

Do presente para o passado: os caminhos da constituição do banco de dados do NELP/UEFS em busca da configuração sócio-histórica da língua portuguesa nos sertões baianos

Zenaide de Oliveira Novais Carneiro
(Universidade Estadual de Feira de Santana)

Norma Lucia Fernandes de Almeida
(Universidade Estadual de Feira de Santana)

Mariana Fagundes de Oliveira Lacerda
(Universidade Estadual de Feira de Santana)

Silvana Silva de Farias Araújo
(Universidade Estadual de Feira de Santana)

Priscila Starline Estrela Tuy Batista
(Universidade Estadual de Feira de Santana)

Considerações iniciais

Ao propormos o título acima para este capítulo, tínhamos como objetivo abordar brevemente o processo de constituição de *corpora* orais e escritos, iniciado no âmbito do Núcleo de Estudos da Língua Portuguesa (NELP)¹, na

1 O NELP, criado em 1998 por Norma Lucia Fernandes de Almeida e Zenaide de Oliveira Novais Carneiro, abriga o banco de dados com amostras de fala representativas do semiárido baiano, que são produto do projeto A língua Portuguesa no Semi-Árido Baiano, criado por Dante Lucchesi, em 1993, quem também coordenou a recolha das primeiras amostras. As atividades do projeto se iniciaram efetivamente em 1996 e estenderam-se até 2001. Além dos dados orais, o NELP abriga o banco de dados formado por documentos históricos produzidos nessa região, no âmbito do projeto Contribuições para a constituição de um banco de textos e de um banco de dados para o estudo da história do português no Brasil, do séc. XVII ao XX, pensado e implementado, em 1997, por Ilza Ribeiro, que o filiou ao Projeto para a História da Língua Portuguesa (PROHPOR), da Universidade Federal da Bahia e ao PHPB.

década de 1990. O banco oral do NELP é composto por gravações de fala de comunidades rurais do semiárido baiano e, mais recentemente, da área urbana do município de Feira de Santana. Ao lado dessas amostras orais, o NELP disponibiliza também textos escritos, a maior parte documentação epistolar; essa prospecção de fontes escritas de séculos passados produzidas nos sertões baianos iniciou-se igualmente na década de 90, tendo em vista uma das agendas de trabalho do Projeto para a História do Português Brasileiro (PHPB): a constituição de *corpora* diacrônicos.

Os sertões baianos – cujas fronteiras começam a ser estabelecidas a partir de 1640 e que englobam atualmente 62% dos municípios do Estado, a maior parte, portanto, de suas regiões econômicas² – apresentam uma ocupação luso-brasileira característica do período colonial, marcado por amplo contato com línguas indígenas e africanas.

Mattos e Silva (2009), ao discutir os meios para a reconstrução histórico-diacrônica do português brasileiro (doravante PB), a partir de uma visão que inclui, pelo menos, duas origens distintas do PB – português culto *versus* português popular³ –, defende que os caminhos são distintos; enquanto aquele, para a autora, poderá ser recuperado “pela via tradicional desse tipo de trabalho para as línguas que estão representadas pela escrita, ou seja, pelo escrutínio das fontes documentais remanescentes”, o percurso para a reconstrução do passado do *português brasileiro popular/falado* seria um “percurso análogo, *mutatis mutandis* e *modus in rebus*, ao da reconstrução do chamado ‘latim vulgar’” (p.14). Percurso espinhoso este, segundo Mattos e Silva, mas não impossível.

De forma resumida, podemos afirmar que, para definir a constituição do chamado tradicionalmente de português *culto* brasileiro, faz-se imprescindível a realização de pesquisas com dados escritos por informantes escolarizados do passado, quer brasileiros, quer portugueses. Por outro lado, para o escrutínio da constituição histórica das variedades faladas do PB por classes socialmente desprestigiadas e com pouco contato com a escrita formal, é fundamental a realização de estudos feitos com base em textos escritos por informantes poucos escolarizados, fazendo, em ambos os casos, a sócio-história dessas variedades. Outra via de investigação que pode ser realizada, no caso das variedades fala-

2 As regiões econômicas do semiárido baiano são: Nordeste, Paraguaçu, Sudoeste, Baixo Médio São Francisco, Piemonte da Diamantina, Irecê, Chapada Diamantina, Serra Geral e Médio São Francisco (SEI, 2011).

3 De acordo com Lucchesi (1994, 2001), que vê o português do Brasil como um sistema não apenas heterogêneo e variável, mas plural, um diassistema formado por dois subsistemas, por sua vez, igualmente heterogêneos e variáveis, definidos como “normas”.

das, é a busca do presente para o passado, por meio de gravações da fala de pessoas analfabetas ou parcialmente escolarizadas, tanto em comunidades urbanas como em comunidades rurais.

Assim, indo na direção das ideias expostas por Mattos e Silva (2009), Lobo (2001, p. 109) discorre sobre a constituição de *corpora* diacrônicos do PB, definindo o que seria um *corpus* linguístico ideal:

- (i) **Subcorpus 1:** as variedades do português europeu transplantadas para o Brasil;
- (ii) **Subcorpus 2:** as variedades do português falado como segunda língua pelos aloglotas;
- (iii) **Subcorpus 3:** as variedades do português brasileiro que paulatinamente se iam constituindo: 3.1) as variedades cultas⁴ – supostamente mais unitárias e descendentes diretas das variedades do português europeu – e 3.2) as variedades populares⁵ – supostamente mais diversificadas e descendentes diretas das variedades do português como segunda língua.

É desafiadora a composição de *corpora*, sobretudo aqueles formados por fontes documentais históricas que visem à recuperação da vertente amplamente falada nos sertões e pouco documentada nos textos escritos. Em relação às amostras de escrita, indaga-se o quanto cada amostra pode efetivamente evidenciar a gramática que escapa do caráter homogeneizador da norma de prestígio que ronda os textos escritos. Desse ponto de vista, a variação e a mudança observadas nos textos são, por definição, variação e mudança instanciadas nos textos. Entretanto, a linguística histórica tem, desde a sua gênese, buscado meios para lidar, à moda de arqueólogos, com essa aparente impossibilidade. Atualmente, um dos caminhos é o de interpretar a variação encontrada nos textos como parte de uma dinâmica de mudança em que uma gramática entre em competição com outra, como proposto por Kroch (1994, 2001), uma competição captada mais facilmente em *corpora* extensos, em que elementos da gramática vernacular conseguem inserir-se.

No que tange à coleta de fontes amealhadas ao longo das últimas duas décadas, no âmbito do NELP – a maior parte dessas nos vastos sertões baianos –, foram sempre submetidas a uma contextualização sócio-histórico-demográfica e ao controle das condições de produção da documentação escrita, a fim de os *corpora* permitirem recuperar minimamente variedades diatópicas e sociais, no presente, se houver, e no passado.

