

OS TRABALHADORES DO MAR: ANÁLISE DA TRADUÇÃO DE MACHADO DE ASSIS SOB A PERSPECTIVA DA LINGÜÍSTICA DE *CORPUS*.

Estela Maria Faustino de Carvalho
Universidade Federal de Santa Catarina
emfcarvalho@yahoo.com.br

RESUMO

Este artigo explorará a análise da escolha de correspondências lexicais, empregada por Machado de Assis, na tradução do romance *Les Travailleurs de la Mer* (Os Trabalhadores do Mar), de Victor Hugo. Esta análise será fundamentada nos estudos da lingüística com base em *corpora* computadorizados, ou seja, legíveis por computador. O objetivo deste artigo é demonstrar como a Lingüística de *Corpus* pode ser útil na análise de obras literárias e suas respectivas traduções.

PALAVRAS-CHAVE: Lingüística de *corpus*. Tradução.

ABSTRACT

This article will explore the analysis of the lexical correspondences chosen by Machado de Assis in the translation of the novel *Les Travailleurs de la Mer* (The Workers of the Sea), from Victor Hugo. This analysis is founded on linguistics studies based on computerized corpora, that is, machine-readable corpora.

KEYWORDS: *Corpus Linguistics*. Translation.

Uso de corpora computadorizados em pesquisas lingüísticas

Entende-se por *corpora* computadorizados aqueles cujo formato é passível de leitura por máquina. A análise de *corpora* computadorizados é feita através de programas computacionais específicos que executam operações, tais como listagem de palavras e concordância.

Segundo Tony Berber Sardinha (2004), para que a análise de *corpora* computadorizados tenha validade, algumas condições devem ser respeitadas:

1. Os textos que formam o *corpus* devem ser autênticos, provenientes de linguagem natural e não produzidos com o propósito de serem analisados através de investigações lingüísticas. No caso da análise feita neste artigo, os textos foram obtidos através de pesquisa na Internet.

2. Os textos devem ser autênticos, ou seja, produzidos por falantes nativos, a não ser que o estudo tenha como objetivo a análise de *corpora* de aprendizes, mais conhecido como *learner corpora*. No caso de análise contrastiva, utilizam-se os dois tipos, originais e traduções. Neste estudo, foram utilizados textos autênticos produzidos por falantes nativos: *Les Travailleurs de la Mer*, texto autêntico produzido pelo francês Victor Hugo, e Os Trabalhadores do Mar, tradução produzida em autêntico português do Brasil pelo brasileiro Machado de Assis.
3. O conteúdo deve ser criteriosamente escolhido, seguindo condições de naturalidade e autenticidade, além de obedecer a regras estabelecidas por seus criadores, de forma a corresponder às características pretendidas. Os textos analisados neste trabalho são autênticos, pois se tratam da cópia eletrônica de livros publicados no século XIX.
4. O conteúdo deve ser representativo de um idioma ou de uma variedade lingüística. A variedade lingüística aqui analisada refere-se ao século XIX.

Observam-se algumas das possibilidades de comparação de línguas através do uso de *corpora* computadorizados, organizadas por Aijimer e Altenberg (1996, p. 12) e comentadas por Stig Johansson (1998):

- a) Podem dar uma compreensão clara sobre as línguas comparadas, cujas particularidades podem passar despercebidas caso sejam estudadas individualmente. Adiante serão apresentadas as particularidades das escolhas semânticas feitas por Machado de Assis na tradução da obra de Victor Hugo.
- b) Podem ser utilizadas para objetivos de comparação, de forma a aumentar o conhecimento sobre diferenças culturais, tipológicas e específicas de cada língua, assim como características universais. Neste contexto, esta análise poderá ainda ser base para estudos diacrônicos da tradução (século XIX versus século XXI).
- c) Fazem ressaltar as diferenças entre textos originais e traduções e entre textos produzidos por falantes nativos e não-nativos. Esta análise apresentará as diferenças entre a obra original de Victor Hugo e a tradução de Machado de Assis.
- d) Podem ser usados para várias aplicações práticas como, por exemplo, lexicografia, ensino de línguas e tradução.

Ainda segundo Johansson (1998), dois tipos principais de *corpora* podem ser utilizados em análise contrastiva: (a) *corpora* formados por textos originais produzidos em

duas ou mais línguas e (b) *corpora* formados por textos originais e respectivas traduções. Eles também podem ser chamados de *corpora* comparáveis e *corpora* de tradução.

