

## APRESENTAÇÃO

Os Estudos Lingüísticos com Corpus vêm, paulatinamente, conquistando seu espaço no cenário acadêmico brasileiro, principalmente no âmbito da tradução, onde, até há pouco, eram praticamente desconhecidos. Para difundir essa área editei, em 2003, um número especial dos *Cadernos de Tradução* do Núcleo de Tradução da Universidade Federal de Santa Catarina sob o título “Tradução e Corpora”.

O presente volume vem na esteira daquele e pretende apresentar uma amostragem do que se tem feito com corpora na área da tradução, em especial, mas não apenas, no Brasil. Traz artigos que enfocam corpora já construídos ou em construção, que discutem os usos a que se prestam, ou que relatam pesquisas com eles realizadas.

O primeiro artigo, de Sara Laviosa, docente da Universidade de Bari (Itália) e criadora do ECC – English Comparable Corpus –, em 1996, apresenta-nos um panorama detalhado dessa parceria. A autora traça a evolução dos Estudos da Tradução Baseados em Corpora (ETCs) desde o artigo seminal de Baker (1993), passando por um período em que fica marcante o interesse comum entre os Estudos da Tradução e a Lingüística de Corpus, chegando aos dias de hoje, em que se vislumbra uma penetração dos Estudos Culturais nos ETCs. Assim, o artigo examina as relações entre os Estudos Descritivos da Tradução, os ETCs e a Lingüística de Corpus, no intuito de estabelecer quais propostas do passado permanecem válidas e quais áreas de investigação revelam-se como mais promissoras dentro dos ETC.

A esse texto, seguem-se artigos que discutem a construção de vários corpora, como, por exemplo, o de Stig Johansson, do Departamento de Estudos Britânicos e Americanos da Universidade de Oslo (Noruega). O autor é um dos criadores do Corpus LOB (Lancaster-Oslo/Bergen Corpus), em 1978, e um dos primeiros pesquisadores europeus a construir um corpus paralelo multilíngüe bidirecional, o ENPC (English-Norwegian Parallel

TRADTERM, 10, 2004, p. 7-17

Corpus), em 1994.<sup>1</sup> Em seu artigo, Johansson apresenta o modelo desse corpus, a metodologia seguida para sua construção, além de um relato detalhado das formas como pode o corpus ser explorado pelo programa Translation Corpus Explorer, especialmente criado para esse propósito. Por fim, sempre com ilustrações bastante ricas, o autor discute como é possível estudar uma língua por meio da tradução e a tradução por meio de corpora.

Uma das áreas de investigação apontadas por Johansson – a de estudos lingüísticos contrastivos – é explorada por Josef Schmied, diretor do Centro REAL (Research in English and Applied Linguistics), ligado ao Departamento de Língua Inglesa e Lingüística da Universidade Tecnológica de Chemnitz (Alemanha), e pesquisador especialmente interessado no uso da linguagem natural em diversos contextos: históricos, sociais, regionais e estilísticos. O corpus em que Schmied se baseia é o Chemnitz English-German Translation Corpus, também um corpus paralelo bidirecional. O artigo enfoca a Chemnitz Internet Grammar, uma ferramenta de aprendizado que permite aos usuários investigarem partes da gramática da língua inglesa de forma indutiva ou dedutiva. A ferramenta foi criada em 1999 e atualizada em 2003 e se destaca por utilizar, em suas análises, material fornecido pelo referido corpus. Neste artigo, então, Schmied salienta o uso do corpus como ferramenta de descoberta, principalmente por parte de falantes não-nativos, uma vez que seus resultados são baseados na análise de grande quantidade de dados, portanto, independem da introspecção do pesquisador. O autor salienta, ainda, que os corpora paralelos bidirecionais são de grande valia para os estudos contrastivos, por permitirem detectar preferências lingüísticas de uma língua ou analisar fenômenos que indicam gradação, tais como a escala verbo auxiliar-de ligação-pleno. Sob uma perspectiva contrastiva, Schmied ilustra essas possibilidades com estudos de caso sobre o verbo auxiliar *help*, os verbos de ligação *appear/seem* e seus equivalentes em alemão, e o verbo modal *may/might*, com seus sentidos mais co-

