

USO DO CONJUNTO LÉXICO POR UM TRADUTOR LITERÁRIO EM
THE WAR OF THE SAINTS

Diva Cardoso de Camargo (UNESP)

RESUMO: Por meio da razão forma/item (type/token ratio), observamos algumas evidências de um estilo próprio de um tradutor literário em termos de variação individual e recorrente de vocabulário (Baker 2000; Berber Sardinha 1997). Se considerarmos a razão forma/item e a razão forma/item padronizada como indicações do emprego que o tradutor faz da linguagem, podem-se destacar, apesar da influência de possíveis variáveis, a diferença menor de 5,64 e a de 5,93 apresentadas por Gregory Rabassa em relação a Jorge Amado como marcas significativas da utilização de padrões estilísticos preferenciais e distintivos desse tradutor, as quais revelam o impacto da extensão dessas diferenças em *The War of the Saints* em comparação com *O Sumiço da Santa*.

PALAVRAS-CHAVE: Léxico e Tradução, Léxico e Linguística de Corpus, Estudos da Tradução Baseados em Corpus, Estilo do tradutor, Literatura Brasileira traduzida.

Tomando-se as pesquisas realizadas no âmbito da disciplina dos estudos da tradução, a maior parte dos projetos estava, até recentemente, direcionada para a avaliação da qualidade do texto traduzido (TT), não levando em conta o que realmente diferencia a tradução de outros tipos de comunicação nem procurando distinguir suas características inerentes. Admitindo-se que, atualmente, muitos fatores externos estão sendo considerados e não simplesmente se as palavras do texto original (TO) foram reproduzidas no TT, ainda existe uma lacuna no tocante a investigações que explorem a natureza da tradução e evidenciem o uso do conjunto léxico e marcas estilísticas deixadas pelo tradutor na obra traduzida. Por essa razão, cresce a necessidade de estudos descritivos na área, e é precisamen-

te para esse enfoque que os estudos da tradução baseados em corpus têm-se voltado ultimamente.

A utilização de grandes corpora de tradução em formato eletrônico pode constituir-se numa abordagem teórica e metodológica para a disciplina não apenas por estar fundamentada em um novo conceito teórico de pesquisas descritivas baseadas na língua em uso mas também por aceitar os princípios da análise de corpus como metodologia. Os primeiros trabalhos que se valeram de corpora de TTs para investigações da tradução como um fenômeno *sui generis* foram propostos por Baker em 1993. A autora enfatiza que corpora extensos oferecem

uma oportunidade única para observar o seu objeto de estudo e explorar o que o diferencia de outros objetos de estudo, como a língua em geral ou outros tipos de interação cultural. Também permitirão, em uma escala maior do que era possível anteriormente, explorar os princípios que governam o comportamento tradutório e as amarras em que opera! (BAKER, 1993, p. 235).

Embora seja recente a aplicação de técnicas de corpus para o estudo da tradução, várias dissertações e teses já foram defendidas nesse campo, e também um número considerável de publicações tem-se dedicado a esse assunto. Dentre elas, a tese de Laviosa-Braithwaite (1996) mostra que aplicações práticas de pesquisas baseadas em corpus podem ser úteis para tradutores profissionais, alunos e professores de tradução, bem como podem oferecer subsídios para pesquisadores e estudiosos da tradução, da lexicografia e da terminologia. De acordo com Laviosa (2002), são úteis na medida em que fornecem conhecimentos para a utilização de técnicas e de ferramentas que podem contribuir para o desenvolvimento profissional; são inspiradoras porquanto revelam fatos novos e consistentes, baseados em fundamentação empírica, sobre o processo e o produto tradutório e sobre o uso dos vocábulos nos TOs e TTs. A respeito de publicações sobre o tema, o conhecido periódico *Meta* dedicou um número especial intitulado *The Corpus-based approach*. Dentre os artigos contidos nesse número, o de Tymoczko (1998, p. 652) afirma que essa nova abordagem é importante para a sustentação e o desenvolvimento da disciplina Estudos da Tradução neste século. A autora ressalta que um corpus computadorizado é uma forma eficiente para armazenar, acessar e pesquisar

uma vasta quantidade de informações, muito maior do que seria possível a um ser humano examinar manualmente, sem auxílio de ferramentas eletrônicas (ibid., p. 653).

