

## Para um estudo da formação do português brasileiro: descrição, representatividade e potencialidades do corpus colonial do CE-DOHS

For a study of Brazilian Portuguese formation: description, representativeness and potentialities of the CEDOHS colonial corpus

DOI: <https://doi.org/10.24206/lh.v7iespec.41843>

*Lara da Silva Cardoso*

Mestre em Estudos Linguísticos pela Universidade Estadual de Feira de Santana (UEFS). Professora de Língua Portuguesa da Secretaria de Educação da Bahia (SEC). Integrante do projeto Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS).

E-mail: [laracardooso@hotmail.com](mailto:laracardooso@hotmail.com)

ORCID: <https://orcid.org/0000-0003-1149-2007>

*Zenaide de Oliveira Novais Carneiro*

Doutora em Linguística pela Universidade Estadual de Campinas (Unicamp). Professora Plena da Universidade Estadual de Feira de Santana. Coordenadora do projeto Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS).

E-mail: [zoncarneiro@uefs.br](mailto:zoncarneiro@uefs.br)

ORCID: <https://orcid.org/0000-0001-5990-4854>

*Mariana Fagundes de Oliveira Lacerda*

Doutora em Linguística pela Universidade Federal da Bahia (UFBA). Professora Titular da Universidade Estadual de Feira de Santana. Coordenadora do Núcleo de Estudos de Língua Portuguesa (NELP) e co-coordenadora do projeto Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS).

E-mail: [marianafag@gmail.com](mailto:marianafag@gmail.com)

ORCID: <https://orcid.org/0000-0003-4335-3458>

## RESUMO

Este artigo tem por objetivo apresentar o *corpus* colonial formado pelo CE-DOHS (Corpus Eletrônico de Documentos Históricos do Sertão). O CE-DOHS, ao longo de sete anos (2012-2018), elaborou um *corpus* diacrônico dos séculos XIX e XX, oferecendo dados em formato eletrônico para o estudo linguístico do português brasileiro (PB). A partir do *corpus* colonial, o CE-DOHS busca oferecer aos pesquisadores da história do Português Brasileiro um pequeno conjunto de textos representativos do cenário multilíngue e multiétnico do Brasil Colônia, mais especificamente dos séculos XVII e XVIII que, somado aos outros *corpora*, oferece, em um único banco, quatro séculos de dados linguísticos do português em solo brasileiro. O *corpus* em questão contém 151 documentos, escritos por 16 escreventes, todos nascidos no Brasil, e possui mais de dez gêneros textuais diferentes, que foram agrupados em pequenos acervos, conforme as etnias dos escreventes e os possíveis processos de aquisição da língua portuguesa. O objetivo da constituição do *corpus* é investigar a difusão do português na América e a dinâmica dessa língua no processo de surgimento do PB. Neste trabalho, serão enfatizadas: as estratégias para o encontro dos manuscritos, uma vez que a delimitação temporal estabelecida para o *corpus* é, ainda, obscura dentro dos estudos sobre o PB; a representatividade dos manuscritos, mostrando que, apesar de em pequena quantidade, são significativos; as potencialidades do material organizado, indicando alguns dos possíveis trabalhos linguísticos possíveis de serem realizados, explorando o *corpus* em diferentes áreas e/ou constituindo *corpora* maiores, capazes de fornecer mais dados sócio-históricos e linguísticos.

**Palavras-chave:** Linguística Histórica. Constituição de corpus. Português Brasileiro. CE-DOHS. Brasil Colonial.

## ABSTRACT

This article is intended to present the colonial corpus formed by CE-DOHS. This group, over a period of seven years, drew up a diachronic corpus two XIX and XX secles, offering data in electronic format for the Brazilian Portuguese language study. Starting from the colonial corpus, or CE-DOHS seeks to offer researchers of the Brazilian Portuguese history a small set of representative texts of the multilingual and multi-ethnic group of Brazil Colônia, more specifically two seventeenth and eighteenth centuries that, based on years of other corporations, offer, In a single bank, four secles

of Portuguese linguistic data in only Brazilian. The corpus contains 151 documents, written by 16 skeptics, all born in Brazil, and has more than ten different textual genres, which are grouped in small collections, according to the two skeptical ethnic groups and the possible Portuguese language acquisition processes. O objective of the constitution of the corpus and to investigate the diffusion of Portuguese in America and the dynamics of the language without the process of emergence of PB. It will be emphasized: as strategies for or found two manuscripts, one time that a temporal delimitation established for the corpus and, furthermore, obscure within two studies on or PB; the representation of two manuscripts, showing that, despite their small quantity, they are significant; The potentialities of organized material, indicating some two possible linguistic works that may be carried out, exploring or corpus in different areas and constituting major corpora, capable of providing more socio-historical and linguistic data.

**Keywords:** Historical Linguistics. Corpus composition. Brazilian Portuguese. CE-DOHS. Colonial Brazil.

## Introdução

A compreensão do percurso histórico da formação do português brasileiro (PB) é o foco dos estudos do CE-DOHS - Corpus Eletrônico de Documentos Históricos do Sertão<sup>1</sup> - a partir de três agendas principais: a formação de banco de textos de língua portuguesa, o estudo sócio-histórico e o estudo linguístico (CARNEIRO; LACERDA, 2019). A respeito da primeira agenda, a constituição de um banco de textos, ou de um *corpus*, o CE-DOHS, durante sua primeira etapa (2012-2018), constituiu *corpora* escrito e oral dos séculos XIX e XX, abarcando entrevistas, manuscritos e impressos e privilegiando a escrita de brasileiros natos que mostrasse, direta ou indiretamente, o português falado no sertão baiano. São mais de mil cartas, editadas<sup>2</sup>, datadas e localizadas, representativas de variedades prestigiadas e não prestigiadas do português brasileiro, alcançando, ainda, a escrita de mãos inábeis do século XX (cf. LACERDA; CARNEIRO; SANTIAGO, 2016).

A segunda etapa do CE-DOHS, iniciada em 2019, recua ainda mais no tempo e se concentra no cenário multilíngue generalizado do Brasil Colonial, mais especificamente nos séculos XVII e XVIII. Propõe-se, nessa fase, a investigação do português **do** Brasil e não do português **no** Brasil (BARBOSA, 1999, p. 14), buscando a língua portuguesa aprendida e usada pelos colonos brasileiros, isto é, pelos indivíduos nascidos na colônia após a invasão europeia. Esse objetivo torna-se relevante porque os bancos de textos do período colonial brasileiro, organizados até então, dão conta de apenas um dos dois grupos sociolinguísticos existentes, o de portugueses fixados no Brasil, ou seja, de portugueses que tiveram sua aquisição de língua na Europa e desembarcaram no Brasil depois de concluído o processo de aprendizagem da língua materna (BARBOSA, 2005). Para esse objetivo, formou-se um *corpus*, composto por 151 documentos, com 16 gêneros textuais diferentes, escritos por 16 escreventes de 3 etnias (mulatos, mamelucos e brancos), todos nascidos no Brasil, alcançando mais de cem mil palavras. A proposta de investigação do português da América Portuguesa se estende à tentativa de delinear o percurso histórico das vertentes do português brasileiro (PB), sobretudo a formação da vertente popular, proveniente dos sertões, cuja extensão territorial, naquela época, era maior do que a atual, além de ser marcada como um local de trânsito de sertanistas mamelucos, mulatos e de povos indígenas que ainda conseguiam resistir à invasão europeia.

---

<sup>1</sup> O CE-DOHS é coordenado pelas professoras Zenaide de Oliveira Novais Carneiro (UEFS) e Mariana Fagundes de Oliveira Lacerda (UEFS). Cf. [www.uefs.br/cedohs](http://www.uefs.br/cedohs).

<sup>2</sup> A edição dos documentos foi realizada conforme as regras de transcrição propostas pelo PHPB. A edição eletrônica, anotação e disponibilização do material editado foi realizada conforme padrões de outros projetos de *corpora* eletrônicos, como o Corpus Tycho Brahe (Unicamp).

Para a execução dessa tarefa, se fez necessário levantar os debates e as discussões sobre a constituição de *corpus*, promovidos pela Linguística de Corpus, associando-os às proposições específicas de constituição de *corpora* históricos do português e, sobretudo, do período colonial brasileiro, levantadas inicialmente pelo Projeto Para História do Português Brasileiro (PHPB). Aqui serão discutidos os procedimentos adotados pelo CE-DOHS para a constituição do *corpus* colonial, a descrição quantitativa e qualitativa dos materiais obtidos, bem como a representatividade e as suas potencialidades dentro dos estudos linguísticos, demonstrando que, apesar de pequeno, contempla, na medida do possível, a diversidade étnica existente no Brasil Colônia.