4 Atualmente, a autora usa a terminologia “socialmente prestigiadas”.

5 Atualmente, a autora usa a terminologia “socialmente des prestigiadas”.

1 Os sertões baianos: alguns aspectos de sua sócio-história

Os sertões baianos, na historiografia, são tidos como “verdadeiros mundos perdidos e isolados”, segundo Silva *et al* (1989, p. 94-97). Na Bahia, durante o período colonial, as povoações ou vilas surgiram em pontos de interseção das estradas que serviam de rotas para o gado, o ouro e iam avançando rumo ao interior⁶. Para os autores, a forma de ocupação de terras no Brasil se caracteriza, até o século XVII, pela concessão de sesmarias, por meio de Ordenações do Reino⁷.

Para Santos (2010, p. 403), a ocupação luso-brasileira no sertão “foi marcada por reversões, lacunas e descontinuidades”. Sujeita a grupos dominantes, estava a chamada “gente miúda”, isto é, “curraleiros, pequenos proprietários de lavouras de subsistência, mestiços livres e pobres, escravos índios e escravos negros; nas zonas de indefinição do sistema territorial, no limiar entre o território luso-brasileiro e os espaços exteriores, estavam os chamados “fora-da-lei”, isto é, “quilombolas, índios fugidos e luso-brasileiros perseguidos pela justiça régia. Do outro lado das fronteiras territoriais, mantinham-se “as dezenas e mesmo centenas de povos indígenas que os documentos coloniais reuniram sob a denominação de gentio bárbaro, uma diversificada população ameríndia milenar”. A ocupação dos sertões, distinta da ocupação do litoral, de domínio português, teria sido feita por brasileiros, na visão de Freire (1998).

Esse cenário, em si, representa um contexto que não poderia ser outro senão aquele no qual foi constituído, em longo processo histórico, o português falado em sua origem, como segunda língua, e que aí se instalou ou foi gestado, resultante de amplo contato linguístico, característico do período colonial brasileiro e que o sertão baiano representa bem; parece, portanto, provável considerar que se tenha fixado, nessa região, o chamado português popular, que, “adquirido na oralidade e em situações de aquisição imperfeita”, seria, para Mattos e Silva (p. 298-299),

6 Em geral, o tratamento histórico das estruturas de ocupação em zonas de fronteira do sertão baiano de 1640 a 1750, segundo Santos (2010), são o *caminho, a sesmaria, a povoação, o posto militar e a missão religiosa*. As povoações se caracterizam como formas de concentração de população em lugares distintos dos espaços de produção pecuária, agrícola e extrativista, característicos da economia sertaneja. Elas constituem, segundo o autor, em nucleações pioneiras que, de uma forma ainda muito rudimentar, apresentam alguns dos elementos que normalmente caracterizam os espaços urbanos coloniais, com pequeno comércio, com pousos, pequenas oficinas de artefatos de couro, ferro, pedra e madeira. As nucleações maiores recebiam também rudimentos de algumas funções administrativas, sediando julgados, abrigando registros e tornando-se “locus de vida social e religiosa” e “postos avançados para novos descobrimentos e conquistas” (2010, p. 356).

7 No caso da Bahia, mostram que a ocupação territorial começa com o que denominou de *rush* fundiário, entre 1670 e 1690.

difundido pelo geral do Brasil, “sobretudo pela maciça presença africana e dos afro-descendentes, que perfizeram uma média de mais de 60% da população por todo o período colonial”. Para Carneiro e Almeida (2006), a população de origem africana pode ter exercido um maior papel no sertão baiano, quando parte do contingente de escravos e seus descendentes – tanto a que trabalhou nas minas quanto a que trabalho no recôncavo – passa a migrar para a zona de pecuária, no século XIX, período em que o semiárido baiano é mais densamente repovoado.

Carneiro e Almeida (2006) levantaram alguns vetores que nortearam a composição dos *corpora*, tanto orais quando escritos, refletindo, em linhas gerais, a dinâmica do contato linguístico ocorrido no sertão baiano. Os contextos históricos resumidos pelas autoras são:

- (i) **Séculos XVII e XVIII:** aldeamentos⁸, vilas⁹ e núcleos quilombolas¹⁰.
- (ii) **Séculos XVIII e XIX:** Criação e expansão de vilas e criação de novas vilas a partir de aldeamentos.
- (iii) **Século XX, antes e depois da década de 40:** expansão acentuada das nucleações, crescimentos de núcleos urbanos, migração acentuada para zonas urbanas¹¹.

A língua portuguesa no Brasil, notadamente no semiárido baiano, afetada por processos de transmissão linguística irregular, inicialmente na aprendizagem dos índios e, mais tarde, dos negros, foi incorporando fatos linguísticos comuns em situações de contato entre línguas. Assim, a primeira coisa que se deve ter em mente quando se fala em PB é que essa expressão não recobre uma realidade homogênea; mais ainda: deve-se ter em mente que a realidade linguística é bi-

8 Tiveram um importante papel para a posterior formação de vilas, lugares e morgados, mas a situação linguística no que diz respeito ao português ainda era muito imprecisa, havendo, pelo menos, duas línguas de contato: a do colonizador, uma ou mais línguas gerais de base indígena, e línguas indígenas.

9 São expansões agropecuárias. Com os caminhos das boiadas, o processo da agropecuária deve ter sido um fator importante para a difusão e propagação da língua portuguesa pelos sertões da Bahia. Essa difusão inicialmente deve ter-se dado por meio da boca de brancos pobres e descendentes de índios e escravos brasileiros.

10 Uma situação bastante peculiar é a dos quilombos. Surgidos a partir de grupos de negros fugidos, esses locais eram propícios para o desenvolvimento de línguas crioulas; o sertão, todavia, apresenta uma situação diferente do Recôncavo e do sul do estado, onde está localizada Helvécia, comunidade de afrodescendentes, alvo de importantes estudos de Alan Baxter e Dante Lucchesi (1994; 2001; 2002; 2003), no âmbito do projeto Vestígios de dialetos crioulos em comunidades afro-brasileiras isoladas. No semiárido, como já dito, não há notícias da existência de muitos deles e, algumas vezes, quando formados, parecem ter sido “dizimados”.

11 Os processos de industrialização, escolarização e urbanização em muito vão contribuir para firmar a dita unidade linguística brasileira.

polarizada, como bem expôs Lucchesi (1994, 2001, entre outros). Certamente, essa realidade atual é fruto das condições de formação do Brasil, com “atores” diversificados¹². Haja vista que preconceito social e linguístico correlacionam-se, as marcas características das variedades da população socialmente desprestigiada não demoraram a ser estigmatizadas; aqueles que estudaram e/ou que são favoráveis à europeização da cultura brasileira tiveram acesso às avaliações negativas sobre essas variedades.

