

UNIVERSIDADE DE SÃO PAULO – USP
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
ALESSANDRA OTERO GOEDERT
ORIENTADOR: PROF. TITULAR FRANCIS HENRIK AUBERT

PROJETO DE MESTRADO

**A COLABORAÇÃO DA LINGUÍSTICA DE CORPUS AO OFÍCIO DO
TRADUTOR NA ELABORAÇÃO DE GLOSSÁRIOS EFICAZES – A
TIPOLOGIA DE DOCUMENTOS ESCOLARES E ACADÊMICOS INGLÊS-
PORTUGUÊS**

São Paulo – SP

Maio de 2012

SUMÁRIO

INTRODUÇÃO, PERGUNTAS DA PESQUISA	4
JUSTIFICATIVA	6
OBJETIVOS GERAIS	7
PRESSUPOSTOS TEÓRICOS	8
METODOLOGIA DE PESQUISA	10
CRONOGRAMA	12
BIBLIOGRAFIA	13

RESUMO

Este projeto propõe a realização de uma pesquisa de mestrado a qual se dedicará a estudar e expor o modo como a Linguística de Corpus pode contribuir para o ofício de tradutor na elaboração de glossários bilíngues. Com o avanço tecnológico e fácil acesso a informações, o tradutor se depara com prazos cada vez mais exíguos para realização dos trabalhos, restando pouco tempo para se dedicar a pesquisas detalhadas de termos. Torna-se essencial para sua prática tradutória maior agilidade na armazenagem e recuperação de dados já pesquisados. A Linguística de Corpus, abordagem empirista que vê a língua como um sistema probabilístico, tem se dedicado à identificação de padrões léxico-gramaticais recorrentes na linguagem por meio da observação de textos autênticos, sob a forma de corpora eletrônicos. Nesse contexto, pretendemos utilizar textos extraídos de Livros de Registro de Traduções de Tradutor Público da tipologia de documentos escolares e acadêmicos, segundo os critérios de coleta de corpus que constituirá a base para a identificação das unidades de tradução para posterior exploração desse corpus. O levantamento dos padrões léxico-gramaticais será efetuado mediante o uso do software *WordSmith Tools*. A escolha da tipologia de documentos acadêmicos, foco de nosso estudo, funda-se no crescente incentivo ao intercâmbio internacional entre Brasil e Estados Unidos, permitindo também que seja traçado um paralelo entre os sistemas educacionais dos países abrangidos pelo par de línguas, na tentativa de se elaborar um glossário bilíngue que possa contribuir para a compreensão do original e auxiliar o tradutor na produção textual na língua-alvo.

INTRODUÇÃO, PERGUNTAS DA PESQUISA

Language cannot be invented, it can only be captured.

Sinclair

Diante da rápida evolução tecnológica, como utilizar de forma otimizada as inúmeras ferramentas à disposição do tradutor? Percebe-se a grande disponibilidade de informações nunca antes tão acessíveis e ao mesmo tempo a crescente demanda do mercado por mais agilidade e rapidez na entrega dos trabalhos. Nunca se teve acesso quase que imediato a tanto conhecimento e tão pouco tempo para usufruir e pesquisar mais detalhadamente termos e expressões mais frequentemente utilizados nas respectivas línguas envolvidas no ato tradutório. A evolução deu-se de forma tão vertiginosa que há menos de meio século passou-se da armazenagem de dados em equipamentos pesados (*mainframes*) a um salto quase que quântico literalmente para as nuvens (*cloud computing*), sem falar nas ferramentas *CAT* (*Computer-Assisted Translation*), *MT* (*Machine Translation*) e *TM* (*Translation Memory*). Nesse novo contexto informático, como deve proceder o tradutor? Mais do que se atualizar e acompanhar as novas tecnologias, o tradutor para desempenhar bem seu ofício dentro desse contexto de extremo dinamismo, precisa adequar seu procedimento tradutório e se valer das novas descobertas. A questão em voga refere-se não mais apenas ao *know-how* adquirido na formação acadêmica e durante as longas jornadas de trabalho e pesquisas incessantes em buscas dos termos e expressões adequados na transposição do código fonte para o código de chegada, porém principalmente ao *know-where*, ou seja, saber onde estão as informações, qual o lugar para capturá-las de modo ágil e eficaz. Ademais, o tratamento e organização dos dados capturados para utilização futura torna-

se crucial. A agilidade exigida pelo mercado só pode ser satisfeita mediante a recuperação eficaz de dados suportada por uma armazenagem qualitativa. Eis os problemas que a adaptação à evolução tecnológica nos coloca, problemas estes cujas possíveis soluções são fundamentais para o bom desempenho do ofício do tradutor. Ou seja, tratam-se de problemas a partir dos quais é possível estabelecer uma nova metodologia de trabalho, visando não apenas o momento presente, mas como viabilizar a utilização dos termos e expressões adquiridos e capturados a cada novo trabalho para uso em tarefas futuras.

