

8. PARA O ESTUDO DA VARIAÇÃO E MUDANÇA NO PORTUGUÊS BRASILEIRO: EDIÇÃO XML DE *CORPORA* DO SEMIÁRIDO BAIANO

Matheus Santos Oliveira¹

Mariana Fagundes de Oliveira Lacerda²

RESUMO: Esta comunicação tem, por objetivo, primeiramente, discutir as duas principais teorias sobre a história do Português Brasileiro, e, depois, mostrar como a edição XML de documentos históricos do semiárido baiano pode contribuir para o mapeamento dos fenômenos linguísticos pelos quais os linguistas se interessam no estudo das variações e mudanças linguísticas que fizeram com que o Português Brasileiro tenha as características que conhecemos hoje.

PALAVRAS-CHAVE: Variação, Mudança, Origem do Português Brasileiro, Edição XML

ABSTRACT: This communication aims to discuss the two main theories about the origin of Brazilian Portuguese, and then show how the edition XML of historical documents of Bahia can contribute for the mapping of linguistic phenomena by which the linguistics are interested in study of variations and linguistic changes that made Brazilian Portuguese have the character we know today.

KEYWORDS: Variation, Change, Origin of the Brazilian Portuguese, Edition XML.

INTRODUÇÃO

Este artigo foi apresentado, em forma de comunicação oral, numa das mesas do *III Seminário de Sociolinguística da UNEB*, realizado entre os dias 03 e 04 de outubro de 2013. Na ocasião, tínhamos dois objetivos: apresentar uma revisão crítica da literatura, focando nas duas principais teorias sobre a origem do Português Brasileiro em voga, e

¹ Graduando em Licenciatura em Letras com Língua Espanhola, Universidade Estadual de Feira de Santana, e-mail: matheusuefs@live.com

² Professora do Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, e-mail: marianafag@gmail.com

discutir em que medida as edições XML de documentos históricos do sertão baiano, no âmbito do projeto CE-DOHS (de que falaremos mais adiante), têm contribuído para a reconstrução da história dessa vertente do português na América. Discutiremos aqui, então, a teoria do contato, defendida, principalmente, por Dante Lucchesi (2009) e a teoria da *deriva secular*, apregoada pelos professores Anthony Julius Naro e Maria Marta Pereira Scherre (2009). Feito isso, passaremos a tratar do projeto de pesquisa CE-DOHS, da Universidade Estadual de Feira de Santana, explanando seus objetivos, metodologia, perspectivas futuras e, tal como supramencionado, as contribuições que ele tem trazido aos estudos linguísticos, quer de natureza sincrônica, quer de natureza diacrônica, que versam sobre a heterogeneidade do PB e sobre as mudanças que o fizeram ter as características que conhecemos hodiernamente, sob as mais variadas perspectivas teóricas.

1 A ORIGEM DO PB: DISCUTINDO AS TEORIAS

Há uma numerosa bibliografia que trata da língua portuguesa do Brasil, focalizando, nas palavras da professora Rosa Virgínia Mattos e Silva, “o português, sua transplantação, difusão e implantação no espaço americano” (2004, p. 12). Parece consenso afirmar que falar do PB é falar de, pelo menos, duas histórias. Dante Lucchesi sintetiza esse pensamento, afirmando que há uma *polarização sociolinguística* de nossa língua. Essa polarização aparta

[...] a fala de uma elite que sempre teve os olhos voltados para a Europa, em busca de seus modelos culturais e linguísticos, da fala da grande população que, no cadinho de sua pluralidade étnica, cultural e linguística, forjou os elementos definidores da originalidade cultural e linguística do Brasil, que tanto assombram e encantam o mundo ocidental, desautoriza todos os estudos que apresentam *uma história única para o português brasileiro* (LUCCHESI, 2009, p. 30).

Lucchesi completa que “assim como ‘o **português são dois**’, a sua história é igualmente bifurcada” (*idem*). É exatamente a constituição histórica da vertente popular do PB que tem suscitado discussões, por vezes, calorosas.