Acreditamos que as condições sócio-históricas do PB repercutiram significativamente na estrutura da língua portuguesa, fazendo com que houvesse uma bipolarização de normas linguísticas no Brasil, com um polo que abriga as variedades ditas cultas – essas mais próximas da norma-padrão – e outro que abriga as variedades ditas populares, marcadas por um processo de extrema redução da morfologia flexional. Entendemos, igualmente, que, com as profundas e contínuas mudanças ocorridas no Brasil, a partir do século XX, tais normas se entrecruzam, influenciando-se mutuamente (LUCCHESI, 2001). A respeito disso, fica evidente que as variedades faladas pela população socialmente desprestigiada são as que mais preservam os efeitos do contato entre línguas na história sociolinguística do país; desta forma, consideramos que as pesquisas com dados do PB não devem ignorar a existência de duas histórias sociolinguísticas, realizando análises empíricas de dados com diferentes perfis, de maneira conjunta. As considerações de natureza sócio-históricas norteadoras da seleção de amostras podem evidenciar as vertentes da língua portuguesa em contato e o efeito disso ao longo do tempo.

2 Partindo do presente: o projeto ‘A língua portuguesa no Semiárido Baiano’ e as amostras de fala

Este projeto, financiado pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), começou a ser executado em 1998, tendo como objetivo geral contribuir para o estudo da sócio-história da língua portuguesa, na sua variedade falada, tendo as seguintes frentes de trabalho:

- (i) Formação de banco de dados com amostras de fala de comunidades rurais e urbanas do semiárido baiano;
- (ii) Realização de estudos sócio-históricos sobre a formação da língua portuguesa nas localidades pesquisadas;

12 Mattos e Silva (2009, p.16) destaca a incomensurabilidade da diversidade desses atores letrados e não letrados; nas palavras dela, os quais “não temos como demonstrar rigorosamente, mas que podemos inferir, aproximadamente”.

- (iii) Realização de análises linguísticas, à luz de diferentes teorias linguísticas¹³;
- (iv) Elaboração de materiais didáticos para serem utilizados nas escolas da região¹⁴.

No ano de 2007, o projeto entrou numa nova fase, denominada Fase 3, quando as atenções voltaram-se para a zona urbana de Feira de Santana/BA. Depois de delinear um quadro do português falado em comunidades rurais baianas, cabia aos pesquisadores do projeto a desafiadora e necessária tarefa de procurar traçar uma descrição sociolinguística do português falado em uma cidade tão múltipla como Feira de Santana. Iniciaram-se, então, naquele ano – novamente com o apoio do Departamento de Letras e Artes da Universidade Estadual de Feira de Santana e da FAPESB –, as gravações na sede do município, que é o segundo da Bahia, em termos populacionais, ficando atrás apenas da capital.

No que se refere à agenda (i), as gravações com amostras de fala de comunidades orais e urbanas foram divididas em 3 fases (a serem explicitadas adiante); foram feitas segundo a orientação metodológica da Sociolinguística Variacionista (WEINREICH; LABOV; HERZOG, 2006; LABOV, 1972, 1982, 1994), com gravações do tipo DID (diálogo entre informante e documentador). O critério utilizado para a realização da coleta de dados foi o de que as localidades apresentassem formações sócio-histórico-demográficas diferenciadas, fazendo parte, principalmente, do processo de urbanização do interior do Estado (ALMEIDA; CARNEIRO, 1999), que envolveu os chamados ciclos da agropecuária, cujas origens remontam ao século XVII – neste caso, Jeremoabo e Feira de Santana –, e da mineração, com origens mais densamente vinculadas ao século XVIII; neste caso, Rio de Contas e Caém.

O banco com gravações de fala do semiárido conta com mais de 90 horas de gravação em comunidades representativas da variedade falada, constituída nas regiões, cujos informantes tinham pouca ou nenhuma escolaridade; conta também com mais de 130 horas da área urbana de Feira de Santana. Essa cidade foi escolhida, tendo em vista sua característica de polo rodoviário e que agrega grande número de migrantes e de população circulante, podendo, por conseguinte,

13 Diversas análises linguísticas vêm sendo realizadas, tendo essas amostras como *corpus*, mesmo antes de sua publicação; artigos, monografias, dissertações, teses de doutorado.

14 Das frentes de trabalho expostas, a quarta é a que menos tem sido realizada até o momento. Essa agenda está prevista para ser executada a partir de 2017, com o Projeto *Gramática da Língua Falada no Semiárido Baiano*, a ser coordenado por Mariana Fagundes de Oliveira Lacerda e Silvana Silva de Farias Araújo, com o objetivo principal de fazer chegar às comunidades envolvidas nas pesquisas os resultados das investigações, oferecendo a elas materiais didáticos que subsidiem e amparem a construção de um ensino de português solidário com todas as suas variantes; a ideia de elaboração de uma Gramática da Língua Falada no Semiárido Baiano é, portanto, antiga, existe desde 1993.

contribuir para o conhecimento sobre a difusão do PB, especificamente no que diz respeito ao entrecruzamento das variedades faladas por pessoas com diferentes graus de contato com a escrita formal e do rural e urbano. Esse *corpus*, diferentemente do anterior, faz controle entre feirenses filhos de feirenses, feirenses filhos de não feirenses, migrantes, indivíduos com nível superior, médio e com pouca ou nenhuma escolarização.

As amostras rurais foram compostas em áreas passíveis de diversidade étnica, havendo indícios de que a população fosse formada por remanescentes de quilombo em Casinhas (Jeremoabo); por presença étnica de origem afro-brasileira em Barra/Bananal (Rio de Contas/Região da Chapada Diamantina), Piabas (Caém/Ancelino da Fonseca/Região do Piemonte da Diamantina) e Matinha (Feira de Santana/Região Paraguaçu); por mestiços de brancos de origem portuguesa em Lagoa do Inácio (Jeremoabo) e Mato Grosso (Rio de Contas). Isto implica a existência de variações linguísticas provenientes de contato linguístico entre línguas distintas, no processo de aquisição de língua.

