *frequência de palavras de conteúdo*. E indo mais além, poderíamos aventar a criação de métricas específicas para o PT, como uma métrica de aferição de pronomes oblíquos para o Coh-Metrix-Port.

4. Trabalhos futuros e considerações finais

A partir do que foi verificado, percebe-se que associar métricas de IT entre originais e traduções pode render importantes *insights* tanto para estudos de PLN quanto para estudos linguísticos em geral e estudos de Tradução e de Linguística de Corpus. Além disso, ao reiterar-se a percepção de uma tendência para menor IT quando se traduz textos de literatura para o português do Brasil no índice Flesch, fica o questionamento se isso ocorreria também na tradução de textos de outra natureza, tais como, p.ex., o texto científico e o texto jornalístico.

É importante ressaltar que, no caso do sistema Coh-Metrix, a unidade de processamento é a unidade do texto, destacado em meio a um *corpus*. O movimento analítico *corpus-texto-corpus*, embutido no sistema, é realizado a partir de um determinado *corpus* de treinamento que pode ter características bem diferentes dos textos literários aqui em foco, e essa é uma limitação deste trabalho. Além disso, fica também a perspectiva de avaliar-se o papel das diferentes métricas e contagens já desenvolvidas para textos literários brasileiros, inclusive sem envolver contrapontos com a sua tradução para o inglês, com destaque para aqueles que fossem mais demandados como leitura no Ensino Fundamental e Médio e no programa para *neoleitores* do MEC. Assim se poderia também propor a inserção e a testagem de novas medidas, como, p.ex., a presença de elipses ou omissões de elementos dos textos.

5. Agradecimentos

Ao CNPq e à FAPESP pela concessão de bolsas e de auxílios à pesquisa. Ao NILC-ICMC-USP pelo apoio institucional.