Na opinião de Berber Sardinha (2003, p. 43), os estudos de tradução baseados em corpus estabeleceram-se como área de concentração devido ao impacto causado ao lidar com aspectos cruciais da tradução a partir de uma perspectiva empírica, permitindo um melhor entendimento sobre a constituição da linguagem. Também sugere que a exploração de “corpora eletrônicos possibilita aos estudantes de tradução tornarem-se conscientes tanto da complexidade como da especificidade das escolhas lingüísticas envolvidas no processo de tradução” (BERBER SARDINHA, 2003, p. 43).

Compartilhando com a visão de outros lingüistas de corpus, Hunston afirma que os corpora têm muito a oferecer aos tradutores “não somente porque fornecem evidências do emprego das palavras e de traduções para uma dada palavra ou expressão mas também porque possibilitam uma percepção do processo e da natureza da tradução *per se*” (HUNSTON, 2002, p. 128).

Por essa perspectiva, o cenário internacional hoje no que concerne aos estudos da tradução mostra uma alta demanda por programas avançados de ferramentas de busca e, principalmente, por novas metodologias de pesquisa empregando corpora computadorizados de textos traduzidos de e para várias línguas.

Dessa forma, o presente trabalho não está voltado para o sentido tradicional de examinar se o estilo do autor foi adequadamente transposto ‘na’ tradução, mas, sim, direciona o foco para um estilo ‘de’ tradução e o uso do conjunto léxico no TT. Apesar das dificuldades que uma investigação de um estilo ‘de’ tradução possa acarretar, a importância, para a disciplina, dessa mudança de enfoque aponta para a necessidade de estudos que proponham uma abordagem que possibilite a identificação da presença do tradutor no TT, referente ao uso de padrões estilísticos próprios, independentemente do estilo do autor do TO e do par lingüístico envolvido.

Outrossim, pela abrangência do perfil social de Rabassa, justifica-se o estudo do seu perfil estilístico, uma vez que, no campo da tradução, tem recebido da crítica constantes elogios para mais de quarenta traduções para o inglês de obras literárias escritas em português e espanhol. Aumenta a significância deste estudo o fato de que o tradutor literário selecionado para análise tanto contribuiu não só para a maior repercussão,

em países de língua inglesa, das obras de Jorge Amado, como também contribuiu para maior projeção, no cenário da literatura mundial, de outros autores da literatura brasileira, como Machado de Assis e Osman Lins. Nesse sentido, tal repercussão poderia ser decorrência, justamente, do seu uso individual do conjunto léxico e de padrões de estilo próprios, preferenciais, distintivos e recorrentes.

1. PERSPECTIVA TEÓRICA

O presente trabalho fundamenta-se no arcabouço teórico-metodológico lançado por Baker (1993, 1995, 1996, 2000), nos estudos de Berber Sardinha (1997, 2003, 2004) e na abordagem adotada por Camargo (2005).

Considerando que, mesmo para as disciplinas de crítica literária e de estilística, o conceito de estilo ainda se apresenta de difícil definição, esta investigação sobre o estudo de um tradutor literário centra-se na noção proposta por Baker, que entende

estilo como uma espécie de impressão digital que fica expressa [no TT] por uma variedade de características lingüísticas (...) as quais estão provavelmente mais no domínio do que algumas vezes é chamado de 'estilística forense' que no da estilística literária (Leech e Short 1981, p. 14). Tradicionalmente, a estilística literária focaliza o que se assume serem escolhas lingüísticas conscientes da parte do autor, porque os estilistas literários estão principalmente interessados na relação entre as características lingüísticas e a função artística, em como um dado autor obtém certos efeitos artísticos. Por outro lado, a estilística forense tende a focalizar hábitos lingüísticos razoavelmente sutis e moderados que estão bem acima do controle consciente do autor e que nós, como receptores, registramos, na maioria das vezes, de forma subliminar. Todavia, como ambos os ramos da estilística, estou interessada em padrões de escolha (quer essas escolhas sejam conscientes ou subconscientes) mais do que em escolhas individuais isoladas (BAKER 2000, p. 246-7).