- (i) **Fase 1 (1996-2000):** Amostras de fala das microrregiões da Chapada Diamantina e Piemonte da Diamantina. Foram realizadas gravações nas zonas rurais dos municípios de Caém e de Rio de Contas, zona de Mineração.
- (ii) **Fase 2 (2000-2002):** Amostras de fala das comunidades rurais da região Nordeste do estado da Bahia, zona de agropecuária. Gravaram-se inquéritos nos seguintes povoados pertencentes ao município de Jeremoabo: Lagoa do Inácio, Tapera, Casinhas e zona rural de Feira de Feira de Santana, no município da Matinha.
- (iii) **Fase 3 (2008-2014):** Amostras de fala urbana de Feira de Santana.¹⁵

Esses materiais estão também disponíveis nas seguintes coleções:

- (i) Anselino da Fonseca/Caem/Piemonte da Diamantina/ Piabas
ALMEIDA, Norma Lucia F. (Org.); CARNEIRO, Zenaide O. N. (Org.). Amostras da língua falada na zona rural de Anselino da Fonseca (Piemonte da Diamantina. 2. ed. Feira de Santana/Salvador: Editora da UEFB/Edufba, 2008. v. 1.
- (ii) Jeremoabo/Nordeste – Tapera
CARNEIRO, Zenaide O. N.; SOARES, Adriana (Org.); ALMEIDA, Norma Lucia F. (Org.) Amostras da língua falada na zona rural de Jeremoabo: Nordeste. 2. ed. Feira de Santana: UEFB, 2008. v. 1.

15 A Fase 4 do projeto, cuja ênfase seria em comunidades indígenas, foi prevista, mas ainda não foi executada. Para um trabalho nessa direção, ver Cunha e Souza (2011).

- (iii) Jeremoabo/Nordeste – Lagoa do Inácio
CARNEIRO, Zenaide O. N.; SOARES, Adriana (Org.); ALMEIDA, Norma Lucia F. (Org.). Amstras da língua falada na zona rural de Jeremoabo: Nordeste. 2. ed. Feira de Santana: UEFS, 2008. v. 1.
- (iv) Jeremoabo/Nordeste – Casinhas
CARNEIRO, Zenaide O. N.; SOARES, Adriana (Org.); ALMEIDA, Norma Lucia F. (Org.). Amstras da língua falada na zona rural de Jeremoabo: Nordeste. 2. ed. Feira de Santana: UEFS, 2008. v. 1.
- (v) Rio de Contas/Chapada Diamantina/Barra dos Negros/Bananal
CARNEIRO, Zenaide de O. N.; ALMEIDA, Norma Lucia F. de (Org.). Amstras da língua falada na zona rural de Rio de Contas. 2. ed. Feira de Santana: Editora da UEFS/Edufba, 2008. v. 1.
- (vi) Rio de Contas/Chapada Diamantina/Mato Grosso
CARNEIRO, Zenaide de O. N.; ALMEIDA, Norma Lucia F. (Org.). Amstras da língua falada na zona rural de Rio de Contas. 2. ed. Feira de Santana: Editora da UEFS/Edufba, 2008. v. 1.
- (vii) Feira de Santana/Paraguaçu/Matinha
ALMEIDA, Norma Lucia F. (Org.). Amstras da língua falada na zona rural de Feira de Santana: Paraguaçu – Matinha. 2. ed. Feira de Santana: UEFS, 2008. v. 1.
- (viii) Feira de Santana/Paraguaçu/Feira de Santana (sede): amostras de altamente escolarizados (cultos) e não escolarizados feirenses filhos de feirenses, feirenses filhos de migrantes, migrantes (ARAÚJO, 2014).
Amstras da língua falada na zona urbana de Feira de Santana: Paraguaçu – Feira de Santana (Sede). Inédito.

As responsáveis pela organização do *corpus* da atual fase/Feira de Santana, ainda inédito, foram as professoras Norma Lucia Fernandes de Almeida, Eliana Pitombo Teixeira, Silvana Silva de Farias Araújo e Zenaide de Oliveira Novais Carneiro.

A Figura 1 mostra a localização das sedes dos municípios em que se encontram as comunidades onde foram feitas as gravações na década de 90 do século XX: Chapada Diamantina/Rio de Contas/Mineração (Mato Grosso e Barra-Bananal); Piemonte da Diamantina/Caém/Pecuária (Piabas), Nordeste/Jeremoaba/Pecuária (Lagoa do Inácio, Tapera e Casinhas); Paraguaçu/Feira de Santana/rural (Matinha). Na primeira década do século XXI, Paraguaçu/Feira de Santana/urbana (60 informantes de fala urbana: 48 representantes da norma falada por população pouco escolarizada e 12 da população altamente escolarizada).

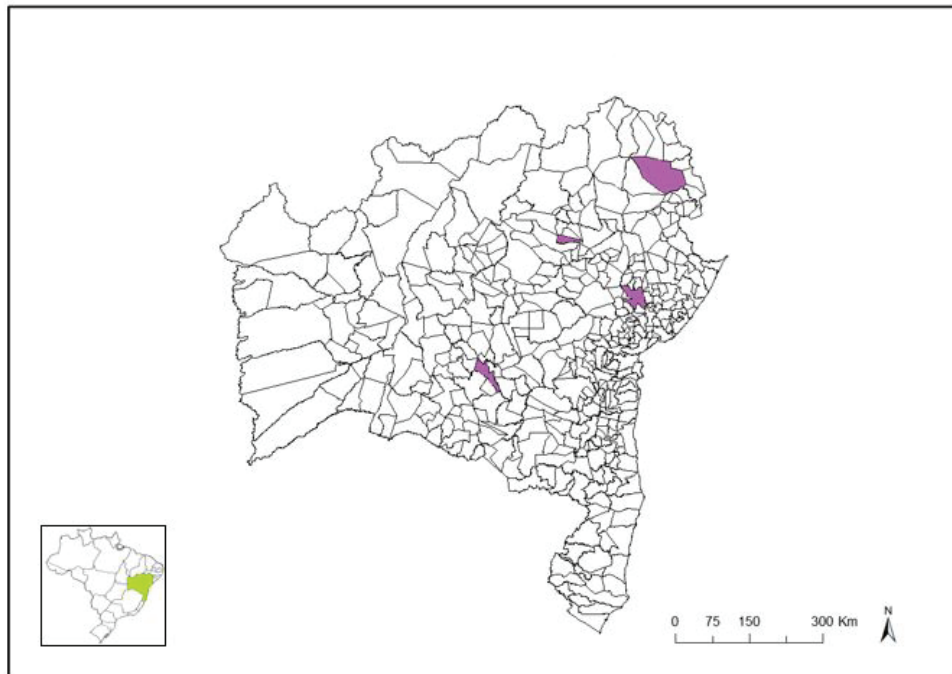


Figura 1 Localização das sedes dos municípios onde foram gravadas as amostras de fala.

Fonte: Adaptado de www.clubdageografia.tk.

