

O LÉXICO DIFERENCIADO DA TRADUÇÃO

THE DISTINCTIVE LEXIS OF TRANSLATION

Ana Frankenberg-Garcia*

RESUMO

Ao lermos uma tradução, é comum termos a clara noção de que estamos perante um texto traduzido, ao contrário do que acontece quando lemos um texto original, produzido sem as limitações impostas por um texto-fonte escrito num idioma diferente. Uma das características que poderá conferir esta sensação de diferença em relação às traduções é a frequência fora do comum de determinado léxico. Pesquisas existentes comparando a frequência de certas palavras em traduções e em textos que não são traduções revelam diferenças de distribuição significativas. A maioria dos estudos em causa tem como base uma abordagem de baixo para cima. Seleciona-se uma dada palavra específica, cuja frequência é então comparada em textos traduzidos e não traduzidos. No presente estudo, invertemos essa metodologia e adotamos uma abordagem exploratória de cima para baixo. Começamos com um *corpus* de textos literários em português traduzido e português não traduzido e, a partir daí, procuramos identificar palavras sobre e sub-representadas nas traduções. Os resultados obtidos não só reforçam a nossa intuição relativamente ao léxico característico da tradução, como também revelam uma série de contrastes inesperados, que provavelmente não teriam sido detectados se não se tivesse utilizado a presente metodologia.

Palavras-chave: tradução, léxico, *corpora*

ABSTRACT

It is a well-known fact that translated texts read differently from texts that have been written without the constraints imposed by source texts from another language. One of the features that can confer a distinctive feel to translations is the frequency with which certain lexical items are represented in them. Previous research has compared the frequency of specific words in translations and in texts that are not translations and unveiled substantial differences in their distributions. Most of these studies adopt a bottom-up approach. Their starting point is a given word whose frequency in translated and non-translated texts is then compared. In the present study, we adopt an explorative, top-down approach instead. We begin with a Portuguese language corpus of translated and non-translated literary texts

*. Universidade Nova de Lisboa (FCSH-UNL), Portugal. ana.frankenberg@gmail.com

and attempt to identify lemmas which are markedly over and under-represented in the translations. Our results not only appear to support existing bottom-up intuitions regarding distinctive lexical distributions, but also disclose a number of unexpected contrasts that would not have been discernible without recourse to corpora.

Keywords: translation, lexis, *corpora*

INTRODUÇÃO

Uma das grandes vantagens das análises baseadas em *corpora* é o fato de permitirem identificar convenções linguísticas difíceis de observar a olho nu. Tal como previu Baker (1993), a recente utilização de *corpora* nos estudos de tradução teve um grande impacto sobre o nosso conhecimento das especificidades do texto traduzido. Contudo, algumas características das traduções têm recebido bem mais atenção do que outras. A hipótese da explicitação, por exemplo, segundo a qual a informação implícita nos originais tende a ser mais explícita nas traduções (VINAY e DARBELNET, 1958, BLUM-KULKA 1986), já foi corroborada por diversas análises quantitativas baseadas em *corpora* (por exemplo, ØVERÅS 1998; OLOHAN E BAKER, 2000; PÁPAI, 2004 E FRANKENBERG-GARCIA, 2009a).

Um fenômeno bem menos explorado é a distribuição diferenciada do léxico dos textos traduzidos e não traduzidos. Num dos poucos estudos existentes, Shama'a (1978, citada em BAKER, 1993) constatou que as palavras inglesas *day* e *say* eram duas vezes mais frequentes em inglês traduzido do árabe do que em inglês original, fazendo com que as traduções soassem diferentes e contribuindo para a identificação desses textos como traduções. Num trabalho mais recente, baseado num *corpus* paralelo, pude notar que o advérbio inglês *already* é duas vezes mais frequente em inglês traduzido do português do que em inglês original (FRANKENBERG-GARCIA, 2004).

No entanto, a sobre-representação de determinado léxico nas traduções não é necessariamente a regra. Em outro estudo baseado num *corpus*, Tirkkonen-Condit (2004) centra a sua análise nos verbos tipicamente finlandeses de suficiência e nota que estes são bem menos frequentes nas traduções do que em textos originalmente escritos em finlandês.

Tanto a sub-representação como a sobre-representação de certos itens lexicais podem afetar a impressão que os leitores têm de um texto traduzido, ou seja, se aparenta ou não ser uma tradução. Num estudo anterior, Tirkkonen-Condit (2002) pediu para um grupo de falantes nativos de finlandês decidir se certos excertos de texto eram originais ou traduções e chegou à conclusão de que tanto as

avaliações corretas como as incorretas baseavam-se principalmente na presença ou ausência de palavras tipicamente finlandesas. Conforme sugerido por Baker (1993), a distribuição diferenciada de certos itens lexicais nas traduções pode dever-se ao confronto existente entre o código-fonte e o código-alvo. Para Toury (1995), a tradução, não sendo nem língua fonte nem propriamente língua alvo, é uma espécie de terceiro código. Embora *a priori* não se associe nenhum juízo de valor ao terceiro código, nas traduções de má qualidade, a distorção do texto devido a uma influência despropositada da língua fonte é o que Baker (1993) chama de *translationese*, ou seja, “tradutês”.