Com o propósito de observar padrões de escolha estilística de um tradutor literário em dois TTs, o termo "estilo" é definido no âmbito

deste estudo como o perfil de seus hábitos lingüísticos individuais, recorrentes, preferenciais e distintivos, referentes à variação e diversidade de vocabulário, a qual pode ser medida em termos da razão forma/item (*type/token ratio*). Dentre as diferentes concepções de estilo oferecidas pela literatura e pela lingüística, proponho esta noção de estilo focalizada em padrões de variação vocabular empregados pelo tradutor por mostrar-se a mais adequada às necessidades da presente investigação.

Com base na comparação de padrões lexicais de elementos coesivos, Berber Sardinha (1997) mostra que a tradução pode ser considerada um texto diferente do original. Em seu estudo, foram registradas pelo computador diferenças no uso de padrões de coesão no TT em inglês em relação ao TO em português, levantados a partir de um relatório publicado pela Companhia Metropolitana de São Paulo para o ano fiscal de 1991. De um lado, os resultados sugerem, se considerado o texto como um todo, mudanças aparentemente isoladas que acabam por criar diferenças consideráveis no modo como o TT foi organizado lexicalmente. De outro lado, todas essas diferenças também deixam entrever uma importante similaridade entre o TT e o TO, dado que ambas as versões apresentam, em sua maioria, as mesmas orações principais (Berber Sardinha, 1997, p. 153).

Com referência à linguagem do TT, a utilização de corpora eletrônicos paralelos ou comparáveis possibilita maior amplitude e funcionalidade para estudos da natureza da tradução e do uso do léxico. Investigações, realizadas no Centre for Translation and Intercultural Studies - CTIS, têm detectado certas características recorrentes (Baker, 1996, p. 180-184) que se apresentam tipicamente na tradução.

Dentre os traços recorrentes, um dos que mais especificamente se relacionam com este trabalho é a simplificação, que pode ser identificada como uma tendência em tornar mais simples e de mais fácil compreensão a linguagem empregada na tradução, como, por exemplo, a utilização de uma quantidade maior de repetições em relação à obra original. Uma medida possível de traços de simplificação é fornecida pela razão forma/item (FI), por permitir o exame da variação e diversidade de vocabulário empregadas pelo tradutor e pelo escritor num dado corpus ou corpora. São contadas todas as palavras corridas (*running words* ou *tokens*) nos textos, e cada forma ou vocábulo (*type*) é contado apenas uma vez a fim de identificar padrões de repetição nos TTs e nos TOs². Na maneira tradicional, a razão é obtida dividindo-se o total de formas pelo total de itens. No

WordList Statistics, porém, transforma-se esse valor em porcentagem: divide-se o total de formas pelo total de itens dividido por cem (Berber Sardinha, 1999; Scott, 1999).

Outro cálculo gerado pelo WordList Statistics é a razão forma/item padronizada (*standardised type/token ratio*), apropriada para observação em textos de tamanhos diferentes. Ao contrário da razão FI, acima, a forma FI padronizada calcula FI em intervalos regulares, ou seja, faz este mesmo cálculo por partes do texto e, depois, tira a média dos valores FI entre os vários trechos.

Com referência ao *material*, o corpus consiste de um subcorpus de TO escrito por Jorge Amado: *O Sumiço da Santa* (1988, 1992), Rio de Janeiro: Editora Record, 126.443 palavras; e do respectivo subcorpus de TT: *The War of the Saints* (1993), traduzido por Gregory Rabassa, Toronto/ Nova York: Bantam Books, 147.341 palavras.

Para uma observação do perfil estilístico do tradutor literário em questão, procurei identificar seus usos lingüísticos característicos e individuais, ou seja, traços preferenciais e recorrentes de seu comportamento lingüístico relacionados à variação vocabular, efetuando *comparações internacorpora* - na obra traduzida (variação intratextual), - na obra traduzida em relação à respectiva obra original (variação intercorpora da pesquisa); e ainda estabelecendo comparações *externacorpora* - na obra traduzida com os corpora de referência Translational English Corpus - TEC, e British National Corpus - BNC; bem como - na obra original com o corpus de referência Banco de Português - BP (variações *interna-* e *externacorpora*).