3 Em direção ao passado: o projeto 'Vozes do sertão em dados' e os textos escritos

O Projeto Vozes do Sertão em Dados: história, povos e formação do português brasileiro, financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), executado entre julho de 2009 e julho de 2011 (<http://www.uefs.br/nelp/zenaide-nelp/index.html>), foi também organizado em fases.

A Fase I teve como propósito fundamental organizar os resultados de pesquisas desenvolvidas no sertão baiano, entre 1997 e 2009, em três campos de investigação, previstos no PHPB, a saber:

- i) O campo histórico-filológico – voltado à constituição de *corpora* diacrônicos de documentos de natureza vária, escritos no Brasil, a partir do século XV;
- ii) O campo gramatical – voltado ao estudo de mudanças linguísticas depreendidas na análise dos *corpora* constituídos;
- iii) O campo da história social linguística – voltado à reconstrução mais ampla da história social linguística do Brasil e, em particular, do PB.

O projeto Vozes do Sertão em Dados colabora com o PHPB, por meio da prospecção, da edição de documentos e da formação de *corpora* representativos de demandas histórico-sociais da região semiárida baiana – com repercussões sobre o processo de formação histórica do PB, com amplo contato linguístico de populações de origem portuguesa, indígena e africana –, bem como com projetos temáticos de análise linguística.

Produto de pesquisas desenvolvidas no campo histórico-filológico, *Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português* (CARNEIRO, 2012) está organizada em três volumes: a apresentação de cada volume e a descrição dos acervos estão em formato impresso, e as edições fac-similadas das cartas (além das fichas dos remetentes) acompanham os livros em formato de CD-ROM, na forma como segue:

(i) VOLUME 1 (1809-1904)¹⁶

Zenaide de Oliveira Novais Carneiro (Organização)

CD-ROM 1. Cartas para vários destinatários (1809-1904): edição fac-similada/Zenaide de Oliveira Novais Carneiro

CD-ROM 2. Cartas para Severino Vieira, governador da Bahia (1901-1902): edição fac-similada/Zenaide de Oliveira Novais Carneiro

CD-ROM 3. Cartas para Cícero Dantas Martins, Barão de Jeremoabo (1880-1903): edição fac-similada/Zenaide de Oliveira Novais Carneiro

(ii) VOLUME 2 (1902-1993)

Zenaide de Oliveira Novais Carneiro; Mariana Fagundes de Oliveira; Norma Lucia Fernandes de Almeida (Organização)

CD-ROM 1. Cartas do acervo Dantas Jr. (1902-1962): edição fac-similada/ Zenaide de Oliveira Novais Carneiro; Marta Carvalho Ferreira Lisboa; Mônica Araújo Cruz; Denise Branco Cerqueira; Eliane Santos Leite

CD-ROM 2. Cartas baianas: o acervo de João da Costa Pinto Victoria (1911-1958): edição fac-similada/ Zenaide de Oliveira Novais Carneiro; Maria Rosane Passos; Priscila Tuy Batista; Anderléia Mascarenhas

CD-ROM 3. Correspondências amigas: o acervo de Valente, Bahia (1980-1993): edição fac-similada/Mariana Fagundes de Oliveira; Maiany Soares de Oliveira; Adilson Silva de Jesus

(iii) VOLUME 3 (1906-2000)

Huda da Silva Santiago; Zenaide de Oliveira Novais Carneiro; Klebson Oliveira (Organização)

CD-ROM 1. Cartas em Sisal: Riachão do Jacuípe, Conceição do Coité e Ichu (1906-2000): edição fac-similada/Huda da Silva Santiago.

16 Volume 2 da tese de Carneiro (2005).

Com relação aos remetentes das cartas, pode-se observar, na Figura 2, a distribuição por grau de escolaridade:

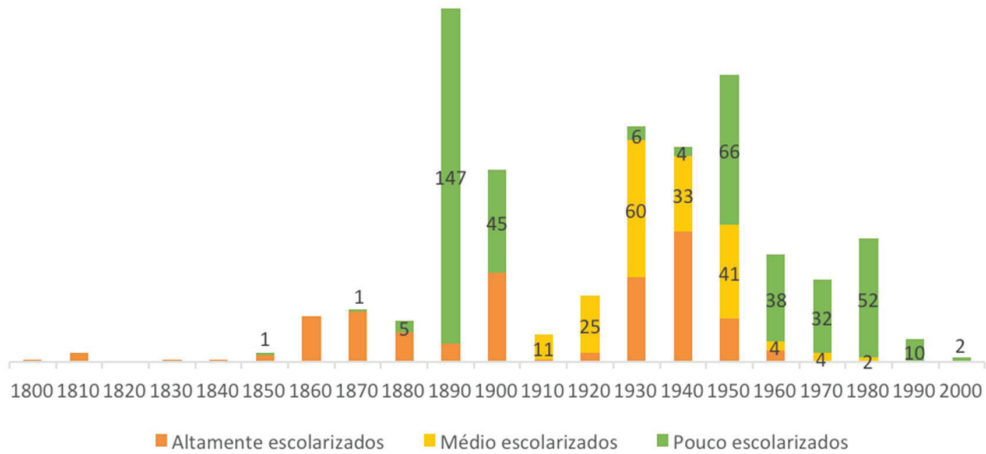
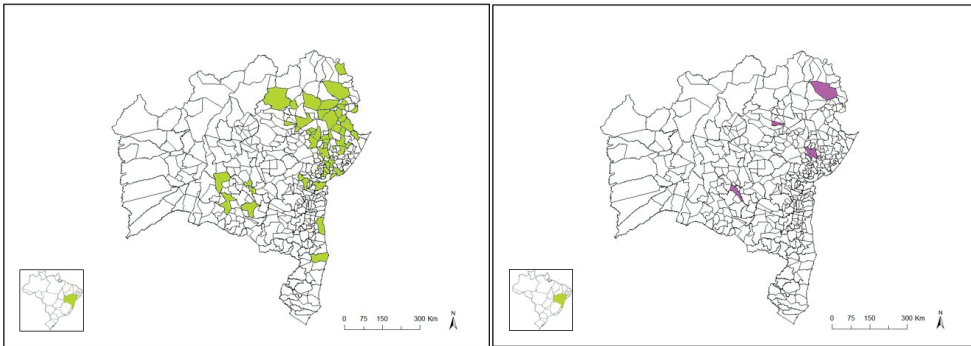


Figura 2 Distribuição das cartas baianas por década de escrita e por escolaridade.

Fonte: Elaborado pelos autores.

Olhando as amostras, tanto escritas quanto orais, por local de nascimento dos autores/informantes, podemos verificar que abrangem grande parte da Bahia, como se pode observar nas Figuras 3 e 4, respectivamente.