A primeira grande teoria para a explicação da origem do português brasileiro popular é a da *deriva secular*, apregoada por Anthony Julius Naro e Maria Marta Pereira Scherre (1993; 2003; 2009 etc.), segundo a qual o português brasileiro nada mais é do que uma continuação do português arcaico, “com pequenas alterações”, uma vez que não se consegue, até hoje, ainda segundo eles, “identificar nenhuma característica do português do Brasil que não tenha um ancestral claro em Portugal” (2007, p. 13). Ademais, Naro e Scherre afirmam que o PB possui uma expansão de “estruturas e variações” acelerada por uma confluência de motivos, dentre os quais o contato entre línguas. Há que ressaltar que esses autores utilizam a noção de deriva do linguista estadunidense Edward Sapir, para quem a linguagem “não é apenas uma coisa que cresça no espaço [...], [mas] move-se pelo tempo em fora num curso que lhe é próprio. Tem uma deriva” (p. 121).

O principal oponente da visão supracitada é Dante Lucchesi, que tem dedicado seus estudos às análises linguísticas da vertente popular do português brasileiro no Estado da Bahia. O Projeto *Vertentes do Português Popular do Estado da Bahia*, do qual é coordenador, tem, por objetivo,

“[...] sistematizar evidências empíricas da relação historicamente motivada entre certas características morfossintáticas da norma popular brasileira e o maciço contato do português com as línguas indígenas e africanas ocorrido nos primeiros séculos da formação da sociedade brasileira” (LUCCHESI, 2009, p. 251)

Segundo esse autor, “a aquisição precária do português como segunda língua por parte de milhões de índios brasileiros e africanos escravizados” leva a uma simplificação morfológica, que é “característica geral das situações de contato massivo entre línguas” (2009, p. 252).

A hipótese de o PB ser um *crioulo* de base portuguesa (cujo léxico seria do português europeu e sua gramática, das línguas africanas com que teve contato) não se comprova, na medida em que não houve a formação de uma gramática original. O que teria havido, nas palavras de Holm (2004), seria a formação de uma *variedade vernácula parcialmente reestruturada*. À formação dessa variedade, Dante Lucchesi atribui o processo *transmissão linguística irregular* (doravante TLI), o qual seria responsável pela simplificação morfológica do PB em sua vertente popular. Por TLI,

entendem-se os processos de mudança decorrentes do maciço contato entre línguas, sem que as alterações ocorridas aí “cheguem a configurar a emergência de uma nova entidade linguística qualitativamente distinta” (LUCCHESI, 2009, p. 255).

A redução na estrutura gramatical da língua a que nos referimos, ainda segundo Lucchesi, deve-se:

“(i) à dificuldade de acesso dos falantes das outras línguas aos modelos da LA [língua-alvo], sobretudo nas situações em que o grupo dominante é numericamente muito inferior aos falantes das outras línguas;

(ii) ao fato de os falantes das outras línguas serem, em sua grande maioria, adultos, o que faz com que esses falantes não possam contar com os dispositivos inatos que atuam naturalmente no processo de aquisição da língua materna;

(iii) à ausência de uma ação normatizadora, ou seja, de uma norma ideal que oriente e restrinja o processo de aquisição da LA, já que esse processo tem seus objetivos circunscritos aos intentos comunicativos” (LUCCHESI, 2009, p. 102)

A teoria cunhada por Lucchesi é, para Naro e Scherre, uma excrescência teórica. Isso porque, segundo os autores, fazendo uma referência ao frade franciscano do século XVI, Guilherme de Ockham, atribuir as características do PB popular ao contato linguístico e à reestruturação parcial da morfossintaxe dessa língua seria explicar erroneamente *o sumiço dos carneiros*. Segundo a estória do frade, um fazendeiro, numa certa manhã, dá pela falta de alguns carneiros de seu rebanho.

“[O fazendeiro] conclui que os carneiros devem ter sido comidos pelos lobos que ele vira na floresta vizinha à sua propriedade no dia anterior. Mas ele poderia também pensar que ouvira dizer que há leões no zoológico que fica na capital. Talvez tenham escapado alguns leões e estes poderiam ter comido os carneiros. [...] Ou então poderia ter havido uma invasão despercebida de extraterrestres, casualmente famintos exatamente por carneiros. A Navalha de Ockham determina que a primeira teoria é a que deve ser preferida, por envolver menos construtos teóricos do que a teoria da fuga dos leões ou a da existência de extraterrestres”. (LUCCHESI, 2009, p. 259)

Isso porque, segundo a doutrina do frade, é preciso cortar qualquer excrescência teórica que não contribua para explicar os dados relevantes. Entretanto, sabiamente, Lucchesi afirma que, na verdade, a excrescência teórica seria acreditar numa *deriva secular*, o que nos remete à Linguística imanentista de Saussure, não mais aplicável à realidade das línguas.

