

LÉXICO HISTÓRICO DO PORTUGUÊS BRASILEIRO

Vandererci de Andrade AGUILERA ¹

(Universidade Estadual de Londrina)

RESUMO: O Projeto Para a História do Português Brasileiro (PHPB), em comemoração aos seus quinze anos de atividade, pretende publicar em 2012 uma série de livros com os estudos já realizados sobre o português do Brasil. Dentre os volumes projetados, o quarto será dedicado ao Léxico Histórico do Português Brasileiro, composto a partir dos dados dos *corpora* coletados e organizados pelas diferentes equipes regionais do Projeto. Neste artigo, fazemos um breve retrospecto sobre obras lexicográficas semelhantes, publicadas no Brasil, sobre alguns projetos dessa natureza em andamento, tratamos dos objetivos, da metodologia adotada e dos primeiros passos para a consecução de tão ousada empresa.

PALAVRAS-CHAVE: Léxico Histórico; Português Brasileiro; Primeiros Passos.

GLOSSARY HISTORY OF BRAZILIAN PORTUGUESE

ABSTRACT: The project for the History of Brazilian Portuguese (PHPB), in celebration of its fifteen years of activity, intends to publish in 2012 a series of books with the previous studies on the Portuguese of Brazil. Among the projected volumes, the fourth one will be dedicated to the history of Brazilian Portuguese lexicon, made from data corpora collected and organized by the different regional teams of the Project. In this article, we present a brief review on similar lexicographical works, published in Brazil, on some ongoing projects of this nature; we also treat the objectives, the adopted methodology and the first steps towards achieving such a bold undertaking.

KEYWORDS: Historical Lexicon; Brazilian Portuguese; First Steps.

1 INTRODUÇÃO

O Projeto Para a História do Português Brasileiro (PHPB), em comemoração aos seus quinze anos de atividade, pretende publicar em 2012 uma série de livros com os estudos já realizados sobre o português do Brasil. Dentre os volumes projetados, o quarto será dedicado ao Léxico Histórico do Português Brasileiro, composto a partir dos dados dos *corpora* diacrônicos coletados e organizados pelas diferentes equipes regionais² do Projeto.

Acreditamos que o *Léxico Histórico do Português Brasileiro* trará uma grande contribuição aos estudiosos da língua portuguesa, principalmente aos que se dedicam aos

¹ Pós-Doutora pela Universidad de Alcalá de Henares (Espanha). Docente da Universidade Estadual de Londrina. E-mail: vandererci.aguilera@terra.com.br.

² Compõem o Projeto Para a História do Português Brasileiro pesquisadores da Universidade de São Paulo (USP), Universidade Federal da Bahia (UFBA), Universidade Federal do Rio de Janeiro (UFRJ), Universidade Federal de Minas Gerais (UFMG), Universidade Federal da Paraíba (UFPB), Universidade Estadual de Feira de Santana (UEFS), Universidade Estadual de Londrina (UEL) e Universidade Estadual de Campinas (UNICAMP).

estudos diacrônicos, dado que vimos observando, nas últimas décadas, um interesse muito grande dos pesquisadores brasileiros, não só pela consulta a dicionários, glossários e léxicos históricos do português, como também pela elaboração de obras dessa natureza. Uma das grandes referências é o *Dicionário histórico das palavras portuguesas de origem tupi*, de Cunha, publicado em 1978.

Outro grande projeto, mais recente e também orientado por Cunha, é o *Vocabulário Histórico-Cronológico do Português Medieval*, idealizado para ser o primeiro levantamento exaustivo do léxico da língua portuguesa nos séculos XIII, XIV e XV. Iniciado em janeiro de 1979, dispunha, no final de 1983, *de um acervo de mais de 120 mil fichas com passagens abonatórias de vocábulos recolhidos em cerca de uma centena de textos medievais, nos quais foi feito, sempre que possível, o levantamento integral do léxico*. Veio à luz em 2006, sete anos após o falecimento de Cunha, em versão acompanhada de CD ROM.

Em Londrina-Paraná, Silva (2008) publicou o *Dicionário da língua portuguesa medieval*, projeto gestado durante mais de uma década na escolha paciente e criteriosa de 17.000 verbetes coletados junto a dezenas de fontes medievais já transcritas, como as *Cantigas de Santa Maria*, as *Décadas*, *O Cancioneiro da Ajuda*, *As Cantigas D'El rei Dom Dinis*, entre outras.