Anexos – Dados dos textos do corpus

BLOCO 1.1 – EDGAR A. POE E TRADUÇÕES PARA O PORTUGUÊS BRASILEIRO									
TÍTULO ORIGINAL	NÚMERO DE CARACTERES		FONTE	TÍTULO TRADUÇÃO	NÚMERO DE CARACTERES		TRADUTOR	FONTE	
The oval portrait	7.204		Projeto Gutenberg	O retrato oval	7.081		Marcelo Bueno	http://www.bestiario.com.br/18_arquivos/poe.html	
Mesmeric revelation	21.615			11.058	Revelação mesmeriana	22.262		Oscar Mendes	POE, EDGAR ALLAN POE. FICÇÃO COMPLETA. 7ª. ED. RIO DE JANEIRO: NOVA FRONTEIRA, 1981.
	10.529	11.071				11.143			
The black cat	21.749			10.991	O gato preto	21.749		Bernardo Carvalho	SÃO PAULO: COSAC & NAIFY, 2004.
	10.798	10.610				10.645			
The imp of the perverse	13.741			14.135	O demônio da perversidade	14.135		Oscar Mendes	POE, EDGAR ALLAN POE. FICÇÃO COMPLETA. 7ª. ED. RIO DE JANEIRO: NOVA FRONTEIRA, 1981.
The tell-tale heart	11.147					11.780		Celina Portocarrero	HTTP://WWW.RELEITURAS.COM/EAPOE_CORACAO.ASP
Berenice	18.635			9.887	Berenice	20.300		Oscar Mendes	POE, EDGAR ALLAN POE. FICÇÃO COMPLETA. 7ª. ED. RIO DE JANEIRO: NOVA FRONTEIRA, 1981.
	9.978	8.545				10.227			
Eleonora	13.375			13.400		13.400		Oscar Mendes	POE, EDGAR ALLAN POE. FICÇÃO COMPLETA. 7ª. ED. RIO DE JANEIRO: NOVA FRONTEIRA, 1981.
The masque of the red death	13.683			13.213		13.213			
The cask of amontillado	12.757			12.800		12.800		Dorothee de Bruchard	HTTP://WWW.BESTIARIO.COM.BR/12_ARQUIVOS/O%20HOMEM%20DA%20MULTIDAO.HTML
The Man of the Crowd	20.375		20.697		20.697				
	9.928	10.440	http://classclit.about.com	O homem da multidão	10.165	10.532			
BLOCO 1.2 – CONTOS DE LITERATURA EM LÍNGUA INGLESA E TRADUÇÕES PARA O PORTUGUÊS BRASILEIRO									
AUTOR	TÍTULO ORIGINAL	NÚMERO DE CARACTERES		FONTE	TÍTULO TRADUÇÃO	NÚMERO DE CARACTERES		TRADUTOR	FONTE
James Joyce	Araby	12.240		http://fiction.eserv er.org/short/araby.html	Arábia	12.182		Roberto Schmitt-Prym	http://www.bestiario.com.br/18_arquivos/arabia.html
Virginia Woolf	Monday or Tuesday	1.884		http://www.bartleby.com/85/3.html	Segunda ou terça-feira	1.998			http://www.bestiario.com.br/1_arquivos/woolf.html
O. Henry	The ransom of the Red Chief	21.778		http://fiction.eserv er.org/short/ransom_of_red_chief.html	O resgate do Chefe Vermelho	22.960		Bianca Pasqualini	Arte e Letra, Estórias K, 2010
		12.376	9.402			13.133	9.827		
Nathaniel Hawthorne	Wakefield	19.737		http://classclit.about.com	Wakefield	19.693		Zaida Maldonado	http://www.bestiario.com.br/5_arquivos/Wakefield.html
		10.544	9.193			10.556	9.146		
BLOCO 2.1 – MACHADO DE ASSIS E TRADUÇÕES PARA O INGLÊS									
TÍTULO ORIGINAL	NÚMERO DE CARACTERES		FONTE	TÍTULO TRADUÇÃO	NÚMERO DE CARACTERES		TRADUTOR	FONTE	
A cartomante	18.203		http://machado.mec.gov.br	The fortune-teller	22.170		Isaac Goldberg	Projeto Gutenberg	
	10.362	7.617			12.675	9.338			
Viver!	14.209		http://machado.mec.gov.br	Life	15.450		Francis Johnson	http://www.brazilianstories.com	
	8.204	5.788			8.939	6.384			
Cantiga de esponsais	8.066		http://www.releituras.com	Wedding song	8.844		Francis Johnson	http://www.brazilianstories.com	
O enfermeiro	17.911		http://www.brazilianstories.com	Looking after	19.914				
	8.797	9.022			9.627	10.157			
Marcha fúnebre	14.431		http://www.brazilianstories.com	With muffled drum	15.415		Francis Johnson	http://www.brazilianstories.com	
	7.552	6.811			8.116	7.202			
A vida eterna	27.737		http://www.brazilianstories.com	Life eternal	28.622		Francis Johnson	http://www.brazilianstories.com	
	14.796	12.881			14.780	13.721			
BLOCO 2.2 – CONTOS BRASILEIROS E TRADUÇÕES PARA O INGLÊS									
AUTOR	TÍTULO ORIGINAL	NÚMERO DE CARACTERES		FONTE	TÍTULO TRADUÇÃO	NÚMERO DE CARACTERES		TRADUTOR	FONTE
Lima Barreto	O único assassinato de cazuza	9.800		http://pt.wikisource.org/wiki/O_%C3%BAnico_assassinato_de_Cazuza	Killer	10.300		Francis Johnson	http://www.brazilianstories.com
Humberto de Campos	A promessa	14.930		http://pt.wikisource.org/wiki/A_Promessa	Light of my life	15.200			
Coelho Netto	Firmo, o vaqueiro	8.478		http://peregrinacultural.wordpress.com	Christmas corral	8.971		Francis Johnson	http://www.brazilianstories.com
Lima Barreto	O homem que sabia javanês	18.251		http://www.brazilianstories.com	The man who spoke javanese	18.387			
Humberto de Campos	O diálogo das Caveiras	7.311			http://www.brazilianstories.com	Fish and Filosofy	7.393		
Coelho Netto	O duplo	7.774		http://www.brazilianstories.com		Me Too	8.268		
Lima Barreto	O número da sepultura	18.067			http://www.brazilianstories.com	Late bet	18.684		
Humberto de Campos	Vingança	12.556		http://www.brazilianstories.com		In the forests of the night	12.525		
		10.320					10.603		
		7.747				8.081			