Independentemente de se tratar de um terceiro código ou de uma má tradução, nem sempre é fácil avaliar se existem palavras sobre ou sub-representadas nas traduções. Uma maneira de o fazer seria através de uma abordagem de baixo para cima, tomando uma determinada palavra como ponto de partida e depois comparando a sua distribuição em textos traduzidos e não traduzidos. No entanto, para tal, em primeiro lugar seria preciso decidir que palavras são potencialmente dignas de tal comparação. Na pesquisa de Tirkkonen-Condit (2004), o léxico em análise eram palavras sem equivalentes na língua dos textos que deram origem às traduções finlandesas. Seguindo uma abordagem equivalente de baixo para cima, observei que o verbo inglês *nod*, que não tem uma tradução direta para português, apresenta uma frequência significativamente mais elevada num *corpus* de textos originais em inglês do que num *corpus* de inglês traduzido do português (FRANKENBERG-GARCIA, 2007). Contudo, também poderão existir palavras com equivalentes diretos nas línguas do original e da tradução, mas cuja distribuição em textos traduzidos e não traduzidos seja totalmente distinta. Como já foi referido, a sobre representação das palavras *day* e *say* no inglês traduzido do árabe (SHAMA’A, 1978) e do advérbio *already* no inglês traduzido do português (FRANKENBERG-GARCIA, 2004) são exemplos disso.

Por vezes, os tradutores e os professores de línguas estrangeiras, por trabalharem justamente com línguas em contato, são intuitivamente capazes de identificar este gênero de distribuição lexical diferenciada. Por exemplo, numa discussão informal anterior à realização do presente estudo, uma tradutora profissional portuguesa referiu que tinha a impressão de que os advérbios terminados em *mente* tinham uma frequência excepcionalmente elevada no português traduzido do inglês se comparado com o português original (BASTOS, 2008). Da mesma forma, Stella Tagnin, professora de tradução na Universidade de São Paulo, comentou que tinha a impressão de que o verbo *poder* era utilizado em demasia no português das traduções (TAGNIN, 2008). Para a presente autora, antes da realização deste estudo, os adjetivos *diferente* e *possível* e os advérbios *simplesmente*, *exatamente*, *perfeitamente*

e *absolutamente* pareciam ser involuntariamente frequentes no português traduzido do inglês. É preciso reconhecer, no entanto, que as abordagens de baixo para cima tais como as descritas acima podem ser limitadas, uma vez que poderão existir palavras com distribuições díspares que escapam à nossa percepção.

O presente estudo baseado em *corpus* é uma tentativa de investigar o léxico diferenciado da tradução a partir de uma perspectiva inversa, que vai de cima para baixo. Pretende-se verificar se as nossas impressões sobre palavras com distribuições distintas na tradução se confirmam e, ao mesmo tempo, procurar descobrir outros fenômenos de distribuição lexical marcada que escapam à observação direta.

1. MÉTODO

Numa abordagem de baixo para cima, começaríamos por selecionar uma palavra e a seguir compararíamos a sua frequência em *corpora* comparáveis de textos traduzidos e não traduzidos. Na abordagem de cima para baixo adotada no presente estudo, partimos de um *corpus* comparável de textos traduzidos e não traduzidos para então estabelecer quais as palavras sub e sobre representadas nas traduções. O *corpus* utilizado na presente análise foi o COMPARA versão 10.0, um *corpus* paralelo bi-direcional com um total de três milhões de palavras em inglês e português provenientes de textos literários (FRANKENBERG-GARCIA e SANTOS, 2003)¹. Como já referi em Frankenberg-Garcia (2009b), uma das muitas vantagens dos *corpora* paralelos bi-direcionais é o fato de possibilitarem a comparação não só de duas línguas, mas também dos subconjuntos traduzidos e não traduzidos das línguas em questão. O presente estudo teve em conta 39 excertos de textos originalmente escritos em português no *corpus* (634.601 palavras) e 32 excertos comparáveis em português traduzido do inglês (733.282 palavras)². Encontram-se representados na amostra 22 autores de Portugal, Brasil, Angola e Moçambique e 25 tradutores portugueses e brasileiros³.

1. O *corpus* COMPARA encontra-se disponível em <http://www.linguateca.pt/COMPARA/>

2. Três dos originais do corpus encontram-se alinhados com duas traduções cada: EBDL1, EBDL3 e PBJA1. Para evitar distorções causadas pela contagem em duplicado de textos-fonte com alinhamentos múltiplos, apenas uma das traduções destes textos foi contabilizada no presente estudo. No caso do EBDL1 e do EBDL3, utilizou-se a tradução brasileira, a fim de dar um maior equilíbrio às variantes do português representadas nas traduções. No caso do PBJA1, preferiu-se o alinhamento com a tradução inglesa de 2000 à tradução de 1865, que se demarca completamente das restantes traduções contemporâneas representadas no *corpus*.

3. Nenhuma distinção foi feita em relação às diferentes variantes do português representadas no *corpus*, embora, como se verá adiante, isso poderá ter influenciado alguns dos resultados obtidos.

Os textos em língua portuguesa do COMPARA foram automaticamente anotados com o analisador gramatical PALAVRAS (BICK, 2000) e, no momento deste estudo, a anotação automática estava sendo submetida a uma revisão manual (SANTOS e INÁCIO, 2006). Desta forma, alguns ajustes resultantes dessa revisão posterior à realização do presente estudo poderão eventualmente reverter em pequenas variações em relação aos dados que serão aqui apresentados, mas sem que isso altere os resultados globais obtidos de forma significativa.

O ponto de partida da presente análise consistiu em obter uma distribuição de lemas nos sub-*corpora* de textos traduzidos para português e de textos originais em português do COMPARA (doravante traduções e originais). Selecionaram-se para um exame mais detalhado todos os lemas classificados de acordo com as categorias gramaticais genéricas para nomes (excluindo os nomes próprios), adjetivos, verbos e advérbios (ver tabela 1)⁴. Não foram incluídas na análise palavras gramaticais, tais como artigos, preposições e conjunções.