Dentre os projetos em andamento, o mais amplo e ambicioso diz respeito ao *Dicionário Histórico do Português do Brasil* (séculos XVI, XVII e XVIII), que inclui 10 universidades, 17 doutores e 17 alunos de Graduação e de Pós-graduação, inicialmente coordenado pela inesquecível professora Dr^a Maria Tereza Camargo Biderman e, a partir de agosto de 2008, pela Dr^a Clotilde Murakawa. A proposta inicial seria reunir mais de 10 mil termos do vocabulário falado no país entre os séculos XVI e XVIII a partir de manuscritos, de autoria de brasileiros ou de portugueses radicados no país, que retratam a linguagem usada no início da formação do Brasil. Biderman explica que o novo dicionário deverá ser um pouco diferente do padrão tradicional, pois “a partir da grafia contemporânea de cada palavra, serão descritos todos os significados e usos que a palavra teve durante o período analisado”.

A palavra ‘inhame’ é um dos exemplos que ilustram bem o espírito da obra. “Nos manuscritos do período analisado, encontramos diferentes grafias para essa mesma palavra. Isso porque os portugueses não entendiam a pronúncia dos indígenas e acabavam escrevendo de formas distintas”, conta Maria Tereza, que é autora de outras obras de referência: *Dicionário Contemporâneo de Português* (Editora Vozes, 1982), *Dicionário Didático de*

Português (Editora Ática, 1998), *Dicionário Ilustrado de Português* (Editora Ática, 2005) e *Dicionário do Estudante* (Editora Globo, 2005).

Dentre os projetos de menor abrangência, participamos, juntamente com pesquisadores integrantes do Projeto Para a História do Português Paranaense (PHPP), da Universidade Estadual de Londrina, da organização de duas obras de orientação filológica: *Scripturae nas villas de São Luiz de Goaratuba e Antonina: manuscritos setecentistas e oitocentistas* (AGUILERA; ALMEIDA-BARONAS³, 2007) e *Scripturae na villa de Pernagoa: manuscritos setecentistas* (AGUILERA; VASCONCELOS⁴, 2007). Ambas trazem, na primeira parte, o fac-símile dos manuscritos, a lição (ou edição) correspondente a cada documento e, ao final, um glossário das palavras que apresentam diferenças ortográficas e/ou semântico-lexicais em relação ao português atual. O *corpus* constitui-se, pois, de manuscritos depositados no Arquivo Público de São Paulo, referentes a documentos oficiais emanados das primeiras vilas da 5ª Comarca da Província de São Paulo, atual estado do Paraná. Dentre os 734 fólios, selecionamos, para estudo, os pertinentes aos séculos XVII, XVIII e meados de XIX, mais especificamente até o ano da emancipação da 5ª Comarca, que se deu no dia 19 de dezembro de 1853.

O primeiro livro traz 20 fólios de Guaratuba e 30 de Antonina, escritos entre os anos de 1791 e 1803; o glossário compõe-se de 59 verbetes. O segundo livro traz a edição semidiplomática ou lição de vinte e cinco documentos, num total de trinta e quatro fólios, selecionados dentre os redigidos nos períodos de 14 de novembro de 1721 a 27 de dezembro de 1743. Ao final, também, incluímos um glossário⁵, com 106 verbetes, que foram construídos de acordo com as seguintes normas:

- a) Entrada das lexias em ordem alfabética (os vocábulos em caixa alta apresentam o mesmo registro dos manuscritos; os que se encontram entre colchetes representam a variante padrão-culta lexicalizada nos dicionários de Língua Portuguesa do Brasil);
- b) Forma atual e significado no texto;
- c) Abonação com extratos dos textos manuscritos (constam dos exemplos: número do documento, recto ou verso do fólio e indicação das linhas);

³ Organizadoras.

⁴ Organizadoras.

⁵ A primeira versão desse glossário consta da Dissertação de Vasconcelos, *Edição de manuscritos de Paranaguá do século XVIII: Um percurso filológico*, defendida em 2006, sob minha orientação.