Nesse contexto se enquadra a Linguística de Corpus, abordagem empirista que vê a língua como um sistema probabilístico, permitindo a coleta e análise de dados autênticos passíveis de serem compilados em glossários para um uso mais eficaz das informações. Sendo assim, o modo pelo qual a Linguística de Corpus pode contribuir para o ofício do tradutor, para que este profissional não se sinta continuamente recomeçando a mesma pesquisa ou em um ciclo de retrabalho, constituirá o cerne deste projeto de pesquisa.

O foco de nosso estudo para a aplicação da Linguística de Corpus será a área acadêmica, através de diplomas, certificados e históricos escolares, tipologia textual normalmente muito solicitada aos tradutores e que nos últimos tempos vem demonstrando um crescimento acelerado. O referido tema de estudos que constituirá o corpus da pesquisa apresenta características peculiares, uma vez que abrange sistemas educacionais de diferentes países, reforçando, portanto, a necessidade de análise criteriosa dos dados.

JUSTIFICATIVA

O presente projeto de pesquisa justifica-se, em primeiro lugar, por se propor a analisar de que forma a Linguística de Corpus pode contribuir na organização e análise de dados acumulados no decorrer da realização do ofício de tradutor, de modo a que seja incorporada ao processo tradutório como uma ferramenta de apoio na compilação e constituição de glossários bilíngues comparáveis. A proposta de tratamento de dados obtidos a partir de traduções juramentadas de diplomas, certificados e históricos escolares do inglês para português ou vice-versa, mostra-se pertinente em virtude da grande abertura para intercâmbio acadêmico demonstrada pelos Estados Unidos através das palavras do Sr. Thomas J. Dowling, *Public Affairs Officer*, Adido Cultural do Consulado dos Estados Unidos em São Paulo, em entrevista concedida ao programa JustTv –Programa de Negócios e Carreira datada de 4/10/2010, reiterando o interesse dos Estados Unidos em “manter as portas abertas aos brasileiros – “*our doors are wide open*”. O fomento a este estreitamento de relações acadêmicas certamente ocasionará um fluxo crescente de estudantes brasileiros aos Estados Unidos, bem como de estrangeiros para o Brasil interessados em compartilhar essa troca cultural e melhores oportunidades de trabalho, sem mencionar o desenvolvimento de pesquisas acadêmicas.

Destaca-se aqui a relevância da presente pesquisa no auxílio ao ofício do tradutor, visto que a partir da utilização da Linguagem de Corpus na organização de dados visando a elaboração de glossários, o tradutor irá otimizar seu tempo de pesquisa estando apto a desse modo atender a crescente demanda por traduções na área acadêmica.

Em segundo lugar, justifica-se na medida em que se encaixa no perfil do tradutor que irá utilizar seus conhecimentos acadêmicos e da pesquisa em si adquiridos no

Mestrado para seu aperfeiçoamento profissional, bem como a possibilidade de interação produtiva entre a referida vivência profissional e o contexto da pesquisa.

OBJETIVOS GERAIS

O objetivo principal da pesquisa que se pretende realizar a partir deste projeto é descrever a forma pela qual a Linguística de Corpus pode auxiliar os tradutores na tarefa de coleta criteriosa e comparação de dados de traduções já realizadas (os corpora) através do uso de softwares, como por exemplo, o *WordSmith Tools* de autoria de Mike Scott (1999) para a elaboração de glossários bilíngues. O principal benefício oriundo da incorporação no processo tradutório da Linguística de Corpus que, conforme aponta Sardinha (2000, p.349) “(...) apresenta uma abordagem empirista e uma visão da linguagem enquanto sistema probabilístico (...)”, será a obtenção de glossários mais confiáveis e adequados devido ao fato de considerarem quaisquer padrões associativos entre palavras. Desse modo, pretendemos levar a cabo uma interpretação dos dados coletados de traduções realizadas por tradutor público para a compilação de um glossário bilíngue da área acadêmica.