Tabela 1 - Distribuição de lemas por categoria gramatical no COMPARA 10,0

CATEGORIA GRAMATICAL	Tipos	<i>Tokens</i>	<i>Sub-corpus</i>
ADJETIVO	4215	33,697	originais
	4118	41,042	traduções
VERBO	6857	130,816	originais
	6256	149,042	traduções
NOME	11,465	137,114	originais
	10,517	150,858	traduções
ADVÉRBIO	1025	52,318	originais
	1099	62,136	traduções

Obtidas essas distribuições, consideraram-se insuficientemente representados e conseqüentemente descartaram-se do estudo os lemas com frequência inferior a 10 ocorrências por 100 mil palavras nos sub-*corpora* em análise⁵. No entanto, preservaram-se os lemas que atingiram ou superaram este limiar num dos sub-*corpora* (por exemplo, nos originais), mas não no outro (por exemplo, nas traduções). A amostra resultante engloba 1003 lemas ao todo, distribuídos de acordo com as seguintes categorias gramaticais: 482 nomes, 113 adjetivos, 309 verbos e 99 advérbios.

Palavras com grafias alternativas foram consideradas como um único lema, uma vez que as diferenças ortográficas não são relevantes para um estudo centrado sobre o léxico. Assim sendo, ignoraram-se diferenças ortográficas entre o português

4. Ver Inácio & Santos, 2008, para uma descrição detalhada dessas categorias gramaticais.

5. Em números absolutos, o ponto de corte equivale a um mínimo de 73 ocorrências no *corpus* de traduções e 63 ocorrências no *corpus* de originais.

do Brasil e o português europeu, como *direção e direcção*⁶ e outras grafias alternativas, tais como *loiro e louro*, ainda que no COMPARA estas formas sejam tratadas como lemas distintos. Os empréstimos que não fazem oficialmente parte da língua portuguesa, tal como o pronome de tratamento inglês *sir*, também foram excluídos do estudo⁷.

A seguir, calculou-se a frequência por 100 mil palavras de cada um dos lemas acima nos originais e nas traduções, a fim de os comparar e determinar o grau de diferença entre cada caso. Consideraram-se sobre representados na tradução os lemas que ocorressem pelo menos duas vezes mais do que ocorreram no *corpus* dos originais. Seguindo esta mesma lógica, consideraram-se sub-representados na tradução os lemas com uma frequência igual ou menor do que a metade no *corpus* de originais. Os lemas sobre e sub-representados foram então inspecionados em mais detalhe.

Tendo em conta que alguns autores estão mais representados do que outros no COMPARA, foi necessário determinar se estariam em causa distorções dos resultados devidas ao estilo próprio de um autor específico. Se mais de um terço das ocorrências de um dado lema provinha de um único autor, os resultados para este lema deixaram de ser considerados no estudo. Não se julgou necessário proceder a um controle deste tipo de distorção em relação aos tradutores representados no *corpus*, uma vez que estes se encontram razoavelmente bem distribuídos. Os resultados obtidos encontram-se descritos nas próximas quatro seções.

2. NOMES DIFERENCIADOS

Dos 482 lemas nominais com uma frequência suficiente que justificasse uma análise mais pormenorizada, 137 foram inicialmente considerados sobre ou sub-representados nas traduções. Contudo, 46 lemas foram excluídos do estudo, já que mais de um terço das suas ocorrências provinham de um único autor. Na tabela 2 encontram-se os 42 nomes sobre representados, enquanto a tabela 3 contém os 49 nomes sub-representados que restaram.

6. Os textos do corpus são anteriores à recente implementação do Acordo Ortográfico de 1990.

7. Contudo, ver FRANKENBERG-GARCIA, 2009c, para um estudo aprofundado sobre a utilização de empréstimos na tradução literária.

Tabela 2 - Lemas nominais sobre-representados nas traduções⁹

NOME	ORIGINAIS		TRADUÇÕES		Dif. T/O
	F	F rel.	F	F rel.	
g(ê)nero	20	3,15	108	14,73	4,7
fa(c)to ¹⁰	76	11,98	377	51,41	4,3
Plástico	15	2,36	74	10,09	4,3
bocado	35	5,52	145	19,77	3,6
Membro	25	3,94	100	13,64	3,5
problema	54	8,51	215	29,32	3,4
escola	44	6,93	168	22,91	3,3
medida	31	4,88	118	16,09	3,3
bebida	20	3,15	74	10,09	3,2
aspecto	50	7,88	180	24,55	3,1
maioria	24	3,78	86	11,73	3,1
início	21	3,31	74	10,09	3,0
altura	103	16,23	359	48,96	3,0
quadro	34	5,36	116	15,82	3,0
emprego	30	4,73	100	13,64	2,9
possibilidade	30	4,73	95	12,96	2,7
Procura	36	5,67	110	15,00	2,6
recordação	27	4,25	82	11,18	2,6
casaco	29	4,57	87	11,86	2,6
aldeia	26	4,10	74	10,09	2,5
oportunidade	35	5,52	99	13,50	2,4
peça	52	8,19	145	19,77	2,4
cozinha	81	12,76	224	30,55	2,4
espécie	106	16,70	293	39,96	2,4
lista	28	4,41	77	10,50	2,4
sítio	57	8,98	148	20,18	2,2
discussão	33	5,20	83	11,32	2,2
semana	117	18,44	289	39,41	2,1
rapariga	122	19,22	301	41,05	2,1
ajuda	33	5,20	81	11,05	2,1
ombro	123	19,38	297	40,50	2,1
inglês	39	6,15	94	12,82	2,1
segurança	34	5,36	81	11,05	2,1
grupo	84	13,24	200	27,27	2,1
Modo	159	25,06	376	51,28	2,0
tom	88	13,87	207	28,23	2,0
questão	74	11,66	173	23,59	2,0
tipo	103	16,23	240	32,73	2,0
difficuldade	44	6,93	102	13,91	2,0
expressão	73	11,50	167	22,77	2,0
atitude	35	5,52	80	10,91	2,0
cortina	37	5,83	84	11,46	2,0