---

<sup>1</sup> Devido à falta de uniformização terminológica, cabe esclarecer que o autor entende por corpus paralelo aquele que contém textos originais e suas respectivas traduções.

muns e as estruturas que lhe correspondem em alemão. Como conclusão, o autor discute as limitações dos corpora, salientando, contudo, seu papel nos estudos contrastivos, uma vez que podem apresentar “more grammar than meets the eye”, ou seja, evidenciarem aspectos gramaticais que, de outra forma, passariam despercebidos.

Stella E. O. Tagnin é docente do programa de pós-graduação Estudos Lingüísticos e Literários em Inglês e do Curso de Especialização em Tradução da Universidade de São Paulo e tem publicações nas áreas dos Estudos de Corpora, fraseologia inglesa e tradução literária,<sup>2</sup> entre outros. Neste artigo, ela descreve o projeto COMET – Corpus Multilíngüe para Ensino e Tradução – em desenvolvimento naquela universidade. O COMET é composto de três subcorpora: o CorTeC (Corpus Técnico-Científico), o Corpus de Tradução e o Corpus de Aprendizes. O CorTeC é um corpus bilíngüe inglês-português, composto de subcorpora de diversas áreas técnicas e científicas, dentre as quais quatro que estão recebendo especial atenção, constituindo-se conjuntos textuais ampliados de forma criteriosa e sistemática: Informática, Ortodontia, Meio Ambiente e Direito Comercial. A autora procura, então, esclarecer que se trata de um corpus comparável, ou seja, que os textos desses subcorpora são sempre os originais nas duas línguas, o que permite detectar as formas naturais de expressão em cada língua e em cada área, ou seja, sua fraseologia e terminologia típicas. O Corpus de Tradução, por sua vez, é subdividido em um corpus paralelo e um comparável. O primeiro contempla textos literários e técnico-científicos e contém originais e respectivas traduções, possibilitando pesquisas sobre estratégias ou normas de tradução, bem como estudos léxico-gramaticais contrastivos. Já o corpus comparável de traduções juramentadas é composto de traduções em diversas línguas, produzidas por tradutores juramentados os quais, por aposentadoria ou fa-

<sup>2</sup> Tagnin, S.E.O. *Expressões idiomáticas e convencionais*. São Paulo, Ática, 1989.

\_\_\_\_\_. *Contos Canadenses*. São Paulo, Olavobrás, 2002.

\_\_\_\_\_. (ed.) “Tradução e Corpora” número especial de *Cadernos de Tradução*. Florianópolis, Universidade Federal de Santa Catarina, vol. IX/2002/1.

lecimento, têm seu trabalho depositado na Junta Comercial, neste caso, do Estado de São Paulo. Esse corpus é detalhado no artigo de Aubert & Tagnin abaixo. O Corpus de Aprendizes, por não contemplar a tradução, não é abordado no artigo.

Outro corpus em construção no Brasil é o CORDIAL, apresentado no artigo de Adriana Pagano, Célia Maria Magalhães e Fábio Alves, todos da Faculdade de Letras da Universidade Federal de Minas Gerais. Adriana Pagano atualmente coordena os projetos “Corpora, cognição e discurso: uma proposta interdisciplinar para os estudos da tradução a partir de bancos eletrônicos de dados” e “Corpora, gênero e (re)textualização: interfaces nos estudos da tradução”. Célia Magalhães é professora adjunta de língua inglesa e tradução. Seu principal interesse de pesquisa é o estudo de redes coesivas e seu papel discursivo em corpora paralelos do par lingüístico inglês-português. Também é autora do livro *Os Monstros e a Questão Racial na Literatura Modernista Brasileira*<sup>3</sup> e organizadora de *Reflexões sobre a Análise Crítica do Discurso*.<sup>4</sup> Tem artigos publicados nas revistas *Cadernos de Tradução*, *DELTA*, *CROP* e *TradTerm*, além de capítulos de livros publicados e traduzidos no país. Por fim, Fábio Alves, que é professor adjunto de língua alemã e tradução. Seu principal interesse de pesquisa é o estudo do processo de tradução e de aspectos cognitivos relacionados à aquisição da competência tradutória. É também autor de numerosas publicações no Brasil e no exterior.<sup>5</sup> Juntos escreveram *Traduzir com Autonomia: estratégias para o tradutor em formação* (2000).<sup>6</sup>