Corpora comparáveis são formados por textos originais de cada língua, selecionados por época de criação, domínio, gênero, público-alvo, etc. Eles devem representar a língua natural de cada idioma analisado. Exemplos deste tipo de *corpus* são as leis de países que consideram mais de uma língua como oficial (Suíça, Canadá), ou os princípios reguladores da União Européia, cujos documentos são considerados originais nos idiomas de cada país-membro.

Comentando citações de Carl James (1980, p. 178), Stig Johansson (1998) identifica a equivalência de traduções como a melhor forma de comparação para análise contrastiva. Ele explica que *corpora* de tradução contém textos cujo objetivo é expressar os mesmos significados e ter as mesmas funções discursivas nas línguas em questão. Toma-se como ponto de partida a língua de origem ou a língua alvo para que as correspondências possam ser estabelecidas.

Seleção do corpus de análise

As análises baseadas em *corpora* computadorizados devem seguir alguns critérios básicos:

- *Origem*: Alguns autores afirmam que uma das formas de se buscar controlar a qualidade dos textos é utilizar apenas publicações, visto que textos publicados normalmente correspondem a um certo padrão de aceitabilidade no idioma em que foi escrito. Nesta análise foram utilizados textos de livros originais, acessíveis na Internet.
- *Direitos Autorais*: Paralelamente à escolha do tipo de texto, deve-se ter em mente que, caso a análise tenha como objetivo a publicação de um artigo técnico-científico ou a elaboração de um glossário técnico, deve ser solicitada a permissão dos respectivos autores para a reprodução do material consultado. A maior parte dos autores especialistas no assunto afirma que é mais fácil conseguir permissão para utilização de trechos de documentos do que textos integrais. Neste estudo, foram analisados apenas trechos de obras literárias, mas de qualquer maneira tratam-se de clássicos da literatura já categorizados como domínio público.
- *Organização*: Os textos devem ser organizados e salvos em formato legível por máquina. Os formatos legíveis pela maior parte dos concordanciadores são .txt (arquivos de texto puro sem

formatação) ou .HTML (linguagem de marcação padrão usada em documentos na Internet). A análise descrita a seguir foi elaborada a partir de textos digitalizados em formato .pdf (em inglês, *portable document format*) e transformados em formato texto.

Importância da análise de corpora computadorizados para os estudos da tradução

A *tradutologia* (ou estudos da tradução) envolve, entre outros, a análise contrastiva de traduções através da comparação do texto na língua de partida com o texto correspondente na língua de chegada.

É importante citar os critérios mencionados por Mona Baker (1995) em relação aos seis critérios de seleção de tipos de *corpora*:

1. Linguagem genérica ou de domínio restrito;
2. Escrito ou falado;
3. Sincrônico (relativo a uma mesma época a ser determinada) ou diacrônico (relativo a épocas distintas);
4. Origens (escritores / falantes) e gêneros (mídia impressa ou eletrônica, literatura, audiências de tribunais);
5. Limites geográficos (inglês americano x inglês britânico, ou português do Brasil x de Portugal);
6. Monolíngüe, bilíngüe ou multilíngüe.

Exemplos práticos de aplicação do concordanciador WordSmith Tools para a análise de traduções

Conforme já citado, algumas ferramentas computacionais auxiliam o analista nos estudos da tradução. Para este estudo, utilizou-se o concordanciador *WordSmith Tools*, de autoria de Mike Scott, obtido através da Internet.

Representatividade do Corpus

A obra *Os Trabalhadores do Mar*, em francês, é formada por um *corpus* de 148.046 palavras. No texto correspondente em português, foram computadas 116.204 palavras.

De acordo com a classificação sugerida por Berber Sardinha (2004), cada *corpus* aqui analisado, considerando idioma fonte e idioma alvo, tem um tamanho pequeno-médio (80 a 250 mil palavras).

Origem dos dados

- *Les Travailleurs de La Mer* (original em francês). Hugo, Victor (1802-1885). Documento eletrônico, datado de 1997, disponível no site *Gallica / Centre National de la Recherche Scientifique*, <http://gallica.bnf.fr>. Documento extraído da base de dados textuais *Frantext*, realizado pelo Instituto Nacional da Língua Francesa (*INaLF*), acessado em 12 de março de 2005.

- *Os Trabalhadores do Mar* (tradução para o português por Machado de Assis). Copyright © 2000, www.virtualbooks.com.br – Editora Virtual Books Online. M&M Editores Ltda., acessado em 05 de março de 2005.