- d) Definição dos vocábulos encontrados nos dicionários (a ordem de citação dos dicionaristas segue o critério cronológico). A sigla N.C. significa “Não Consta”, uma vez que a lexia não está dicionarizada, pelo menos nos principais dicionários (Ferreira, Nascentes, Houaiss, Cunha, Morais Silva).

A propósito de ambos os glossários, os autores verificaram que: (i) a maioria dos vocábulos listados já caiu em desuso, pelo menos na linguagem paranaense oral corrente; (ii) outros podem ser encontrados somente na escrita; (iii) outros, ainda, apresentam um novo valor semântico adquirido com o tempo e, finalmente, (iv) outros se apresentam sob uma grafia diferente da usada na atualidade.

Essa primeira experiência nos levou a aceitar o desafio proposto pelo Dr. Ataliba de Castilho: compor uma equipe para elaborar o *Léxico Histórico do Português Brasileiro*.

2 O LÉXICO HISTÓRICO DO PORTUGUÊS BRASILEIRO: DEFINIÇÕES QUE INTERESSAM À LEXICOGRAFIA

Antes de tratarmos dos procedimentos metodológicos para a elaboração do *Léxico histórico do português brasileiro*, faz-se necessário definir os conceitos de dicionário, glossário, léxico e vocabulário assumidos neste texto. Para tal, buscamos as definições constantes de Jota (1981) e de Dubois *et al.* (1998).

Para dicionário:

- Registro alfabético do léxico de uma língua com os respectivos significados (JOTA, 1981).
- Objeto cultural que apresenta o léxico de uma ou mais línguas sob a forma alfabética, fornecendo sobre cada termo certo número de informações (pronúncia, etimologia, categoria gramatical, definição, construção, exemplo de emprego, sinônimos, idiotismos) (DUBOIS *et al.*, 1998).

Para Glossário:

- Dicionário especial que consigna vocábulos sobre os quais pode o leitor comum ter dificuldades (JOTA, 1981).
- Dicionário que dá, sob a forma de simples traduções, o sentido de palavras raras ou mal conhecidas. (DUBOIS *et al.*, 1998)

Para Vocabulário:

- Conjunto de *vocábulos* de uma língua. Via de regra, opõe-se *vocabulário*, mero repositório de vocábulos (palavras sem as respectivas significações), a *dicionário*, conjunto das palavras e significações (JOTA, 1981)

Para Léxico:

- Conjunto de vocábulos de uma língua. Se ordenado, constitui o dicionário. Há quem prefira *léxico* como o repositório da língua, reservando *vocabulário* para a fala, dando o nome de *lexema* à unidade do léxico e *vocábulo* ao lexema atualizado (JOTA, 1981).
- Livro que compreende a lista dos termos utilizados por um autor, por uma ciência ou por uma técnica (DUBOIS *et al.*, 1998).

Como vimos, tanto Jota quanto Dubois *et al.* definem *léxico* como um conjunto de vocábulos, uma lista de termos. Assumimos essa definição para o Léxico Histórico do Português Brasileiro, que será um conjunto o mais exaustivo possível dos vocábulos usados em documentos manuscritos oficiais, pessoais ou familiares, datados dos séculos XVII ao XIX e coletados pelas equipes da USP, UFRJ, UFBA, UEMS, UFMG, UFPB, UNICAMP e UEL para o Projeto Para a História do Português Brasileiro.

3 O LÉXICO HISTÓRICO DO PORTUGUÊS BRASILEIRO: PRIMEIROS PASSOS

Em 2000, a convite da Dr^a Sonia Cyrino, passamos a integrar, na qualidade de dialetóloga, o projeto PHPB. No recente encontro dos pesquisadores, em Salvador (julho de 2009), o professor Ataliba Teixeira de Castilho nos deu a incumbência de organizar uma equipe para, nos moldes do *Léxico histórico del español de México* (COMPANY; MELIS, 2002), compor o *Léxico histórico do português brasileiro*.

