

# LINGÜÍSTICA DE CORPUS E ESTUDOS DA TRADUÇÃO: O ESTADO-DA-ARTE

Adail Sebastião RODRIGUES JÚNIOR<sup>1</sup>  
Universidade Federal de Minas Gerais  
Brasil  
[adail.sebastiao@terra.com.br](mailto:adail.sebastiao@terra.com.br)

## Sinopse

O objectivo principal deste ensaio é discutir o interesse crescente que os Estudos da Tradução tem expressado acerca da Lingüística de Corpus e seu importante papel para a investigação do fenómeno tradutório. O ensaio apresenta os primeiros trabalhos na interface entre corpora e tradução e como a disciplina dos Estudos da Tradução de base em corpus evoluiu na última década. Uma atenção mais pontual é dada à noção de *estudos em corpus de pequena dimensão* e sua influência no campo das pesquisas em tradução no Brasil. Além disso, o ensaio explora como o conceito de *representatividade* tem recebido definições diferentes nos estudos dos fenómenos tradutórios de base em corpus.

**Palavras-Chave:** Estudos da Tradução, Lingüística de Corpus, Estudos de Corpus de Pequena Dimensão, Representatividade de Corpus, Estudos da Tradução no Brasil.

## Abstract

The main aim of this essay is to discuss the growing interest Translation Studies has expressed in Corpus Linguistics and the important role its methods play on the investigation of translation phenomenon. The essay presents the first works on the interface between corpora and translation and how the Corpus-based Translation Studies discipline evolved in the last decade. A more close attention is given to the notion of

---

<sup>1</sup> Núcleo de Estudos da Tradução (NET) da Faculdade de Letras da Universidade Federal de Minas Gerais, Brasil.

*small corpus studies* and its influence on the field of translation research in Brazil, as well as how the concept of *representativeness* has received different conceptualisations in corpus-based studies of translation phenomena.

**Key words:** Translation Studies, Corpus Linguistics, Small Corpus Studies, Corpus Representativeness, Translation Studies in Brazil.

*Corpus linguistics sees language as a social phenomenon. Meaning is, like language, a social phenomenon. It is something that can be discussed by the members of a discourse community. There is no secret formula, neither in natural language nor in a formal calculus, that contains the meaning of a word or phrase. There is no right or wrong. What I call a weapon of mass destruction differs probably a lot from what President George W. Bush calls a weapon of mass destruction. What I call a baguette is not the same as what many supermarkets sell as a baguette. What I call love may not be what my partner calls love. Different people paraphrase words or phrases in different ways. They do not have to agree. In a democracy, everyone's opinion is as good as anyone else's.*

Wolfgang Teubert, 'Language and Corpus Linguistics', p.97-8.

O comentário de Teubert (2004) expressa as duas faces da Linguística de Corpus: analisar estatisticamente e probabilisticamente manifestações lingüísticas a partir de um *corpus* representativo, embora o conceito de representatividade seja bastante abstrato diante das incontáveis formas de uso da linguagem por falantes e escritores nativos. No caso, por exemplo, da “colocação” (*collocation*) em língua inglesa *a weapon of mass destruction*, suas intenções de uso e, conseqüentemente, seus significados diferenciam-se sobremodo quando usada por um falante nativo qualquer e quando empregada pelo actual Presidente dos Estados Unidos da América, tendo como pano de fundo sua relação com os eventos internacionais nos quais ele tem sido personagem política central. Este aspecto é fundamental para as pesquisas na interface Corpora e Tradução. Ou seja, “[p]ara

responder *representativo do quê?*, deve-se olhar para a questão da amostragem” (Berber Sardinha 2004:23). No que se segue, apresento a evolução dessa discussão e suas conseqüências teóricas e metodológicas para os Estudos da Tradução. Meu objectivo neste ensaio, pois, é demonstrar, sucintamente, como o interesse pela Lingüística de Corpus nasceu entre os teóricos da Tradução, a evolução das teorias e suas aplicações apresentadas em algumas pesquisas internacionais e, por fim, a viragem dos corpora de pequenas dimensões e seu uso em pesquisas internacionais e brasileiras nessa linha de interesse até o presente momento.