Processamento dos dados

Itens lexicais mais freqüentes na obra *Os Trabalhadores do Mar* de Victor Hugo (original em francês e tradução para o português, realizada por Machado de Assis):

Obs.: Até 100 ocorrências.

WordSmith Tools -- 12/3/2005 19:18:41

N	Word	Freq.	%
1	DE	6.311	4,29
2	LA	5.089	3,46
3	LE	3.793	2,58
4	L	3.651	2,48
5	ET	3.197	2,17
6	IL	3.058	2,08
7	À	2.934	1,99
8	LES	2.434	1,65
9	UN	2.164	1,47
10	D	1.807	1,23
11	ÉTAIT	1.779	1,21
12	UNE	1.762	1,20
13	EST	1.654	1,12
14	DES	1.649	1,12
15	DANS	1.550	1,05
16	EN	1.549	1,05
17	DU	1.236	0,84
18	AVAIT	1.230	0,84
19	QUE	1.207	0,82
20	ON	1.202	0,82
21	QUI	1.165	0,79
22	CE	1.134	0,77
23	SE	1.051	0,71
24	QU	998	0,68
25	S	944	0,64
26	PAS	894	0,61
27	N	884	0,60
28	NE	870	0,59
29	C	857	0,58
30	Y	849	0,58
31	AU	797	0,54
32	SUR	758	0,52
33	CETTE	736	0,50
34	GILLIATT	696	
35	POUR	690	0,47
36	ELLE	686	0,47
37	PAR	682	0,46
38	PLUS	668	0,45
39	A	647	0,44
40	SON	556	0,38
41	VOUS	505	0,34
42	LUI	501	0,34
43	CES	497	0,34
44	TOUT	493	0,34
45	COMME	480	0,33
46	SA	471	0,32
47	DEUX	438	0,30
48	LÀ	430	0,29
49	MER	419	0,28
50	AVEC	409	0,28
51	MAIS	346	0,24
52	FAIT	342	0,23
53	HOMME	340	0,23
54	ÊTRE	329	0,22
55	OÙ	317	0,22
56	ILS	313	0,21
57	EAU	311	0,21
58	JE	310	0,21
59	SES	306	0,21
60	AUX	292	0,20
61	MÊME	274	0,19
62	AUTRE	271	0,18
63	PEU	255	0,17
64	NOUS	250	0,17
65	SONT	250	0,17
66	POINT	249	0,17
67	LETHIERRY	245	
68	VENT	245	0,17
69	SI	244	0,17
70	DIT	242	0,16

WordSmith Tools -- 12/3/2005 19:21:20

N	Word	Freq.	%
1	A	4.779	4,04
2	DE	4.358	3,69
3	O	3.443	2,91
4	E	3.034	2,57
5	QUE	2.970	2,51
6	SE	2.510	2,12
7	UM	1.659	1,40
8	DO	1.594	1,35
9	DA	1.434	1,21
10	UMA	1.369	1,16
11	NÃO	1.283	1,09
12	OS	1.247	1,06
13	ERA	1.113	0,94
14	EM	1.109	0,94
15	É	1.079	0,91
16	AS	1.011	0,86
17	NO	906	0,77
18	NA	792	0,67
19	GILLIATT	770	
20	COM	769	0,65
21	POR	748	0,63
22	PARA	733	0,62
23	AO	613	0,52
24	COMO	503	0,43
25	À	501	0,42
26	MAIS	486	0,41
27	ESTAVA	452	0,38
28	TINHA	441	0,37
29	MAR	421	0,36
30	DAS	404	0,34
31	LHE	384	0,32
32	DOS	370	0,31
33	MAS	368	0,31
34	ELE	339	0,29
35	HOMEM	320	0,27
36	ÁGUA	298	0,25
37	LETHIERRY	281	
38	HÁ	279	0,24
39	FOI	254	0,21
40	HAVIA	253	0,21
41	DEPOIS	247	0,21
42	JÁ	229	0,19
43	VENTO	228	0,19
44	CLUBIN	218	0,18
45	TEM	216	0,18
46	DÉRUCHETTE	214	
47	OU	213	0,18
48	QUANDO	212	0,18
49	SEM	212	0,18
50	CASA	209	0,18
51	DURANDE	205	
52	NEM	202	0,17
53	SÃO	202	0,17
54	NAS	197	0,17
55	DUAS	195	0,17
56	SUA	191	0,16
57	TUDO	191	0,16
58	TEMPO	183	0,15
59	SER	181	0,15
60	POUCO	177	0,15
61	GRANDE	176	0,15
62	DIA	174	0,15
63	ENTRE	174	0,15
64	MESS	173	0,15
65	NOITE	173	0,15
66	NADA	169	0,14
67	ONDE	168	0,14
68	TODOS	164	0,14
69	COISA	163	0,14
70	NAVIO	162	0,14