Para a consecução da incumbência de construir o Léxico Histórico do Português Brasileiro (LHiPB), propusemos os seguintes passos:

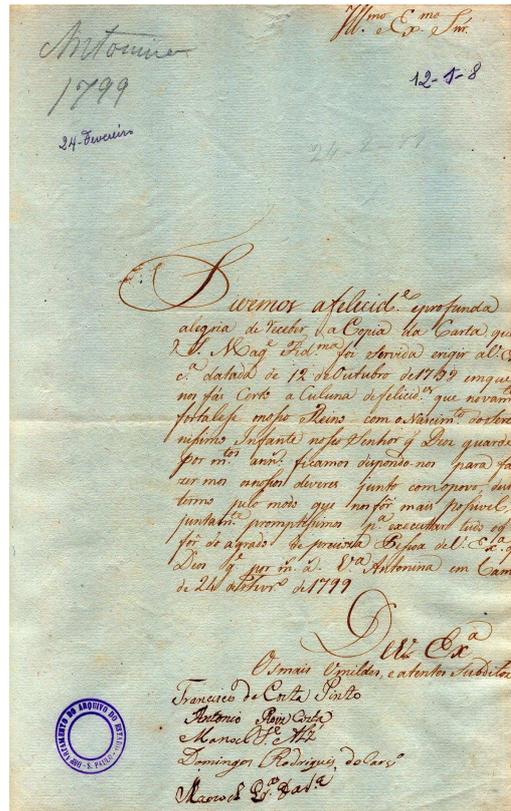
- a) organização do banco de dados a partir dos documentos coletados e transcritos pelas regionais de São Paulo, Minas Gerais, Rio de Janeiro, Paraíba, Bahia e Paraná.
- b) preparação das transcrições (lições) dos manuscritos, que consiste em eliminar toda informação inserida *a posteriori*, tais como observações apenas de arquivistas, pesquisadores, historiadores.
- c) junção de todos os documentos num só *corpus* no formato de bloco.

- d) aplicação do programa de análise linguística, Léxico 3, que trará a frequência de cada variante, para posterior lematização. Ilustramos, abaixo, com o lema Excelência.
- e) construção do verbete segundo o modelo do Léxico histórico *del español* de México.

O banco de dados organizado pela equipe do Projeto Para a História do Português Paranaense (PHPP), UEL, conta com os fac-símiles e a edição semidiplomática de 734 documentos selecionados junto ao Arquivo Público de São Paulo. Inicialmente buscamos nos catálogos do Arquivo todos os manuscritos referentes às primeiras vilas da então 5ª Comarca de São Paulo: Paranaguá, Antonina, Morretes, Castro, Guaratuba, Vila Nova do Príncipe e Curitiba, entre outros, datados até 1853, ano em que o Paraná se emancipou da Província de São Paulo. Recentemente a equipe do PHPP conseguiu recolher mais 750 novos documentos, na Casa da Memória e no Arquivo Público do Estado de Paraná (a grande maioria). Vinte desses documentos são do livro de batismo de Paranaguá. Todos eles estão sendo transcritos e revisados para serem incluídos no Léxico Histórico do Português Brasileiro (LHiPB).

Ao contrário do *Léxico histórico del español de México*, não faremos distinção entre texto de natureza oficial e texto de caráter coloquial (familiar, pessoal), uma vez que a leitura dos manuscritos do *corpus* do PHPP tem indicado que, apesar de oficiais, a ortografia ali usada está longe de ser considerada a padrão ou de estar rigorosamente dentro da gramática normativa, como têm discutido Almeida (2008) e Toniolo (2007). 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.

Segue abaixo, a título de ilustração, um fac-símile de manuscrito oficial paranaense que compõe o banco de dados do PHPP.



A edição de cada manuscrito tem o seguinte formato: uma ficha descritiva que traz o nº do documento e, entre colchetes, o nº na versão digitalizada. Na segunda linha, no item composição, registra-se o nº do fólio e se é frente (recto) ou verso. A terceira linha traz o assunto ou resumo para orientar o leitor sobre o conteúdo de cada documento. Embora a maioria dos documentos seja composta de um só fólio, há outros com até oito fólios. A quarta linha é destinada ao nome do(s) autor(es), isto é o(s) que assina(m) o documento, não sendo necessariamente o escriba.

Nº Doc: 07 [426]
Composição: fólio 1 recto
Assunto/resumo: Recebimento da cópia de uma carta informando o nascimento do infante.
Datação: 24/fevereiro/1799
Autores: Francisco da Costa Pinto, Antonio Rodrigues Costa, Manoel José Álvares, Domingos Rodrigues de Carvalho, Manoel Pereira da Silva.