2. RESULTADOS E DISCUSSÃO

A razão de formas em relação aos itens (ocorrências) indica a variedade de vocábulos (formas) em um determinado texto ou corpus de tradução. De acordo com o programa WordSmith Tools, se for empregado um grande número de repetições, pode-se esperar uma razão FI mais baixa; em decorrência, havendo maior diversidade de vocabulário a razão FI tenderia a ser mais alta.

Contudo, é preciso ter cuidado no uso da razão forma/item. Esse cálculo estatístico é extremamente sensível ao comprimento do texto, uma

vez que as palavras têm maior probabilidade de se repetirem em textos mais extensos, conseqüentemente levando a uma razão FI mais baixa. Por esse motivo, com textos ou corpora de tamanhos diferentes, costuma-se utilizar a razão FI padronizada.

2.1. ALGUNS PADRÕES DISTINTIVOS DE VOCABULÁRIO ENTRE TT E TO (COMPARAÇÕES INTERNACORPORA)

A fim de examinar a distribuição de itens e formas no corpus de TO e TT, foram extraídas as *Tabelas 1-3*:

Tabela 1.

Estatística do corpus contendo o TT por Rabassa e TO de Jorge Amado

	Subcorpus de TT - Rabassa/ Jorge Amado <i>The War of the Saints</i>	Subcorpus de TT - Rabassa/ Jorge Amado <i>O Sumiço da Santa</i>
Itens (<i>tokens</i>)	147.341	126.443
Formas (<i>types</i>)	14.365	19.465
Razão FI (<i>tt ratio</i>)	9,75	15,39
Razão FI padronizada (<i>Std tt ratio</i>)	49,06	54,99

Tabela 2.

Diferença da razão FI entre o TT e o TO

Resultado do TT por Rabassa	Razão FI	9,75	Diferença da razão FI
Resultado do TO de Jorge Amado	Razão FI	15,39	5,64

Tabela 3.

Diferença da razão FI padronizada entre o TT e o TO

Resultado do TT por Rabassa	Razão FI	49,06	Diferença da razão FI padronizada
Resultado do TO de Jorge Amado	Razão FI	54,99	5,93

Procedi a comparações, separadamente, do TT por Rabassa e do TO de Jorge Amado, a fim de examinar se o uso de vocabulário na obra traduzida para o inglês é mais ou menos variado do que o da respectiva obra originalmente escrita em português. Algumas evidências puderam ser encontradas nas variações intratextual.

Para as comparações do par de obras *The War of the Saints* com *O Sumiço da Santa*, consideradas individualmente, a distribuição dos itens e das formas (*Tabelas 1*) aponta que, em termos absolutos, esse TT registra, como esperado com base no princípio de explicitação, um número mais alto de itens (147.341) e um menor número de formas (14.365) em relação ao respectivo TO (126.443 itens e 19.465 formas). Nesse par de textos, o tradutor apresenta uma razão FI de 9,75 em contraste com a do autor de 15,39, o que indica maior incidência de repetições no TT do que no TO, confirmando-se o princípio de simplificação. Esses resultados também confirmam a hipótese levantada pelo software WordSmith Tools ao gerar uma razão FI mais baixa no TT em decorrência da menor quantidade de vocábulos/formas em relação ao respectivo TO.

Por sua vez, foram tomados conjuntamente o TT e o TO, a fim de verificar padrões estilísticos em termos de variedade de vocábulos empregados pelo tradutor em relação ao estilo do autor (variação intertextual). Os resultados indicam que a obra traduzida por Rabassa apresenta uma razão FI menor que a obtida na respectiva obra originalmente escrita por Jorge Amado, com uma diferença na ordem de 5,64 (cf. *Tabela 2*). Essa medida, ao registrar uma frequência mais baixa de variação vocabular, evidencia uma tendência desse profissional em tornar mais simples e de mais fácil compreensão a linguagem empregada na tradução, ratificando a hipótese de simplificação.

Outra indicação de que haveria um uso menos variado de vocabulário no TT é fornecida pelo exame da razão FI padronizada ou densidade lexical simples e em intervalos regulares. Para *The War of the Saints* tomada conjuntamente em relação a *O Sumiço da Santa*, a razão FI padronizada gerada no TT (49,06) é menor em relação ao respectivo TO (54,99). A diferença de 5,93 (cf. *Tabela 3*) é significativa, porquanto existe, na verdade, menos palavras 'diferentes' na obra traduzida, o que mostra que há mais repetições nas escolhas efetuadas por Rabassa. O comportamento lingüístico observado parece evidenciar uma maneira de tornar a

tradução mais fácil de ser processada pelo leitor de língua inglesa, novamente confirmando o princípio da simplificação.