## 1. Sobre a importância de *corpora* para o estudo do português no período colonial brasileiro

Para Mattos e Silva (2004), a formação polarizada do português brasileiro, com as vertentes culta e popular, é oriunda dos diferentes antecedentes históricos que influenciaram a formação do PB: enquanto a vertente culta, que contempla as variedades prestigiadas, teve como antecedente histórico o português europeu, a vertente popular teve como atores fundamentais as línguas gerais indígenas e o português geral brasileiro, usado, em massa, pelos africanos e seus descendentes, os quais eram obrigados a aprender, na oralidade e como língua segunda e língua primeira, respectivamente, o português em um processo de transmissão linguística irregular, a partir do *input* dos colonos portugueses que obtinham mais contato, os da baixa classe, falantes, segundo Lucchesi (2006), da “fala rude”. Portanto, o período colonial tem forte importância para a formação do PB, uma vez que, a chegada desses diferentes povos e sua adaptação ao contexto social brasileiro foram movimentos ocorridos durante esse período, marcando a colônia brasileira com um multilinguismo generalizado (MATTOS E SILVA, 2004).

O processo de formação do PB aconteceu junto à formação de outras gramáticas, como as de base em línguas indígenas, as línguas gerais, e as de base no português rudimentar, como o português geral. A língua portuguesa foi, por mais de dois séculos, preterida, frente às demais línguas generalizadas no contexto colonial. Esse multilinguismo, em meados do século XVIII, foi visto como negativo para a integração da colônia com a metrópole, fazendo com que Marquês de Pombal decretasse, em 1758, o uso e ensino obrigatório da língua portuguesa. É nesse terceiro quartel do século XVIII que Mattos e Silva (2004) considera o início da virada linguística em favor do português, fazendo com que o Brasil alcançasse, a passos lentos, o multilinguismo localizado dos dias atuais.

Além disso, foi no período colonial que houve as primeiras movimentações que transformaram a dinamização da difusão do português brasileiro. A economia voltada aos engenhos de açúcar, nos

primeiros séculos após a invasão portuguesa, permitiu uma grandiosa concentração populacional no Nordeste. A invasão dos sertões, em busca de escravos e/ou terras, apesar de ter sido tardia em relação à invasão da costa, contribuiu para que o português começasse a se difundir no interior brasileiro. Em seguida, a busca pelo ouro levou milhares de portugueses e africanos ao sudeste brasileiro, reforçando a entrada nos sertões, agora de outras regiões, e constituindo, por todo o Brasil, uma miscigenação étnica e linguística. Todos esses fatos históricos aconteceram no Brasil Colonial e foram preponderantes para demarcar o processo de formação do PB.

Assim, a constituição de *corpora* do período colonial tem a tarefa de, por ser um período que antecede a formação do português brasileiro, tentar trazer a compreensão sobre (i) qual gramática portuguesa chegou nas caravelas, (ii) quais foram as influências, no português, das línguas maternas indígenas e africanas da maioria da população brasileira, que aprendeu o português em situação de aquisição imperfeita e (iii) se houve contato e/ou atuação significativa da língua portuguesa com/ as línguas gerais indígenas e o português geral brasileiro. A ausência de informações precisas a respeito das gramáticas que chegaram a solo brasileiro e a escassez de indivíduos letrados no Brasil são aspectos que dificultam essa tarefa, tornando o trabalho de constituição de *corpus* desse período um problema “arqueológico” e de “cabouqueiro” (CASTRO, 1996).

Embora os documentos de portugueses que chegaram ao Brasil no século XVI já tenham sido bem explorados (cf. MATTOS E SILVA, 2002), os séculos XVII e XVIII ainda se configuram como obscuros no cenário linguístico brasileiro. O PHPB, em sua fase 01 (2009 a 2015), constituiu, com o trabalho de 11 equipes estaduais, *corpora* de cartas, oficiais e particulares, a partir da segunda metade do século XVIII. Apenas duas equipes constituíram *corpora* da primeira metade desse século, e o período anterior, os anos seiscentistas, não foi objeto central de investigação do grupo (CASTILHO, 2007). Desse modo, o *corpus* colonial do CE-DOHS busca contribuir com os estudos já feitos sobre o século XVIII, oferecendo cartas e outros gêneros textuais diversos desde o início do século, com a identificação dos escreventes preenchida, permitindo a divisão do material por vertentes. Além disso, o *corpus* formado procura, também, abrir os caminhos da pesquisa linguística para o século XVII, período essencial para se entender a base do português brasileiro.

## 2. A descrição do *corpus* colonial do CE-DOHS

O *corpus* formado contém 151 documentos de 16 gêneros textuais diferentes, limitados, temporalmente, por data de escrita, aos séculos XVII e XVIII. Escritos por 16 indivíduos, esses materiais alcançam, por local de nascimento, uma boa parte do Brasil Colonial: há escreventes nascidos em São Paulo, Paraíba, Bahia (estado majoritário), Pernambuco e Pará. Os textos desses

séculos, porém, não são vistos de forma homogênea ou uniforme: ao selecioná-los, a organização foi feita de forma a capturar a formação distinta do PB a partir de duas grandes macro-vertentes heterogêneas e plurais, agrupando os textos em dois conjuntos, a saber:

**Conjunto 1:** Textos escritos por indivíduos cujo *input* de aquisição do português venha em uma transmissão de língua primeira para língua primeira.

**Conjunto 2:** Textos escritos por indivíduos de populações não portuguesas, mas já mestiças de diversas etnias, sobretudo, indígenas e africanas, para quem o português se configura como língua primeira, a partir de um *input* de aquisição, que venha de uma língua de segunda para a primeira língua.

Seguindo esse critério, classificamos os documentos de brancos, filhos de portugueses, como representativos do conjunto 1, enquanto dentro do conjunto 2 estão os descendentes de portugueses e/ou brancos brasileiros com, geralmente, mulheres mamelucas ou mulatas, que adquiriram o português como segunda língua. Tal classificação seguiu a hipótese de transmissão linguística proposta por Baxter (1995), uma vez que depoimentos a respeito do processo de aquisição de língua (oral e escrita) dos escreventes são praticamente inexistentes.

É importante ressaltar que, devido à prática da escrita delegada (em que os que possuíam habilidades na escrita escreviam documentos para aqueles que não dominavam tal tecnologia), torna-se imprescindível, sobretudo para os documentos do conjunto 2, uma investigação paleográfica a respeito da conferência da letra do texto e da assinatura.

Dentro desses conjuntos, ainda foi possível distribuir os documentos em pequenos acervos, conforme a contextualização histórica dos escreventes e/ou dos manuscritos, dando a possibilidade, ao usuário, de observar o comportamento linguístico de determinado informante em relação ao comportamento linguístico da comunidade em que ele vivia. A organização em pequenos acervos é importante para, sobretudo, ampliar as possibilidades de investigação de estudos linguísticos dentro do *corpus*: como muitos dos escreventes são parentes e/ou conviveram em um mesmo lugar, tiveram ofícios semelhantes e/ou participaram dos mesmos eventos históricos, é possível a realização de estudos em tempo real de longa ou curta duração, com as duas estratégias propostas por Labov (1994), o estudo de painel e o estudo tendências.

A seguir, são descritos os acervos organizados dentro dos conjuntos 1 e 2.



## 2.1 Acervos do conjunto 1: as famílias e os grupos administrativos coloniais brasileiros

**Acervo da Família Ravasco – ramo brasileiro:** constitui-se de 30 documentos<sup>3</sup>, digitalizados e disponibilizados pelo Projeto Resgate Barão do Rio Branco, com dez gêneros textuais diferentes, escritos por Bernardo Vieira Ravasco e Gonçalo Ravasco Cavalcante de Albuquerque, irmão e sobrinho do Padre Antônio Vieira, respectivamente. A produção desses documentos é resquício do papel importante que teve essa família no Brasil. Em virtude da boa relação do Padre Vieira com a metrópole, foi dado a seu irmão, Bernardo, o ofício de secretário do Estado do Brasil, que, após a morte dele, passou a seu filho, Gonçalo Ravasco. A esse ofício, conforme Puntoni (2005), se destinavam a preservação e a produção de documentos, como emissão de patentes, alvarás e provisões, além da cópia de documentos escritos por diversos representantes da elite portuguesa e de documentos escritos pelo próprio secretário, em virtude da necessidade de circulação documental dentro e fora da colônia e da ausência de imprensa no Brasil, gerando, assim, uma produção extensa de material escrito.