Ilustríssimo e Excelentíssimo Senhor
[[Antonina]]
[[1799]]
[[24-fevereiro]]

[[12-1-8]]

Tivemos a *felicidade* e profunda
alegria de receber a Cópia da Carta que
Sua Magestade Fidelíssima foi Servida erigir a *Vossa Ex=*
celencia datada de 12 de Outubro de 1798 em que
nos fás Certo a Culuna de *felicidades* que novamente
fortalese o nosso Reino com o Nascimento do Sere=
níssimo Infante nosso Senhor *que* Deos guarde
por *muitos annos* ficamos dispondo-nos para fa=
zer-mos os nossos deveres junto com o povo, deste
termo pelo modo que nos fôr mais possível, e
juntamente promptísimos *para* executar tudo o *que*
fôr do agrado de preciosa Pessoa de *Vossa Excelencia* que
Deos *guarde* por *muitos anos* Vila Antonina em Camara
de 24 de Fevereiro de 1799
De *Vossa Excelência*
Os mais Umildes, e atentos Subditos
Francisco da Costa Pinto
Antonio Rodriguez Cozta
Manoel Jose Alvarez
Domingos Rodriguis de Carvalho
Manoel Pereira da Silva

De posse dos textos editados, segundo o modelo acima, procede-se à preparação de cada documento, isto é: (i) eliminam-se a ficha catalográfica e as inserções posteriores, como informações dadas por editores, historiadores, funcionários, que não pertencem ao texto original; (ii) retiram-se as palavras rasuradas ou incompletas; (iii) separam-se as palavras coladas, (iv) juntam-se as separadas, de acordo com a grafia atual e (v) grava-se no formato bloco.

Illustríssimo e Excelentíssimo Senhor
Tivemos a *felicidade* e profunda
alegria de receber a Cópia da Carta que
Sua Magestade Fidelíssima foi Servida erigir a *Vossa Excelencia*
datada de 12 de Outubro de 1798 em que
nos fás Certo a Culuna de *felicidades* que novamente
fortalese o nosso Reino com o Nascimento do Sereníssimo
Infante nosso Senhor *que* Deos guarde
por *muitos annos* ficamos dispondo-nos para fazermos
os nossos deveres junto com o povo, deste
termo pelo modo que nos fôr mais possível, e
juntamente promptísimos *para* executar tudo o *que*
fôr do agrado da preciosa Pessoa de *Vossa Excelencia* que
Deos *guarde* por *muitos anos* Vila Antonina em Camara
de 24 de Fevereiro de 1799

De Vossa Excelência
Os mais Umildes, e atentos Subditos
Francisco da Costa Pinto
Antonio Rodriguez Cozta
Manoel Jose Alvarez
Domingos Rodriguis de Carvalho
Manoel Pereira da Silva

Os documentos serão todos reunidos e formarão um único *corpus*-base, sobre o qual será aplicada a ferramenta Léxico 3. Isso feito, parte-se para a lematização das variantes para eger as entradas para o LHiPB. O verbete terá 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; MELIS, 2002, p. vii).

Exemplificamos com a entrada ou lema *excelência*:

Excelência recebemos a carta de Vossa *Excelencia*, com a qual nos fazia a honrra (6, 2); Joaõ Neto Mendes João da ujga *Excelentissimo* Senhor com a entrega, que desta (210, 5); comseda a Divinã a vossa *Excelentissima* o premio com a vida que devemos (243, 5); A Muito Ilustre Poesa de Vosa *Exselencia* Guarde Deos muitos annos (128, 5); emformar a Vossa *Excelência* o qual partio desta Villa (537, 1v); asignei Manoel Pereira do ô *Excelentissimo* Senhor Em trinta de novembro (78, 52v; recebemos a ordem de Vosa *exselencia*, em que nos ordenaua (96, 1); recebemos a de Vossa *excelencia* de dezassete do mesmo mes (61, 1); Matheus Luis graces *Excelentissimo* Senhor recebemos (192, 3); decoroso estado. Vossa *Exceletissima* pello sangue de Iesu Christo (44, 1); Gaspar Gonçalvez de Morais *Exceletissimo* Senhor Aos 17 de Outubro (177, 1); 1723 annos Muito de Vosa *Exselensia* Obsequioso , Senados (5, 4); da que devemos desejar a vossa *Exselenssia* Deos Guarde muitos annos (33, 1); João da uejgua de Siqujra *Exselentissimo* Senhor, recebemos a ordem; (9, 1); Antonio Igínjo lustrissimo e *excelentissimo* Senhor General Naô pouco temos; apresentar a Vossa Ilustrissima *exselenssia* guarde os Reverendos Padres (335, 1); Senado recebemos huã carta do *exselentissimo* General Rodrigo Sesar Governador (100, 1); rogamos a Vossa Ilustrissima e *exselinçia* a prouidencia sobre esta matéria (65, 6).

