

## Apresentação

Este número especial da revista *Crop* pretende dar um panorama das possibilidades de pesquisa baseadas e/ou direcionadas por corpora eletrônicos. A maioria dos artigos resultou de trabalhos elaborados pelos autores para alguma disciplina por mim ministrada na Universidade de São Paulo e na Universidade Estadual do Ceará, tendo por foco a Lingüística de Corpus.

A Lingüística de Corpus é uma abordagem que privilegia a observação de uma grande quantidade de dados autênticos com o intuito, entre outros, de identificar as ocorrências mais **prováveis** na produção de falantes nativos, em detrimento das ocorrências apenas **possíveis**. Em outras palavras, interessa à Lingüística de Corpus o que de fato o falante diz, não apenas o que ele poderia dizer, já que muitas dessas ocorrências possíveis jamais são realmente produzidas.

A observação de um grande número de dados a partir de corpora eletrônicos – coletâneas de textos reunidos com um objetivo específico de pesquisa e analisáveis com ferramentas computacionais próprias – também permite a visualização de padrões recorrentes, quer lexicais ou gramaticais, padrões que na maioria das vezes passariam despercebidos sem o uso desses recursos eletrônicos. Conhecer esses padrões recorrentes é essencial para o aprendizado de uma língua, pois vai fornecer ao aprendiz elementos para tornar sua produção mais natural, mais fluente, mais próxima da produção de um falante nativo.

Ainda nesse campo, os corpora de aprendizes – constituídos de redações não corrigidas de alunos – são extremamente úteis para identificar áreas problemáticas de aprendizagem, o que, por sua vez, pode servir de base para um ensino mais bem direcionado, assim como para a elaboração de material didático mais eficaz.

Uma das áreas que também se beneficia muito de estudos baseados em corpora é a da tradução. Corpora paralelos, isto é, textos originais juntamente com suas respectivas traduções, permitem ao pesquisador estudar as estratégias empregadas por tradutores profissionais, tanto num nível macrotextual de or-

## Apresentação

ganização do texto, de acréscimos ou omissões norteadas ou não pela função da tradução no contexto de chegada, quanto num nível microtextual como, por exemplo, a tradução de nomes próprios, de termos culturalmente marcados, de termos técnicos ou mesmo do vocabulário de determinado campo semântico.

Outro tipo de *corpus* essencial para a identificação de vocabulário técnico bilíngüe é o *corpus* comparável – composto de textos originais nas duas línguas em estudo –, pois fornece ao pesquisador os termos empregados em contexto natural naquelas línguas, sem a possível ocorrência do que se convencionou denominar, em inglês, de *translationese*, ou, em português, de tradutês.

Todas essas áreas de pesquisa estão representadas neste número.

O artigo que abre o volume, *Carving Language at its Joints – Os Corpora como Ferramentas de (Des)Trinchar a Língua*, de Inacio Abdulkader, serve de introdução à Lingüística de Corpus, ou melhor, relata a primeira incursão do autor no mundo dos corpora ao fazer uma pesquisa sobre os verbos ingleses *have* e *go*. O autor chama atenção para a relevância da sintaxe da expressão de busca, isto é, a formulação da expressão que se pretende investigar. Sua pesquisa é feita no *corpus* COMPARA, um *corpus* paralelo bidirecional inglês-português de textos literários, que permite uma sintaxe de busca bastante elaborada. Abdulkader salienta também que o potencial dos corpora não se restringe a pesquisas lexicais, podendo oferecer aspectos de interesse para a Filosofia da Ciência e a Filosofia da Linguagem.

Segue-se *Idéias que Cruzam o Oceano*, uma apresentação de projetos similares e até complementares que estão sendo desenvolvidos nos dois lados do Oceano Atlântico: na Universidade do Porto, em Portugal, por Belinda Maia e Luís Sarmiento, e na Universidade de São Paulo, por Sandra Maria Aluísio e Stella E. O. Tagnin. Enquanto na Europa o enfoque é na criação de ferramental para a construção de corpora e glossários, como o *Corpógrafo*, ou para a avaliação da tradução automática, como o *TrAva*, no Brasil o objetivo é a própria construção de corpora para diversos fins, desde pesquisas lexicográficas até o desenvolvimento de corretores gramaticais e ortográficos, a saber, o portal Lácio-Web, que disponibiliza, entre outros, um *corpus* de português brasileiro com 10 milhões de palavras, e o Projeto COMET, composto de três corpora, um Corpus Técnico, um Corpus de Traduções e um Corpus de Aprendizizes.

Os dois artigos seguintes, *Os Corpora como Ferramentas para a Solução de Problemas de Tradução de Colocações Verbais*, de Márcia Epstein Fiker e Stela Foley, e *Traduzir Colocações Adverbiais. Tarefa fácil? Parece, mas não é*, de Helmara Febeliana Real de Moraes, abordam, como já se depreende dos títulos, a tradução de unidades de duas categorias convencionais, a das colocações verbais e a das colocações adverbiais. Enquanto no primeiro as autoras estudam detalhada-

mente as colocações com os verbos *make* e *perform*, e o substantivo *attention*, no segundo a autora detém-se em critérios quantitativos e qualitativos para identificar as informações que devem integrar os verbetes de um glossário de colocações verbais. Para tanto, a título de exemplo, apresenta o verbete para a colocação *deeply hurt*.

A análise do vocabulário de determinados campos semânticos é o tópico dos próximos três artigos.