Os manuscritos desse acervo foram escritos entre a segunda metade do século 17 e a primeira metade do século 18, na cidade de Salvador. Bernardo e Gonçalo nasceram, respectivamente, no ano de 1617 e 1639, localizando os manuscritos, por data de nascimento, como parte representativa da primeira metade do século XVII. O material aqui reunido representa a escrita da vertente culta do português brasileiro, uma vez que ambos os escreventes tiveram acesso à escolarização e estudaram no Colégio dos Jesuítas. Além disso, estão presentes nos manuscritos informações históricas cruciais para a compreensão da sociedade colonial brasileira.

**Acervo da Família Ferrão Castelo Branco:** constituído por mais de 30 fólios de rascunhos de cartas, escritos por Antonio Gomes Ferrão Castelo Branco, e 1 lista e 1 procuração, escritas no século XVIII por Maria Cardozo de Oliveira, representantes de uma das famílias mais importantes da Bahia Colonial e mais influentes dentro do sertão baiano. Antonio, filho de Maria Cardozo, nasceu, aproximadamente, em 1728 e, em virtude de sua linhagem culta, adquiriu muitos privilégios na colônia, sobretudo o acesso à educação da mais alta qualidade, sendo fluente, inclusive, em francês. O

---

<sup>3</sup> A investigação em distintos arquivos levou à prospecção de 135 documentos escritos no século XVII e século XVIII, os quais compõem um *corpus* com mais de vinte gêneros textuais, classificados conforme Bellotto (2002). Neste acervo, consideramos, por ora, apenas os documentos originais, atingindo um total de 30 manuscritos, de dez gêneros textuais diferentes. Os demais documentos, que correspondem a cópias de manuscritos escritos por portugueses e brasileiros, ainda estão em fase de organização.

gosto pelas Letras, a importância de seu sobrenome e a necessidade de comunicação foram alguns dos fatores que, possivelmente, levaram Antonio a manter um borrador, livro de rascunho das suas cartas, enviadas a partir de 1749, para familiares e para a administração portuguesa, que compõe majoritariamente os documentos desse acervo. Os documentos estão depositados na Biblioteca Brasileira Guita e José Midlin (USP).

O assunto principal das cartas se volta às tentativas de Antonio e de Maria em manter as riquezas da família após as inúmeras dívidas ocasionadas por Alexandre Gomes respectivamente pai e esposo dos escreventes. Há, neste acervo, importantes fatos para o estudo da formação do português brasileiro. Considerando a hipótese de Mattos e Silva (2004), de que o português culto brasileiro começou a delinear-se na segunda metade do século XVIII, os escritos de Maria Cardozo de Oliveira e Antonio Gomes Ferrão Castelo Branco, datados exatamente nesse período, podem revelar aspectos linguísticos profícuos para a confirmação dessa hipótese. O fato de haver, dentro do acervo, documentos escritos por uma mulher brasileira, sinaliza que o acesso à escolarização para mulheres, naquele período, era extremamente difícil, mas não impossível, abrindo caminhos para uma investigação a respeito da escolarização de mulheres cultas no Brasil Colonial. Sendo grande parte do material constitutivo de um borrador, a presença dos rascunhos das cartas pode sinalizar importantes aspectos linguísticos que ficam, muitas vezes, escondidos em documentos oficiais, por pressões externas à língua. Além disso, o teor do rascunho das cartas de Antonio Gomes tem um valor sócio-histórico de grande notoriedade: é possível conhecer melhor não só a Feira do Capuame, importante evento econômico que marcou a constituição de muitas cidades Baianas, tais quais Feira de Santana e Dias D'Ávila (HENRIQUE, 2014), como o dia a dia dos fazendeiros e, por depoimentos indiretos, dos escravos que viviam nos sertões baianos.

**Documentos Avulsos de Brancos da Elite Portuguesa:** este acervo é composto por 41 documentos escritos <sup>4</sup> por cinco indivíduos brancos brasileiros que não necessariamente compartilharam uma vida ou ações em comum, mas tiveram postos administrativos e militares que lhes oportunizaram escrever manuscritos, solicitando ou informando decisões ao Rei. Os escreventes selecionados para compor o acervo foram aqueles que participaram da invasão do sertão brasileiro ou aqueles que tiveram influência nas decisões políticas cometidas ao sertão baiano (como os governadores de determinadas capitanias, por exemplo).

---

<sup>4</sup> Os documentos desse acervo foram digitalizados pelo Projeto Resgate e pelo Arquivo Nacional da Torre do Tombo (ANTT). O CE-DOHS está organizando um site ([www.corpus2cedohs.com.br](http://www.corpus2cedohs.com.br)) em que é possível ver a localização específica de cada documento, bem como o local original de depósito.

Há, neste acervo, documentos escritos por três governadores do Nordeste. André Vidal de Negreiros, paraibano, foi nomeado para o governo do Maranhão e, em seguida, para assumir o Governo de Pernambuco, no século XVII. Antonio da Silva Barbosa, nascido na primeira metade do século XVII, em Pernambuco, serviu à metrópole como capitão-mor, no terço do Mestre de Campo Pedro Gomes, por 28 anos, e, em virtude de seus grandes serviços, tornou-se governador das capitanias do Rio Grande do Norte e da Paraíba. Bernardo Vieira de Melo, também pernambucano, nascido em 1658, teve uma progressiva carreira militar e assumiu, entre tantos postos militares, o governo da capitania do Rio Grande do Norte, em 1695. Tais escreventes participaram ativamente da Guerra dos Mascates, ocorrida na época do governo de Vidal de Negreiros, podendo se formar, então, uma comunidade linguística entre eles.

Para além dele, há, ainda, documentos escritos por João de Gois e Araújo<sup>5</sup>, desembargador e procurador baiano, possuidor de muitas terras que alcançavam o sertão da Bahia, e de João de Moraes Montesinhos, também baiano, mas com uma posição social inferior, sendo mercador que atuava nas rotas entre Bahia e Minas Gerais (SEVERS, 2001). Por ser filho de cristão-novo, foi denunciado por um de seus inimigos, fato que o levou a ser preso em Lisboa. Durante a viagem, foi muito maltratado, motivo que o fez, após chegar em Portugal, escrever, a próprio punho, uma carta relatando o ocorrido.

## 2.2 Acervos do conjunto 2: os mamelucos e mulatos letrados

**Documentos de Mamelucos – Tomo I:** este acervo, ainda em fase de construção, constitui-se de manuscritos e de indícios de manuscritos, considerados, pela Inquisição, como feitiçaria, produzidos por mamelucos nascidos entre o século XVI e XVIII. Os relatos da inquisição, presente no Brasil, no século XVII, afirmaram existir alguns mamelucos que escreveram “cartas de tocar” que destinavam, àqueles que as possuísem, qualquer pessoa que desejasse. Temos, nesse acervo, apenas uma única “carta de tocar” registrada, escrita por Adrião Ferreira, no século XVIII, que foi anexa ao seu processo inquisitório. É um acervo voltado à busca de depoimentos indiretos e, quando existir, diretos, da existência de tais cartas, que comprovam a escolarização, ainda que mínima, de mamelucos no Brasil Colonial. Como tais cartas eram rudimentares, espera-se encontrar dados mais aproximativos da

---

<sup>5</sup> João Goes de Araújo também tinha uma relação familiar próxima com escreventes do Acervo da Família Vieira Ravasco: era casado com Catarina de Sousa, filha de Rui de Carvalho Pinheiro e concunhada de Padre Antonio Vieira (FONTES, 1994).

gênese do português popular. O manuscrito desse acervo está depositado no Arquivo Nacional da Torre do Tombo (ANTT).

**Documentos de Mamelucos – Tomo II:** o acervo contém 23 documentos (cartas e memória), digitalizados pelo Projeto Resgate, de três escreventes; dois são parentes do Caramuru e resgatam uma história linguística dos primeiros passos após a invasão portuguesa. Um dos mais famosos portugueses que viveu no Brasil, Diogo Álvares, conhecido também como Caramuru, conforme Amado (2000), chegou à América aproximadamente no início do século XVI, após um naufrágio, e permaneceu na Bahia por três a seis décadas, quase sem contato com seus conterrâneos. Casou-se com a índia Paraguaçu e formou, a partir daí, senão a primeira, uma das primeiras famílias brasileiras desenvolvidas a partir do processo de miscigenação.