2.2. ALGUNS PADRÕES DISTINTIVOS DE VOCABULÁRIO ENTRE O TT E O TEC, E ENTRE O TT E O BNC (COMPARAÇÃO EXTERNA CORPORA)

A fim de distinguir entre a variação de vocabulário empregada pelo tradutor em análise e a encontrada no inglês traduzido, foi examinado um corpus de referência: o Translational English Corpus - TEC. Nesse corpus, os textos estão compilados na íntegra, englobando quatro tipologias textuais: ficção, revistas de bordo, biografia e artigos de jornais. O tipo de texto predominante é o ficcional, que abarca 82% do total do corpus e 84 arquivos por ocasião da coleta de dados efetuada para este trabalho; como os outros tipos de texto constituem apenas uma pequena porção do TEC, somente o subcorpus de textos de ficção foi aqui considerado para fins estatísticos. As traduções armazenadas no TEC são realizadas por tradutores falantes nativos de língua inglesa, e a maior parte desses TTs foi feita a partir de 1983. Os dados extraídos do TEC constam da tabela abaixo:

Tabela 4.
Estatística do corpus de referência TEC

TEC	Subcorpus de ficção
Itens	5.848.203
Formas	70.700
Razão FI	1,21
Razão FI padronizada	44,53

No TEC, há 5.848.203 itens e 70.700 formas, e uma razão FI de 1,21. Desse modo, o tradutor selecionado para estudo mostra, nos resultados gerais, a utilização de uma linguagem mais rica e variada (razão FI: 9,75) do que a representada pelos textos do TEC (razão FI: 1,21).

No que concerne ao cálculo da razão FI padronizada, o TT, ao gerar um resultado de 49,06, é maior do que a razão FI padronizada de 44,53, apresentada pelo TEC. Esses dados revelam, de um lado, uma diferença significativamente mais alta para Gregory Rabassa, indicando maior

variação no padrão vocabular do TT; de outro, apontam uma tendência para um uso de padrões lexicais mais restritos por parte dos tradutores representados no TEC.

Por sua vez, com o propósito de observar a linguagem empregada pelo tradutor em relação à linguagem normalmente usada em textos originalmente escritos em inglês, utilizei o corpus de referência British National Corpus - BNC. É considerado um marco histórico por ter sido o primeiro corpus eletrônico a conter 100 milhões de palavras. De modo análogo ao TEC, os TOs compilados para o BNC são produzidos por autores falantes nativos de língua inglesa. Diferentemente do TEC, porém, os tipos de texto predominantes são os de língua geral; o subcorpus de textos de ficção do BNC contava, no momento da coleta para este estudo, com 485 arquivos e alguns dos textos são fragmentos - ainda que com a extensão de 40.000 palavras. Por essa razão, tanto o BNC como o subcorpus de textos de ficção do BNC foram tomados em consideração na tabela abaixo:

Tabela 5.
Estatística do corpus de referência BNC

	BNC	BNC Ficção
Itens	90.748.880	19.444.150
Formas	377.784	101.577
Razão FI	0,42	0,52
Razão FI padronizada	44,04	41,54

No BNC, há 90.748.880 itens e 377.784 formas, o que corresponde a uma razão FI de 0,42. No subcorpus de ficção do BNC, os dados levantados mostram 19.444.150 itens e 101.577 formas, gerando uma razão FI de 0,52.

A comparação desses resultados gerais indica que o tradutor literário em questão utiliza uma diversidade maior de vocabulário (razão FI: 9,75) do que em relação aos textos do BNC (razão FI: 0,42) e mesmo em relação aos textos de ficção do BNC (razão FI: 0,52).

No tocante ao cálculo da razão FI padronizada, a tradução em análise mostra que a razão FI padronizada de 49,06 é mais alta do que a levantada para o BNC (44,04); também é ainda mais elevada em relação à razão FI padronizada obtida para o subcorpus de ficção do BNC (41,54).

