



SALÃO DE INICIAÇÃO CIENTÍFICA XXVIII SIC

paz no plural



Evento	Salão UFRGS 2016: SIC - XXVIII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2016
Local	Campus do Vale - UFRGS
Título	CONTRIBUIÇÃO PARA UM DICIONÁRIO DE HOMÓGRAFOS PORTUGUÊS-ESPANHOL RIO-PLATENSE BASEADO EM CORPUS: BUSCANDO REGIONALISMOS DO BRASIL
Autor	BIANCA RABAIOLI
Orientador	MARIA JOSE BOCORNY FINATTO

CONTRIBUIÇÃO PARA UM DICIONÁRIO DE HOMÓGRAFOS PORTUGUÊS-ESPAANHOL RIO-PLATENSE BASEADO EM *CORPUS*: BUSCANDO REGIONALISMOS DO BRASIL

Apresentadora: Bianca Rabaioli (PIBIC-UFRGS),

Orientadora: Maria José Bocorny Finatto (UFRGS, CNPq).

Este estudo associa-se a uma investigação da área de Processamento de Linguagem Natural (PLN), que pretende obter um dicionário computacional de formas homógrafas extraídas de *corpora* de jornais populares e regionais, tendo-se em mente o espanhol rio-platense e o português do Brasil. Uma vez alinhadas as formas de grafia idêntica nas duas línguas, buscam-se descrever, formalmente, os padrões e valores de uso desse vocábulos. Exemplos de palavras homógrafas, vemos grifadas no segmento a seguir, extraído de um jornal de Córdoba (Argentina): *El gobierno **nacional** oficializó ayer la decisión **de** introducir nuevas modificaciones en las subas dispuestas en las **tarifas** de servicios **públicos**, al disponer **que** el **tope** de aumento para el gas a nivel **residencial** será del 400% en **todo** el país y del 500% para los comercios, pymes y establecimientos hoteleros*. Nessa tarefa, destaca-se, como um desafio especial, identificar possíveis regionalismos de cada uma das variedades de língua conforme empregadas na Argentina, Uruguai e Brasil, o que se poderia associar, por exemplo, ao substantivo português **tope**, que no RS é usado como o valor de *laço de fita*, mas também pode corresponder à flexão do verbo **topar**, no sentido de **tropeçar em algo** ou de **aceitar-se algo**. Na parte do Brasil, foram tomados como *corpora* dois jornais destinados a leitores de menor escolaridade e pouco hábito de leitura: o *Diário Gaúcho*, jornal porto-alegrense, e o *Massa!*, um periódico baiano. A partir desses *corpora*, gerou-se a lista de contextos para as unidades homógrafas espanhol-português previamente identificadas pelos colegas da Computação/PLN. A lista de contextos de cada jornal brasileiro foi feita com o apoio da ferramenta AntConc. Feito isso, analisaram-se os contextos de cada uma com apoio de obras lexicográficas e de ferramentas da internet, buscando-se identificar regionalismos, o que, embora pareça, não é uma tarefa trivial por diferentes razões. Os regionalismos identificados serão marcados como palavras específicas e isso deve facilitar uma série de procedimentos computacionais associados ao futuro dicionário, como, por exemplo, a tradução automática de notícias.

Palavras-chave: análise contrastiva, dicionários, jornal popular, regionalismos, português, espanhol.