O CORDIAL é composto de quatro subcorpora. O primeiro é um corpus paralelo com textos originais e suas traduções nos pares lingüísticos português-espanhol, português-alemão,

<sup>3</sup> Belo Horizonte: Editora UFMG, 2003.

<sup>4</sup> Belo Horizonte: POSLIN/FALE/UFMG, 2001.

<sup>5</sup> Entre outros:

Alves, F. *Zwischen Schweigen und Sprechen: Wie bildet sich eine transkulturelle Brücke?* Hamburg, Dr. Kovac, 1995.

Alves, F. (ed.) *Triangulating translation: perspectives in process oriented research*. Amsterdam, John Benjamins, 2003.

<sup>6</sup> Alves, F.; Magalhães, C. & Pagano, A. *Traduzir com autonomia: estratégias para o tradutor em formação*. São Paulo, Contexto, 2000.

português-inglês e espanhol-inglês. Destina-se a análises comparativas sobre as decisões e estratégias dos tradutores em suas retextualizações, assim como ao exame de parâmetros históricos na produção de traduções. O segundo é um corpus comparável de português do Brasil com textos originais e traduzidos nessa língua, constituição que permite, em especial, detectar peculiaridades da língua traduzida. O terceiro, denominado CORPRAT, é um corpus processual multilíngüe, cujo objetivo é fornecer material para se investigar padrões de inferência, planejamento estratégico, solução de problemas e tomada de decisões durante o processo tradutório. O quarto subcorpus, um corpus especializado, não está relacionado especificamente aos estudos da tradução, já que foi construído com vistas aos estudos de padrões retóricos e aspectos lexicais. É composto de textos associados aos gêneros jornalísticos e acadêmicos. Enfim, neste artigo, os autores discutem o embasamento teórico para a construção desses diferentes corpora, juntamente com projetos em andamento dentro de suas principais linhas de pesquisa, os quais se caracterizam por uma “abordagem interdisciplinar que congrega subsídios dos estudos de corpora, dos estudos da tradução, dos estudos da cognição, da análise do discurso e dos estudos culturais”. Salientam, também, que o CORDIAL é constantemente alimentado e reestruturado e que privilegia estudos baseados em corpora menores, pois permitem uma abordagem discursiva e cognitiva a partir de metodologias desenvolvidas para os corpora em estudo.

Um corpus bastante peculiar é apresentado por Francis Henrik Aubert e Stella. E. O. Tagnin, ambos do Departamento de Letras Modernas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Aubert é tradutor juramentado e tem produção bibliográfica significativa na área de tradução e terminologia. Tagnin, como já foi dito antes, é coordenadora do projeto COMET, onde se insere o corpus em questão. Trata-se de um corpus de traduções juramentadas, produzidas num período de 100 anos (1902-2002) e registradas em livros próprios que, em muitos casos, por serem muito antigos, encontram-se em precário estado de conservação. São, na maior parte dos casos, traduções de uma língua estrangeira para o português e abrangem uma tipologia textual bastante diversificada,

TRADTERM, 10, 2004, p. 7-17

incluindo desde documentos pessoais, como certidões de nascimento e históricos escolares, até estatutos de empresas e contratos de cooperação internacional. Devido aos numerosos problemas técnicos envolvidos na sua digitalização, os autores decidiram iniciar a construção do corpus com as traduções mais recentes (1972-2002), nas cinco línguas que compõem o Departamento onde atuam (alemão, espanhol, francês, inglês e italiano), embora mais de vinte línguas estejam representadas no corpus. Por sua vez, as informações detalhadas do "cabeçalho", como se verá, permitirão ao pesquisador-usuário diversos recortes do corpus (a seleção por língua, por tipo textual, por assunto, por período, por tradutor etc.), adequando-o a suas pesquisas. Os autores salientam, ainda, a possibilidade de estudos interdisciplinares, por exemplo, com a História, a Antropologia, a Economia ou a Política.