Figuras 3 e 4 Distribuição das amostras baianas, dos séculos XIX e XX (década de 90), escritas e orais, respectivamente.

Fonte: adaptado de www.clubdageografia.tk.

O trabalho com os acervos continua a crescer; recentemente, incorporou-se ao banco o Acervo da Família Estrela Tuy (1930-1980), composto por 107 cartas pessoais, escritas durante o século XX, entre as décadas de 1930 e 1980, por

remetentes baianos de origem interiorana e rural, em sua maioria com baixa escolaridade. As cartas são oriundas da Bahia, de pequenos municípios e de fazendas pertencentes às imediações desses municípios, e também de Salvador¹⁷.

Há que se mencionar a edição, em curso, de documentos dos séculos XVIII e XIX: O Livro do Gado e o Livro de Razão¹⁸. Esses dois livros manuscritos do final do período colonial que se conservaram no arquivo do Sobrado do Brejo, da família Pinheiro Canguçu, na fazenda de criação do Brejo do Campo Seco, no povoado de Bom Jesus dos Meiras – que pertenceu à Comarca de Rio de Contas –, hoje denominado Brumado, na região da Serra Geral, no sertão baiano – nos chegam às mãos por meio da leitura e análise de Santos Filho (1956), segundo o qual “Já no século XVIII os criadores do Nordeste baiano possuíam livros para seus próprios assentamentos.” (p. 109) Esses livros eram uma prática cultural de escrita naquela época, nas fazendas. Segundo o autor (p. 109), “talvez alguns livros ainda existam, encafuados em arquivos de família nas demais propriedades da região.”

Essa ampliação do *corpus* “favorece essencialmente uma Linguística descritiva, fortemente apoiada pelas novas tecnologias, e permite tomar como ponto de partida da descrição a análise de quantidade significativa de dados autênticos, à semelhança do que se faz noutros domínios científicos. O uso de *corpora* permite a realização de descrições linguísticas de base empírica e promove, com isso, a discussão de questões teóricas solidamente fundamentadas” (BACELAR DO NASCIMENTO, 2004, p. 1).

Todos os materiais aqui referidos – orais e escritos – estão disponíveis, em versão eletrônica, no CE-DOHS – *Corpus* Eletrônico de Documentos Históricos do Sertão, coordenado por Zenaide de Oliveira Novais Carneiro e Mariana Fagundes de Oliveira Lacerda, o qual, por meio de parceria tecnológica com o projeto

17 O acervo está organizado da seguinte maneira: Cartas de Amor - trata-se de 29 cartas trocadas entre Antonio Carneiro da Silva Tuy e Maria de Souza Estrela [Maria Estrela Tuy]; dessas, 24 foram trocadas durante seu noivado, escritas entre 1952 e 1953, e 5 cartas trocadas entre 1956 e 1968, depois de casados; Cartas para Antonio e Maria - são 52 cartas enviadas a Antonio Carneiro da Silva Tuy e Maria Estrela Tuy por familiares e amigos; Cartas para outros destinatários - são 26 cartas e rascunhos de cartas escritas por Antônio Carneiro da Silva Tuy, Maria Estrela Tuy e outros familiares para amigos e demais entes da família. Esse acervo também é composto por 15 bilhetes e 2 cartões trocados entre os mesmos remetentes. Os documentos foram depositados ao longo dos anos em um baú construído pelo próprio Antonio Carneiro da Silva Tuy, durante sua juventude, na Fazenda Bom Jardim, por volta de 1940. Após seu casamento com Maria de Souza Estrela em 1953, o baú foi transportada para a fazenda Mucambo, no município de Biritinga/BA, onde passaram a residir. Os documentos estão, em sua maioria, em um bom estado de conservação, graças ao armazenamento, que, apesar de não ser o ideal, minimizou o desgaste temporal.

18 Ver Lacerda e Carneiro (2016).

Corpus Histórico do Português Tycho Brahe (www.tycho.iel.unicamp.br), traz um conjunto de documentos originados sobretudo da grande área do semiárido baiano, editados em linguagem XML, com o uso do eDictor, desenvolvido por Paixão de Sousa, Kepler e Faria (2009; 2010), um editor de textos especialmente voltado ao trabalho filológico e à análise linguística automática, o qual combina um editor de XML e um etiquetador morfossintático e permite a geração automática de versões correspondentes a edições diplomáticas, semidiplomáticas e modernizadas (em HTML), e de versões com anotação morfossintática (em texto simples e XML). O banco está sendo preparado para a anotação morfossintática, que manterá a maioria das características do padrão de anotação existente e permitirá a busca automática de dados, o que facilitará o estudo linguístico dos acervos, no que consiste o principal objetivo do CE-DOHS. O material disponível no banco atende, entretanto, não somente a pesquisadores interessados em análises de aspectos linguísticos, mas em aspectos da difusão da escrita, da leitura, das transmissões textuais, históricas, políticas, econômico-sociais, entre outras.

Considerações finais

Os materiais orais e escritos foram organizados de maneira a permitir que pesquisadores interessados na história do PB possam constituir *corpora* específicos de acordo com seus objetivos; cada pesquisador, pode, para tanto, valer-se dos metadados disponíveis nas coletâneas e *sites*.

Do ponto de vista da documentação, é possível a sua organização por ordem cronológica, tanto por data de escrita – como é tradicional na Linguística Histórica –, quanto por data de nascimento do autor (individual) – no caso de autores que possuem acervos mais significativos –, ou por grupos de autores.

A base documental aqui apresentada pode ajudar nos estudos sobre variação e mudança do PB; no que concerne à documentação epistolar, estudos do século XIX ao século XX, se se considerar a data de escrita das cartas; a partir do século XVIII, se a cronologia for feita por data de nascimento de seus remetentes. As informações sobre o grau de escolaridade dos escreventes, aliadas a possíveis análises do tipo de escrita, permitem também a separação das cartas representativas de variedades *standard* e *não-standard* do PB, em uma perspectiva histórica.

É importante salientar que as vertentes culta e popular do PB – constituídas ao longo do período colonial brasileiro – têm apresentado, atualmente, na língua falada, uma tendência à aproximação; por exemplo, quanto à colocação dos clíticos, prevalece a próclise nas duas vertentes; na escrita de altamente escolarizados, todavia, há, ainda, modulações, como, por exemplo, a manutenção do uso da ênclise do português moderno. Essas modulações decorrem do fato de fala e escrita constituírem-se como modalidades distintas.

Não se verificaram, nas amostras coletadas no semiárido baiano – à época ainda com pouco contato com variedades urbanas –, diferenças significativas em relação ao que se observa em outras regiões rurais do Nordeste brasileiro; se existiram, essas diferenças foram dissipadas no tempo. Considerar o conceito de *continuum* dialetal nos parece ser o mais adequado no tratamento dessa realidade.