8. F= frequência no corpus; F rel. = frequência relativa por 100 mil palavras; Dif. T/O = frequência relativa nas traduções dividida pela frequência relativa nos originais (e nas tabelas subsequentes, Diff O/T= frequência relativa nos originais dividida pela frequência relativa nas traduções).
9. Embora *fato* signifique *terno* em português europeu, quando a palavra é grafada sem a letra C só se contabilizaram as ocorrências provenientes do português do Brasil, equivalentes a *facto* em Portugal. Para todos os outros lemas em apreço, não se levou em conta a polissemia.

Tabela 3 – Lemas nominais sub-representados nas traduções

NOME	ORIGINAIS		TRADUÇÕES		DIF O/T
	F	F Rel	F	F Rel	
sobrinho	65	10,24	6	0,82	4,2
lembrança	84	13,24	10	1,36	4,1
moço	79	12,45	8	1,09	4,1
menino	206	32,46	34	4,64	3,9
velha	108	17,02	18	2,45	3,7
soldado	110	17,33	24	3,27	3,6
crime	154	24,27	36	4,91	3,5
saudade	86	13,55	22	3,00	3,3
remédio	67	10,56	16	2,18	3,3
praça	101	15,92	26	3,55	3,1
prédio	87	13,71	27	3,68	3,0
fogo	124	19,54	37	5,05	3,0
português	68	10,72	19	2,59	3,0
cavalo	120	18,91	46	6,27	2,9
diabo	99	15,60	35	4,77	2,9
menina	141	22,22	47	6,41	2,9
o(u i)ro	177	27,89	57	7,77	2,9
prata	82	12,92	33	4,50	2,8
arma	108	17,02	42	5,73	2,8
velho	278	43,81	107	14,59	2,8
padre	201	31,67	82	11,18	2,6
colégio	67	10,56	27	3,68	2,6
pedra	289	45,54	119	16,23	2,5
povo	93	14,65	41	5,59	2,4
alma	273	43,02	114	15,55	2,4
sangue	200	31,52	99	13,50	2,3
mistério	89	14,02	44	6,00	2,3
tristeza	67	10,56	33	4,50	2,3
cheiro	137	21,59	67	9,14	2,3
senhora	320	50,43	154	21,00	2,3
fome	78	12,29	36	4,91	2,3
dono	90	14,18	40	5,45	2,3
senhor	486	76,58	249	33,96	2,2
alto	71	11,19	36	4,91	2,2
varanda	85	13,39	43	5,86	2,2
unha	72	11,35	36	4,91	2,2
sonho	182	28,68	98	13,36	2,1
rua	403	63,50	215	29,32	2,1
graça	108	17,02	57	7,77	2,1
primo	65	10,24	34	4,64	2,1
coração	289	45,54	149	20,32	2,1
dente	160	25,21	92	12,55	2,0
ordem	159	25,06	91	12,41	2,0
seda	86	13,55	49	6,68	2,0
rei	67	10,56	38	5,18	2,0
carne	106	16,70	59	8,05	2,0
contrário	105	16,55	58	7,91	2,0
boca	296	46,64	162	22,09	2,0
letra	87	13,71	47	6,41	2,0

Os resultados na tabela 2 mostram que a maioria dos lemas nominais sobre-representados consiste de nomes abstratos. O mais notável de todos é *gé(ê)nero*, que, por sinal, pode ser considerado sinônimo de mais dois nomes sobre representados: *espécie* e *tipo*. Vários dos lemas nominais constantes nessa tabela também transmitem a ideia geral de modo (por exemplo, *tom*, *modo*, *expressão*, *aspecto* e *atitude*), e muito deles são usados para classificar e agrupar (por exemplo, *membro*, *grupo*, *lista* e *maioria*). Em contrapartida, como mostra a tabela 3, a maioria dos nomes sub-representados refere-se a seres humanos. Sem surpresas, há também vários nomes sub-representados na tradução que facilmente se associam à psique que permeia a língua portuguesa: *lembrança*, *saudade*, *tristeza* e *alma*.

Também é interessante notar a presença de *quasi*-sinônimos em extremidades opostas das distribuições: *rapariga* (sobre representado) e *menina* (sub-representado), *recordação* (sobre representado) e *lembrança* (sub-representado), *escola* (sobre-representado) e *colégio* (sub-representado).¹⁰

3. ADJETIVOS DIFERENCIADOS

Dos 113 adjetivos selecionados na amostra, apenas dois lemas tiveram de ser excluídos do estudo devido ao fato de mais de um terço das suas ocorrências provirem de um único autor. Restaram 11 adjetivos com pelo menos o dobro de ocorrências nas traduções (ver tabela 4), e 10 adjetivos com pelo menos o dobro de ocorrências nos originais (ver tabela 5).