O pioneirismo de Caramuru favoreceu seu domínio territorial e político na América Portuguesa. A sua efetiva aproximação com determinadas tribos indígenas e o domínio de suas línguas, tornando-se, inclusive, intérprete durante a chegada do primeiro governador geral, Tomé de Souza, em 1549, permitiu que Diogo obtivesse grandes prestígios perante a elite portuguesa que desembarcava. Sendo assim, apesar de seus descendentes serem todos mamelucos, era mais fácil pertencerem à classe alta da colônia. O acervo, aqui apresentado, contém documentos do bisneto de Caramuru, Lourenço de Brito Correa, e seu filho, Lourenço de Brito Figueiredo.

Lourenço de Brito Correa nasceu aproximadamente em 1590. Filho de Sebastião de Brito Correa (branco) e Maria de Figueiredo Mascarenhas (brasileira, mameluca), seu parentesco com Diogo Álvares lhe permitiu ter acesso à educação e a ofícios administrativos destinados à elite colonial (SANTANA, 2012). Estudou no Colégio da Companhia de Jesus e, após concluir os estudos, tornou-se Provedor-mor da Fazenda Real do Estado do Brasil<sup>6</sup>. Após sua morte, o ofício de provedor-mor passou a seu filho, Lourenço de Brito Figueiredo, nascido na primeira metade do século XVII. Teve um caminho acadêmico e político semelhante a seu pai. Atuou fortemente durante a invasão holandesa e recebeu o título de Fidalgo da Casa Real e cavaleiro da ordem de Cristo. Seus escritos se relacionam a esses ofícios e à invasão holandesa, estando, nesses textos, importantes informações acerca da História da Bahia.

Os documentos de Lourenço Figueiredo e Lourenço Correa são importantes fontes para perceber a relação entre a aquisição de língua imperfeita e a força da norma da escolarização, que os

---

<sup>6</sup> Lourenço Correia governou interinamente a Bahia em 1641 e, por questões políticas, brigou com seu contemporâneo Bernardo Vieira Ravasco (escrevente do acervo da Família Vieira Ravasco – ramo brasileiro), motivo que o fez ser preso e enviado a Portugal. Durante o tempo que administrou a Bahia e, em virtude de seu ofício de comendador da Guerra de reconquista da cidade de Salvador, precisou escrever muitas cartas ao Rei, sobretudo para queixar-se de Bernardo.

dois obtiveram. Ambos são exemplos de escreventes mamelucos que foram capazes de ascender socialmente e seus textos podem elucidar um novo olhar sobre os mamelucos dentro do Brasil Colonial, merecendo aprofundamentos. A fim de estabelecer um contraste com os manuscritos dos mamelucos escolarizados, este acervo também contém um único documento de Domingos Jorge Velho. Mameluco paulista, foi o sertanista responsável pela captura de Zumbi de Palmares. Não há notícias sobre sua escolarização; a origem humilde de sua família, entretanto, faz levantar a hipótese de que não houve acesso regular à escola. Há, ainda, registros históricos que atestam ser o Domingos Jorge Velho falante de uma língua “tupi”, não dominando a língua portuguesa (MATTOS E SILVA, 2004) e que podem ser tema de investigação. Por ser sertanista, a sua presença dentro do interior baiano se configura um dado importante para entender que português chegava no interior, se havia diferenças para a língua portuguesa situada na costa brasileira, e o contraste entre o acesso ou não da escolarização por indivíduos que tiveram um processo de aquisição linguística semelhante.

**Escrita de Pretos e Pardos:** este acervo é composto por quatro documentos (três cartas e um atestado), digitalizados pelo Projeto Resgate, escritos por participantes do “Terço da gente Preta” ou “Terço de Henrique Dias”. Surgido em Pernambuco, nos primeiros anos da invasão holandesa, esse terço era uma espécie de pequena tropa de infantaria, formada por indivíduos pretos nascidos no Brasil, que participou ativamente da defesa da cidade de Salvador. Seu principal e primeiro líder, Henrique Dias, natural de Pernambuco, recebeu a promessa de foro de fidalgo, em virtude de seus bons serviços, e escreveu cartas ao Rei, solicitando o cumprimento das promessas feitas e relatando o preconceito sofrido por ele. Após sua morte, o terço se manteve vivo, tendo outros sete capitães negros, até o final do século XVIII. Encontramos registros escritos pelo último líder do Terço, Brás de Brito Souto. Escravo forro, nascido em Pernambuco (1684), filho de africanos escravizados alforriados, foi soldado na companhia do mestre de campo Domingos Rodrigues Carneiro. A partir de 1720, teve uma ascensão em sua carreira militar e tornou-se sargento-mor das ordenanças do Regimento da Gente Preta da Jurisdição do governo de Pernambuco.

O último escrevente desse acervo é João da Silva Guimarães. Ivo (2009), em sua tese, afirma não saber nada a respeito de sua nacionalidade: encontramos, entretanto, no Dicionário de Sertanistas Brasileiros, a sua classificação como paulista. Pardo, atuou no Terço de Gente Preta e na conquista do norte da Bahia. Foi muito próximo de João Gonçalves da Costa, ex-escravo português que repovoou o sertão da Ressaca, fundando Vitória da Conquista. João Guimarães escreveu, a partir de 1755, durante seu percurso no norte baiano, algumas cartas, principalmente dando notícias ao Rei sobre a descoberta das minas de diamantes na Chapada Diamantina e em algumas localidades do Recôncavo.

Este é o menor acervo já concluído, contendo apenas quatro documentos. Apesar disso, são cartas relevantes sócio-historicamente e linguisticamente, uma vez que esses documentos são

considerados, por alguns historiadores, como Camargo (1987), registros dos primeiros afro-brasileiros letrados. É possível perceber, nesses manuscritos, o sofrimento dos negros em conseguir suas mercês e, contrariamente, a importância deles para a manutenção das terras brasileiras, a favor da metrópole portuguesa.

### 3. A representatividade do *corpus* colonial do CE-DOHS

Delimitar o que seria um *corpus* representativo é uma questão difícil dentro dos estudos linguísticos, uma vez que não há como ter conhecimento da dimensão total da população. A noção de representatividade é vista por Berber Sardinha (2000) a partir da extensão do *corpus*, em três aspectos:

- (i) quanto ao número de palavras: quanto maior é o *corpus*, maior é a probabilidade do encontro de palavras com baixa produtividade;
- (ii) quanto aos gêneros textuais: quanto mais gêneros textuais, mais representatividade a língua em questão encontrará;
- (iii) quanto à quantidade de textos em certos gêneros textuais: quanto maior é o número de textos de certos gêneros, mais representativo o *corpus* será.

Em se tratando do período colonial brasileiro, há que se propor mais um critério relevante: o da diversidade de aquisição do português, associado ao fator étnico. Isso porque, segundo Mattos e Silva (2004), as variedades prestigiadas e não prestigiadas do PB tiveram atores de formação distintos: enquanto estas foram construídas a partir do português geral brasileiro, falado pelos africanos e seus descendentes, e das línguas gerais indígenas, predominantes no cenário colonial até o século XVIII, aquelas se constituíram a partir do português europeu e daqueles que o adquiriram em condições regulares de aquisição, os denominados brancos brasileiros por Mussa (1991). Sendo assim, propomos o seguinte aspecto:

- (iv) diversidade étnica: quanto mais diversificados etnicamente forem os escreventes dos manuscritos integrados ao *corpus*, maior é a probabilidade de alcançar os diferentes processos de aquisição do português que influenciaram a formação das vertentes do PB.

