

ESPECIFICIDADES DO VOCABULÁRIO DA OBRA *CONTOS GAUCHESCOS*, DE SIMÕES LOPES NETO: UM ESTUDO EXPLORATÓRIO

Nathalia Marcon e Maria José Finatto

Universidade Federal do Rio Grande do Sul (UFRGS)

ntolivam@gmail.com, mariafinatto@gmail.com

INTRODUÇÃO: O objetivo deste trabalho é apresentar uma análise inicial, com apoio de recursos da Linguística de Corpus, do vocabulário empregado na obra literária *Contos Gauchescos* (1912), de Simões Lopes Neto (1865-1916), importante autor da literatura sul-riograndense, visto que o acesso desse livro tende a ser considerado como bastante difícil para leitores atuais (conforme Fischer, 2000, p. 7-8). Ao descrever o vocabulário da obra, apontam-se possíveis especificidades do gênero literário e do autor frente a autores seus contemporâneos (em um *corpus* de contraste) e frente a dois *corpora* gerais, não literários, um do mesmo período da produção da obra e um *corpus* atual do português brasileiro (*corpora* de referência). **MATERIAIS E MÉTODOS:** A observação do vocabulário empregou a ferramenta AntConc 3.4.1m. mediante os recursos lista de palavras, relação *types/tokens* e lista de palavras-chave. O livro *Contos Gauchescos* apresenta 32.879 *tokens* e 6.428 *types*; e teve seu vocabulário contrastado com aqueles das seguintes obras: *Triste Fim de Policarpo Quaresma* (1915): 66.683 *tokens* e 10.713 *types*, *Os Sertões* (1902): 154.735 *tokens* e 22.134 *types* e *Iracema* (1865): 29.898 *tokens* e 5.563 *types*; como *corpora* de referência utilizamos: (a) cartas de diferentes remetentes escritas entre os anos de 1800 e 18995, disponíveis no Corpus Histórico do Português Tycho Brahe; (b) listas de palavras (da escrita e da fala) do

Banco de Português. Inicialmente, foram feitas listas de palavras dos *corpora* de estudo e de contraste. Em seguida, foi criada uma *keyword list*, contrastando esses dois *corpora*. Dessa lista, as primeiras 1750 palavras (31%) foram comparadas com os *corpora* de contraste e de referência, identificando-se os itens de frequência zero em ambos os *corpora*. **RESULTADOS PARCIAIS:** Os resultados da tabela 1 sugerem que o texto dos *Contos Gauchescos* não possui uma grande riqueza lexical, mas é composto por vocábulos específicos: os 5.647 vocábulos da *keyword list*, que representam as palavras cuja frequência é estatisticamente diferente entre os *corpora* de estudo e de contraste, correspondem a 87,8% dos *types* do *corpus* de estudo. Essa idéia é reforçada pelos resultados da tabela 2, que destaca a quantidade de vocábulos de frequência zero na comparação entre os *corpora*. A análise realizada não pretendeu esgotar a investigação sobre a especificidade lexical encontrada, merecendo o presente trabalho um maior aprofundamento. Identificamos, por exemplo, palavras como ‘china’, que, embora conste na lista de palavras do *corpus* de referência, é plausível supor-se que com significado diverso daquele empregado por S. Lopes Neto, a saber: mulher, especialmente a mestiça ou a índia. Assim, a possibilidade de lematização e análise semântica dos vocábulos deve ser levada em consideração em uma análise futura.