Tabela 4 - Lemas adjetivais sobre-representados nas traduções

ADJETIVO	ORIGINAIS		TRADUÇÕES		DIF T/O
	F	F Rel	F	F Rel	
sentado	31	4,88	172	23,46	4,8
calmo	18	2,84	91	12,41	4,4
maravilhoso	18	2,84	85	11,59	4,1
evidente	23	3,62	90	12,27	3,4
familiar	22	3,47	83	11,32	3,3
pessoal	21	3,31	74	10,09	3,0
especial	36	5,67	119	16,23	2,9
horível	26	4,10	82	11,18	2,7
jovem	54	8,51	147	20,05	2,4
suficiente	37	5,83	94	12,82	2,2
principal	47	7,41	108	14,73	2,0

10. A presença marcada de *rapariga* no corpus de traduções pode ser em parte explicada pelo fato de que as traduções do *corpus* provêm principalmente de Portugal, onde esta palavra é muito mais

Tabela 5 - Lemas adjetivais sub-representados nas traduções

ADJETIVO	ORIGINAIS		TRADUÇÕES		DIF T/O
	F	F Rel	F	F Rel	
gordo	73	11,50	26	3,55	3,2
grosso	98	15,44	39	5,32	2,9
igual	110	17,33	45	6,14	2,8
nu	127	20,01	58	7,91	2,5
doce	72	11,35	35	4,77	2,4
raro	72	11,35	36	4,91	2,3
triste	153	24,11	79	10,77	2,2
rico	103	16,23	53	7,23	2,2
alegre	71	11,19	42	5,73	2,0
morto	71	11,19	41	5,59	2,0

Como mostram os resultados, o adjetivo mais sobre-representado nas traduções é *sentado*, e o mais sub representado é *gordo*. Curiosamente, a maioria dos adjetivos sobre-representados (com exceção de *sentado* e *jovem*) são adjetivos que refletem opiniões pessoais e sentimentos. Em contrapartida, a maioria dos adjetivos sub-representados parece transmitir descrições concretas.

4. VERBOS DIFERENCIADOS

Nenhum dos 309 verbos com frequências iguais ou acima do limiar de análise pré-estabelecido teve de ser descartado por distorções devido à influência de um único autor. Ao contrário dos nomes e adjetivos até agora analisados, entre os verbos parece haver uma tendência maior para a sobre representação do que para a sub-representação. Apresentam-se os 32 verbos com pelo menos o dobro das ocorrências nas traduções e os 19 verbos com pelo menos o dobro das ocorrências nos originais nas tabelas 6 e 7, respectivamente.

comum do que em português do Brasil. O fato de não se ter separado as diferentes variantes do português pode ter constituído uma variável interveniente na distribuição de alguns lemas.

Tabela 6 - Lemas verbais sobre-representados nas traduções

VERBO	ORIGINAIS		TRADUÇÕES		DIF T/O
	F	F Rel	F	F Rel	
encontrar-se	10	1,58	87	11,86	7,5
acenar	18	2,84	101	13,77	4,9
constituir	16	2,52	83	11,32	4,5
inclinarse	19	2,99	92	12,55	4,2
sentir-me	32	5,04	135	18,41	3,7
tornar-se	46	7,25	186	25,37	3,5
replicar	24	3,78	95	12,96	3,4
abandar	28	4,41	100	13,64	3,1
sentir-se	57	8,98	203	27,68	3,1
comentar	37	5,83	128	17,46	3,0
regressar	55	8,67	168	22,91	2,6
sugerir	35	5,52	104	14,18	2,6
dirigir-se	45	7,09	130	17,73	2,5
preocupar	33	5,20	94	12,82	2,5
baixar	30	4,73	85	11,59	2,5
virar-se	27	4,25	76	10,36	2,4
apanhar	67	10,56	183	24,96	2,4
compreender	115	18,12	311	42,41	2,3
conseguir	302	47,59	764	104,19	2,2
fazê-lo	82	12,92	207	28,23	2,2
apoiar	36	5,67	90	12,27	2,2
manter	80	12,61	193	26,32	2,1
lamentar	34	5,36	82	11,18	2,1
exclamar	76	11,98	183	24,96	2,1
provocar	38	5,99	91	12,41	2,1
arranjar	77	12,13	183	24,96	2,1
permitir	75	11,82	178	24,27	2,1
revelar	49	7,72	116	15,82	2,0
tentar	225	35,46	525	71,60	2,0
verificar	34	5,36	78	10,64	2,0
representar	48	7,56	110	15,00	2,0
voltar-se	50	7,88	114	15,55	2,0

Tabela 7 - Lemas verbais sub-representados nas traduções

VERBO	ORIGINAIS		TRADUÇÕES		DIF
	F	F Rel	F	F Rel	O/T
vencer	78	12,29	22	3,00	4,1
cuidar	134	21,12	44	6,00	3,5
sonhar	124	19,54	42	5,73	3,4
morar	136	21,43	53	7,23	3,0
recolher	77	12,13	31	4,23	2,9
fugir	230	36,24	91	12,41	2,9
roubar	93	14,65	41	5,59	2,6
beijar	69	10,87	33	4,50	2,4
entender	262	41,29	131	17,86	2,3
inventar	68	10,72	34	4,64	2,3
amar	168	26,47	89	12,14	2,2
faltar	144	22,69	74	10,09	2,2
chorar	220	34,67	123	16,77	2,1
quebrar	68	10,72	38	5,18	2,1
bastar	107	16,86	59	8,05	2,1
conversar	171	26,95	93	12,68	2,1
confessar	100	15,76	59	8,05	2,0
cantar	113	17,81	66	9,00	2,0
cumprir	89	14,02	51	6,96	2,0

Como se pode constatar nas tabelas 6 e 7, a anotação morfossintática do *corpus* COMPARA classifica como lemas separados os verbos seguidos de pronomes clíticos diferentes. Por exemplo, *sentir-se* e *sentir-me* são analisados separadamente. Ao contrário das diferenças ortográficas, decidiu-se manter esta distinção dos clíticos na medida em que as consideramos relevantes para o presente estudo.