Essas estatísticas revelam que Rabassa apresenta, na obra traduzida, uma variação de palavras maior tanto em relação aos textos representados pelo corpus de referência BNC como em relação aos textos do seu subcorpus de ficção. Desse modo, os resultados da razão FI padronizada mostram um comportamento diferente do tradutor, rejeitando a hipótese de simplificação para este TT em inglês em relação aos TOs em inglês do BNC.

2.3. ALGUNS PADRÕES DISTINTIVOS DE VOCABULÁRIO ENTRE O TO E O BP (COMPARAÇÃO EXTERNA CORPORA)

Com a intenção de obter um parâmetro similar a respeito da linguagem utilizada pelo autor em relação à linguagem habitualmente empregada em textos originalmente escritos em português, foram efetuados os mesmos procedimentos entre o TO e o corpus de referência Banco de Português - BP. Esse corpus de TOs do português brasileiro está sediado na Universidade Católica de São Paulo - LAEL-PUC/SP. Os textos compilados para esse corpus são produzidos por vários autores, e os tipos de texto predominantes são os de língua geral e artigos de jornais. Como o conjunto de textos literários constitui apenas uma pequena parte do BP, foram utilizados os dados totais para o levantamento estatístico, os quais podem ser observados na tabela abaixo:

Tabela 6.
Estatística do corpus de referência BP

	BP
Itens	230.460.560
Formas	607.392
Razão FI	0,26
Razão FI padronizada	46,08

No BP, de acordo com os resultados obtidos por ocasião da coleta, há 230.460.560 itens e 607.392 formas, e uma razão FI de 0,26.

De modo análogo, o autor recorre a uma maior variação vocabular (razão FI: 15,39) do que a encontrada, em geral, nos TOs escritos em português armazenados no BP (razão FI: 0,26).

Quanto à razão FI padronizada, a variação apresentada pelo autor correspondente a 54,99 é mais elevada do que a gerada para o BP, de 46,08, apontando uma diversidade maior de vocábulos na obra original em relação aos textos representados nesse corpus de referência.

De certa maneira, esses resultados mais altos encontrados no TO de Jorge Amado já eram esperados, em virtude de tratar-se de um romance de cunho regionalista, o qual, pela sua natureza, passa a requerer uma diversidade de termos culturalmente marcados. Também o estilo do autor, normalmente associado ao pitoresco e mesmo ao exotismo, mostra-se como um fator que contribui para o uso de uma linguagem mais rica e variada.

Desse modo, o acesso ao subcorpus de tradução desta pesquisa permitiu detectar que, embora Gregory Rabassa apresente uma razão FI e uma razão FI padronizada mais elevada em relação aos corpora de referência de TTs em inglês e de TOs em inglês, o seu comportamento em termos de utilização de vocábulos/formas apresenta uma variação menor que a empregada por Jorge Amado.

Em contrapartida, se a simplificação é geralmente vista como uma característica da linguagem de tradução devido a estratégias do tradutor para tornar o TT mais simples e de mais fácil compreensão, o uso de um corpus paralelo em formato eletrônico possibilitou observar que Rabassa apresenta um emprego bem mais diversificado de padrões lingüísticos e um menor padrão de repetições do que os encontrados no subcorpus literário do TEC, bem como na variedade de tipos de texto do BNC e do subcorpus de ficção do BNC.

3. À GUIA DE CONCLUSÃO

Como já comentado acima, a forma padronizada é empregada para neutralizar a influência do tamanho do texto na computação da razão FI, já que esta é muito sensível à extensão do material textual. Assim, é importante enfatizar que há restrições quanto à confiabilidade dos dados obtidos por meio de comparações automáticas nas listas e estatísticas processadas pelo programa WordSmith Tools. Normalmente, tamanhos diferentes de textos autênticos, mormente em se tratando de obras literárias ou de textos armazenados em corpora de referência, assim como limita-

ções de ferramentas de busca são problemas com que se depara o pesquisador em estudos dessa natureza.