Referências

ALMEIDA, Norma Lucia Fernandes de; CARNEIRO, Zenaide de Oliveira Novais (Org.). *Varição linguística no semiárido baiano*. Feira de Santana: UEFS Editora, 2014.

_____. *Coleção amostra da língua falada no semi-árido baiano*. Feira de Santana: UEFS EDITORA, FAPESB, 2008. v. 4. 450 p.

_____. *Amostras da língua falada na zona rural de Anselino da Fonseca: Piemonte da Diamantina*. 2. ed. Feira de Santana/Salvador: Editora da UEFS/ Edufba, 2008. v. 1.

ALMEIDA, Norma Lucia Fernandes de (Org.). *Amostras da língua falada na zona rural de Feira de Santana: Paraguaçu: Matinha*. 2. ed. Feira de Santana: UEFS, 2008. v. 1.

ARAUJO, Silvana Silva de Farias. *A concordância verbal no português falado em Feira de Santana-Ba: sociolinguística e sócio-história do português brasileiro*. 2014. 342 f. Tese (Doutorado em Língua e Cultura) – Universidade Federal da Bahia, Salvador, 2014.

ANDRADE, Aroldo Leal de; CARNEIRO, Zenaide de Oliveira Novais. A posição e a colocação de clíticos em predicados complexos: o português brasileiro visto a partir de duas vertentes. In: *Revista de Filologia e Linguística Portuguesa*. 16 (número especial). São Paulo: USP, 2014. p. 125-161.

BACELAR DO NASCIMENTO, Maria Fernanda. O lugar do *corpus* na investigação linguística. In: MENDES, A. et alli (Org.). *Actas do XVIII Encontro da Associação Portuguesa de Linguística*. Lisboa, Associação Portuguesa de Linguística e Edições Colibri, 2002.

CARNEIRO, Zenaide de Oliveira Novais. *Cartas Brasileiras (1809-1904): um estudo linguístico-filológico*. Feira de Santana: UEFS Editora, 2016. 410p.

- _____; OLIVEIRA, Mariana Fagundes de (Org.). *Publica-se em Feira de Santana (1908-2006)*. 1. ed. Feira de Santana: UEFS Editora, 2012. v 1. 55p.
- _____; _____ (Org.). Cartas de Leitores e Redatores em o Progresso e na Folha do Norte (1908-1997). In. *Publica-se em Feira de Santana (1908-2006)*, CD-ROM 1. 1. ed. Feira de Santana: UEFS Editora, 2012. v. 1. 221p.
- _____; _____ (Org.). Anúncios na Folha do Norte (1910-2006). In. *Publica-se em Feira de Santana (1908-2006)*, CD-ROM 2. 1. ed. Feira de Santana: UEFS Editora, 2012. v. 1. 224p.
- _____. *Cartas brasileiras (1809-2000)*: coletânea de fontes para o estudo do português. Feira de Santana: UEFS Editora, 2011. v 1.
- _____. CD-ROM 1. Cartas para vários destinatários (1809-1904): edição fac-similada. In: *Cartas brasileiras (1809-2000)*: coletânea de fontes para o estudo do português. Feira de Santana: UEFS Editora, 2011. v. 1. 549p.
- _____. Cartas para Severino Vieira, governador da Bahia (1901-1902): edição fac-similada. In: *Cartas brasileiras (1809-2000)*: coletânea de fontes para o estudo do português, CD-ROM 2. Feira de Santana: UEFS Editora, 2011. v. 1.
- _____. Cartas para Cícero Dantas Martins, Barão de Jeremoabo (1880-1903): edição fac-similada. In. *Cartas brasileiras (1809-2000)*: coletânea de fontes para o estudo do português, CD-ROM 3. Feira de Santana: UEFS Editora, 2011. v. 1. 755p.
- _____; OLIVEIRA, Mariana Fagundes de; ALMEIDA, Norma Lúcia Fernandes (Org.). *Cartas brasileiras (1809-2000)*: coletânea de fontes para o estudo do português. Feira de Santana: UEFS Editora, 2011. v 2.
- _____; LISBOA, Marta Carvalho Ferreira; CRUZ, Mônica Araújo; CERQUEIRA, Denise Branco; LEITE, Eliane Santos (Org.). Cartas do acervo Dantas Jr. (1902-1962): edição fac-similada. In: *Cartas brasileiras (1809-2000)*: coletânea de fontes para o estudo do português, CD-ROM 1. Feira de Santana: UEFS Editora, 2011. v. 2. 437p.
- _____; SANTOS, Maria Rosane Passos dos; BATISTA, Priscila Starline Estrela Tuy; MASCARENHAS, Anderléia de Jesus. *Cartas baianas (1911-1958)*: o

acervo do João da Costa Pinto Victorio. In: *Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português*, CD-ROM 2. Feira de Santana: UEFS Editora, 2011. v. 2. 301p.

_____; SOARES, Adriana; ALMEIDA, Norma Lucia Fernandes de (Org.) . *Amostras da língua falada na zona rural de Jeremoabo: Nordeste*. 2. ed. Feira de Santana: UEFS, 2008. v. 1.

_____; ALMEIDA, Norma Lucia F. de (Orgs.). *Amostras da língua falada na zona rural de Rio de Contas*. 2. ed. Feira de Santana: Editora da UEFS/Edufba, 2008. v. 1.

_____; _____. Elementos para uma sócio-história do semi-árido baiano. In: Jânia M. Ramos; Mônica A. Alkmim. (Org.). *Para a História do Português Brasileiro: volume 5, estudos sobre mudança linguística e história social*. 1ª ed. Belo Horizonte: Faculdade de Letras da UFMG, 2007. v. 1. p. 423-442.

_____; _____. A criação de escolas a partir de critérios demográficos na Bahia do século XIX: uma viagem ao interior. In: LOBO, Tânia; RIBEIRO, Ilza; CARNEIRO, Zenaide; ALMEIDA, Norma Lúcia (Org.). *Para a história do português brasileiro: novos dados, novas análises*. Salvador: Edufba, 2, 2006. v. 2. p. 649-674.

_____. *Cartas brasileiras (1808-1904): um estudo linguístico-filológico*. 2005. 4v. 2.329f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, São Paulo, 2005.

CE-DOHS. *Corpus eletrônico de documentos históricos do sertão*. Disponível em: <www.uefs.br/cedohs>. Acesso em: 10 ago. 2016.

Corpus Histórico do Português Tycho Brahe. Disponível em: <<http://www.tycho.iel.unicamp.br/~tycho/corpus/>>. Acesso em: 10 ago.2016.