### **Lingüística de Corpus e Tradução**

O estudo pioneiro de Mona Baker (1993) acerca da relação entre Lingüística de Corpus e Tradução inaugurou, por assim dizer, o ‘casamento’ entre esses dois ramos de pesquisa, delimitando, assim, sua interdisciplinaridade. Para Baker, os textos traduzidos devem ser considerados textos autênticos, passíveis de investigação da mesma forma que os textos originais são estudados pelos lingüistas aplicados. Quase uma década após a publicação do artigo de Baker (1993), surge o *Reader* dos Estudos da Tradução no circuito editorial internacional (Venuti 2000), como “uma espécie de ritual de passagem” que traz a “compilação de textos ‘fundacionais’ para a área de estudo em questão” (Pagano 2001:7). Todavia, o *The Translation Studies Reader* reserva pouco espaço para a relação entre Lingüística de Corpus e Tradução, tecendo comentários breves sobre ‘normalização’, ‘explicitação’, ‘sanitização’, dentre outros, considerados ‘universais’ em fenómenos tradutórios (cf. Baker 1997; Kenny 1998, 2001). Tal facto, no entanto, já indica que o próprio *Reader* legitima essa vertente asseverando sua importância histórica, teórica e prática para os Estudos da Tradução.

A meu ver, a Lingüística de Corpus lança luz sobre a própria conceituação de texto apresentada pelos analistas do discurso enquanto unidade lingüística avaliada com base em seu contexto de produção e recepção (cf. Iñiguez 2004; Jaworski & Coupland 2000; Wetherell, Taylor & Yates 2001), ao possibilitar maneiras diferenciadas de abordar os textos conforme os objectivos analíticos do(a) investigador(a). Em outras palavras, apenas um conjunto significativo de textos representativos em formato electrónico (um

corpus), enquanto instância discursiva que expressa em cálculos probabilísticos as formas de uso da linguagem empregadas pelos usuários da língua, pode revelar realizações lingüísticas que materializam fenómenos culturais, sociais e políticos dos contextos de onde elas surgem (cf. Stubbs 1996, 2001; Teubert 2004; Teubert & Čermáková 2004). Conseqüentemente, esse corpus deve ser avaliado no todo, e não em partes, com uma visão mais abrangente de vários textos ao mesmo tempo e de partes desses mesmos textos, através de ferramentas computacionais que a Lingüística de Corpus oferece, visto que a avaliação manual de um corpus de grande dimensão é totalmente inviável. Para Berber Sardinha (2004) e os lingüistas de corpus, os textos de interesse da Lingüística Aplicada de base em corpus são textos naturais, em formato electrónico, produzidos para fins de comunicação em contextos variados, agrupados com o fito de constituírem um corpus representativo. Em virtude disso, com o advento da Lingüística de Corpus os textos passaram a ser compilados, formando grandes bancos de dados, para fins de estudos lingüísticos, de acordo com critérios definidos conforme os objectivos de investigação (cf. Sinclair 1991).

De igual modo, traduções são mais facilmente investigadas a partir da abordagem de corpus, a fim de delimitar como as realizações lingüísticas do texto original se re-materializaram no texto traduzido em busca de significação. Venuti (2000:336) sintetiza essas considerações ao indicar que

*... análises computacionais podem elucidar padrões tradutórios significativos em um corpus paralelo de textos originais e suas traduções, sobretudo se esses padrões são avaliados em contraste com um amplo corpus de 'referência' tanto na língua-fonte quanto na língua-alvo.<sup>2</sup>*

Delimita-se, pois, uma metodologia para a investigação de traduções, mais especificamente as escolhas lingüísticas que o(a) tradutor(a) realiza em comparação com

---

<sup>2</sup> Minha tradução de: "... computer analysis can elucidate significant translation patterns in a parallel corpus of foreign texts and their translations, especially if the patterns are evaluated against large 'reference' corpora in the source and target languages."

as escolhas lingüísticas do original e, concomitantemente, a verificação se tais escolhas são freqüentes em corpora naturalmente produzidos tanto na língua original quanto na língua traduzida.

Com efeito, descortina-se um campo de investigação de textos traduzidos com o advento da Lingüística de Corpus, através do qual Mona Baker (1993, 1995, 1996, 1998, 1999) tem estabelecido uma agenda teórica e metodológica para os Estudos da Tradução. A proposta desta teórica é aliar os estudos descritivos em Tradução, com base na teoria de Gideon Toury (2000), às descobertas do lingüista John Sinclair (1991), “que conseguiu, através da coleção de *corpora* computadorizados e do desenvolvimento de uma metodologia de pesquisa relevante, superar as limitações humanas do pesquisador e minimizar sua dependência da intuição” (Magalhães 2001:97). Dentro do escopo dessa metodologia, Kenny (1998, 2001) destaca a importância de se estudar colocações (*collocations*) em textos traduzidos, no âmbito da estrutura lexical, como forma de identificação das “forças culturais” presentes nos textos e refletidas no mercado editorial literário. Para tanto, esta teórica propõe uma metodologia que necessita de (i) um corpus paralelo (texto original e texto traduzido), (ii) um corpus de referência de grande dimensão na língua do original e (iii) um outro corpus de referência de grande dimensão na língua da tradução. A partir daí, faz-se a comparação entre texto original e texto traduzido, em seguida, texto original e corpus de referência na língua do original e, por fim, texto traduzido e corpus de referência na língua da tradução. Esse tipo de metodologia comparativa, segundo Kenny, inspirando-se no trabalho de Stubbs (1996), possibilita a investigação de escolhas lexicais carregadas de características culturais.

