

LÉXICO HISTÓRICO DO PORTUGUÊS BRASILEIRO: estágio atual

Vanderci de Andrade Aguilera/ Universidade Estadual de Londrina
Fabiane Cristina Altino/ Universidade Estadual de Londrina

O Léxico Histórico do Português Brasileiro – LHisPB está sendo construído a partir do material manuscrito coletado e transscrito (edição semidiplomática) pelas equipes regionais que hoje integram o Projeto Histórico do Português Brasileiro, projeto de pesquisa interinstitucional coordenado pelo professor Dr. Ataliba Teixeira de Castilho.

O objetivo do Léxico Histórico do Português Brasileiro é apresentar, sob tratamento lexicográfico, todo o conteúdo lexical dos documentos manuscritos, datados dos séculos XVII a XIX, que foram coletados, transcritos e editados pelas equipes regionais dos estados do Rio de Janeiro, da Bahia, de Minas Gerais, da Paraíba e do Paraná. Como objetivos específicos, propomos: (i) oferecer, sob a forma de banco de dados informatizado, um vasto material lexicográfico para estudos sincrônicos e diacrônicos do léxico registrado em documentos manuscritos durante os três séculos mencionados; (ii) proporcionar um instrumento de consulta rápida do uso de vocábulos vigentes na época e em várias regiões do Brasil Colônia e Brasil Império; (iii) apresentar, com a frequência de uso, os itens lexicais e suas variantes gráficas registrados em cerca de 2500 documentos emanados das então províncias e vilas sediadas nos atuais estados da Bahia, Rio de Janeiro, São Paulo, Paraná, Minas Gerais, Paraíba, Ceará, Pernambuco e Rio Grande do Sul.

Diante da diversidade de gêneros coletados por essas equipes, do grande volume de material disponível, e tentando compor um *corpus* geral menos heterogêneo, selecionamos documentos manuscritos oficiais (correspondências eclesiásticas, notariais, cartas ao governo, por exemplo), e de caráter coloquial (familiar, pessoal, petições, entre outros).

Para a constituição do *corpus*, os textos estão sendo tratados e submetidos à ferramenta denominada Lexico3, software desenvolvido na Université de la Sorbonne Nouvelle – Paris 3, pela equipe CLA2T. Trata-se de um programa de aplicação lexicométrica que, por sua versatilidade, permite: i) balizar livremente o texto a ser analisado, determinando e distinguindo as partes do texto previamente digitalizado; ii) determinar o tamanho do contexto de um segmento a ser pesquisado; iii) fazer o levantamento das ocorrências do segmento, indicando a distribuição das palavras dentro do texto; iv) exibir as concordâncias; v) apresentar, por meio de gráficos, as frequências relativa e absoluta de uma palavra.

O LHisPB será constituído nos moldes da obra do *Léxico Histórico del Español de México*, de Company y Melis (2002). Pretendemos, pois, oferecer um léxico lematizado que corresponde a um dicionário de construção e uso em que as definições semânticas dos lemas não serão incluídas. No LHisPB estarão contidos registros de alguns dos usos da língua portuguesa com exemplificação dos diversos contextos gramaticais e semânticos extraídos dos manuscritos, datados dos séculos XVII, XVIII e XIX, editados pelas equipes estaduais do PHPB.

Para a composição do LHisPB, reconhecemos que não cogitamos, nem poderíamos, abranger tudo o que foi escrito entre os séculos XVII e XIX. Company e Melis (2002), para o *Léxico Histórico del Español de México*, trabalharam com 320 documentos do período colonial mexicano, procedentes de duas fontes documentais: o Archivo General de Indias en Sevilla e o Archivo General de la Nación en la Ciudad de México. É certo que Company e Melis afirmam terem dado preferência, *quase absoluta*, a documentos informais, tais como cartas pessoais, denúncias e testemunhos em juízo, petições e informes de particulares, com o fim de buscar uma aproximação maior da língua falada do período colonial hispânico. Ao contrário do *Léxico histórico del español de México*, no LHisPB, não haverá distinção entre texto de natureza oficial e texto de caráter coloquial (familiar, pessoal). Esta escolha se justifica pelo fato de, observando o *corpus* constituído de um dos projetos regionais que fará parte do LHisPH, o *Para a história do português paranaense: estudos diacrônicos em manuscritos dos séculos XVII a XIX*, a leitura desses documentos tem indicado que, apesar de oficiais (documentos notariais), a ortografia e a gramática ali aplicadas estão longe de ser consideradas a padrão ou

estar rigorosamente dentro da gramática normativa, como têm discutido Almeida-Baronas (2006, 2009) e Toniolo (2008). Pelo contrário, muitos deles, estão muito próximos de uma modalidade de língua provavelmente falada por boa parte dos brasileiros no tempo da Colônia e do Império.