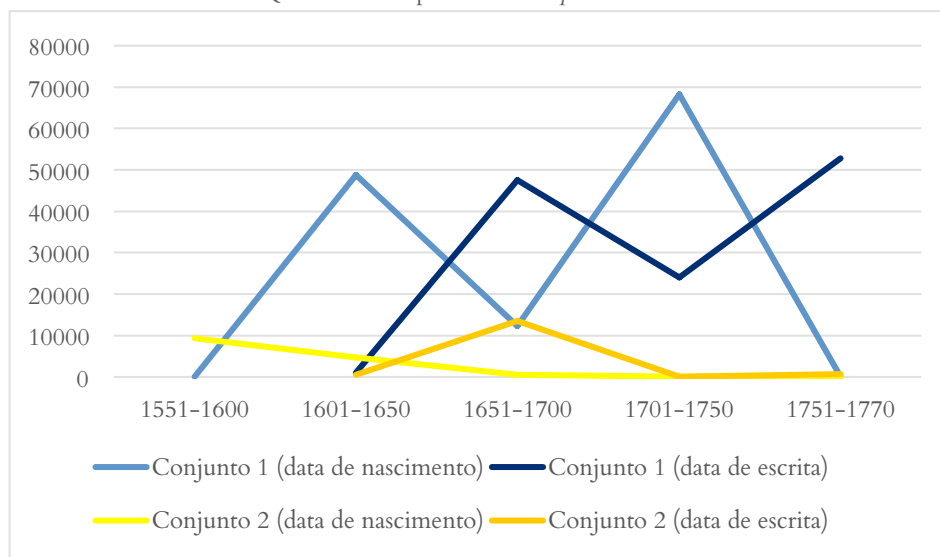
Vejamos, agora, a representatividade do *corpus* nesses quatro critérios.

### 3.1 Quanto ao número de palavras

Esse aspecto pode ser compreendido a partir de dois pontos de vista diferentes: a partir da identificação dos documentos por data de escrita ou a partir da identificação dos documentos por data de nascimento de seus escreventes. Uma das vertentes de estudo da mudança linguística, a gerativista diacrônica, considera que a mudança de uma língua ocorre de forma abrupta durante o processo de aquisição, ainda na infância (LIGHTFOOT, 1991). A localização dos documentos por data de nascimento torna-se, então, necessária para entender em que momento a mudança surgiu e a localização dos documentos por data de escrita é relevante para entender em que momento a mudança linguística transparece nos textos escritos.

O *corpus* em discussão tem 151 documentos que contêm, aproximadamente, 143.984 palavras, caracterizando-o, segundo uma abordagem histórica (BERBER SARDINHA, 2000), como um *corpus* pequeno-médio. Por data de produção do documento, o *corpus* integra, temporalmente, a primeira metade do século XVII até a segunda metade do século XVIII. Por data de nascimento, ele é estendido, compreendendo, também, o final do século XVI. Dividindo o total de palavras nos conjuntos supracitados e, temporalmente, a cada meio século, por data de escrita do documento e por data de nascimento dos escreventes, tem-se o seguinte panorama<sup>7</sup>:

Gráfico 1 – Quantidade de palavras do *corpus* colonial do CE-DOHS



Fonte: autoria própria.

<sup>7</sup> Os documentos sem datação (que totalizam 3.794 palavras) foram retirados da contagem.

Observa-se, inicialmente, uma quantidade pequena de dados e uma grande disparidade entre os conjuntos organizados, em número de palavras, diferença já esperada, tendo em vista o contexto colonial brasileiro. Considerando que boa parte da população naquele período era desconhecida e sua história não foi lembrada pelas genealogias históricas, o encontro de documentos e da possibilidade de reconhecimento biográfico de seus escreventes é um achado para a época. Além disso, segundo Houaiss (1985), até o século XVIII, apenas 0,5% dos portugueses e brasileiros brancos que viviam na Colônia eram letrados. Essa pequena porcentagem ainda é diminuída ao considerar o efeito tempo, que rareia significativamente a sobrevivência de muitos manuscritos produzidos na América Portuguesa. Desse modo, os dados do Conjunto 1, apesar de poucos, mostram-se significativos e relevantes para o contexto do período.

Se entre os portugueses e seus descendentes habitantes do Brasil, considerados de uma classe social elitizada, menos de 1% sabia ler e escrever, essa porcentagem chega a ser ainda mais próxima do zero ao considerar os descendentes de africanos e indígenas – os chamados mamelucos e mulatos por Mussa (1991) – escreventes do conjunto 2. A estigmatização e a escravização foram fatores que os fizeram ser excluídos da sociedade. No caso dos descendentes africanos, raríssimos são os que conseguiram sobressair, a partir de fugas, de vitórias em guerras ou do apreço de seus donos, e tinham oportunidade de aprender a ler e escrever. Assim, o pequeno número de dados do Conjunto 2 representa fielmente o escasso contato desses povos com a escrita.

Entre os documentos selecionados, por data de escrita do documento, a segunda metade do século XVIII é mais significativa para o conjunto 1 e, para o conjunto 2, a segunda metade do século XVII é que tem um maior volume de dados. Já por data de nascimento, o conjunto 2 é mais produtivo na segunda metade do século XVI e o conjunto 1 na primeira metade do século XVII. Na comparação entre os dois conjuntos, pode-se observar que, por data de escrita, o período em que há maior aproximação do número de palavras em ambos os grupos é a segunda metade do século XVII. Isso pode ser justificado pelo fato de as caravelas portuguesas terem chegado em grande quantidade, apenas no final do século XVI, fazendo com que a primeira grande leva de colonos brasileiros (foco da constituição do *corpus*) tenha nascido a partir do século seguinte. Sendo assim, pondo em comparação os dois conjuntos organizados, os anos seiscentistas são um período temporal produtivo, não só em relação aos dados, mas também em relação à história da difusão do português no Brasil. O século XVII constitui-se, historicamente, como o período de difusão do português que se transformaria, mais tarde, no português brasileiro, importante, portanto, para entender a partir de qual gramática o português brasileiro mudou.

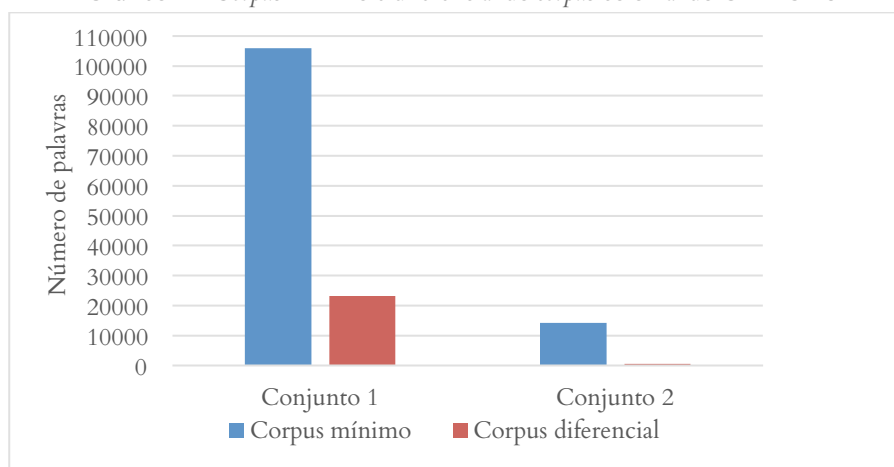


### 3.2 Quanto aos gêneros textuais

A diversificação tipológica não só é um desejo para aqueles que trabalham com a Linguística Histórica, como uma necessidade. O fato de lidar com gramáticas pretéritas, recorrendo unicamente aos documentos escritos do passado, necessita levar em consideração as tradições discursivas que permeiam a escrita dos gêneros textuais, constituindo, assim, em certas partes de um texto, uma gramática solidificada, que não necessariamente é representativa do período em que se escreveu o documento. Um estudo linguístico-histórico ideal deve se basear em textos com gêneros diversificados, que forneçam um trabalho sobre a língua e não sobre o gênero textual analisado (MATTOS E SILVA, 2004).

Sabendo disso, o PHPB, ao longo dos seus vinte anos de existência, dividiu os *corpora* constituídos em dois: o *corpus* mínimo e o *corpus* diferencial. Enquanto este é composto por diferentes gêneros textuais, escritos por brasileiros ou portugueses, aquele é composto apenas por cartas, particulares e oficiais, de brasileiros. O gênero textual carta foi o escolhido para a composição de um *corpus* mínimo, em consequência, entre outros fatores, de ser mais comum à escrita pretérita e da chance de proximidade com uma gramática vernacular. No nosso *corpus*, o gênero carta também foi altamente produtivo. Entretanto, por ser a maioria de nossos manuscritos de circulação pública, e, portanto, notarial, foi possível encontrar uma diversidade de gêneros textuais. Desse modo, observando-se o gênero textual, conseguimos organizar, tal qual o PHPB, os manuscritos em dois *corpora*: um *corpus* mínimo, formado por cartas, e um *corpus* diferencial, composto por outros quinze gêneros textuais, classificados conforme Bellotto (2002). A grande maioria das cartas é de circulação pública. Cartas de circulação privada só aparecem a partir do século XVIII.