Referências

- Berber Sardinha, T. (2004) “Linguística de Corpus”, Barueri-SP: Manole.
- Dias-da-Silva, B. C. (1996) “A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais”, Tese de Doutorado. Faculdade de Ciências e Letras, Universidade Estadual Paulista – UNESP, Araraquara. 285p.
- DuBay, W. (2004). “The principles of readability”, Costa Mesa, CA: Impact Information: <http://www.impact-information.com/impactinfo/readability02.pdf>.
- Dyer, S., Martin, J. and Zulauf, J. (1995) “Motion Capture White Paper”, http://reality.sgi.com/employees/jam_sb/mocap/MoCapWP_v2.0.html, December.
- Fischer, L. A. (2010) “Professor Luís Augusto Fischer fala sobre projeto para neoleitores”, Editora L&PM. Disponível em: <http://www.lpm.com.br>. Acesso em 18 de junho de 2011.
- Fish, Stanley. (1976) Interpreting the Variorum. *Critical Inquiry*, v. 2.
- Fulgêncio, L. and Liberato, Y. (2004) “Como facilitar a leitura: como se processa a leitura; orientação para textos didáticos; aspectos discursivos”, São Paulo: Contexto.
- Gasparin, G. (2011) “Classe C ganha 39,5 milhões de pessoas, diz FGV”. Disponível em: <http://g1.globo.com/economia>. Acesso em 27 de junho de 2011.
- Graesser, A. C., McNamara, D.S., M. Louwerse, and Cai. Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. In: *Behavioral Research Methods, Instruments, and Computers*, v. 36, p. 193-202.
- Henry, O. (2010) “O resgate do Chefe Vermelho”. Tradução de Bianca Pasqualini. Porto Alegre: Arte e Letra, Estórias K.
- Kleiman, A. (1987) Aprendendo palavras, fazendo sentido: o ensino de vocabulário nas primeiras séries. In: *Trabalhos em Linguística Aplicada 9*. Campinas, SP: Universidade Estadual de Campinas. p. 47-81.
- Kleiman, A. (1989) “Leitura: Ensino e Pesquisa”, Campinas, SP: Pontes, 1989.
- Kleiman, A. (1993) “Oficina de Leitura: teoria e prática”, Campinas, SP: Pontes, 1993.
- Kleiman, A. (1997) “Texto e leitor: aspectos cognitivos da leitura”, 5.ed. Campinas: Pontes, 1997.
- Leffa, V. J. (1996) “Fatores da compreensão na leitura”. Projeto ELO, Ensino de línguas online. Disponível em: www.leffa.pro.br.
- Pasqualini, B.; Evers, A. and Finatto, M. J. B. (2011) Medidas de Complexidade Textual entre Traduções Brasileiras e Originais de Literatura Inglesa: um Estudo-Piloto Baseado em *Corpus*. *Anais do IX Encontro de Linguística de Corpus*, Porto Alegre-RS, PUC-RS, 2010.
- Perini, M. A. (1988) A Leitura Funcional e a Dupla Função do Texto Didático. In: *Leitura - Perspectivas Interdisciplinares*, v. 27, n. 3, p. 215-219.

- Perini, M. A. (2005) “Pelos caminhos da perplexidade: uma receita para ler sem entender”. In: Mari, H., Walty, I. and Versiani, Z. (Org.). *Ensaio sobre leitura*. Belo Horizonte: Editora Pucminas, v. 1, p. 33-46.
- Rajagopalan, K. (2008) “A Linguística de Corpus no tempo e no espaço: visão reflexiva”, In: *Um percurso para pesquisas com base em corpus*, Edited by Marcia Regina Gerber e Vera Vasilévski. Editora da UFSC, Florianópolis – SC.
- Reiss, K. and Vermeer, H. J. (1984) “Grundlegung einer allgemeinen Translationstheorie, Tübingen: Niemeyer of Difference”, London & New York: Routledge.
- Scarton, C. E. and Aluísio, S. (2010) Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. *Linguamática* (Revista para o Processamento Automático das Línguas Ibéricas), v. 2, p. 45-61.
- Smith, A. and Jones, B. (1999). On the complexity of computing. In *Advances in Computer Science*, p. 555–566. Publishing Press.
- Tagnin, E. and Vale, O. A. (Org.) (2008). “Avanços da Linguística de Corpus no Brasil”, Humanitas, São Paulo – SP.