71	SANS	242	0,16	71	NOS	161	0,14
72	SOUS	225	0,15	72	QUEM	161	0,14
73	RIEN	224	0,15	73	ALI	152	0,13
74	CLUBIN	218	0,15	74	AQUELE	152	0,13
75	ÉTAIENT	218	0,15	75	SÓ	149	0,13
76	OU	216	0,15	76	DISSE	146	0,12
77	DÉRUCHETTE	208		77	OUTRO	143	0,12
78	ENTRE	200	0,14	78	ÀS	141	0,12
79	DURANDE	196		79	DOIS	140	0,12
80	LEUR	192	0,13	80	SEU	140	0,12
81	SAINT	190	0,13	81	MESMO	138	0,12
82	TEMPS	188	0,13	82	PODE	138	0,12
83	EÛT	186	0,13	83	SOBRE	138	0,12
84	ÉTÉ	178	0,12	84	ROCHEDO	137	
85	TOUTE	176	0,12	85	AOS	136	0,12
86	MESS	170	0,12	86	PODIA	136	0,12
87	FAIRE	169	0,11	87	QUASE	136	0,12
88	CELA	164	0,11	88	MÁQUINA	133	
89	PEUT	163	0,11	89	PANÇA	131	0,11
90	BIEN	159	0,11	90	PELA	129	0,11
91	PUIS	157	0,11	91	TER	129	0,11
92	CHOSE	154	0,10	92	VEZ	129	0,11
93	CET	153	0,10	93	SAINT	128	0,11
94	FAISAIT	150	0,10	94	MEIO	126	0,11
95	JOUR	144	0,10	95	AQUELA	125	0,11
96	NUIT	143	0,10	96	ESSA	125	0,11
97	TOUS	140	0,10	97	MUITO	125	0,11
98	ENCORE	138	0,09	98	IA	124	0,10
99	QUAND	138	0,09	99	LO	124	0,10
100	QUELQUES	138		100	AINDA	122	0,10
101	PANSE	137	0,09	101	DEBAIXO	120	
102	PRÈS	137	0,09	102	ISTO	120	0,10
103	PRESQUE	136		103	ESSE	118	0,10
104	MACHINE	133		104	ESTÁ	118	0,10
105	QUELQUE	131		105	PARECIA	118	0,10
106	CÔTÉ	127	0,09	106	PELO	118	0,10
107	ÉCUEIL	127	0,09	107	EU	117	0,10
108	ONT	127	0,09	108	OUTRA	115	0,10
109	TRÈS	127	0,09	109	ERAM	114	0,10
110	DOUVRES	123		110	TODA	112	0,09
111	TOUTES	122	0,08	111	VER	112	0,09
112	DIRE	121	0,08	112	OLHOS	109	0,09
113	TROIS	121	0,08	113	VEZES	109	0,09
114	MAISON	118	0,08	114	DELE	108	0,09
115	AVOIR	117	0,08	115	ESCOLHO	108	
116	ROCHER	117	0,08	116	ESPÉCIE	107	0,09
117	J	116	0,08	117	FEZ	107	0,09
118	COUP	113	0,08	118	LADO	106	0,09
119	AIR	110	0,07	119	ATÉ	105	0,09
120	FOIS	110	0,07	120	ALGUMA	104	0,09
121	MOINS	108	0,07	121	APENAS	104	0,09
122	NI	107	0,07	122	BEM	104	0,09
123	POUVAIT	105	0,07	123	TODAS	104	0,09
124	AUSSI	104	0,07	124	MAL	103	0,09
125	BON	104	0,07	125	ME	101	0,09
126	APRÈS	103	0,07	126	DIZIA	100	0,08
127	FUT	103	0,07	127	MENOS	100	0,08
128	GUERNESEY	101		128	PÉ	100	0,08
129	GRAND	100	0,07				
130	NAVIRE	100	0,07				

Por meio do levantamento acima, realizado a partir da opção *WordList* do concordanciador *WordSmith Tools*, pode-se analisar a relação direta entre determinadas palavras no idioma de partida (francês) e sua correspondência no idioma alvo (português). Assim sendo, é possível demonstrar como a Lingüística de Corpus pode ser útil na análise de obras literárias e de suas respectivas traduções.

