

Este estudo associa-se a uma investigação da área de Processamento de Linguagem Natural (PLN) - o Projeto RITA - para um dicionário computacional de formas homógrafas extraídas de *corpora* de jornais populares e regionais, considerando-se o espanhol rio-platense e o português do Brasil.

1. OBJETIVO

Examinar uma lista de palavras desconhecidas por um dicionário computacional pré-existente (dicionário DELA, sistema UNITEX) e identificar possíveis regionalismos em textos do português popular escrito de dois jornais: o Diário Gaúcho, jornal porto-alegrense, e o Massa!, um periódico baiano. Além da observação de uma lista genérica de palavras desconhecidas pelo DELA, analisam-se os contextos de cada uma com apoio de obras lexicográficas e de ferramentas da internet.

3. MATERIAIS E MÉTODOS

Todo o *corpus* DG e Massa - na forma de listas de unidades de palavras - foi comparado com o acervo de itens dicionarizados do dicionário computacional DELA. Os itens desconhecidos foram listados por ordem de frequência nos jornais (equipe do projeto RITA).

DG: 11.616 palavras desconhecidas.
Massa: 2.162 palavras desconhecidas.

Nosso trabalho: cada item "desconhecido" pelo DELA foi provisoriamente categorizado (err. = erro de ortografia/ err#. = erro especial/ estr. = estrangeirismo/ gir. = gíria, entre outros, vide item 4.) com base no exame de:

- seus contextos nos jornais-fonte;
- registro em dicionários (gerais e específicos);
- usos na internet;
- estudos de Lexicologia e Lexicografia.

Para verificar prováveis regionalismos: estudos linguísticos + dicionários específicos/ regionais da Bahia e do RS + dicionários gerais + observação de usos.

2. BASE TEÓRICA

Referenciais da Linguística de Corpus de BERBER SARDINHA (2000). Envolve *corpora* textuais coletados criteriosamente para fins de pesquisa ou descrição de uma língua.

Trabalhos de Lexicologia e de Lexicografia sobre regionalismos.

4. CATEGORIAS DE PALAVRAS DESCONHECIDAS

- a) Classes de palavras: em azul
- b) Marcas de uso: em vermelho
- c) Operacionais: em verde

abv. = abreviação
adj. = adjetivo
adv. = advérbio
art. = artigo
comp. = palavra composta
conj. = conjunção
cont. = contração
err. = erro de ortografia
err#. = erro especial
estr. = estrangeirismo
gir. = gíria
interj. = interjeição
neol. = neologismo
n.p. = nome próprio
num. = numeral
onom. = onomatopeia
out. = outro
p.p. = participio passado
pref. = prefixo
prep. = preposição
pron. = pronome
reg. = regionalismo
sgl. = sigla
v. = verbo



SALÃO DE INICIAÇÃO CIENTÍFICA
XXVIII SIC



5. REGIONALISMOS: RESULTADOS

Conceito operacional de regionalismo: palavra usada em dado local (RS ou BA) com um sentido específico válido para as pessoas daquela região (verificado no *corpus*).

a) DG: 11616 palavras.

Regionalismos = 57 palavras (0,49%).

Exemplos:

- Para comemorar a data, hoje tem **bailanta** com o conjunto Os Monarcas.
- Antes de entrar na padaria neste final de semana, pergunte o preço do **cacetinho**.
- A **piazada**, nervosa, pedia o significado e a aplicação das palavras, umas bem difíceis, e ele respondia todo irritadinho.

b) Massa: 2162 palavras.

Regionalismos = 15 palavras (0,69%).

Exemplos:

- Após três tentativas fracassadas de Marcos para o motorista abrir a porta do **buzu**, o ladrão pulou a catraca e deu o primeiro golpe no cobrador.
- Motôs** e cobradores elogiam a estrutura e as reformulações nos ônibus, pois trouxeram mais segurança e melhoraram a comunicação entre os dois.
- Aí, rapaz, o pessoal ficou **retado**, querendo o dinheiro de volta, uns devolvendo fantasias.

REFERÊNCIAS

- BOSSLE, J. B. A. Dicionário gaúcho brasileiro. Porto Alegre: Artes e Ofícios. 2003.
Dicionário de Baianês *online* (<http://www.folderpark.net/baianes/>)
Dicionário Caldas Aulete *online* (<http://www.aulete.com.br/index.php>)
FISCHER, L.A. Dicionário de porto-alegrês: ampliado. Porto Alegre: Artes e Ofícios, 10ed., 2004.
NUNES, Z.C.; NUNES, R.C. Dicionários de regionalismos do Rio Grande do Sul. Porto Alegre: Martins Livreiro, 3.ed., 2009.
SANTOS, O. J. A música dos pampas em uma perspectiva lexical: milongando entre o espanhol e o português. Dissertação (Mestrado) – Universidade de Caxias do Sul, Programa de Pós-Graduação do Mestrado em Letras, Cultura e Regionalidade, 2014.
SARDINHA, T. B. Linguística de Corpus: Histórico e Problemática. DELTA, vol.16, no.2, São Paulo, 2000.
VALADARES, F.B.; SANTOS, M.F. Baianês: regionalismo na perspectiva sociolinguística e cultural. Web-Revista Sociodialeto, UEMS/Campo Grande, v.6, nº 16, jul. 2015.