Construídos os corpora, nos textos que se seguem temos a apresentação de exemplos práticos de como esses materiais podem ser utilizados e explorados.

A primeira pesquisa aqui relatada insere-se no campo dos estudos cognitivos, com um claro vínculo com as questões pedagógicas. Fabio Alves e Célia Maria Magalhães, docentes da Faculdade de Letras da Universidade Federal de Minas Gerais e já apresentados acima, partem da necessidade de mapear a interface processo-produto na tradução e propõem pesquisar o desempenho de tradutores por meio da triangulação de técnicas de elicitação de dados (como TAPs, o software Translog, que registra todos os movimentos feito pelo tradutor no teclado do computador durante o processo tradutório, a ferramenta de análise WordSmith Tools, entre outros), aplicadas a corpora de pequenas dimensões. O corpus usado é o CORPRAT (descrito no artigo de Pagano, Magalhães e Alves, acima), que permite o estudo dos aspectos cognitivos e discursivos do desempenho dos tradutores. O objetivo da pesquisa em questão é distinguir características do desempenho de tradutores novatos em relação ao de profissionais, no intuito de se obter dados que possam servir de subsídios para reflexões pedagógicas sobre o ensino da tradução. Dentre os tópicos de interesse estão a preocupação em elucidar padrões de orientação, rascunho e revisão dos traduto-

res, suas estratégias de solução de problemas e tomada de decisão, seu “ritmo cognitivo”, bem como as características cognitivas e discursivas ligadas a uma conscientização crítica da língua. A pesquisa envolveu a tradução de um trecho, em inglês, de 63 palavras por 17 sujeitos usando o Translog. Os autores postulam que as fases de orientação-rascunho-revisão seriam muito variadas entre os tradutores novatos e que o ritmo cognitivo seria errático. A análise minuciosa dos dados, como veremos, revela que não há uma co-relação entre o ritmo cognitivo errático de tradutores novatos e a qualidade dos textos traduzidos por eles. Observa-se também um baixo nível de conscientização da língua e a produção de textos menos duradouros. Além disso, um rígido processamento linear e a falta de administração cognitiva adequada e de uma consciência lingüística crítica parece impedir que os novatos melhorem efetivamente seu texto e produzam textos mais estáveis. Esses resultados levam os autores a sugerirem formas de preparar melhor os tradutores novatos de modo que seu desempenho se aproxime daquele dos tradutores profissionais.

Embora o recurso a corpora já venha se disseminando no ensino da tradução, o mesmo não ocorre no âmbito profissional. Para investigar as razões dessa discrepância, Lynne Bowker, docente da Faculdade de Tradução e Interpretação da Universidade de Ottawa e autora de dois livros sobre o uso de corpora no ensino da tradução,<sup>7</sup> faz um levantamento da literatura publicada por docentes de tradução, de um lado, e por associações canadenses de tradutores, de outro. Esse último conjunto é complementado por um corpus de classificados de empregos, construído com o intuito de se analisar quantas empresas buscam tradutores com conhecimento de recursos baseados em corpora. A literatura acadêmica, por sua vez, é discutida tanto do ponto de vista das pesquisas com corpora quanto dos usos de corpora no treinamento dos aprendizes da tradução, incluídas aí as memórias de tradução. A literatura das associações acadêmicas abrange

<sup>7</sup> Bowker, L. *Computer-aided translation technology: a practical Introduction*. Ottawa, University of Ottawa Press, 2002.