Todavia, no caso de pesquisas na área desenvolvidas em universidades brasileiras, Berber Sardinha (2002) esclarece que um dos motivos do relacionamento não muito estreito entre lingüistas de corpus e pesquisadores em Tradução se deve pelo facto do difícil acesso à tecnologia, ou seja, o acesso aos corpora de grandes dimensões propriamente ditos e aos programas computacionais para exploração desses corpora voltados para a Tradução. Segundo Berber Sardinha (2002), a maioria dos textos investigados por pesquisadores em Tradução são impressos, o que exige um trabalho custoso e metódico de digitalização desses textos e seu preparo até atingir o formato electrónico, sem falar dos problemas com os direitos autorais dos mesmos. Tal facto, por

si só, já inviabiliza a aplicação *in toto* da metodologia da Lingüística de Corpus aos Estudos da Tradução.

Um outro ponto corriqueiro, conforme esboça Berber Sardinha (2002), é a deficiência de manuseio de software para investigação lingüística apresentada pelos pesquisadores em Tradução, como bem expresso, por exemplo, no II Encontro Internacional de Tradutores, realizado em Belo Horizonte, na Universidade Federal de Minas Gerais, Brasil, em que o tema foi “Translating the New Millennium: Corpora, Cognition and Culture”. Berber Sardinha (2002:21) assim se expressa sobre esse facto:

*Em suma, o pesquisador ou tradutor que deseje fazer incursões na exploração de corpora para a investigação da tradução enfrentará o problema de maior escassez de recursos para sua área, da necessidade de aprender a utilizar software especializados, além de necessitar executar algumas tarefas comuns da Lingüística de Corpus, como a organização, formatação e exploração de corpora.*

Com efeito, o próprio escopo da Lingüística de Corpus, que seja a compilação de um grande banco de dados de linguagem autêntica, usada em contexto, para fins de pesquisa dos traços lingüísticos proeminentes e peculiares de cada língua (cf. Baker 1995; Laviosa 1998), ganha matizes diferenciados em pesquisas tradutórias no contexto brasileiro. Como bem salienta Magalhães (2001), os objectivos de pesquisa em corpus e Tradução devem se desmembrar em sub-projetos que interdependentemente desenvolvam pesquisas com interesses análogos, com o intuito de criação de um banco de dados em língua portuguesa e línguas estrangeiras que sirva de base para um projeto mais amplo. Neste sentido, Pagano, Magalhães e Alves (no prelo) desenvolveram o projeto CORDIAL (Corpus Discursivo para Análises Lingüísticas e Literárias) do Núcleo de Estudos da Tradução (NET) da Faculdade de Letras da Universidade Federal de Minas Gerais, Brasil, cujo objectivo central é compilar, sistematizar e disponibilizar para uso corpora monolíngües e bilíngües<sup>3</sup> para pesquisa, e cujo escopo seja investigar a

---

<sup>3</sup> Corpora bilíngües consistem em textos de uma língua de partida A e suas respectivas traduções para uma língua de chegada B: são também conhecidos como *corpora paralelos*. Corpora monolíngües são corpora

linguagem de textos originais e traduções enquanto realizações lingüísticas dentro do sistema lingüístico de cada língua sob análise e suas comparações. Para a efectivação desse projeto, Magalhães (2001:113) sublinha a importância da expansão dessa vertente de pesquisa “de modo a acolher um número maior de alunos pesquisadores (...), além de interagir com outros grupos acadêmicos, nacionais ou internacionais, com interesses de pesquisa similares (...)”. Assim, o projeto CORDIALL se instalou, de início, com o propósito de compilar corpus de pequena dimensão (Ghadessy, Henry & Roseberry 2001) e investigar como as estruturas sintagmáticas dos textos originais e traduções materializavam o sistema paradigmático donde esses textos se originaram. Para isso, a Gramática Sistêmico-Funcional de M. A. K. Halliday (2004) tem demonstrado ser um referencial teórico eficiente para se levar a termo os objectivos de pesquisa do projeto CORDIALL. No que se segue, apresento, detalhadamente, o que vem a ser o objecto de estudos em corpus de pequena dimensão (*small corpus studies*), sua importância para o CORDIALL, as pesquisas, nacionais e internacionais, realizadas nessa vertente até o momento.