Por exemplo, as palavras francesas *presque* e *machine* têm, respectivamente, o mesmo número de ocorrências em português: 136 para "quase" e 133 para "máquina". Vejamos em algumas passagens como elas foram traduzidas do texto de origem para o texto de partida:

Idiomas:	Origem: Francês	Chegada: Português
Número de ocorrências:	Presque : 136	Quase: 136
Exemplo:	C'étaient deux pointes verticales, aiguës et recourbées, se touchant presque par le sommet.	Eram duas pontas verticais e recurvadas, tocando-se quase no cume.

Idiomas:	Origem: Francês	Chegada: Português
Número de ocorrências:	Machine : 133	Máquina: 133
Exemplo:	La partie de la quille à laquelle se superposait la machine était coupée carrément et prête à glisser avec la machine en la soutenant.	A parte da quilha onde estava a máquina foi cortada em quadro e estava pronta para resvalar com a máquina sustentando-a.

Verificam-se alguns exemplos de frequências de palavras isoladas que não têm equivalência direta no idioma de origem e no idioma de chegada. Para isso, utilizou-se a opção *Concord / KeyWords* do concordanciador *WordSmith Tools*:

Idiomas:	Origem: Francês	Chegada: Português
Número de ocorrências:	Bien : 150	Bem: 104
Exemplo:	Un navire qui tient bien le vent était " bon boulinier ".	O navio que suporta o vento era bom boulinier (bom de bolina).

No exemplo acima, nota-se que Machado de Assis traduziu "*tenir bien*" por "suportar".

Idiomas:	Origem: Francês	Chegada: Português
Número de ocorrências:	Air = 110	Ar = 71
Exemplo:	Ils ont l'air d'avoir reçu un coup de cognée.	Parecem ter recebido um golpe de machado.

Neste segundo exemplo, Machado de Assis traduziu "*avoir l'air*" por "parecer".

CONSIDERAÇÕES FINAIS

O concordanciador *WordSmith Tools* pode ser utilizado como ferramenta de análise de traduções, tanto de forma individual, para projetos de pesquisa, como em sala de aula, para análise em grupo.

Atualmente, as análises de traduções são feitas por meio da leitura de ambos os textos, além do cotejamento do texto de partida com o texto de chegada. No caso de pequenos trechos analisados, esse método é eficiente.

Entretanto, no caso de grandes textos, como é o caso desta obra de Victor Hugo, com mais de 300 páginas, o *WordSmith* é uma ferramenta de grande valia para o exame realizado pelo analista, diminuindo a ocorrência de erros, tão comuns em trabalhos repetitivos. Além disso, o cálculo estatístico de frequências de palavras ou expressões torna-se possível graças à automatização das tarefas e à rapidez de processamento oferecida pelo *WordSmith*.

REFERÊNCIAS

BAKER, Mona. *Corpora in translation studies: an overview and some suggestions for future research*. Filadélfia / Amsterdã: John Benjamins, 1995 (Target 7.2, 223-243).

BERBER SARDINHA, Tony. *Linguística de Corpus*. Barueri: Manole, 2004.

FERREIRA, Eliane Fernanda Cunha. *Para traduzir o século XIX*: Machado de Assis. Rio de Janeiro: ABL, 2004.

HUGO, Victor. *Les travailleurs de la mer*. Arquivo texto disponível em: <http://gallica.bnf.fr>, Acessado em 12 de março de 2005.

HUGO, Victor. *Os trabalhadores do mar*. Tradução de Machado de Assis. São Paulo: Nova Alexandria, 2002. Arquivo pdf disponível em: http://virtualbooks.terra.com.br/freebook/traduzidos/os_trabalhadores_do_mar.htm. Acesso em: 5 mar. 2005.

JOHANSSON, Stig. On the role of *corpora* in cross-linguistic research. In: Stig Johansson e Signe Oksefjell (eds.) *Corpora and cross-linguistic research: theory, method, and case studies*. Amsterdã: Rodopi, 1998.

SCOTT, Mike. *WordSmith Tools*. Ferramenta computacional. Disponível em: www.liv.ac.uk/~ms2928. Acessado em: 24 de agosto de 2004.

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.