Bowker, L. & Pearson, J. *Working with specialized language: a practical guide to using corpora*. London/NY, Routledge, 2002.

especialmente referências a eventos sobre memórias de tradução, pois que as ferramentas convencionais de análise de corpus foram objeto de apenas um evento no Canadá. Estão surgindo, no entanto, ferramentas híbridas, que combinam uma memória de tradução com ferramentas de análise de corpora. Do corpus construído com os classificados, em que, aliás, figuram, além dos tradutores, alguns outros ofícios relacionados (revisor, especialista em localização e terminólogo), Bowker extrai as habilidades que são exigidas dos candidatos, estabelecendo uma diferença entre conhecimentos gerais de computação e familiaridade com memórias de tradução e outras ferramentas similares. A autora conclui, então, que a introdução de ferramentas híbridas pode aumentar o recurso a corpora por parte dos tradutores profissionais, mas que, devido às exigências da profissão, principalmente em termos de produtividade e eficiência, os corpora nunca alcançarão a popularidade de que gozam no meio acadêmico, onde a ênfase é na pesquisa e na divulgação do conhecimento. As conclusões de Bowker apontam, no entanto, que o conhecimento adquirido pelos aprendizes de tradução através do recurso a corpora desenvolve habilidades que podem ser facilmente transferidas para o contexto profissional.

Os quatro artigos seguintes exemplificam possibilidades de pesquisa com corpora bilíngües. Relatam, além disso, estudos pontuais em diversas línguas, enfocando diferentes níveis lingüísticos e recorrendo tanto a abordagens baseadas em corpora (*corpus-based*), quanto a abordagens direcionadas pelo corpus (*corpus-driven*).

Karin Aijmer, docente do Departamento de Inglês da Universidade de Gotemburgo (Suécia) e autora, juntamente com Bengt Altenberg, do ESPC (English-Swedish Parallel Corpus) com quem escreveu *English Corpus Linguistics* (1991), é uma das mais produtivas pesquisadoras na área da Lingüística de Corpus. Em seu artigo, apresenta-nos um estudo contrastivo em nível semântico sobre a polissemia do verbo “to see” e seus equivalentes na língua sueca. Tratando-se de verbos cognatos nas duas línguas e sendo, ambas, línguas germânicas, era de se esperar que houvesse alto índice de casos em que os verbos fossem traduzidos por seus cognatos. Aijmer verificou, no entanto, que na direção inglês-



sueco, essa correspondência era de 69,3%, ao passo que, na direção inversa, é de apenas 50,7%. Uma análise criteriosa das equivalências revelou que, em grande parte, os vários sentidos de *see* podem ser associados a diferentes contextos gramaticais. Por exemplo, *see* seguido de uma oração introduzida por *that* ou *wh-* tem status metalingüístico e se refere a *ver* como fonte em que o emissor se baseia para tirar uma conclusão; já com verbos de emoção (ex.: *I regret to see*) tem significado factivo. Um aspecto interessante que a pesquisadora detecta é o que se denomina “processo de pragmatização”, como é o caso de *I see, you see*, cujo significado é construído no nível interpessoal ou textual. A autora acredita que as extensões dos significados de *see* sejam de certa forma culturais, de modo que uma análise de outras línguas deve revelar novas extensões semânticas desse verbo.

Sara Laviosa contribuiu com outro artigo, desta vez um estudo sobre o anglicismo *business* na língua italiana. O objetivo do estudo é identificar os contextos em que o termo inglês *business* pode ser traduzido pelo anglicismo *business*, em italiano. Para sua investigação, a autora recorre a dois corpora: o COMIC (COMmercial Italian Corpus) e o SALCA, ambos com textos da área de negócios, sendo o primeiro em italiano e o segundo em inglês. Numa primeira etapa a autora identifica, no COMIC, quatro padrões coligacionais para o termo *business* em italiano, relacionando-os a quatro sentidos distintos. Em seguida, investiga, no SALCA, como esses sentidos se realizam em inglês. Por último, sugere os contextos em que *business* pode ser traduzido pelo empréstimo *business*. Trata-se, na realidade, de estudo preliminar para uma pesquisa mais abrangente que pretende enfocar todos os sinônimos de *business* em italiano, tais como *affari, commercio, trattativa, transazione* etc., numa tentativa de sugerir equivalentes para esse termo de alta recorrência na língua italiana.