Esta análise de correspondências oficiais dos séculos XVII, XVIII e XIX, emanadas das antigas vilas paranaenses tem indicado que, na maioria dos casos, os escribas não parecem ter tanta habilidade gramatical e gráfica como se espera de documentos oficiais. Podemos exemplificar com inúmeros casos de metáteses e hipérteses; oscilação ortográfica no mesmo documento e pelas mesmas mãos; liberdade gramatical na concordância verbal e nominal, e na colocação pronominal.

Por outro lado, diante da diversidade de gêneros coletados pelas seis equipes (São Paulo, Bahia, Rio de Janeiro, Minas Gerais, Paraná e Paraíba), do grande volume de material já disponível e, tentando compor um *corpus* geral menos heterogêneo, optamos por selecionar apenas os documentos manuscritos oficiais (correspondências eclesiásticas, notariais, cartas ao governo, por exemplo), além dos de caráter coloquial (familiar, pessoal, petições, entre outros). É importante observar, neste particular, que nem todas as regionais editaram documentos dos mesmos gêneros de texto. Por exemplo: o Paraná trabalha atualmente com cerca de 1 500 documentos oficiais e apenas uma dezena de cartas pessoais.

Para efeito didático, optamos por tratar os dados por região. Assim, inicialmente, coletamos e procedemos à limpeza do *corpus* apresentado pela equipe do Rio de Janeiro uma vez que já está disponível para acesso no site do PHPB (<http://www.letras.ufrj.br/phpb-rj/>). Em seguida demos-lhe o tratamento lexicográfico a partir da ferramenta Lexico3, software desenvolvido na Université de la Sorbonne Nouvelle – Paris 3, pela equipe CLA2T.

Após a tabulação das entradas de acordo com a frequência apontada pelo Léxico 3, procedemos à construção do verbete, seguindo o modelo adotado por Company y Melis (2002, p. vii), com a seguinte configuração:

As entradas, como no dicionário, são os lexemas que permitem reconhecer todo o paradigma assumido por elas. Para cada entrada aparece em primeiro lugar a seleção de contextos que melhor exemplificam o emprego sintático e o sentido da voz em questão em todas as formas gramaticais que se documentam no *corpus* base. Em seguida, e em fonte menor, enumeram-se as variantes gráficas que recobrem as formas gramaticais do paradigma, assinalando entre parênteses a frequência de uso de cada variante gráfica. A última linha das entradas corresponde ao registro da frequência global da voz no *corpus* base. (COMPANY y MELIS, 2002, p. vii).

Nesta comunicação apresentamos o produto final do Léxico Histórico do Português Brasileiro – relativo ao *corpus* do Rio de Janeiro.

Referências

- ALMEIDA-BARONAS, Joyce Elaine. *Manuscritos paranaenses: fontes para um estudo diacrônico*. In: Anais do V SELISIGNO e VI Simpósio de Leitura da Universidade Estadual de Londrina, Londrina – PR, 2006.
- ALMEIDA-BARONAS, Joyce Elaine; ALMEIDA, Polyana Lucena Camargo de. *A escrita dos séculos XVIII e XIX: em busca de dados diacrônicos*. In: XIX Seminário do CELLIP – Centro de Estudos Linguísticos e Literários do Paraná. UNIOESTE – Universidade Estadual do Norte do Paraná, Cascavel – PR, 2009.
- COMPANY, Concepción; MELIS, Chantal. *Léxico histórico del español de México*. México: Universidad Autónoma de México, 2002.
www.letras.ufrj.br/phpb-rj
- TONILO, Énio José. *Possíveis marcas da oralidade em manuscritos paranaenses*. Filologia e Lingüística Portuguesa, v. 10-11, p. 305-315, 2009.