Os resultados obtidos trazem à tona uma série de tendências. O verbo mais sobre representado nas traduções é o verbo de ligação *encontrar-se*, e há vários verbos de ligação com pelo menos o dobro da frequência no português traduzido: *constituir*, *tomar-se*, *sentir-se*, *sentir-me*, *fazê-lo*, *representar* e *manter*. Dois outros grupos que se destacam entre os verbos sobre representados são os verbos dicendi (*revelar*, *exclamar*, *lamentar*, *sugerir*, *comentar* e *replicar*) e os verbos de movimento (*inclinarse*, *regressar*, *dirigir-se*, *baixar*, *virar-se*, *apanhar*, *apoiar*, *voltar-se*, *acenar* e *abanar*). Entre os lemas verbais com pelo menos o dobro da frequência nas traduções, encontramos também verbos que precedem outros verbos (*tentar*, *conseguir* e *permitir*).

Em contrapartida, a maioria dos lemas verbais sub-representados nas traduções são de elevado teor lexical, sendo que muitos deles são típicos da

linguagem dramática da literatura (*vencer, fugir, beijar, cantar, quebrar, sonhar, amar, roubar, chorar, matar, morrer e nascer*). Mais uma vez, detectaram-se *quasi-sinônimos* com distribuições opostas: *compreender* e *apanhar* (sobre representados), mas *entender* e *recolher* (sub-representados).

5. ADVÉRBIOS DIFERENCIADOS

Nenhum dos 99 advérbios selecionados para este estudo tiveram de ser excluídos devido a distorções por causa de um único autor. No total, 13 foram considerados sobre representados e 10 foram classificados como sub-representados nas traduções. Nas tabelas 8 e 9, respectivamente, apresentam-se os advérbios com pelo menos o dobro das ocorrências nas traduções e os advérbios com pelo menos o dobro das ocorrências nos originais.

Tabela 8 - Lemas adverbiais sobre-representados nas traduções

ADVÉRBIO	Originais		Traduções		Dif T/O
	F	F Rel	F	F Rel	
demasiado	22	3,47	194	26,46	7,6
profundamente	14	2,21	79	10,77	4,9
bastante	22	3,47	112	15,27	4,4
claro	73	11,50	320	43,64	3,8
absolutamente	23	3,62	79	10,77	3,0
completamente	48	7,56	161	21,96	2,9
simplesmente	36	5,67	120	16,36	2,9
perfeitamente	41	6,46	127	17,32	2,7
acima	36	5,67	104	14,18	2,5
imediatamente	54	8,51	148	20,18	2,4
sequer	72	11,35	195	26,59	2,3
exa(c)tamente	57	8,98	153	20,87	2,3
através	91	14,34	230	31,37	2,2

Tabela 9 – Le as adverbiais sub-representados nas traduções

ADVÉRBIO	Originais		Traduções		Dif O/T
	F	F Rel	F	F Rel	
enfim	169	26,63	35	4,77	5,6
logo	567	89,35	215	29,32	3,0
ora	226	35,61	102	13,91	2,6
ontem	98	15,44	44	6,00	2,6
jamais	80	12,61	38	5,18	2,4
amanhã	121	19,07	60	8,18	2,3
porém	297	46,80	162	22,09	2,1
hoje	330	52,00	192	26,18	2,0
toda	76	11,98	43	5,86	2,0
todo	101	15,92	57	7,77	2,0

Como indica a tabela 8, mais de metade dos advérbios sobre representados terminam em *mente* e praticamente todos eles são advérbios de modo. Vale a pena notar que os advérbios sobre representados *absolutamente*, *completamente*, *simplesmente*, *perfeitamente*, *imediatamente* e *exa(c)tamente* são muito semelhantes, do ponto de vista fonético e morfológico, aos seus equivalentes ingleses. Em contrapartida, nenhum dos advérbios sub-representados termina em *mente* e a maior parte deles expressam tempo e frequência. Mais uma vez, verifica-se a ocorrência de *quasi*-sinônimos com distribuições opostas, com a sub-representação de *todo* e *toda* e a sobre-representação do advérbio *completamente*.

6. DISCUSSÃO

Surgem vários pontos de discussão quando olhamos para os resultados globais obtidos relativamente às quatro categorias gramaticais analisadas neste estudo. Em primeiro lugar, é interessante notar que praticamente só os nomes sofreram da influência desproporcionada de um único autor, tendo sido necessário excluir da análise 46 lemas nominais. Apenas dois adjetivos, mas nenhum verbo ou advérbio, tiveram de ser eliminados do estudo. Na realidade, o uso aparentemente excessivo de alguns nomes por parte de certos autores não se deveu apenas a uma questão de estilo e preferência vocabular. Em muitos casos, esses nomes têm estritamente a ver com o enredo das histórias. Por exemplo, 90% das ocorrências do nome *cego* (que por razões já referidas não foi contabilizado no estudo) provêm do romance *Ensaio sobre a Cegueira*, de José Saramago.

Os lemas nominais também se destacam das outras categorias gramaticais em apreço na medida em que são os únicos em que a sub-representação de lemas na tradução foi superior à sobre-representação. Para os adjetivos e advérbios, o número de lemas sobre-representados na tradução é muito semelhante ao número de lemas sub-representados. No entanto, em relação aos verbos, ocorre justamente o contrário: há substancialmente mais lemas sobre-representados do que sub-representados. Estes resultados vão de encontro à ideia generalizada de que o português é uma língua mais nominal, enquanto o inglês, o idioma dos textos-fonte que deram origem ao *corpus* de traduções, é mais verbal.