Outro artigo que contrasta o italiano com o inglês, mas de uma perspectiva sócio-cultural, é o de Elena Tognini Bonelli e Elena Manca, ambas da Universidade de Lecce (Itália). Tognini Bonelli é docente do Departamento de Filologia, Lingüística e Literatura e autora de *Linguistics at Work* (2001), obra que, face à clareza e farta exemplificação que oferece, é de leitura obriga-

TRADTERM, 10, 2004, p. 7-17

tória para todos os interessados em desenvolver pesquisas com corpora. As autoras fazem uso de dois corpora comparáveis sobre a linguagem do turismo – o Farmhols, em inglês, e o Agriturist, em italiano – e descrevem uma metodologia bastante inovadora para chegarem a um equivalente tradutório para *welcome*.<sup>8</sup> A partir das concordâncias e colocados de *welcome* (*children*, *pets*, *dogs* e *guests*) as autoras investigam os contextos de ocorrência dos equivalentes *prima facie* em italiano: *bambini*, *animali* e *ospiti*. A comparação dos padrões sintagmáticos e associações paradigmáticas dos equivalentes tradutórios permitiu-lhes identificar os equivalentes funcionais, os quais, por sua vez, revelaram claras implicações culturais e ideológicas. O estudo também demonstra que, embora palavras isoladas possam ter traduções consagradas em outra língua, o mesmo nem sempre ocorre quando esses formam uma unidade lexical. Nesses casos, é o contexto maior que poderá sugerir traduções funcionalmente equivalentes.

O artigo seguinte, de Stella E. O. Tagnin, já apresentada acima, e Elisa Duarte Teixeira, doutoranda do programa de Estudos Lingüísticos e Literários da Universidade de São Paulo, propõe-se a apresentar dados empíricos que justificam a resposta que, costumeiramente, dão à pergunta freqüente de estrangeiros: “O português do Brasil é muito diferente do português de Portugal?”, ou seja, a resposta de que “A diferença é maior do que entre o inglês americano e britânico”. Para tanto, as autoras construíram um corpus comparável quadricultural (português do Brasil (PB), português europeu (PE), inglês americano (IA) e inglês britânico (IB)), composto de receitas culinárias. A análise das palavras-chave de cada corpus e de cada subdivisão dos corpora, de seus colocados e dos agrupamentos lexicais maiores revelaram significativas diferenças nos níveis lexical, sintático e discursivo. Enquanto, no nível lexical, as variantes do português apresentam 20 palavras do PE que não correm no PB e 7 pala-

---

<sup>8</sup> Isso porque *welcome*, de alta freqüência no corpus em inglês (324 ocorrências) parece não ter correspondente cognato em italiano, pois *benvenuto*, seu equivalente *prima facie*, apresenta apenas 4 ocorrências no corpus italiano.

bras do PB que não ocorrem no PE, as diferenças entre as variantes da língua inglesa restringem-se a 6 palavras relativas a medidas (kg, l, ml, etc.) e apenas 4 palavras de conteúdo (*cornflour*, *prawns*, *courgettes* do IB, que não ocorrem no IA, e *ham* do IA, que não ocorre no IB). No nível sintático, há claras diferenças entre o PB e o PE, em especial o uso da estrutura pronominal impessoal *deixa-se*, *junta-se* no PE, em oposição ao imperativo *deixe*, *junte*, no PB. No nível discursivo, observou-se que as receitas no PB costumam ter duas seções: “ingredientes” e “modo de fazer”, configurando-se uma estrutura não tão usual no PE. Os dados também permitiram às autoras levantar hipóteses culturais a partir da observação de semelhanças entre o português do Brasil e o inglês americano, por um lado, e entre o português europeu e o inglês britânico, de outro. Cientes das limitações desses corpora, propõem aumentá-los para aprofundar essa investigação.

Esperamos que a seleção dos artigos apresentados seja mais uma contribuição para difundir os Estudos de Corpus relacionados à Tradução em nosso país, e que sirva para incentivar os pesquisadores, que ainda não estejam familiarizados com essa metodologia a se aventurarem nessa interface tão promissora e reveladora.

São Paulo, março de 2004

*Stella E. O. Tagnin*