

**Clique aqui para escolher seu português: uma análise de corpus de sites
disponíveis em duas versões da língua portuguesa**

Caio Cesar Christiano (FORELL, Université de Poitiers/PG)

Desde o século XIX, os debates sobre as diferenças entre as variedades brasileira e europeia do português têm sido acirrados. Muito já se afirmou e teorizou a respeito, buscando-se sempre exemplos que mostrassem o quanto as variedades são semelhantes ou diferentes. No dealbar do século XXI, as discussões não se mostram menos intensas e eventos como a implementação do novo código ortográfico demonstram o quanto as opiniões divergem nos diversos países lusófonos.

Foi só em anos recentes, com a utilização massiva da informática na análise de corpora que antigamente representariam imenso desafio ao serem tratados manualmente, que respostas mais satisfatórias sobre o real grau destas diferenças têm surgido. Usando das novas tecnologias, TAGNIN & TEIXEIRA (2004), por exemplo, confirmaram que as diferenças entre o português europeu e o brasileiro são maiores que as existentes entre o inglês americano e britânico nos níveis sintático e lexical no gênero das receitas culinárias.

As empresas de tecnologia da informação, setor da indústria cultural com maior crescimento nas últimas décadas, sem esperar pelas decisões de autoridades linguísticas, resolveram tomar partido na discussão e dar seu veredito: há duas línguas - o Português Brasileiro e o Português Europeu. Em domínios populares como Google, Facebook e MSN o usuário pode escolher em qual dos dois « idiomas » prefere acessar seus conteúdos.

Mas quais seriam as verdadeiras diferenças entre as duas versões do português de acordo com estes sites? Limitar-se-iam ao léxico? As mudanças sintáticas estariam restritas a aspectos deveras conhecidos como a colocação pronominal e a utilização da segunda pessoa?

Esta comunicação procura responder a essas questões.

Primeiramente, recolheu-se um corpus comparável formado por versões nas duas variedades do português APENAS em sites que oferecem conteúdos semelhantes nas duas versões a seus usuários. A pequena extensão do corpus reunido (40.000 palavras) é explicada pela escassez de páginas na internet apresentadas nas duas variedades.

Em seguida, o corpus etiquetado morfo-sintaticamente foi tratado com auxílio do programa WordSmith Tools v5. Usou-se o concordanciador do programa para a análise

das estruturas sintáticas. A ferramenta Wordlist foi utilizada para a comparação do léxico utilizado nas duas variedades. As estatísticas e resultados obtidos estão sendo analisados.

Os resultados preliminares mostram que, apesar de diferenças serem encontradas em todos os níveis, elas são mais abundantes no campo do léxico. As conclusões preliminares apontam também para o fato de as respostas poderem variar de acordo com o site pesquisado: enquanto em algumas páginas web há uma clara redação em uma das variedades com posterior adaptação à outra, outros sites demonstram soluções radicalmente diferentes para a expressão de uma mesma idéia. Concluímos que o tratamento das variedades como duas línguas distintas é totalmente justificado do ponto de vista comercial, já que a opção por uma ou outra norma nacional poderia tornar um produto inacessível a seu público-alvo.

Trata-se de um estudo pioneiro, pois as questões aqui levantadas ainda não receberam, sob a ótica da lingüística de corpus, a merecida atenção. Nosso esforço visa também se somar a outros estudos contrastivos das duas variedades que, juntos, poderão, um dia, trazer mais esclarecimentos sobre a unidade ou não na lusofonia.

Referências bibliográficas

NOLL, Volker. Tradução de Mário Eduardo Viaro. O Português Brasileiro: formação e contrastes. São Paulo: Globo, 2008. 397 p.

PINTO, Edite Pimentel. O Português do Brasil: textos críticos e teóricos. São Paulo: Editora da Universidade de São Paulo, 1978. 510 p.

SARDINHA, Tony Berber. Lingüística de Corpus. São Paulo : Manole, 2004. 410 p.

SCOTT, Mike. Wordsmith Tools. Oxford University Press, 2008. Disponível em: <<http://www.liv.ac.uk/~ms2928>>

TAGNIN, S. E. O.; TEIXEIRA E. D. British vs. American English, Brazilian vs. European Portuguese - how close or how far apart? A corpus-driven study. In: LEWANDOWSKA-TOMASZCZYK, B. (Org.). Lodz Studies in Language vol. 9: Practical Applications in Language and Computers - PALC 2003. Frankfurt am Main: Peter Lang, 2004. p.193-208..

WITTMANN, L.H.; PÊGO, T.R.; SANTOS, D. 1995. Português Brasileiro e Português de Portugal: algumas observações. In: ENCONTRO NACIONAL DA APL, XI., 1995. Lisboa.