Outro resultado assinalável é a presença de *quasi*-sinônimos com distribuições opostas. Com a exceção do par *menina-rapariga*, que, como já foi referido, provavelmente tem a ver com as diferenças entre o português do Brasil e de Portugal, os restantes pares sinônimos servem para chamar a atenção para o fato de que uma das palavras é mais típica da literatura não traduzida do que a outra. Vale a pena lembrar que os pares sinônimos destacados até agora são apenas aqueles em que uma das palavras do par foi pelo menos duas vezes mais frequente em português original ou duas vezes mais frequente em português traduzido. Se baixarmos esse limiar para um valor menos restritivo, encontraremos muito mais pares sinônimos com distribuições marcadamente distintas. Na tabela 10 encontram-se pares sinônimos em que uma das palavras do par era pelo menos 1,5 vezes mais frequente do que a outra. Repare-se na presença de um certo contraste estilístico entre alguns dos pares, já que, de modo geral, os lemas preferidos nas traduções parecem ser mais formais do que os seus correspondentes nos originais, nomeadamente, *recordação* vs *lembrança*, *edifício* vs *prédio*, *compreender* vs. *entender* e *recordar* vs. *lembrar*. Isto poderia ser interpretado como um sinal de que os tradutores portugueses favorecem uma linguagem mais formal do que seria natural em português original.

Tabela 10 – Quasi-sinônimos com distribuições contrastivas em originais e traduções

CATEGORIA GRAMATICAL	Lemas pelo menos 1,5 x mais frequentes nas traduções	Lemas pelo menos 1,5 x mais frequentes nos originais
Nome	rapariga	menina
	recordação	lembrança
	escola	colégio
	edifício	prédio
Adjetivo	enorme	imenso
Verbo	compreender	entender
	Recordar	lembrar
	reparar	notar
	observar	examinar
	decidir	resolver
	obrigar	mandar
	manter	guardar
Advérbio	apanhar	recolher
	completamente	todo/toda
	finalmente	enfim/afinal

Os resultados desta análise também chamam a atenção para uma série de contrastes semânticos. Nas traduções, prevalecem os nomes abstratos (por exemplo, *g(ê|ê)nero* e *fa(c)to*), enquanto os nomes alusivos a seres humanos ficam sub-representados (por exemplo, *sobrinho* e *menino*). Também predominam os adjetivos de opinião (por exemplo, *calmo* e *maravilhoso*), ao passo que os adjetivos concretos são comparativamente escassos (por exemplo, *gordo* e *grosso*). Em relação aos advérbios, destacam-se nas traduções os advérbios de forma e destacam-se pela negativa os advérbios de tempo e frequência. Quanto aos verbos, os verbos de ligação, os verbos *dicendi*, os verbos de movimento e os verbos que precedem outras formas verbais se destacam nas traduções, enquanto os verbos de elevado grau lexical e os verbos de significados proposicionais dramáticos marcam as traduções pela sua escassez, revelando, desta forma, possíveis diferenças culturais entre a literatura original e traduzida.

Além das constatações acima, o presente estudo reforça algumas das impressões intuitivas relativamente a certas especificidades do léxico do português traduzido. Confirmando as impressões de Bastos (2008), e da presente autora, há vários advérbios terminados em *mente* sobre-representados nas traduções. Os adjetivos *diferente* e *possível*, que também tinham sido alegadamente identificados como sendo excessivos nas traduções do inglês, não chegam a atingir o dobro da frequência proposto neste estudo, mas verificou-se que o primeiro é 1,7 vezes e o segundo 1,9 vezes mais frequente nas traduções do que nos originais. Talvez isso seja até um indício de que o critério de considerar apenas os lemas duas vezes

mais frequentes nas traduções ou duas vezes mais frequentes nos originais tenha sido excessivamente rigoroso. Embora o verbo *poder*, que Tagnin (2008) supôs ser utilizado em excesso nas traduções, não tenha sido considerado sobre-representado (tendo ocorrido apenas 1,2 vezes mais nas traduções), os seus sinônimos *conseguir* e *permitir* apresentaram o dobro da frequência nos textos traduzidos.

Apesar de não se ter antecipado nenhuma palavra em particular que pudesse *a priori* soar escassa nas traduções, a abordagem exploratória e de cima para baixo adotada neste estudo revelou um elevado número de lemas pelo menos duas vezes mais frequentes nos originais do que nas traduções. Enquanto alguns destes eram completamente imprevisíveis, outros, como os lemas notoriamente associados à alma e à psique portuguesa (*saudade, triste, tristeza, alma, lembrança, sonbar* e *sonbo*) fazem todo o sentido. Da mesma forma, olhando retrospectivamente enquanto tradutora e professora de tradução e de língua inglesa, a relativa escassez dos advérbios *hoje, ontem* e *amanhã* nas traduções para português não surpreende, uma vez que intuitivamente essas palavras parecem ser usadas com mais frequência no português.

CONCLUSÕES E PERSPECTIVAS PARA O FUTURO

Este estudo procurou analisar o léxico diferenciado da tradução a partir de uma perspectiva exploratória de cima para baixo. A análise tomou como ponto de partida dois *corpora* comparáveis de português original e português traduzido do inglês, para então identificar quais eram os lemas nominais, verbais, adjetivais e adverbiais mais sobre- e sub-representados nas traduções.

Os resultados obtidos neste estudo não só reforçam impressões existentes relativamente a palavras sobre-representadas na tradução, como também revelam uma rede imbricada de contrastes linguísticos e culturais entre o português original e o português traduzido que não teria sido possível detectar a olho nu.