Como objetivo secundário, tendo em vista o recorte delimitador do corpus escolhido, notadamente a área acadêmica, procederemos a uma análise comparativa dos sistemas educacionais do Brasil, Inglaterra e Estados Unidos, que complementarará e fornecerá subsídios e esclarecimentos para as definições a serem inseridas no glossário bilíngue, resultado final da pesquisa realizada (glossário este que poderá ser incorporado ao corpus COMET – Corpus Multilíngue para o Ensino e para a Tradução).

PRESSUPOSTOS TEÓRICOS

Conforme colocado por Fillmore (1992, p. 35-60): “Todo corpus (...) por menor que fosse, ensinou-me fatos que nunca poderia sequer imaginar descobrir de outra forma.”¹ Esta assertiva é a explicação sobre a mudança de paradigma que o advento do trabalho com corpora proporcionou para a Linguística, mudança que vai ao encontro do comentário de Izabella dos Santos Martins em sua resenha do livro *Linguística de Corpus* de Berber Sardinha:

(...) o racionalismo é colocado em xeque e o empirismo é apontado como novo modelo e forma de pensar na Linguística. Muitos achados, e uma quantidade surpreendente de evidências linguísticas, só são possíveis de obter pela observação e o trabalho com a linguagem em uso, autêntica, pressupostos para a pesquisa em Linguística de Corpus (LC). O advento do computador, em virtude de sua grande memória e capacidade de armazenamento, teve e tem um papel central nessa mudança de olhar sobre a linguagem, já que proporciona sistematização aos fatos e os evidencia (MARTINS, 2007, p. 384).

Quando se fala que a Linguística de Corpus é uma abordagem empirista da linguagem, o que se quer dizer é dar prioridade aos dados obtidos oriundos da observação da linguagem, sendo a teorização feita posteriormente. Esta visão se choca com os modelos mentalistas, segundo os quais a linguagem deve ser estudada por introspecção e que o conhecimento provém de princípios estabelecidos inicialmente. Nesse modelo, o que está em foco é a competência linguística, enquanto que para os empiristas o foco está no desempenho, no uso linguístico.

Entende-se por Corpus:

(...) um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador,

¹ “Every corpus [...] however small, has taught me facts that I couldn’t imagine finding out about in any other way.”

com a finalidade de propiciar resultados vários e úteis para a descrição e análise (SANCHEZ, 1995, p. 8-9).

A pertinência da escolha de traduções juramentadas extraídas dos Livros de Registro de Traduções de um Tradutor Público devidamente nomeado pela Junta Comercial do Estado de São Paulo deve-se ao fato de atender aos critérios necessários para que um conjunto de dados linguísticos possa ser considerado um corpus, a saber, critério de (a) origem, por serem dados autênticos; (b) formatação, por serem legíveis por computador; e (c) extensão, por ser vasto, conferindo representatividade. Os critérios de (d) propósito, no tocante à finalidade de ser um objeto de estudo linguístico e (e) composição, por ser criteriosamente escolhido, são atendidos pela escolha da área de documentos escolares e acadêmicos, abrangendo diplomas, históricos escolares, certificados.²

A escolha do software *Wordsmith Tools* mostra-se como sendo a mais adequada, dado que, segundo Sardinha, este programa é: “Ainda hoje, depois de muitas versões, o mais completo e versátil conjunto de ferramentas para Linguística de Corpus”. (SARDINHA, 2000, p. 335)

É importante salientar que para a consecução da pesquisa delineada neste projeto serão realizadas leituras de renomados autores e pesquisadores tais como Sinclair (1996), trabalho pioneiro na área de léxico que traçou os caminhos da maioria das pesquisas em Linguística de Corpus até hoje, Biber (1988), trabalho excepcional de descrição da composição linguística de gêneros da língua inglesa a partir de dois dos mais famosos corpora (LOB e London-Lund), Kjllmer (1994), primeiro dicionário de

² Conforme aponta Sardinha (2000, p.325): (...) a Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.”

colocações baseado em corpus, uma vez que seu predecessor, o dicionário BBI de colocações não fornece indicações claras de ter sido criado seguindo o mesmo princípio (BENSON et al., 1986), Stella Ester O. Tagnin, Antonio Paulo Berber Sardinha, dentre outros.