### **Corpus de pequena dimensão**

Para Sinclair (2001), o tamanho do corpus certamente influencia o tipo de pesquisa a ser realizada, mas isso em nada compromete a qualidade da pesquisa: para ele, tudo depende dos objectivos do(a) pesquisador(a). Com essa afirmação, Sinclair inicia o prefácio da coletânea *Small Corpus Studies and ELT: theory and practice* (Ghadessy, Henry & Roseberry 2001), considerado um marco histórico para a Lingüística de Corpus, em especial as pesquisas que utilizam corpus de pequena dimensão para demonstrar sua aplicabilidade e validade para a Lingüística Aplicada (cf. Berber Sardinha 2004). Continuando suas explanações, Sinclair expressa considerações interessantes sobre o conceito exato de *dimensão* de um corpus, uma vez que as pesquisas baseadas em corpus vêm se aprimorando progressivamente, à medida que o mercado da informática

---

independentes, de uma língua-fonte, usados para fins de averiguação de padrões léxico-gramaticais de usuários daquela língua. (cf. Baker 1995; Magalhães 2001).

disponibiliza novos softwares para análises lingüísticas. Por exemplo, na década de 60, Nelson Francis e Henry Kuçera, na Brown University em Rhode Island, EUA, compilaram um corpus de ‘grande dimensão’ com cerca de um milhão de palavras. Nos dias de hoje, esse corpus é considerado de pequena dimensão, devido ao aprimoramento de técnicas de coleta de dados que a Lingüística de Corpus tem desenvolvido nos últimos anos, principalmente em universidades britânicas, dando origem a imensos Bancos de Dados para estudos lexicográficos, gramaticais e discursivos.

A diferença básica entre um corpus de pequena dimensão e um de grande dimensão é o facto de que neste podem-se observar repetições, ou uma variada gama de escolhas léxico-gramaticais recorrentes, enquanto que naquele isso não se torna metodologicamente viável, afirma Sinclair (2001). Neste sentido, ao investigar um corpus de pequena dimensão, o(a) pesquisador(a) adota uma postura que Sinclair (2001:xi) denomina “intervenção humana antecipada” (*early human intervention*), isto é, o(a) pesquisador(a), com uma meta bem delimitada, seleciona um corpus de pequena dimensão para investigar ocorrências peculiares àquele corpus. À medida que esse tipo de metodologia é empregada por um grupo de pesquisadores(a)s afiliado(a)s de preferência a um Grupo de Pesquisa, pouco a pouco, os corpora de pequenas dimensões vão sendo compilados até se ter um corpus representativo, a partir de pequenas amostras. Tal metodologia tem lançado luz sobre pesquisas discursivas aplicadas aos Estudos da Tradução em que pesquisadore(a)s tencionam averiguar fenómenos lingüísticos típicos do corpus sob investigação, sem generalizações que somente um corpus de grande dimensão poderia permitir.

Estudos nessa vertente têm demonstrado a utilidade de ferramentas computacionais para a investigação de fenómenos peculiares ao corpus sob análise, o estilo dos autores e tradutores dos textos, suas formas de construção discursiva e suas escolhas léxico-gramaticais (cf. Baker 2000). Neste ramo de pesquisa temos o estudo de Munday (1998a), de base sistêmico-funcional, em que este teórico analisa traduções para o inglês de romances do escritor colombiano Gabriel García Márquez. Sustentando-se numa abordagem baseada em corpus, Munday analisa tanto o produto quanto o processo social da tradução, através de métodos computacionais automáticos do software *WordSmith Tools* (Scott 1996, 1997, 1999), mais especificamente os alinhadores

(*Aligner*) e as linhas de concordância (*Concord*). Em suas análises, Munday salienta a importância das ferramentas computacionais da Linguística de Corpus e como estas podem ser utilizadas heurísticamente para desvelar estruturas linguísticas não facilmente visualizadas por mera análise manual. Essa visão heurística, indica Munday, verte a atenção para as escolhas lexicais individuais do(a) tradutor(a), na tentativa de demonstrar como tais escolhas influenciam o texto como um todo.