Enquanto alguns dos resultados aqui apresentados, tais como a preferência por certas palavras (algumas das quais apontam também para diferenças culturais e estilísticas) no português não traduzido, poderão ter um impacto imediato sobre o desenvolvimento de ferramentas multilíngues, tradução automática, auxiliares de tradução e formação de tradutores, muitos dos contrastes identificados requerem mais investigação no futuro. Em particular, deve-se ter em conta que uma análise baseada em lemas é muito genérica, e é preciso explorar mais a fundo possíveis contrastes entre palavras flexionadas e os diferentes significados de alguns lemas polissêmicos. Além disso, seria importante desenvolver a pesquisa lexical isolada e desprovida de co-texto efetuada neste estudo com um maior detalhamento em

termos de análises colocacionais e fraseológicas. Os verbos sobre-representados *acenar* e *abandar*, por exemplo, co-ocorrem quase sempre com *cabeça*. Se examinarmos estes lemas verbais em contexto, chega-se à conclusão de que não são apenas os verbos em si que estão sobre-representados no português das traduções, mas sim as unidades fraseológicas nas quais estes verbos se inserem, nomeadamente, *acenar a cabeça* e *abandar a cabeça*.

Embora a pesquisa de formas flexionadas, polissemia e fraseologia contrastiva ultrapasse o âmbito deste trabalho, espera-se que a metodologia desenvolvida e os resultados obtidos possam servir de estímulo para estudos futuros, bem como para estudos análogos a partir de *corpora* de gêneros textuais e línguas diferentes.

REFERÊNCIAS BIBLIOGRÁFICAS

- BAKER, M. (1993) *Corpus linguistics and translation studies. Implications and applications*. In: Baker, M., Francis G. e Tognini Bonelli, E. (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam e Philadelphia: John Benjamins, p. 233-250.
- BASTOS, A. (2008) Comunicação pessoal. Fevereiro de 2008.
- BICK, E. (2000) *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Århus University. Århus: Århus University Press.
- BLUM-KULKA, S. (1986) Shifts of cohesion and coherence in translation. In: House, J. e Blum-Kulka, S. (eds.) *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, p. 17-35.
- FRANKENBERG-GARCIA, A. (2004) Lost in Parallel Concordances. In: Aston, G., Bernardini, S. e Stewart, D. (eds.) *Corpora and language learners*. Amsterdam e Philadelphia: John Benjamins, p. 213-229.
- FRANKENBERG-GARCIA, A. (2007) *Building a parallel corpus for translation research and much more*. Apresentação convidada no seminário de mestrado em Estudos de Tradução, Universitat Jaume I, Castellón, Espanha, novembro de 2007.
- FRANKENBERG-GARCIA, A. (2009a) Are translations longer than source texts? A corpus-based study of explicitation In: Beeby, A., Rodríguez P., & Sánchez-Gijón, P. (eds.) *Corpus use and learning to translate (CULT): An Introduction*. Amsterdam & Philadelphia: John Benjamins, p. 47-58.
- FRANKENBERG-GARCIA, A. (2009b) Compiling and using a parallel corpus for research in translation. *International Journal of Translation*, v.21, n. 1-2, p. 57-71.
- FRANKENBERG-GARCIA, A. (2009c) Utilização de empréstimos na tradução literária. *Tradução & Comunicação – Revista Brasileira de Tradutores*, n.18, p. 43-61.
- FRANKENBERG-GARCIA, A. e SANTOS, D. (2003) Introducing COMPARA, the Portuguese-English Parallel Corpus. In: Zanettin, F., Bernardini, S. e Stewart, D. (eds.) *Corpora in Translator Education*. Manchester: St. Jerome, p. 71-87.

- INÁCIO, S. e SANTOS, D. (2008) Documentação da anotação morfossintáctica da parte portuguesa do COMPARA. Disponível em:
<http://www.linguateca.pt/COMPARA/DocAnotacaoPortCOMPARA.pdf>
Acesso em: 6 jul. 2011.
- OLOHAN, M. e BAKER, M. (2000) Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures* v.1, n. 2, p.141-158.
- ØVERÅS, L. (1998) In Search of the Third Code: an investigation of norms in literary translation. *Meta*, v. 43, n.4, p. 557-570.
- PÁPAI, V. (2004) Explicitation: A universal of translated text? In: Mauranen, A. e Kujamäki, P. (eds.) *Translation Universals: Do They Exist?* Amsterdam e Philadelphia: John Benjamins, p. 143-164.
- SANTOS, D. e INÁCIO, S. (2006) Annotating COMPARA, a grammar-aware parallel corpus. In: Calzolari, N., Choukri, K., Gangemi, A., Maegaard, B., Mariani, J., Odjik, J. e Tapias, D. (eds.) *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)*, p. 1216-1221.
- SHAMA'A, N. (1978) *A linguistic analysis of some problems of Arabic to English translation*. D. Phil thesis, Oxford University.
- TIRKKONEN-CONDIT, S. (2004) Unique items – over – or under-represented in translated language? In: Mauranen, A. e Kujamäki, P. (eds.) *Translation Universals, Do They Exist?* Amsterdam e Philadelphia: John Benjamins, p. 177-184.
- TIRKKONEN-CONDIT, S. (2002) Translationese, a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target*, v.14, n.2, p. 207-220.
- TAGNIN, S. (2008) Comunicação pessoal, fevereiro de 2008.
- VINAY, J. P. e DARBELNET, J. (1958) *Stylistique Comparée du Français et de l'Anglais: Méthode de Traduction*. Paris: Didier.

Recebido: 06/07/2011

Aceito: 09/11/2011