CUNHA e SOUZA, Hirão Fernandes. *O português kiriri: aspectos fônicos e lexicais na fala de uma comunidade do sertão baiano*. 2011. 1 v. 200f. Dissertação (Mestrado em Letras e Linguística) - Universidade Federal da Bahia, Salvador.

FARIA, Pablo Picasso Feliciano de; KEPLER, Fabio Natanael; PAIXÃO DE SOUSA, Maria Clara. An Integrated Tool for Annotating Historical Corpora. In: *Fourth Linguistic Annotation Workshop (LAW IV), 48th Annual Meeting*

of the ACL, 2010, Uppsala, Sweden. Proceedings of the Fourth Linguistic Annotation Workshop, 2010. p. 217-221.

FREIRE, Felisbello. *História territorial do Brasil* - Edição fac-similar. Salvador: Secretaria de Cultura e Turismo, Instituto Geográfico e Histórico da Bahia, 1998.

KROCH, Anthony. Morphosyntactic variation. In: BEALS, K. *et al.* (Eds.). *Papers from the 30th regional meeting of the Chicago linguistics society: parasession on variation and linguistic theory*, 1994. v. 2. p. 180-201.

KROCH, Anthony. Syntactic change. In: BALTIN, M.; COLLINS, C. (eds.). *The handbook of contemporary syntactic theory*, Oxford: Blackwell Publishers Inc., 2001. p. 699-729.

LABOV, William. Building on empirical foundation. In: LEHAMANN, W. P.; MAKIEL, Y (eds.). *Perspectives on historical linguistics*. Philadelphia: John Benjamins Publishing Company, 1982.

_____. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.

_____. *Principles of linguistics: internal factors*. Cambridge: Blacwel, 1994. v.1.

LACERDA, Mariana Fagundes de; CARNEIRO, Zenaide de Oliveira Novais. Edição filológica e digital do Livro do Gado e do Livro de Razão do arquivo do Sobrado do Brejo (Bahia setecentista e oitocentista). *LaborHistórico*, n. 2, v. 1, p. 151-163, 2016.

LOBO, Tânia. *Para uma sociolinguística histórica do português do Brasil*: edição filológica e análise linguística de cartas particulares do recôncavo da Bahia, século XIX. 2001. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. São Paulo.

LUCCHESI, Dante. As duas grandes vertentes da história sociolinguística do Brasil (1500-2000). *DELTA*. v.17, n.1, p. 97-132, 2001.

_____. Norma linguística e realidade social. In: BAGNO, Marcos (Org.). *Linguística da norma*. São Paulo: Edições Loyola, 2002. p. 63-92.

_____. O conceito de transmissão linguística irregular e o processo de formação do português do Brasil. In: Roncarati, Cláudia; Abraçado, Jussara (Org.). *Português brasileiro: contato linguístico, heterogeneidade e história*. Rio de Janeiro: 7 Letras, 2003. p. 272-284.

_____. Variação e norma: elementos para uma caracterização sociolinguística do português do Brasil. *Revista Internacional de Língua Portuguesa*, n.12, p. 17-28, 1994.

MARQUILHAS, R. *A faculdade das letras: leitura e escrita em Portugal no séc. XVII*. Lisboa: Imprensa Nacional-Casa da Moeda, 2000.

MATTOS E SILVA, Rosa Virgínia. *Para a história do português culto e popular brasileiro: sugestão para uma pauta de pesquisa*. Cadernos de Letras da UFF, v. 34. p. 11-30, 2009.

_____. *Caminhos da linguística histórica: ouvir o inaudível*. São Paulo: Parábola Editorial, 2008.

_____. A generalizada difusão da língua portuguesa no território brasileiro. In: MATTOS E SILVA, Rosa Virgínia. *Ensaio para uma sócio-história do português brasileiro*. São Paulo: Parábola Editorial, 2004, p. 91-108.

_____. Variação, mudança e norma (movimentos no interior do português brasileiro). In: BAGNO, Marcos (Org.). *Linguística da norma*. São Paulo: Edições Loyola, 2002. p. 291-316.

_____. De fontes sócio-históricas para a história social linguística do Brasil: em busca de indícios. In: MATTOS E SILVA, R. V. (Org.). *Para a história do português brasileiro: primeiros estudos*. V. II, tomos I e II. São Paulo: Humanitas/FFCHL/USP: FAPESP, v.2, t. 2, 2001, p. 275-302.

PAIXÃO DE SOUSA, Maria Clara; KEPLER, Fabio Natanael; FARIA, Pablo Picasso Feliciano de. E-Dictor: Novas perspectivas na codificação e edição de *corpora* de textos históricos. In: SHEPHERD, Tania; SARDINHA, Tony Berber; PINTO, Marcia Veirano (Org.). *Caminhos da linguística de corpus*. Campinas: Mercado de Letras, 2010.

_____. E-dictor: Novas perspectivas na codificação e edição de *corpora* de textos históricos. In: VIII Encontro de Linguística de *Corpus*, 2009. Rio de Ja-

neiro, *Anais do VIII Encontro de Linguística de Corpus*. Rio de Janeiro: UERJ, 2009. p. 69-105.

SANTIAGO, Huda da Silva. Cartas em Sisal: Riachão do Jacuípe, Conceição do Coité e Ichu (1906-2000): edição fac-similada. In. _____; CARNEIRO, Zenaide de Oliveira Novais; OLIVEIRA, Klebson (Org.). *Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português*, CD-ROM 1. Volume 3. Feira de Santana: UEFS Editora, 2011. v. 3.

_____. *Um estudo do português popular brasileiro em cartas pessoais de “mãos cândidas” do sertão baiano*. 2012. 2v. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-graduação em Estudos Linguísticos, Universidade Estadual de Feira de Santana, Feira de Santana, 2012.

SANTOS FILHO, Lycurgo. *Uma comunidade rural do Brasil antigo: aspectos da vida patriarcal no Sertão da Bahia nos séculos XVIII e XIX*. São Paulo: Nacional, 1956.

SANTOS, M. R. A. dos. *Fronteiras do sertão baiano: 1640-1750*. 2010. 430 p. Tese (Doutorado em História). Faculdade de Filosofia, Letras e Ciências Humanas – Universidade de São Paulo, São Paulo, 2010.

SILVA, Sylvio C. Bandeira de Mello et al. *Urbanização e metropolização no estado da Bahia: evolução e dinâmica*. Salvador: Centro Editorial e Didático da UFBA, 1989.

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA (SEI). Disponível em <<http://www.sei.ba.gov.br>>. Acesso em: 23 nov. 2011.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. *Fundamentos empíricos para uma teoria da mudança linguística*. Tradução de Marcos Bagno. São Paulo: Parábola Editorial, 2006.