Em um outro estudo sobre relações temáticas no discurso do texto original comparado ao texto traduzido, Munday (1998b), com o suporte da Gramática Sistêmico-Funcional hallidayana, sugere que a organização tema-rema difere-se de uma língua para outra, como é o caso do inglês e do espanhol. Ao investigar o conto *El verano feliz de la señora Forbes*, de Gabriel García Márquez, e sua tradução *Miss Forbes' Summer of Happiness*, feita por Edith Grossman, Munday (1998b) afirma que o sistema linguístico da língua inglesa poderia criar uma diferença considerável na progressão temática do discurso do conto de Gabriel García Márquez. Em outras palavras, ao se traduzir o conto do espanhol para o inglês, a liberdade inerente de posicionamentos temáticos comum à língua espanhola pode não ser possível em língua inglesa, visto que nesta os temas são mais não-marcados se comparados aos temas da língua espanhola. Para Munday, essa rigidez temática pode igualmente afetar as relações de espaço e tempo nas narrativas do escritor colombiano.

Esses estudos de Munday culminaram em um ensaio (Munday 2002) cujo enfoque é o desenvolvimento de um modelo de três estágios. Sustentando-se nos Estudos Descritivos da Tradução (Toury 2000), Munday esclarece que os dois primeiros estágios de seu modelo se erigem a partir da identificação sistemática de “mudanças” (*shifts*) na tradução, orientado pela teoria Sistêmico-Funcional de Halliday e auxiliado pelas ferramentas da Linguística de Corpus. O terceiro estágio do modelo localiza os resultados da análise feita nos dois primeiros estágios dentro dos contextos editorial, político e sócio-cultural com o intuito de identificar factores externos que tenham motivado essas “mudanças”. Em seu modelo, Munday explora um texto de Gabriel García Márquez a respeito do naufrágio em terras norte-americanas de um garoto cubano, Elián González, ao final do ano de 1999, traduzido para o inglês em dois jornais de ampla circulação nos Estados Unidos da América, *The Guardian*, de origem britânica, e *The New York Times*,

de origem norte-americana, e uma versão em inglês traduzida pelo grupo cubano *Granna International* e disponibilizada via Internet. Investigando as três metafunções desenvolvidas por Halliday (2004), Munday (2002) demonstra que os editores, e talvez os próprios tradutores, tenham sido motivados por questões ideológicas e políticas ao produzirem versões diferenciadas da narrativa em torno do dilema do garoto cubano, através de mudanças na transitividade dos textos traduzidos (aspecto ideacional) (p.85), de expressões discursivas de modalização (aspecto interpessoal) (p.86), de aumento ou diminuição de citações do nome Elián nas traduções (aspecto textual) (p.87), dentre outros factores. Percebe-se, pois, que a teoria Sistêmico-Funcional e a Linguística de Corpus, aliadas à teoria de Estudos Descritivos da Tradução, oferecem meios analíticos e metodológicos para a investigação de escolhas léxico-gramaticais em traduções motivadas por questões ideológicas e hegemônicas. Munday, portanto, sistematiza adequadamente esse movimento *do-texto-ao-contexto-e-vice-versa*, partindo de análises textuais e apontando para possíveis impactos na cultura receptora que o texto traduzido possa causar.

Um outro estudo, na área da Tradução, baseado em corpus e amparado pela teoria hallidayana é o de Ghadessy & Gao (2001). Estes teóricos analisaram quantitativamente as estruturas temáticas de textos originais em língua inglesa e suas traduções para o chinês. A amostra de textos investigada é de livros-texto usados para o ensino de tradução em universidades chinesas. Após identificar os limiares sintagmáticos das orações para a identificação dos temas, Ghadessy e Gao classificaram-nos em categorias conforme suas funções oracionais, oferecendo um elenco variado de tipos de temas encontrados no corpus paralelo por eles estudado. Os resultados da pesquisa levaram os autores a averiguarem uma semelhança na estrutura temática do texto original em comparação com o texto traduzido, embora, segundo Ghadessy e Gao, tenham ocorrido diferenças relevantes na quantidade e nos tipos de temas em ambos os corpora. Para os autores, esse tipo de pesquisa, longe de responder aos questionamentos sobre tradução e equivalência, “conduz-nos para uma melhor compreensão de um aspecto discursivo significativo dos textos que exercem papel importante em nossas decisões acerca do conceito de equivalência em tradução” (2001:353).