Kátia Hanna e Teresa Cristina S. Hilst fazem uma minuciosa análise contrastiva do verbo *limp* e suas possíveis traduções para o português, com base em dados levantados do *corpus* COMPARA, um *corpus* paralelo bidirecional inglês-português de textos literários.

Os *Verbos de Elocução* são estudados por Laís Cardoso Loffredo, Deborah Grossman, Gladys Bitar e Janice Gonçalves, também de uma perspectiva contrastiva, conforme ocorrem no *corpus* COMPARA. As autoras observam que nos textos originais em língua inglesa há muita repetição do verbo *said*, enquanto nos em língua portuguesa há um empenho em empregar um vocabulário mais variado. Esse padrão, no entanto, não ocorre na mesma proporção nos textos traduzidos, gerando, por vezes, o fenômeno a que aludimos acima, o tradutês.

Nessa mesma linha, o artigo de Josimeire Martins, *The Case of laugh: a parallel-corpus-based research*, explora os verbos do campo semântico de *laugh*, ainda sob uma perspectiva contrastiva baseada em dados do *corpus* COMPARA.

Um *corpus* comparável é utilizado por Alvamar Lamparelli para estudar um problema de cognatos em *Absolutamente e Absolutely – são ou não são absolutamente intercambiáveis?*. Examinando as ocorrências desses vocábulos em corpora de textos jornalísticos originais, em inglês e em português, a autora conclui que, embora alguns itens possam parecer inicialmente equivalentes, diferenciam-se pelos seus padrões de co-seleção.

Adriana Zavaglia, em *Linguística de Corpus e Lexicografia Bilingüe: o caso experimental de como e suas traduções para o francês*, amplia o escopo das pesquisas aqui relatadas. Em primeiro lugar, seu trabalho aborda a língua francesa, em segundo, entra no campo da lexicografia, sugerindo, inclusive, um verbete bilingüe experimental, que privilegia a definição e a contextualização do marcador *como*.

Da lexicografia passamos à terminologia com o trabalho de Guilherme Fromm, *A Construção do Sentido em Vocabulários Técnicos: o uso de corpora e outros procedimentos*. O autor, ao propor um modelo de elaboração de vocabu-

## Apresentação

lários técnicos, discute diversas abordagens de construção do sentido à luz da Semântica e da Lingüística de Corpus.

Ainda dentro dos estudos lexicais, Diana Costa Fortier Silva e Renato Barros da Costa oferecem-nos um estudo dos *Anglicismos no Português do Brasil: uma perspectiva de análise de usos baseada em corpora eletrônicos*. Os autores selecionam quinze anglicismos e examinam seus usos e suas acepções na língua inglesa e as eventuais modificações que sofrem ao passarem para a língua portuguesa.

Uma questão sintática é o tópico de estudo de Benivaldo José de Araújo Júnior, em *Estruturas Passivas no Espanhol Escrito: um estudo de caso utilizando a Lingüística de Corpus*. A partir da observação de que os livros didáticos e manuais de espanhol como língua estrangeira pouca ou nenhuma relevância dão às estruturas passivas – mencionando apenas a voz passiva –, o autor lança mão da Lingüística de Corpus para melhor estudar o fenômeno, partindo da hipótese de que a análise de uma grande quantidade de ocorrências pode revelar dados inesperados que permitirão compreendê-lo melhor.

O artigo seguinte, *A Investigação de Mudança de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*, é um exemplo claro de como a construção de um *corpus* é ditada pelos objetivos da pesquisa a que se destina. Para estudar a alternância de código (*code switching*) de duas crianças bilíngües aprendendo inglês e português simultaneamente, a autora, Catherine Anne Lönngren Sampaio, construiu um *corpus*, gravando vários tipos de interações dos dois sujeitos, ao longo de dois anos. Para transcrever o material, seguiu as normas do projeto CHILDES (Child Language Data Exchange System), mas, para atender às peculiaridades de sua pesquisa, inseriu outros códigos que lhe permitissem identificar os momentos em que ocorria a mudança de código. Trata-se de um dos raros trabalhos – senão o único – que se vale de um *corpus* eletrônico de língua falada para estudar os estágios iniciais do bilingüismo infantil.

Mas a metodologia da Lingüística de Corpus também se aplica aos estudos literários. É o que demonstra o trabalho de Lourdes Bernardes Gonçalves, *Lingüística de Corpus e Análise Literária – uma aplicação em Dubliners de Joyce*. A partir de um *corpus* composto pelos quinze contos irlandeses que compõem a coletânea *Dubliners*, James Joyce, a autora faz um levantamento eletrônico do vocabulário mais recorrente nos contos e compara-o ao vocabulário de outros contos da mesma época, porém escritos na Inglaterra, o que lhe permite identificar o léxico peculiar empregado pelo autor. A análise desse léxico – só identificável por meio dos procedimentos oferecidos pela Lingüística de Corpus – revela o papel da música na construção dos personagens joyceanos.

Acreditamos que a seleção dos artigos que compõem este número especial da *Crop* deixe clara a relevância e abrangência da Lingüística de Corpus nas

*Crop*, 10, 2004

várias áreas dos estudos lingüísticos, como a lexicologia, a terminologia, a tradução, o ensino e a aprendizagem de língua estrangeira e a sintaxe, entre outros, assim como nos estudos literários, em que sua aplicação ainda é incipiente, porém promissora.

Gostaria, ainda, de deixar registrado meu agradecimento a Luciana Carvalho Fonseca Corrêa Pinto, que coordenou a coleta e o acompanhamento dos artigos que integram este volume. Sem a ajuda dela, este trabalho não teria sido publicado no prazo estipulado.