Gráfico 2 – *Corpus* mínimo e diferencial do *corpus* colonial do CE-DOHS



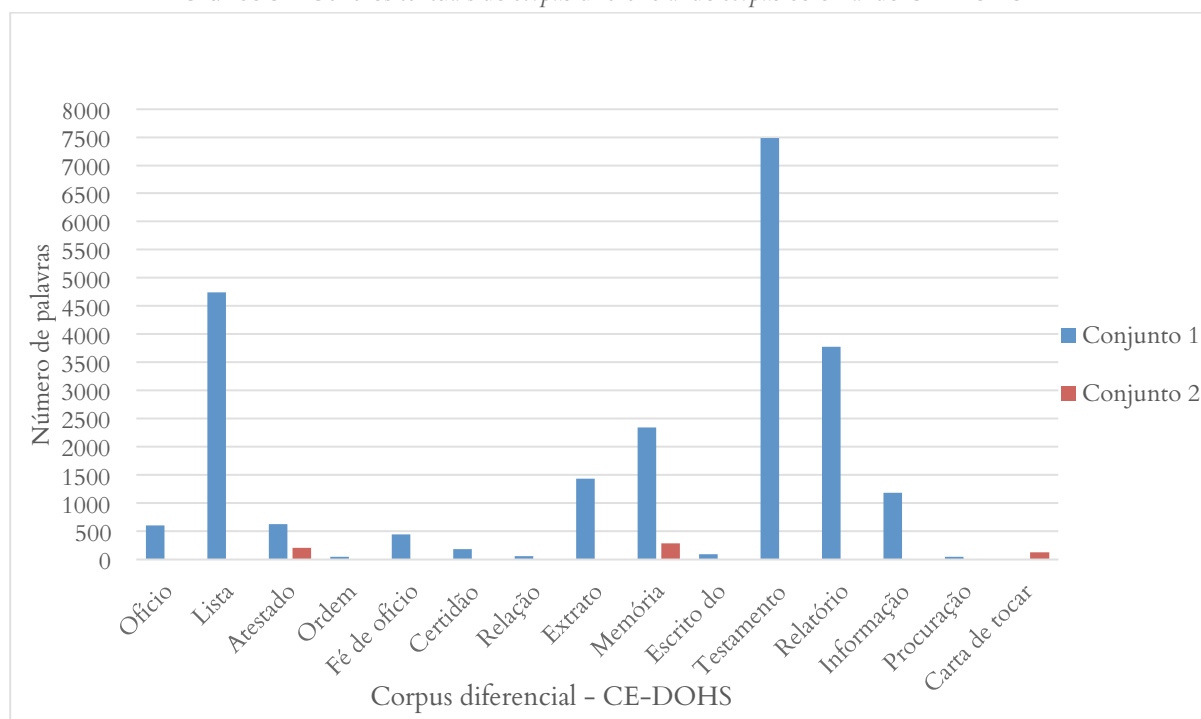
Fonte: autoria própria.

Em número de palavras, o gênero textual carta é nitidamente mais produtivo, em ambos os conjuntos. Isso torna possível a comparação com textos de mesma natureza tipológica, mas escritos por indivíduos que tiveram diferentes processos de aquisição da língua portuguesa. O *corpus* diferencial também é menos produtivo nos dois conjuntos: a quantidade de palavras do *corpus* diferencial em ambos os grupos é quase proporcional ao número total de palavras de cada conjunto, representando, portanto, uma igualdade em distribuição de dados por gênero textual, considerando o volume de palavras.

### 3.3 Quanto à quantidade de textos em certos gêneros textuais

Analisando mais profundamente o *corpus* diferencial, há certos gêneros textuais que alcançam um maior volume de palavras:

Gráfico 3 – Gêneros textuais do *corpus* diferencial do *corpus* colonial do CE-DOHS



Fonte: autoria própria.

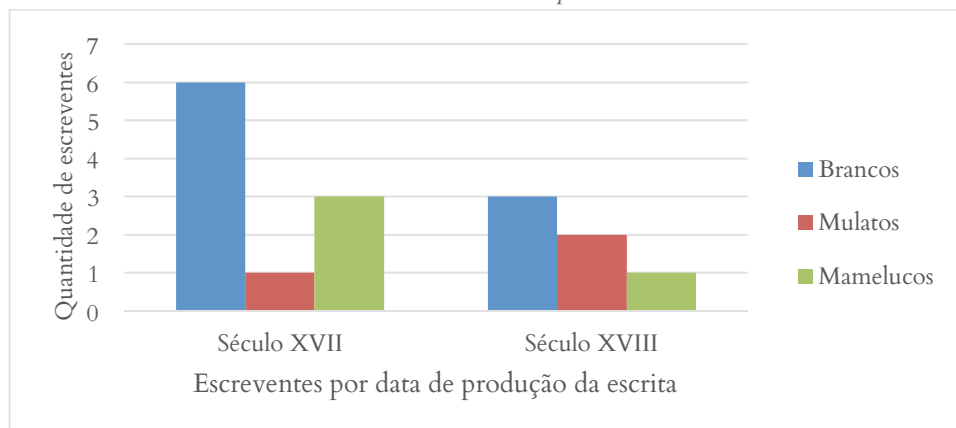
Entre a diversificação tipológica, alguns gêneros se sobressaem: testamento, lista, relatório e memória. A maioria dos gêneros textuais do *corpus* diferencial se concentra no conjunto 1, em consequência do ofício administrativo que os escreventes tiveram, obrigando-os à escrita de textos diversos para informar ao Reino os acontecimentos da metrópole. A escrita mais significativa dos

escreventes do conjunto 2 foi de cartas: sem ter ofícios importantes, seus atos de escrita representavam solicitações e pedidos de mercês não concedidas pelos seus contemporâneos, sendo necessário pedir diretamente ao Rei o cumprimento dessas promessas.

Desse modo, percebe-se que o contexto sócio-histórico dos escreventes não só influencia a tradição discursiva, como o uso restrito de determinados gêneros textuais. Saber sobre o contexto social do escrevente e o contexto social do documento traz contribuições significativas aos estudos linguísticos, principalmente no período colonial, em que a escrita era rara e sua prática era, sobretudo, condicionada a fatores políticos e administrativos. Apesar de a constituição do *corpus* ser voltada à investigação de fenômenos linguísticos, o estudo linguístico dos manuscritos deve ser feito paralelamente ao estudo da História da Cultura Escrita, que, segundo Petrucci (2002, p. 7-8), se volta à história da produção, das características formais e dos usos sociais da escritura e de testemunhos escritos em uma sociedade determinada.

### 3.4 Quanto à diversidade étnica e de aquisição do português no Brasil colonial

O Brasil, nos séculos XVII e XVIII, apresentava uma demografia populacional bem particular. Segundo Mussa (1991), os portugueses e brancos brasileiros não passavam de 30%, enquanto os negros do Brasil, africanos e mulatos correspondiam, nos dois séculos, a 60% da população total. Os indígenas integrados foram drasticamente extintos, alcançando apenas 8% no século XVIII. Foi a partir desse cenário que Mattos e Silva (2004) propôs que os principais difusores do português popular brasileiro seriam os africanos e seus descendentes, enquanto aos europeus e seus filhos coube a difusão das variedades prestigiadas do português. Assim, um *corpus* do período colonial precisa considerar, na medida do possível, essa diversidade existente. Em nosso *corpus*, há 16 escreventes, que podem ser divididos etnicamente da seguinte maneira:

Gráfico 4 – Etnia dos escreventes do *corpus* colonial do CE-DOHS

Fonte: autoria própria.

O *corpus* apresenta um contraste populacional inverso à realidade demográfica da época: há, nos séculos XVII e XVIII, um maior número de escreventes brancos que mamelucos e mulatos. Entretanto, esse é um cenário plausível, considerando os dados sobre escolarização mencionados anteriormente. A disposição, em número de palavras, também segue essa mesma lógica, tendo mais textos escritos por brancos que por outras etnias.