METODOLOGIA DE PESQUISA

O trabalho se dividirá na seleção de traduções pertencentes aos Livros de Registro de Traduções de Tradutor Público devidamente nomeado pela Junta Comercial do Estado de São Paulo, para a exploração da linguagem de documentos escolares e acadêmicos, a saber, diplomas, históricos escolares e certificados, delimitando desse modo o recorte do corpus e, com base nessa seleção criteriosa, na submissão do corpus ao programa *WordSmith Tools*. O referido programa de software irá realizar a seleção de termos e fraseologias pertinentes à área escolhida para estudo, com o objetivo de capturar os termos mais recorrentes para posterior análise. Procederemos à análise individual de cada termo com a finalidade de inclusão no glossário bilíngue de termos acadêmicos, a ser obtido como resultado final deste projeto de pesquisa, servindo de exemplo tanto no aspecto metodológico quanto analítico para a elaboração e compilação de futuros glossários nas mais diversas áreas do conhecimento.

Mediante referência ao trabalho de pesquisa de Iniciação Científica de Adriana Zavaglia (USP) e Carolina Poppi Martins (USP-PG), a saber, um Glossário jurídico-cartorial português-francês, no qual o par de línguas estudadas remete a culturas diferentes, bem como sistemas jurídicos diferentes, observamos que o mesmo ocorre em nossa pesquisa, no tocante aos sistemas educacionais do Brasil, Inglaterra e Estados Unidos. Isto posto, em caráter complementar serão realizadas análises e traçados paralelos através das leituras pertinentes, incluindo, particularmente, a Lei de Diretrizes

e Bases da Educação do Brasil e leis similares nos países objeto da presente pesquisa para a finalidade de nortearmos as análises e definições dos termos do glossário pertinentes a cada sistema educacional. Ademais, no tocante ao tema central serão realizadas leituras adicionais relativas ao funcionamento do software, aos diversos corpora existentes, bem como dos autores e pesquisadores citados nos pressupostos teóricos.

CRONOGRAMA

Na execução da pesquisa de Mestrado, compreendendo o depósito da Dissertação, seguiremos o cronograma abaixo:

ATIVIDADES	PERÍODO		
	Primeiro ano	Segundo ano	Terceiro ano
Leitura da bibliografia principal	X	X	X
Leitura da bibliografia secundária		X	X
Levantamento das traduções para formação do corpus	X	X	
Aplicação do Software <i>WordSmith Tools</i> para coleta e comparação dos dados	X	X	X
Análise dos resultados obtidos com o Software <i>WordSmith Tools</i>		X	X
Elaboração do glossário		X	X
Redação da tese		X	X

REFERÊNCIAS BIBLIOGRÁFICAS

AUBERT, FRANCIS H.; BARROS, Lidia Almeida; BABINI, Maurizio. *Terminologia e Tradução Juramentada: questões de tipologia textual e equivalência terminológica interlinguística Português-francês-italiano*. *Filol. linguíst. port.*, S.1, v. 2, n. 12, p. 233-249, 2010.

BASSNETT, Susan. *Estudos de Tradução: Fundamentos de uma disciplina*. Trad. Vivina de Campos Figueiredo. Lisboa: Calouste Gulbenkian, 2003.

DOWLING, Thomas J. Entrevista concedida a George Niaradi. Trad. Alessandra Goedert. São Paulo, 2010. Disponível em: <http://www.justtv.com.br/carreirasenegocios/>. Acessado em 10/05/12.

LEPRE, Larissa. *A elaboração de glossários bilíngues para interpretar textos em inglês com base em um corpus paralelo*. Florianópolis, 2007. Dissertação (Mestrado) - UFSC.

MARTINS, Izabella dos Santos. *Linguística de corpus*. *DELTA*, São Paulo, v. 23, n.2, p. 383-393, 2007. Resenha de: SARDINHA, Tony Berber. *Linguística de corpus*. Barueri: Manole, 2004. Disponível em: <http://dx.doi.org/10.1590/S0102-44502007000200009>. Acessado em 10/05/12.

SARDINHA, Tony Berber. *Linguística de Corpus: Histórico e Problemática*. *DELTA*, São Paulo, v. 16, n.2, p. 323-367, 2000.

TAGNIN, Estella E. O. *A Linguística de Corpus e o Tradutor: uma relação de futuro Parte I*. São Paulo: Edusp/CIATI, 2004.

TEIXEIRA, Elisa Duarte. *A Linguística de Corpus a serviço do tradutor: proposta de um dicionário de Culinária voltado para a produção textual*. São Paulo, 2008. Tese (Doutorado) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

VIANNA, Vander; TAGNIN, Estela E.O. (Orgs.). *Corpora no Ensino de Línguas Estrangeiras*. São Paulo: HUB, 2010.

ZAVAGLIA, Adriana; MARTINS, Carolina Poppi. *Glossário Jurídico-Cartorial português-francês*. In: VII ENCONTRO DE LINGUÍSTICA DE CORPUS, Rio de Janeiro, 2009.