

## INTRODUÇÃO

O presente número da revista *Diacrítica* é dedicado aos *Corpora nas Humanidades Digitais*. Na última década, temos assistido a grandes avanços no processamento de dados, motivados pelo estudo e pela análise da imensurável quantidade de informação que se encontra dispersa e cuja produção aumenta de dia para dia. Graças às novas tecnologias e ao desenvolvimento da Ciência da Computação, a Linguística de Corpus abriu caminho ao estudo dos fenómenos linguísticos de forma nunca antes concebida, assim como a outras utilizações desenvolvidas à medida de diversos profissionais (professores, tradutores, linguistas, lexicógrafos, terminólogos, informáticos, historiadores, etc.). No âmbito deste volume, pretende-se dar uma visão das diferentes aplicações da Linguística de Corpus, com especial foco na investigação da linguagem em contexto académico.

Em “Identidade e diferenças na terminologia da *fauna* e da *flora*: notas sobre um estudo comparativo entre as línguas portuguesa, inglesa, italiana e espanhola”, Sabrina de Cássia Martins examina o fenómeno da variação denominativa na terminologia da *fauna* e da *flora* nas línguas portuguesa, inglesa, italiana e espanhola, analisando as variantes denominativas de cerca de duzentas espécies.

Em “*Agroquímico, biocida, pesticida, plaguicida e producto fitosanitário*: uma pesquisa com corpus”, Mauren Thiemy Ito Cereser e Cleci Regina Bevilacqua estudam a equivalência do termo *agrotóxico* em espanhol, num *corpus* de textos legais de países hispanofalantes, a partir dos termos referidos no título do artigo, conforme empregados no cenário de leis ambientais do Brasil e dos países hispânicos.

Em “Quando o léxico dá bandeira – aspectos cognitivo-discursivos da mudança semântica na construção de brasileirismos em registros lexicográficos luso-brasileiros”, Anderson Salvaterra Magalhães e Janderson Lemos de Souza analisam duas unidades simbólicas em que constam tensões conceituais no português do Brasil e no português europeu, *bandeira* e *bandeirante*, em dois trabalhos lexicográficos luso-brasileiros dos séculos XVIII e XIX, que são cotejados com trabalho lexicográfico brasileiro e português do século XXI para fins de identificação, descrição e análise da mudança semântica.

Em “Características identificadoras e dificuldades na aplicação de listas para a anotação de entidades geográficas mencionadas”, Afonso Xavier Canosa descreve as dificuldades na utilização de listas de entidades geográficas (índice de topónimos, ou *gazetteers*) no processo de anotação automática da *Peregrinação* de Mendes Pinto, dado que a simples aplicação dos topónimos da lista pode produzir ambiguidades (*Carvalho*, por exemplo, pode ser um topónimo, um antropónimo ou um nome comum).

Em “Uma versão em português europeu do C-test European-Portuguese version of the C-test”, Masayuki Yamada examina a fiabilidade e a validade de uma versão em português europeu do C-test, um teste de preenchimento simples utilizado para medir a proficiência geral. O teste foi desenvolvido seguindo os procedimentos propostos por Raatz & Klein-Braley (2002). Nele participaram 104 alunos que frequentam cursos de português para estrangeiros em universidades portuguesas.

Em “Aplicação de ferramentas para coleta e análise de dados em linguística”, Roberlei Alves Bertucci mostra que o desenvolvimento crescente de ferramentas informáticas específicas facilita a criação e a análise de *corpora* electrónicos. Neste artigo, são apresentadas três ferramentas (*Netvizz*, *Linguakit* e o *Tropes*) que podem contribuir para o fortalecimento da investigação linguística baseada no uso de *corpora*, e mais especificamente de dados provenientes de redes sociais.

Em “Análise diacrónica dos tempos compostos *tinha feito*, *terei feito* e *teria feito* na língua portuguesa”, Jan Hricsina propõe uma análise diacrónica dos tempos compostos referidos no título do artigo. Efetuada no *corpus* linguístico [www.corpusdoportugues.org](http://www.corpusdoportugues.org), essa análise compara a frequência e o emprego desses tempos na evolução da língua portuguesa.

Em “Análise contrastiva das formas de tratamento ao interlocutor no teatro brasileiro e português dos séculos XIX e XX”, Ana Carolina Morito Machado traça um interessante panorama do sistema de tratamento do interlocutor no PB e no PE dos séculos XIX e XX. As estratégias de referên-

cia ao interlocutor são analisadas em duas amostras de textos dramáticos, à luz do modelo das Tradições Discursivas (TDs), da Teoria da Variação (Weinreich, Labov e Herzog 1968) e da Teoria do Poder e Solidariedade (Brown e Gilman 1960).

No artigo “The role of pragmatic markers in academic spoken interlanguage: a corpus-based study of a group of Brazilian EFL university students”, Bárbara Malveira Orfanò, Ana Larissa Adorno Marciotto Oliveira e Spencer Barbosa da Silva descrevem um estudo que incide sobre o uso de marcadores pragmáticos na produção de discurso oral por parte de estudantes universitários brasileiros inscritos na unidade curricular de Inglês para Fins Acadêmicos. O trabalho baseia-se em dados extraídos do corpus *Brazilian Academic Spoken English* (BRASE) e do subcorpus do *British Academic Spoken English* (BASE).

No texto “Corpus Stylistics in translation-oriented text analysis: Approaching the work of Denton Welch from a Functionalist Perspective”, enquadrando-se no modelo de tradução funcionalista de Christiane Nord, Guilherme da Silva Braga explora a aplicabilidade de uma abordagem baseada em *corpus* numa fase pré-tradutória do texto literário, apontando para as vantagens do tradutor complementar os resultados da análise quantitativa com uma análise qualitativa do texto a traduzir.

O artigo “Em vida e na hora da morte também: o que dizem registros de óbito oitocentistas da Freguesia de Nossa Senhora da Penha de Corumbá (1847–1855)”, da autoria de Maria Helena de Paula e Amanda Moreira de Amorim, aborda importantes aspetos da história e da cultura do período escravocrata do Brasil oitocentista com base nos dados de um *corpus* de registros de óbito da Freguesia de Corumbá de Goiás.

Retomando o tema deste número da *Diacrítica*, é legítimo afirmar que as Humanidades se deparam atualmente com inúmeros desafios tecnológicos que exigem de toda a comunidade científica a necessária abertura a novos paradigmas do conhecimento e da investigação. A construção e a análise de *corpora* nas mais variadas áreas de estudo impulsiona a procura de novos modelos interpretativos que possam ajudar à compreensão de fenómenos culturais, linguísticos, literários e sociais.

*A equipa editorial*