A Sociolinguística, já que invoca os fatores extralinguísticos em seus estudos sobre variação e mudança, não deve, para Lucchesi e seus seguidores, deitar fora a importância do contato de línguas no processo de constituição histórica da vertente popular do PB.

2 A EDIÇÃO XML DE *CORPORA DO CE-DOHS* E A CONTRIBUIÇÃO À RECONSTRUÇÃO DA HISTÓRIA DO PB

É inevitável dizer que, na discussão até aqui explanada e em muitas outras que concernem às variações e às mudanças por que passa o PB, a formação de bancos de dados é de irrefutável relevância. Daí a importância da Linguística de *corpus*.

Em 1985, segundo Houaiss, não se havia ainda preenchido os requisitos da pesquisa e conhecimento com que se pudesse elaborar uma história da língua portuguesa no Brasil. No *I Seminário para a História do Português Brasileiro*, na década de 1990, criaram-se as condições de pesquisa e conhecimento para elaborar uma história do PB. Vários pesquisadores, em equipes regionais sediadas em universidades de sete estados brasileiros – Bahia, Minas Gerais, Paraná, Pernambuco, Rio de Janeiro, Santa Catarina e São Paulo (depois, também a Paraíba) –, iniciaram o Projeto Para a História do Português Brasileiro (PHPB), para construir um conjunto significativo de documentação representativa tanto do chamado português culto do Brasil como do português popular do Brasil.

O CE-DOHS (*Corpus* Eletrônico de Documentos Históricos do Sertão), projeto de pesquisa sediado no NELP (Núcleo de Estudos da Língua Portuguesa), do DLA (Departamento de Letras e Artes) da UEFS (Universidade Estadual de Feira de Santana), parceiro do PHPB, tem contribuído em larga escala para os estudos linguísticos que dão suporte às teorias sobre variação e mudança no PB e para descrições em geral da realidade plurilinguística do PB falado e escrito no semiárido baiano.

O trabalho do grupo, formado por bolsistas de IC, pesquisadores e colaboradores externos, é contribuir com a primeira agenda de trabalho do PHPB (Projeto para a História do Português Brasileiro): constituição de *corpora*. No mestrado, os antigos bolsistas de IC contribuem para a segunda agenda de trabalho, que compreende as análises linguísticas a partir dos *corpora* editados.

O projeto em questão tem por afã desenvolver metodologias para formação de grandes bancos de dados eletrônicos para fins linguísticos, com vistas à otimização no acesso aos documentos históricos do sertão da Bahia, por meio do uso da linguagem XML (Extended Markup Language) e da ferramenta integrada de anotação de *corpus*, E-Dictor, desenvolvida por Faria, Kepler e Paixão de Souza (2004-2009). Serão disponibilizadas, portanto, edições fac-similadas em versão semidiplomática para o estudo do português brasileiro, em diferentes perspectivas teóricas, de acordo com critérios estabelecidos pelo Projeto Para a História do Português Brasileiro (PHPB).

A edição de textos feita no âmbito do projeto é controlada, em linguagem anotada XML, e possibilita o controle e mapeamento das intervenções realizadas nos textos, garantindo a recuperabilidade das formas originais. Esse tipo de edição eletrônica, em XML, permite a busca automática de dados para análise linguística, o que a edição tradicional não possibilita ao pesquisador.

A metodologia baseia-se fundamentalmente na metodologia do Projeto *Corpus Histórico do Português Tycho Brahe*, composto por um *corpus* eletrônico anotado de textos em português escritos por autores nascidos entre 1435 e 1845, desenvolvido desde 1998, em <http://www.tycho.iel.unicamp.br/~tycho/corpus/>, onde estão definidos os modelos e as ferramentas que estão subsidiando o Projeto CE-DOHS, um *corpus* voltado a um banco inédito de textos do sertão da Bahia. Esse tipo de banco de dados vem-se mostrando uma tendência mundial, com um grande número de projetos, sendo o pioneiro o projeto *Penn Helsinki Parsed Corpus of Middle English* (<http://www.ling.upenn.edu/hist-corpora/>), coordenado por Anthony Kroch na Universidade da Pensilvânia.