Um outro ponto significativo é o de aquisição da língua portuguesa. A forma de aquisição, regular ou irregular, como língua materna ou como língua segunda, afeta a composição da gramática adquirida. Como já discutido, os brancos tiveram, hipoteticamente, uma aquisição regular. O processo irregular de aprendizagem do português, primeiramente como língua segunda ou terceira e, conforme os descendentes de africanos nasciam, como língua primeira, aliado ao contato linguístico diverso, influenciou a formação de uma variedade do português distante do padrão europeu. Há, no *corpus*, mais escreventes que, hipoteticamente, tiveram uma aquisição regular, diferentemente do cenário colonial.

A identificação dos escreventes por etnia é pioneira nos estudos linguísticos desse período. O acesso a tais informações se deu a partir das genealogias históricas e das pesquisas dos historiadores preocupados com a micro-história. Se os *corpora* que contam somente com autores plenamente identificados são considerados exceções (BARBOSA, 2019, p. 21), ter os autores não só identificados como também separados por etnia, garantindo, assim, uma visão melhor dos processos de aquisição do PB, é um aspecto inovador, na medida em que tenta se aproximar do multi-etnismo generalizado que havia no Brasil Colonial.

#### 4. Potencialidades do *corpus* colonial do CE-DOHS: possibilidades de estudos linguísticos

O material aqui apresentado exhibe distintas possibilidades de estudos linguísticos, sendo possível traçar diferentes *subcorpora* dentro do mesmo *corpus* e/ou constituir novos *corpus* de pesquisa, associando-o a outros *corpora* brasileiros e portugueses. Uma vez que é um material novo, sendo uma parte analisado linguisticamente apenas por Cardoso (2020), os prováveis fenômenos linguísticos para a investigação são diversos. A seguir, estão algumas das potencialidades do *corpus* formado, considerando as possibilidades de organização e/ou temas de pesquisa, podendo ser estudados a partir de qualquer fenômeno linguístico sintático e/ou fonológico:

- (i) estudo de painel, seguindo os procedimentos metodológicos de Rumeu (2008): como muitos dos acervos constituídos são formados por indivíduos de uma mesma família, e que escrevem por um longo período de tempo, como o Acervo da Família Ravasco e o acervo da Família Ferrão Castelo Branco, torna-se possível um estudo de painel, conforme definido por Labov (1994), delineando o comportamento linguístico dos escreventes em extensos lapsos de sincronias passadas, tal qual a metodologia proposta por Rumeu (2008), alcançando as fases da vida adulta e da velhice dos escritores;
- (ii) estudo de tendências: como grande parte dos indivíduos nasceram na Bahia, é possível realizar um estudo diacrônico de tendências (LABOV, 1994) entre os acervos com documentos produzidos no século XVII e no século XVIII, tentando traçar o comportamento linguístico de uma mesma comunidade com diferentes informantes. Como o acesso à escolarização pela elite cabia, principalmente, ao Colégio dos Jesuítas, acompanhar a escrita de diferentes indivíduos também pode fornecer dados relevantes sobre os processos de escolarização da companhia jesuítica no Brasil;
- (iii) estudo comparativo: é possível comparar os dados entre os acervos do conjunto 1, entre os acervos do conjunto 2 e entre acervos de ambos os conjuntos, fornecendo resultados com base no recorte temporal e nos processos de aquisição do português, tentando resgatar indícios que possam servir de base para a construção de uma gramática brasileira;
- (iv) constituição de um novo *corpus*, associado a outros *corpora* coloniais: após a edição do material, a ser realizada conforme as normas do PHPB, será possível a comparação do material com o já obtido pelo PHPB, ampliando, assim, o conjunto de cartas do século XVIII e fornecendo outros gêneros textuais para compor o *corpus* diferencial desse período.

- (v) comparação com *corpora* portugueses do mesmo período: a comparação com documentos escritos por portugueses do mesmo período traz dados relevantes que ajudam a entender a dinâmica do processo de diferenciação do PE e do PB. Com *corpora* já constituídos de documentos escritos por portugueses do século XVII e XVIII – como o Corpus Tycho Brahe e o *corpus* editado e analisado por Martins (1994) – a comparação se torna mais fácil de ser realizada.
- (vi) estudo comparativo entre as gramáticas usadas pelos colonos, pertencentes ou não à elite brasileira, e as normas linguísticas propostas pelos portugueses para o uso da língua portuguesa no século XVIII: segundo Lucchesi (2006), os brasileiros da elite colonial, sobretudo os responsáveis pelas funções administrativas, intencionavam preservar os modelos de cultura e língua metropolitanas. Como muitos dos escreventes estudaram no Brasil e não foram a Portugal, uma comparação entre a gramática usada na escrita e as normas linguísticas portuguesas do século XVIII – a exemplo da gramática do Pe. Jerônimo Contador de Argote, *Regras da língua portuguesa, espelho da língua latina* (1725) – pode indicar em quais aspectos gramaticais o português do Brasil, embora ainda distante de se tornar o português brasileiro contemporâneo, já se afastava da gramática prescritiva do português europeu, refletindo, assim, os primeiros passos de uma transformação mais profunda na língua.
- (vii) estudo de marcas de inabilidade dos manuscritos do *corpus*: considerando o acesso precário à escolarização no Brasil Colonial, o estudo de marcas de inabilidade nesse período constitui um espaço profícuo para entender os graus de habilidade dos escreventes, sobretudo os mamelucos e mulatos do conjunto 2, em que as razões e o processo de escolarização são desconhecidos. O CE-DOHS já tem estudos sobre o tema com *corpus* do século XX (cf. SANTIAGO, 2012 e SANTIAGO, 2019); recuar para o Brasil Colonial pode ser interessante para traçar os critérios de inabilidade que estiveram presentes em documentos brasileiros durante toda a trajetória de formação do PB.
- (viii) estudo de fenômenos linguísticos diversos, visando à identificação de marcas específicas do português do Brasil Colonial: a investigação de determinados fenômenos linguísticos nos documentos do nosso *corpus* pode contribuir para identificar, a partir de marcas linguísticas, as possíveis nacionalidades de escreventes de *corpora* que ainda não possuem tal informação. A identificação biográfica de todos os escreventes do *corpus* colonial é o fator decisivo para a possibilidade de realizar essa investigação. Com o encontro de propriedades específicas da gramática portuguesa do Brasil Colonial, a pesquisa pode seguir por meio de outros documentos, tentando encontrar a gramática correspondente à comumente usada, na escrita, por indivíduos nascidos no

Brasil. Um exemplo disso é a comparação, a partir dos estudos dos pronomes clíticos, com os dados dos séculos XVII e XVIII obtidos por Pagotto (1992), uma vez que o próprio autor afirma não ser possível identificar os redatores desse período como portugueses ou brasileiros. O não encontro dessas propriedades específicas é, também, um dado relevante, apontando a semelhança com o português de Portugal do mesmo período.

## Considerações finais

O Brasil Colonial foi marcado por intensos processos linguísticos e sociais que culminaram em uma realidade brasileira miscigenada e polarizada, tanto socialmente quanto linguisticamente. Para um estudo sobre a formação do português brasileiro, o CE-DOHS buscou construir um *corpus* do período colonial que representasse, mesmo que minimamente, o multilinguismo e multiétnico generalizado da época. O resultado foi um *corpus* com 151 documentos e 16 escreventes, de diferentes etnias, totalizando, aproximadamente, 143 mil palavras. Organizados, etnicamente, em conjuntos e subdivididos em pequenos acervos, conforme a comunidade linguística dos escreventes, esses manuscritos são representativos do uso da língua portuguesa dentro do Brasil Colônia e podem contribuir com informações sócio-históricas e linguísticas relevantes para compreender a partir de qual português o PB se formou e de que maneira o PB se configurou diferentemente do português de Portugal.

Este é um *corpus* que está longe de dar respostas concretas e precisas a respeito das indagações acima e mais distante ainda, em virtude da formalidade dos documentos selecionados, de chegar próximo ao português usado na oralidade, tanto pelos descendentes de portugueses quanto pelos mamelucos e mulatos. Somado a outros *corpora*, entretanto, o *corpus* ganha força e suas potencialidades aumentam. Além disso, em virtude do desconhecimento do uso do português por brasileiros nascidos no século XVII e XVIII, consideramos o *corpus* formado como um entre tantos passos a serem dados em busca da compreensão da formação do PB.