No Brasil, os estudos na interface Estudos da Tradução e Lingüística de Corpus se deram, pioneiramente, no Projeto CORDIALL citado anteriormente, com as pesquisas de Cruz (2003), Mauri (2003), Jesus (2004) e Assis (2004). Cruz (2003) examina o texto original *Harry Potter and the chamber of secrets* e sua tradução para o português brasileiro, *Harry Potter e a câmara secreta*. A instância do discurso que a pesquisadora investiga são os verbos de elocução (*reporting verbs*) e os processos de transitividade que os mesmos configuram. O aspecto principal de investigação são as escolhas que a autora Joanne K. Rowling fez dos verbos de elocução para a construção de suas personagens e como a tradutora, Lia Wyler, retextualiza essa construção através dos verbos de elocução. Os resultados da pesquisa de Cruz apontam para um predomínio, no texto original, de ocorrências de verbos de elocução em torno do processo verbal *say*, diferentemente da tradução, em que a tradutora opta pelo processo verbal *perguntar*, sinalizando um ato de fala das personagens. Além disso, no texto original percebe-se o uso freqüente de processos que sinalizam sons de animais e expressões faciais das personagens, ao passo que na tradução a utilização desses processos é bem reduzida. Por fim, ao utilizar, em número maior, processos comportamentais, a autora, Joanne K. Rowling, constrói personagens conscientes dos comportamentos físicos e psicológicos executados por elas; já na tradução, Lia Wyler, ao preferir processos verbais, “apresenta os personagens como falantes, podendo ser conscientes ou não, sem focalização específica de comportamentos físicos e psicológicos” (Cruz 2003:189).

Mauri (2003), por sua vez, realizou uma análise contrastiva dos contos *Laços de família*, de Clarice Lispector, e sua tradução para o italiano, *Legami familiari*, feita por Adelina Aletti. O objectivo da pesquisa concentrou-se nos aspectos indicadores de semelhanças e diferenças no uso dos verbos de elocução, tanto em língua portuguesa quanto italiana, a fim de observar se os níveis de introspecção das personagens femininas variavam na tradução. Os resultados apresentados por Mauri sinalizam para um número considerado de processos mentais sendo utilizados como verbos de elocução no original. Na tradução, por sua vez, essas mesmas escolhas nem sempre ocorriam, devido às estruturas da língua italiana, ou levando-se em conta o contexto sócio-cultural do texto de chegada, ou o estilo da própria tradutora.

Jesus (2004) investiga padrões de textualização característicos do estilo do escritor e tradutor Érico Veríssimo. Após a publicação de sua tradução, em 1934, da obra do autor inglês Aldous Huxley, *Point counter point*, cujo título foi *Contraponto*, Érico Veríssimo escreve *Caminhos cruzados* em 1935, sofrendo críticas por ter reempregado nesta última as técnicas ou estratégias de textualização presentes em *Point counter point*. Jesus, pois, investiga os três romances, i.e. original e sua tradução e *Caminhos cruzados*, sob a perspectiva do sistema de transitividade de Halliday (2004), mais especificamente focalizando o processo mental *think* e sua tradução *pensar*, com o intuito de avaliar padrões de textualização nas três obras analisadas. Os resultados da pesquisa demonstram que Veríssimo-autor e Veríssimo-tradutor textualizam o processo mental *pensar* de maneiras diferenciadas, apresentando diferenças em ambas as relações de textualização. No dizer de Jesus (2004:115),

*[e]nquanto Huxley enfatiza a representação do conteúdo das personagens através do uso de citação e paráfrase, configurando uma significação condizente com o gênero da obra - um romance de idéias, Veríssimo, por outro lado, focaliza, em Caminhos cruzados, a fragmentação dos pensamentos das personagens, através do uso de sintagma preposicionado e relato descontínuo, aspecto também condizente com o gênero da obra - um romance urbano moderno, entre a crônica de costumes e a notação intimista.*

Por fim, Assis (2004) investiga a representação discursiva da personagem Sethe no romance original *Beloved*, de Tony Morrison, em comparação com a tradução *Amada*, feita por Evelyn Kay Massaro. Privilegiando o sistema de transitividade (Halliday 2004), Assis focalizou os tipos de processos utilizados para representar a protagonista Sethe como *Ator* (processo material), *Experienciador* (processo mental), *Dizente* (processo verbal), *Comportante* (processo comportamental) ou *Portador* (processo relacional), dando origem à nomenclatura AEDCP proposta por ele. A fim de analisar a construção discursiva de Sethe inserida nessa metodologia, Assis selecionou as instâncias discursivas em que Sethe era narradora onisciente e onipresente. Dessa forma, Assis explorou a representação de Sethe no original e em sua tradução, podendo constatar que as

retextualizações dos processos relacionais e verbais ocorreram de forma mais “flexível” (Assis 2004:111), ao passo que os outros processos assumiram padrões de (re)textualização análogos.

Além desses trabalhos pioneiros na área de tradução e Lingüística de Corpus, existem outros do Projeto CORDIAL em andamento, bem como no NUT (Núcleo de Tradução) da Universidade Federal de Santa Catarina, Brasil, principalmente os trabalhos de Morinaka (em andamento) e Paquilin (em andamento). Morinaka investiga como ocorre a coesão lexical no romance *Gabriela cravo e canela*, de Jorge Amado, sobretudo como os padrões coesivos contribuem para a representação da personagem principal do romance, Gabriela, no texto traduzido para o inglês. Paquilin, por sua vez, explora como as relações temáticas são usadas para demonstrar as várias facetas de construção de significados no original *Bridget Jones' diary*, de Helen Fielding, e sua tradução para o português brasileiro.