A perspectiva é que, em breve, todos os textos de que dispõe o Tycho Brahe ganhem anotações morfossintáticas. Os textos já anotados morfossintaticamente têm otimizado sobremaneira a coleta de dados pelo pesquisador.

Quanto ao banco do CE-DOHS, é por 1037 cartas particulares (1808-2000), escritas por 418 remetentes (nascidos entre 1724 e 1980), além de 25 cartões. São acervos extraídos, em sua maior parte, de *Cartas Brasileiras: coletânea de fontes para o*

estudo do português (Fapesb/Processo 1493/2010). Na segunda fase, há uma ampliação do número de cartas, com a inserção de novos acervos manuscritos e impressos.

Além desses acervos já disponibilizados dos séculos XVIII, XIX e XX, há previsão da inserção de novos acervos por alunos da pós-graduação que estão editando documentos históricos inéditos, procedentes da grande área do semiárido, gestada sob forte ambiente de contato linguístico, principalmente com línguas ameríndias e africanas, durante o processo de multilinguismo que caracterizou o interior da Bahia no período colonial brasileiro.

3 CONCLUSÕES

Como a preocupação dos que fazem o CE-DOHS é menos ganhar títulos e notoriedade do que contribuir com a interpretação da realidade plural da língua portuguesa do Brasil, pensamos estar no caminho certo.

Diversos trabalhos de análise linguística da variação e da mudança linguísticas na Bahia, artigos, monografias, dissertações de mestrado, como a de Huda Santiago (2012) e três livros – *Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português*, organizado por Zenaide Carneiro; *Publica-se em Feira de Santana*, organizado por Zenaide Carneiro e Mariana Fagundes de Oliveira Lacerda e *Coleção Amostras da Língua Falada no Semi-Árido Baiano*, de Norma Lúcia Fernandes Almeida e Zenaide Carneiro – são produções nascidas no âmbito do CE-DOHS.

Nenhum trabalho em Sociolinguística ou em Linguística Histórica pode ter validade científica sem uma base empírica bem construída. É pensando nisso que lançamos mão da formação de um grande banco de dados que, sem dúvida, possibilita estudos de diversas perspectivas teóricas importantes para uma coerente descrição da história do PB popular na Bahia. Em outras palavras, é a construção de *corpora* que dá margem à explicação do *sumiço dos carneiros*.

REFERÊNCIAS

ALMEIDA, Norma Lúcia Fernandes de & CARNEIRO, Zenaide de Oliveira Novais. *Notícias sobre o banco de textos de Feira de Santana e Cachoeira*. In: ALKMIM, Tânia M. *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas/FFCHL/USP: FAPESP, 2002, v. 2, p. 465-488.

BAXTER, Alan; LUCCHESI, Dante & RIBEIRO, Ilza (org.). *O Português Afro-brasileiro*. Salvador: EDUFBA, 2009.

CARNEIRO, Z. O. N. (Org.) ; OLIVEIRA, Mariana Fagundes de (Org.) . Publica-se em Feira de Santana (1908-2006). 1. ed. Feira de Santana: UEFS Editora, 2012. v. 1. 55p .

CASTILHO, Ataliba T. de (Org.). *Para a história do português brasileiro: primeiras idéias*. São Paulo: Humanitas Publicações / FFLCH/USP, 1998, v.1.CORPUS DOHS. *Documentos Históricos do Sertão* (disponível em www.uefs.br/dohs), 2010.

SOUZA, Constância Maria Borges de. *A concordância verbal em Salvador: resultado de uma transmissão linguística irregular?* In: LOPES, Norma da Silva; SOUZA, Constância Maria Borges de; SOUZA, Emília Helena Portella Monteiro de (orgs.). *Um estudo da fala popular de Salvador – PEPP*. Salvador: Quarteto Editora, 2009.

FARACO, Carlos Alberto. *Linguística Histórica: uma introdução ao estudo da história das línguas*. São Paulo: Parábola Editorial, 2005.

GUY, Gregory. (2005). *A questão da crioulização no português do Brasil*. In: ZILLES, A. M. S. Estudos de variação linguística e no Cone Sul. Porto Alegre: Editora da UFRGS.