Mesmo que pequeno, o *corpus* colonial complementa o banco de textos e de dados linguísticos do português brasileiro constituído ao longo dos dez anos de existência do CE-DOHS. Junto aos outros acervos, de diferentes séculos, o CE-DOHS passa a conter mais de um milhão de palavras, abarcando quatro séculos (XVII ao XX), configurando-se como um banco capaz de propor um estudo diacrônico com dados linguísticos desde a chegada do português no Brasil.

## Referências bibliográficas

- BARBOSA, A. G. **Para uma história do português colonial: Aspectos linguísticos em Cartas de Comércio**. Tese de Doutorado – Universidade do Rio de Janeiro, 1999.
- BARBOSA, A. G. Demografia histórica e história da língua portuguesa no Brasil-colônia: reflexões sobre o fim dos setecentos. **Revista Linguística da Alfal**, v. 17, p. 75-94, 2005.
- BARBOSA, A. G. **A plataforma de corpora do PHPB: uma apresentação *ad infinitum***. In: CASTILHO (Org.). **História do Português Brasileiro: Corpus diacrônico do português brasileiro**. São Paulo: Contexto, 2019. p. 16-67.
- BAXTER, Alan Norman. Transmissão Geracional Irregular na História do Português Brasileiro: divergências nas vertentes afro-brasileiras. **Revista Internacional de Língua Portuguesa**, v. 14, p. 72-90, 1995.
- BELLOTTO, H. L. **Como fazer análise diplomática e análise tipológica em documento de arquivo**. São Paulo: Arquivo do Estado e Imprensa Oficial do Estado, 2002.
- BERBER SARDINHA, T. B. Linguística de Corpus: histórico e problemática. **D.E.L.T.A.** v. 16, n. 2, p. 323-367, 2000. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S010244502000000200005&lng=en&nrm=isso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010244502000000200005&lng=en&nrm=isso). Acesso em 31 jan. 2021.
- CAMARGO, O. **O negro escrito**. São Paulo: Imprensa Oficial do Estado, 1987.
- CARDOSO, L. **A gramática dos pronomes clíticos no Brasil Colônia: o português clássico na história do português brasileiro**. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual de Feira de Santana, Feira de Santana, 2020. Disponível em: <https://www.researchgate.net/profile/Lara-Cardoso-3>. Acesso em 05 maio 2021.
- CARNEIRO, Z. O. N.; LACERDA, M. F. O. Corpus Eletrônico de Documentos históricos do Sertão: etapa 01 (1750-2000). **Revista Binacional Brasil Argentina, Vitória da Conquista**, v. 8, n. 1, p. 205-221, jul. 2019. Semestral. Disponível em: <http://periodicos2.uesb.br/index.php/rbba/article/view/5588>. Acesso em: 02 out. 2019.
- CASTILHO, A. **Projetos de Pesquisa Linguística**. [S.l.] [2007?]. Disponível em: <https://www.museudalinguaportuguesa.org.br/wp-content/uploads/2017/10/Projetos-depesquisa-linguistica-Ataliba.pdf>. Acesso em 31 jan. 2021.
- CASTRO, I. **Para uma história do português clássico**. In: Actas do Congresso Internacional sobre o português. Lisboa: Colibri/APL, v. II, p. 135-150, 1996.
- FONTES, J.; BOGACIOVAS, M. Os Góes da Bahia. **Asbrap**, [s.l.], v. 1, n. 1, p. 75-90, jan./dez. 1994. Anual. Disponível em: [http://www.asbrap.org.br/documentos/revistas/rev1\\_art10.pdf](http://www.asbrap.org.br/documentos/revistas/rev1_art10.pdf). Acesso em: 09 jun. 2019.



- HENRIQUE, J. **A Feira de Capuame: pecuária, territorialização e abastecimento** (Bahia, século XVIII). Dissertação (Mestrado em História Econômica) – Universidade de São Paulo, São Paulo, 2014.
- HOUAISS, A. **O português do Brasil**. Rio de Janeiro: UNIBRADE/UNESCO, 1985.
- IVO, I. P. **Homens de caminho: trânsitos, comércio e cores nos sertões da América portuguesa – século XVIII**. Tese (Doutorado em História) – Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: [repositorio.ufmg.br/bitstream/1843/VGRO82TM88/1/tese\\_de\\_isnara\\_pereira\\_ivo.pdf](https://repositorio.ufmg.br/bitstream/1843/VGRO82TM88/1/tese_de_isnara_pereira_ivo.pdf). Acesso em 12/11/2019.
- LACERDA, M. F. O. CARNEIRO, Z. O. N. SANTIAGO, H. S. Corpus eletrônico de documentos históricos do sertão: as cartas de inábeis. **Revista A Cor das Letras**, Feira de Santana, v. 17, p. 127-143, 2016.
- LABOV, W. **Principles of Linguistic Change: Internal Factors**. Cambridge: Blackwell Publishers, v. 1. 1994.
- LIGHTFOOT, D. **Howto set parameters**. Arguments from language change. Cambridge: MIT Press, 1991.
- LUCCHESI, D. Parâmetros sociolinguísticos do português brasileiro. **Revista da ABRALIN**, v. 5, n. 1 e 2, p. 83-112, 2006. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/941>. Acesso em 31 jan. 2021.
- MATTOS E SILVA, R. V. Para a história do português culto e popular brasileiro: sugestões para uma pauta de pesquisa. In: ALKMIM, Tânia M (org.). **Para a história do português brasileiro: novos estudos**. São Paulo: Humanitas/FFCHL/USP: FAPESP, v. 2, 2002. p. 443-464.
- MATTOS E SILVA, R. V.; MACHADO FILHO, A. V. L. (Orgs.). **O Português quinhentista: estudos linguísticos**. Feira de Santana: EDUFBA, UEFS, 2002.
- MATTOS E SILVA, R. V. **Ensaio para uma sócio-história do português brasileiro**. São Paulo: Parábola Editorial, 2004.
- MUSSA, A. **O papel das línguas africanas na história do português do Brasil**. Dissertação de Mestrado – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1991.
- PAGOTTO, E. G. **A posição dos clíticos em português: um estudo diacrônico**. Dissertação (Mestrado em Linguística) – Instituto da Linguagem, Universidade Estadual de Campinas, Campinas, 1992.
- PUNTONI, P. Bernardo Vieira Ravasco, secretário do estado do Brasil: poder e elites na Bahia do século XVII. In: BICALHO, M. F.; FERLINI, V. L. A. **Modos de governar: ideias e práticas políticas no império português, séculos XVI a XIX**. 2. ed. São Paulo: Alameda, 2005. p. 157-178.
- RUMEU, M. C. B. **A Implementação do ‘Você’ no Português Brasileiro Oitocentista e Novecentista: Um Estudo de Painel**. Tese (Doutorado em Letras Vernáculas – Língua

Portuguesa) – Programa de Pós Graduação em Letras Vernáculas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.

PETRUCCI, A. **La ciência de la escritura**: primera lección de Paleografía. Buenos Aires: Fondo de Cultura Econômica de Argentina, 2002.

SANTANA, R. G. S. **Lourenço de Brito Correa**: sujeito mais perverso e escandaloso. Conflitos e suspeitas de motim no segundo vice-reinado do Conde de Óbidos. (Bahia 1663-1667). Dissertação (Mestrado) – Curso de História, Programa de Pós-graduação em História, Universidade Estadual de Feira de Santana, Feira de Santana, 2012. Disponível em: <http://www2.uefs.br/pgh/docs/Dissertacao/DISSERTA%C3%87%C3%83ORicardo.pdf>. Acesso em: 09 jun. 2019.

SANTIAGO, H. S. **Um estudo do português popular brasileiro em cartas pessoais de “mãos cândidas” do sertão baiano**. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual de Feira de Santana, Feira de Santana, 2012.

SANTIAGO, H. S. **A escrita por “mãos inábeis”**: uma proposta de caracterização. Tese (Doutorado em Língua e Cultura) – Universidade Federal da Bahia, Salvador, 2019.

BRITO, P. **Cartas Marienses (séc. XX)**: edição fac-similar e semidiplomática e estudo da concordância nominal. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual de Feira de Santana, Feira de Santana, 2012.

SEVERS, S. M. S. S. “Sapatos ao mato”: o sentimento de “um triste homem que vem preso” pelo Santo Ofício. **Politeia**: História e Sociedade, Vitória da Conquista, v. 11, n. 1, p. 105-125, jun. 2011. Semestral. Disponível em: <http://periodicos2.uesb.br/index.php/politeia/article/view/3795>. Acesso em: 09 jun. 2019.