Dentre as limitações do presente trabalho, o requisito de compilação de um corpus extenso é inerente à abordagem baseada no uso de corpora de TTs como uma ferramenta para investigar e descrever traduções e o uso do conjunto léxico, bem como o comportamento de tradutores dentro dos estudos da tradução. Por essa razão, Olohan e Baker atestam que as “investigações levadas a efeito até o momento podem ser descritas como exploratórias e em pequena escala” (OLOHAN & BAKER 2000, p. 141).

Além do problema do tamanho do corpus, existem variáveis difíceis de serem controladas, referentes à poética da narrativa brasileira, aos diversos recursos estilísticos do autor, normas de tradução, coerções do mercado editorial, escolha pelo tradutor do tipo de material a ser traduzido. No caso do presente estudo, uma variável, a ser contornada futuramente, decorre do par lingüístico envolvido, em virtude do número consideravelmente maior de flexões do português comparado ao do inglês. Nesse sentido, tenciono, nas minhas próximas investigações, utilizar, para o cálculo estatístico, uma razão FI móvel e também o lematizador VISL, criado por Bick, a fim de efetuar novas observações para estudos de padrões individuais, recorrentes e preferenciais de Rabassa bem como de outros tradutores de obras de Jorge Amado.

Por outro lado, com o suporte teórico-metodológico de Baker (1993, 1995, 1996, 2000, 2004) e com a abordagem sugerida por Camargo (2005), os dados obtidos por meio da razão FI e da razão FI padronizada permitiram, ainda que em apenas um par de obras, a identificação de escolhas estilísticas do tradutor Gregory Rabassa, feitas consciente ou inconscientemente, as quais evidenciam frequências moderadas no uso de padrões de repetição vocabular. Esse uso de padrões estilísticos individuais, recorrentes e preferenciais, encontrado na tradução analisada, sugere que o tradutor em questão vale-se, com certa regularidade, de estratégias simplificadoras para conferir fluência ao texto de chegada e facilitar a compreensão do estilo amadeano para leitores de língua inglesa.

Foi mencionada na introdução a existência de uma similaridade entre o TT e o respectivo TO, notadamente devido ao par de textos apresentar, em geral, as mesmas orações principais no corpus compilado para a pesquisa desenvolvida por Berber Sardinha (1997, p. 153). Contudo, se tomarmos o nível dos textos como um todo, os resultados das *Tabelas 1-*

3 sugerem que padrões de uso do vocabulário acabam por criar uma diferença considerável no modo como o TT é organizado lexicalmente em relação ao respectivo TO.

Desta feita, se considerarmos a razão FI e a razão FI padronizada como indicações do emprego que o tradutor faz da linguagem, podem-se destacar, apesar da influência de possíveis variáveis, a diferença de 5,64 e a de 5,93 apresentadas por Gregory Rabassa em relação ao autor Jorge Amado como marcas significativas da utilização de padrões estilísticos preferenciais e distintivos desse tradutor, as quais revelam o impacto da extensão dessas diferenças, respectivamente na obra traduzida para o inglês *The War of the Saints* em comparação com a obra originalmente escrita em português *O Sumiço da Santa*.

ABSTRACT: By means of generating type/token ratio, it is possible to observe some evidence of a literary translator's own use of style viewed as an individual and recurring range of variation of vocabulary (Baker 2000; Berber Sardinha 1997). Despite all possible variables, if we consider type/token ratio and standardized type/token ratio as indications of a translator's language use, the lower 5.64 and 5.93 differences showed by Gregory Rabassa in relation to the Brazilian author Jorge Amado may be pointed out as strong marks of this translator's distinctive and preferred stylistic patterns, revealing the impact of his extent of difference in the translated text *The War of the Saints* in relation to the original text *O Sumiço da Santa*.

KEY-WORDS: Lexicon and Translation, Lexicon and Corpus Linguistics, Corpus-Based Translation Studies, Translator's style, Translated contemporary Brazilian Literature.

REFERÊNCIAS BIBLIOGRÁFICAS

AMADO, J. *O sumiço da santa*. Rio de Janeiro: Record, 1988.

_____. *The war of the saints*. Trad. Gregory Rabassa. Toronto/Nova Iorque: Bantam Books, 1993.

BAKER, M. Corpus Linguistics and Translation Studies: implications and applications. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (eds.).

Text and technology: In honour of John Sinclair. Amsterdã/Filadélfia: John Benjamins, 1993, p. 233-250.