GUY, Gregory. (1981). *Linguistic variation in Brazilian Portuguese: aspects of the phonology, syntax and language history*. Philadelphia: PH.D. Dissertation.

HOUAISS, Antônio. *O português no Brasil*. Rio de Janeiro: UNIBRADE, 1985.

LOBO, Tânia. *Arquivos, acervos e a reconstrução histórica do português brasileiro*. In: OLIVEIRA, Klebson; CUNHA E SOUZA, Hirão F.; SOLEDADE, Juliana (Org.). *Do português arcaico ao português brasileiro: outras histórias*. Salvador: EDUFBA, 2009.

LOBO, Tânia & OLIVEIRA, Klebson (org.). *África à vista: dez estudos sobre o português escrito por africanos no Brasil do século XIX*. Salvador: EDUFBA, 2009.

LOBO, T. C. F. (Org.) ; CARNEIRO, Z. O. N. (Org.) ; COELHO, J. S. B. (Org.) ; ALMEIDA, Ariadine (Org.) ; RIBEIRO, Silvana (Org.) . ROSAE: linguística histórica, história das línguas e outras histórias. 1. ed. Salvador: Edufba, 2012. v. 1. 918p .

LUCCHESI, Dante. *As duas grandes vertentes da história sociolinguística do do Brasil*. Delta [online], 2001. Disponível em http://www.scielo.br/scielo.php?pid=S010244502001000100005&script=sci_arttext

LUCCHESI, Dante.(2003). O conceito de “transmissão linguística irregular” e o processo de formação do português do Brasil. In: RONCARATI, Cláudia; ABRAÇADO, Jussara (Org.). *Português brasileiro: contacto linguístico, heterogeneidade e história*. Rio de Janeiro: 7 Letras. p. 272-284.

LUCCHESI, Dante. *A deriva secular na formação do português brasileiro: uma visão crítica*. In: LOBO, T. C. F. (Org.) ; CARNEIRO, Z. O. N. (Org.) ; COELHO, J. S. B. (Org.) ; ALMEIDA, Ariadine (Org.) ; RIBEIRO, Silvana (Org.) . *ROSAE: linguística histórica, história das línguas e outras histórias*. 1. ed. Salvador: Edufba, 2009.

LUCCHESI, Dante; BAXTER, Alan; RIBEIRO, Ilza. *O Português Afro-Brasileiro*. Salvador: EDUFBA, 2009.

MATTOS E SILVA, Rosa Virgínia. *Para a história do português culto e popular brasileiro: sugestões para uma pauta de pesquisa*. In: ALKMIM, Tânia M. *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas/FFCHL/USP: FAPESP, 2002, v. 2, p. 443-464.

MATTOS E SILVA, Rosa Virgínia. *Ensaio para uma sócio-história do português brasileiro*. São Paulo: Parábola, 2004.

NARO, Anthony Julius & SCHERRE, Maria Marta Pereira. *Origens do português brasileiro*. São Paulo: Parábola Editorial, 2007.

OLIVEIRA, Matheus Santos & OLIVEIRA, Mariana Fagundes. *Edição XML de Cartas de Sisal: contribuições à reconstrução da história do Português Brasileiro*. In: Caderno de Resumos do VIII Congresso Internacional da ABRALIN. EDUFRN: Natal, 2013.

PAIXÃO DE SOUSA, M. C.; KEPLER, F. N.; FARIA, P. *E-dictor: Novas perspectivas na codificação e edição de corpora de textos históricos*. In: VIII Encontro de Linguística de *Corpus*, 2009, Rio de Janeiro. Resumos, 2009. (a sair em: Shepherd, T., Berber Sardinha, T. e Veirano Pinto, M. (Org.). *Linguística de Corpus: Sínteses e Avanços*, 2009. Anais do VIII Encontro de Linguística de *Corpus*, realizado na UERJ, 13 a 14 de novembro de 2009. Rio de Janeiro, 2009. p. 69-105.

PROJETO CORPUS ELETRÔNICO DE DOCUMENTOS HISTÓRICOS DO SERTÃO (disponível em www.uefs.br/cedohs), 2011.

PROJETO VOZES DO SERTÃO EM DADOS (disponível em: ww.uefs.br/nelp), 2011.