### **Considerações finais**

As pesquisas apresentadas neste ensaio retrataram o potencial da Lingüística de Corpus para as investigações em Tradução, sobretudo a utilidade de ferramentas computacionais para a viabilização de análises tradutórias em comparação com originais. O enfoque principal recaiu sobre o interesse crescente que os estudiosos em Tradução têm apresentado acerca do papel metodológico que a Lingüística de Corpus exerce para a viabilização de suas pesquisas. Neste sentido, o advento dos corpora de pequenas dimensões abriu novas perspectivas teóricas e metodológicas de análise de textos originais e suas respectivas traduções, preenchendo a lacuna que o conceito de *representatividade* apresentava quando da investigação de corpora menores. A noção de representatividade, pois, torna-se flexível ao se abrir para as especificidades dos objectivos de pesquisa, para os interesses do(a) pesquisador(a) e, notadamente, para os problemas de pesquisa exequíveis ao seu tamanho. A questão da representatividade, neste enfoque, passa a depender dos interesses que o(a) pesquisador(a) tenciona perseguir em sua pesquisa, levando sempre em consideração que os resultados da mesma serão cada

vez mais genéricos e pontuais quanto maior o corpus, tanto em número de ocorrência de palavras quanto de gêneros textuais.

## Referências

- ASSIS, R. C. *A Transitividade na Representação de Sethe no Corpus Paralelo Beloved-Amada*. Belo Horizonte: FALE/UFMG, 2004. (Dissertação de Mestrado)
- BAKER, M. “Corpora in Translation Studies: an overview some suggestions for future research”, *Target*, 7, 1995. Pp.223-243.
- BAKER, M. “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/Philadelphia: John Benjamins, 1993. Pp.233-250.
- BAKER, M. “Corpus-based Translation Studies: the challenges that lie ahead”. In: SOMERS, H. (org.). *Terminology, LSP and Translation: studies in language engineering, in honour of Juan C. Sager*. Amsterdam: John Benjamins, 1996. Pp.175-186.
- BAKER, M. “Réexplorer la langue de la traduction: une approche sur corpus”, *Meta*, 43, 1998. Pp.480-485.
- BAKER, M. “The role of corpora in investigating the linguistic behaviour of professional translators”, *International Journal of Corpus Linguistics*, v. 4, n.2, 1999. Pp.281-298.
- BAKER, M. “Towards a Methodology for Investigating the Style of a Literary Translator”, *Target*, 12 (2), 2000. Pp.241-266.
- BERBER SARDINHA, T. “Corpora Eletrônicos na Pesquisa em Tradução”, *Cadernos de Tradução*, n.2, 2002. Pp.15-59. (Número especial sobre Tradução e Corpora organizado por Stella Tagnin.)
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri, SP: Manole, 2004.
- CRUZ, O. M. de S. e S. da. ‘Harry Potter and the Chamber of Secrets’ e sua tradução para o português do Brasil: uma análise dos verbos de elocução com base na lingüística sistêmica e nos estudos de corpora. Belo Horizonte: FALE/UFMG, 2003. (Dissertação de Mestrado.)
- GHADESSY, M. e GAO, Y. “Small Corpora and Translation: comparing thematic organization in two languages”. In: GHADESSY, M., HENRY, A. e ROSEBERRY, R. L. *Small Corpus*

*Studies and ELT: theory and practice*. Amsterdam/Philadelphia: John Benjamins, 2001. Pp.335-359.

GHADESSY, M., HENRY, A. e ROSEBERRY, R. L. *Small Corpus Studies and ELT: theory and practice*. Amsterdam/Philadelphia: John Benjamins, 2001.

HALLIDAY, M. A. K. *An Introduction to Functional Grammar*. 3<sup>rd</sup> edition, rev. ampl. por Christian M. I. M. Matthiessen. UK: Arnold, 2004.

IÑIGUEZ, L. (Ed.). *Manual de Análise do Discurso em Ciências Sociais*. Trad. Vera Lúcia Joscelyne. Petrópolis: Vozes, 2004.

JAWORSKI, A. e COUPLAND, N. "Introduction: perspectives on discourse analysis". In: JAWORSKI, A. e COUPLAND, N. (eds.). 2<sup>nd</sup> print. *The Discourse Reader*. London e New York: Routledge, 2000.