FORMAS DOCUMENTADAS: *Excelencia* (72), *Excelentissimo* (11), *Excelentissima* (10), *Excelência* (5), *Exselencia* (5), *Excelentissimo* (4), *exselencia* (3), *excelencia* (2), *Excelentissimo* (1), *Exceletissima* (1), *Exceletissimo* (1), *Exselensia* (1), *Exselenssia* (1), *Exselentissimo* (1), *excelentissimo* (1), *exselenssia* (1), *exselentissimo* (1), *exselinçia* (1).

FREQUENCIA GLOBAL: **122**

Nos moldes da obra de Company y Melis (2002), ofereceremos um léxico lematizado, que corresponde, em sua essência, a um dicionário de construção e uso no qual não se incluem definições semânticas dos lemas.

4 ESTÁGIO ATUAL DA CAMINHADA

Estamos trabalhando, inicialmente, com os manuscritos paranaenses para, até meados de 2010, estarmos com o *corpus* definitivo organizado, após a junção dos manuscritos das demais instituições participantes do projeto do PHPB. Quanto à equipe executora, esta ainda está se consolidando. Atualmente, sob a minha coordenação, estão se dedicando ao tratamento dos manuscritos editados o graduando Jonas Passos da Silva e o doutorando Cláudio de Assis, que também é o responsável pela aplicação do programa de análise linguística Léxico 3, que, como já vimos, trará a frequência de cada variante. A última fase, antes da publicação, será a da composição de cada um dos verbetes a partir das informações dadas por esse programa de análise do léxico.

5 REFERÊNCIAS

AGUILERA, V. de A.; ALMEIDA-BARONAS, J. E. (Orgs.). *Scripturae nas villas de São Luiz de Goaratuba e Antonina: manuscritos setecentistas e oitocentistas*. Londrina: EdUEL, 2007.

_____; VASCONCELOS, C. (Orgs.). *Scripturae na villa de Pernagoa: manuscritos setecentistas*. Londrina: EdUEL, 2007.

ALMEIDA, J. E. de; LINI, V. *Apontamentos sobre a escrita sob um perfil diacrônico*. Polifonia (UFMT), v. 15, p. 137-149, 2008.

COMPANY, C.; MELIS, C. *Léxico histórico del español de México*. México: Universidad Autónoma de México, 2002.

CRYSTAL, D. *Dicionário de Linguística e Fonética*. [Trad. e adapt. – da 2. ed. inglesa ver. e ampl., publicada em 1985 – Maria Carmelita Pádua Dias]. Rio de Janeiro: Jorge Zahar Ed.,

CUNHA, A. G. da. *Dicionário histórico das palavras portuguesas de origem tupi*. 5. ed. São Paulo: Companhia Melhoramentos; Brasília: Universidade de Brasília, 1978.

DUBOIS, J. *et al. Dicionário de Linguística*. Rio de Janeiro: Cultrix, 1998.

JOTA, Z. dos S. *Dicionário de Linguística*. 2. ed. Rio de Janeiro: Presença/INL-MEC, 1981.

SILVA, J. C. da. *Dicionário da Língua Portuguesa Medieval*. Londrina: Eduel, 2008.

TONIOLO, Ê. J. Aspectos ortográficos de manuscritos paranaenses. In: XVIII Seminário do CELLIP, 2007, Ponta Grossa. *Caderno de Resumos*. Ponta Grossa: Editora da UEPG, 2007. v. 1. p. 68-68.

VOCABULÁRIO Histórico-cronológico do Português Medieval. Disponível em: <<http://www.casaruibarbosa.gov.br/medieval/cadastronline.html>>. Acesso em: 15 out. 2009.