_____. Corpora in Translation Studies: An overview and some suggestions for future research. *Target*, v. 7.2, p. 223-243, 1995.

_____. Corpus-based translation studies: the challenges that lie ahead In: SOMERS, H. (ed.). *Terminology, LSP and Translation Studies in language engineering*, In honour of Juan C. Sager. Amsterdã/ Filadélfia: John Benjamins, 1996, p. 175-186.

_____. Towards a methodology for investigating the style of a literary translator. *Target*. v. 12.2, p. 241-266, 2000.

_____. A corpus-based view of similarity and difference in translation. In ARDUINI, S.; HODGSON, R. (eds.) *Translating similarity and difference*. Manchester: St. Jerome. 2004.

BANCO DE PORTUGUÊS: corpus do Português escrito, de língua geral. São Paulo: PUC/SP.

BERBER SARDINHA, A. P. Patterns of lexis in original and translated business reports: textual differences and similarities. In: SIMMS, K. *Translating sensitive texts: Linguistic aspects*. Amsterdã/Atlanta: Rodopi, 1997, p. 147-153.

_____. Ferramentas de busca e de exploração de corpora. In: I SEMINÁRIO DE ESTUDOS DE CORPUS, 1999, São Paulo. Anais... São Paulo: USP, 1999.

_____. Lingüística de Corpus: histórico e problemática. *D.E.L.T.A.*, p. 323-367, 2000.

_____. Uso de corpora na formação de tradutores. *D.E.L.T.A.* 19(Especial), p. 43-70, 2003.

_____. *Lingüística de Corpus*. São Paulo: Manole. 2004.

BICK, E. *VISL - Visual interactive system learning*. Morphological analyser. Projeto desenvolvido no Institute of Language and Communication - ISK, University of Southern Denmark, Syddansk Universitet - SDU, câmpus de Odense, Dinamarca, desde set. 1996. Disponível em: <<http://visl.hum.sdu.dk/visl/pt/parsing/automatic/upload.php>>. Acesso em: 18 nov. 2005.

BRITISH National Corpus: Written corpus design specification. OUP Promotional Document, Oxford: Oxford University Press. 02 set. 1991.

CAMARGO, D. C. *Padrões de Estilo de Tradutores*: Um estudo de semelhanças e diferenças em corpora de traduções literárias, especializadas e juramentadas. 2005. Tese (Livro-Docência em Estudos da Tradução) - UNESP/IBILCE, São José do Rio Preto.

HUNSTON, S. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press. 2002.

- LAVIOSA-BRAITHWAITE, S. *The English comparable corpus (ECC): A resource and a methodology for the empirical study of translation*. 1996. Tese (Doutorado em Tradução) - Department of Language Engineering, UMIST, Manchester.
- LAVIOSA, S. *Corpus-based Translation Studies: Theory, findings, applications*. Amsterdã/Atlanta: Rodopi. 2002.
- LEECH, G. & SHORT, M. H. *Style in fiction: A linguistic introduction to English fictional prose*. Harlow: Longman. 1981.
- OLOHAN, M. & BAKER, M. Reporting *that* in translated English: Evidence for subconscious processes of explication?. *Target*. 1:2, p. 141-158, 2000.
- SCOTT, M. *WordSmith Tools*. Software for Windows 3.1, 95 e 98. Oxford: Oxford University Press. Versão demo com funções limitadas, disponível em: <http://www.liv.ac.uk/~ms2938>. Acesso em: 18 nov. 2205.
- _____. *WordList Manual*. Oxford: Oxford University Press. 1999, p. 22-26.
- TOGNINI-BONELLI, E. *Corpus Linguistics at work*. Amsterdã/Atlanta: John Benjamins. 2001.
- TYMOCZKO, M. Computerized corpora and the future of Translation Studies. *Meta* 43(4), p. 652-659, 1998.

NOTAS

- ¹ Com exceção da nota de rodapé no. 2 traduzida por Berber Sardinha, a tradução das citações são de minha responsabilidade.
- ² Por exemplo, o fragmento: "As the minute hand moves, the hour hand also moves correspondingly" contém 11 itens (*tokens*), mas somente 8 formas (*types*), porquanto há 2itens para as formas: *the*, *hand* e *moves*.