JESUS, S. M. *Representação do Discurso e Tradução: padrões de textualização em corpora paralelo e comparável*. Belo Horizonte: FALE/UFMG, 2004. (Dissertação de Mestrado.)

KENNY, D. "Creatures of habit? What translators usually do with words", *Meta*, v.43, 1998. Pp.515-523.

KENNY, D. *Lexis and Creativity in Translation: a corpus-based study*. Manchester, UK: St. Jerome Publishing, 2001.

LAVIOSA, S. "The Corpus-based Approach: a new paradigm in translation studies", *Meta*, 43 (4), 1998. Pp.631-651.

MAGALHÃES, C. M. "Pesquisas Textuais/Discursivas em Tradução: o uso de corpora." In: PAGANO, A. (org.). *Metodologias de Pesquisa em Tradução*. Belo Horizonte: FALE/UFMG, 2001. Pp.93-116.

MAURI, C. *Um Estudo da Tradução Italiana de 'Laços de Família', de Clarice Lispector, a partir da Abordagem em Corpora: a construção da introspecção feminina através dos verbos de elocução*. Belo Horizonte: FALE/UFMG, 2003. (Dissertação de Mestrado.)

MORINAKA, E. M. *Shifts of Cohesion in 'Gabriela Cravo e Canela' and its English Translation: a case-study based on systemic-functional linguistics and corpus linguistics*. Florianópolis: UFSC, em andamento. (Projeto de Dissertação de Mestrado.)

MUNDAY, J. "A Computer-Assisted Approach to the Analysis of Translation Shifts", *Meta*, 43 (4), 1998a. s/p.

MUNDAY, J. “Problems of Applying Thematic Analysis to Translation between Spanish and English”, *Cadernos de Tradução*, n. 3, 1998b. Pp.183-213.

MUNDAY, J. “Systems in Translation: a systemic model for descriptive translation studies”. In: HERMANS, T. (ed.). *Crosscultural Transgressions: research models in translation studies II, historical and ideological issues*. Manchester, UK e Northampton MA: 2002. Pp.76-92.

PAGANO, A. “Apresentação.” In: PAGANO, A. (org.). *Metodologias de Pesquisa em Tradução*. Belo Horizonte: FALE/UFMG, 2001. Pp.7-11.

PAGANO, A., MAGALHÃES, C. e ALVES, F. “Towards the Construction of a Multilingual, Multimodal Corpus: factors in the design and applications of CORDIAL”, *Revista TradTerm* (no prelo).

PAQUILIN, V. The Various Facets of a message: an analysis of the thematic structure in *Bridget Jones’s Diary* in the light of the systemic functional grammar, corpus linguistics and translation studies interface. Florianópolis: UFSC, em andamento. (Projeto de Dissertação de Mestrado.)

SCOTT, M. *WordSmith Tools Manual - version 3.0*. Oxford: Oxford University Press, 1996, 1997, 1998.

SINCLAIR, J. M. “Preface”. In: GHADESSY, M., HENRY, A. e ROSEBERRY, R. L. (eds.). *Small Corpus Studies and ELT: theory and practice*. Amsterdam/Philadelphia: John Benjamins, 2001. Pp.vii-xv.

SINCLAIR, J. M. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

STUBBS, M. *Text and Corpus Analysis; computer-assisted study of language and culture*. Oxford: Blackwell, 1996.

STUBBS, M. *Words and Phrases: corpus studies of lexical semantics*. Oxford: Blackwell, 2001.

TEUBERT, W. & ČERMÁKOVÁ, A. “Directions in corpus linguistics”. In: HALLIDAY, M. A. K., TEUBERT, W., YALLOP, C. & ČERMÁKOVÁ, A. (Eds.). *Lexicology and Corpus Linguistics*. London & New York: Continuum, 2004. Pp. 113-165.

TEUBERT, W. “Language and corpus linguistics”. In: HALLIDAY, M. A. K., TEUBERT, W., YALLOP, C. & ČERMÁKOVÁ, A. (Eds.). *Lexicology and Corpus Linguistics*. London & New York: Continuum, 2004. Pp. 73-112.

TOURY, G. "The Nature and the Role of Norms in Translation". In: VENUTI, L. e BAKER, M. (eds.). *The Translator Studies Reader*. London e New York: Routledge, 2000. Pp.198-211. (originalmente publicado em 1978 e revisto em 1995.)

VENUTI, L. e BAKER, M. *The Translation Studies Reader*. London e New York: Routledge, 2000.

WETHERELL, M., TAYLOR, S. & YATES, S. J. (Eds.). *Discourse Theory and Practice: a reader*. London, Thousand Oaks & New Delhi: The Open University & Sage, 2001.