



Reeditar é preciso? As cartas oficiais norte-rio-grandenses e os corpora diacrônicos

Is it necessary to reedit? The Rio Grande do Norte official letters and the diachronic corpora

Felipe Morais de Melo

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN),
Natal, Rio Grande do Norte / Brasil

felipemorais_m@yahoo.com.br

<https://orcid.org/0000-0003-1881-6262>

Resumo: São bastante estreitas e profícuas as relações entre a Linguística Histórica e a Linguística de *Corpus*, sobretudo a partir da década de 90, quando a primeira ganha fôlego renovador no Brasil, tendo na edição de *corpora* diacrônicos uma das motrizes basilares para seu progressivo desenvolvimento. Sob o signo desse entrelaçamento entre as duas áreas, este trabalho, através da análise de conteúdo, apresenta um caso de reedição de *corpus* diacrônico, as *cartas oficiais norte-rio-grandenses* (1713-1950), e, por meio de pesquisa bibliográfica, propõe algumas reflexões acerca do contato entre esses dois campos de estudo no Brasil e dos impactos, positivos e negativos, do meio digital nessa conexão. O artigo revela como alterações nas orientações de uma investigação podem justificar um processo frutífero de reedição e evidencia tanto riscos éticos quanto aprimoramentos funcionais a que esses *corpora* estão sujeitos devido a fatores como a tecnologia e a internet, propondo, ao final, que haja maior visibilidade das questões levantadas no estudo.

Palavras-chave: reedição; cartas oficiais; corpus diacrônico; linguística histórica.

Abstract: The relations between Historical Linguistics and Corpus Linguistics are quite close and fruitful, especially from the 90's, when the former gains renewed breath in Brazil, having in the edition of diachronic corpora one of the basic drivers for its progressive development. Under the sign of this intertwining between the two areas, this work, through content analysis, presents a case of reedition of a diachronic corpus, the *Rio Grande do Norte official letters* (1713-1950), and, through bibliographic research, proposes some reflections about the contact between these two fields of study in Brazil

and the impacts, positive and negative, of the digital medium in this connection. The paper reveals how changes in research guidelines can justify a fruitful process of reedition and highlights both ethical risks and functional improvements to which these corpora are subject due to factors such as technology and the internet, proposing, in the end, a greater visibility of the issues raised in the study.

Keywords: re-edition; official letters; diachronic corpus; historical linguistics.

Recebido em 13 de outubro de 2020

Aceito em 22 de dezembro de 2020

1 Considerações iniciais

La biblioteca es ilimitada y periódica. Si un eterno viajero la atravesara en cualquier dirección, comprobaría al cabo de los siglos que los mismos volúmenes se repiten en el mismo desorden (que, repetido, sería un orden: el Orden). Mi soledad se alegra con esa elegante esperanza¹

(BORGES, 2009, p. 566).

editar é preciso

(CAMBRAIA, 1999, p. 14).

Enxergamos – na mecânica borgiana pela qual se antepõe ao espelho descoberto o próprio infinito e da qual extraímos essas linhas em que a engrenagem ficta do escritor argentino ressuma em corredores de estantes – uma alegoria do próprio fazimento de se (re)editar um *corpus*. É quase como se ouvíssemos uma tradução filosófica (claramente circunstancial e assaz pragmática dentre as tantas e tão profundas expressões que os cantos de Borges podem ativar e evocar) do que escreve Almeida Cabrejas: “es imposible en las arduas tareas de transcripción y edición no cometer errores; la única manera de reducir su aparición es la

¹ “A Biblioteca é ilimitada e periódica. Se um eterno viajante a atravessasse em qualquer direção, comprovaria ao fim dos séculos que os mesmos volumes se repetem na mesma desordem (que, reiterada, seria uma ordem: a Ordem). Minha solidão alegra-se com essa elegante esperança” (BORGES, 1998, p. 523).

revisión”² (ALMEIDA CABREJAS, 2013, p. 15). Ainda que revisitado, revisado, repassado, recuidado, no cabo dos dias, num próximo lance de olhos, para além do reconhecimento das melhorias, constatar-se-á a repetição da desordem: outros erros que ficaram, velhos deslizes que sobraram, e outros bruxuleios, inda que só de ilusórias imprecisões. Se a ciência não nos dá – a história sim – a paciência da eternidade como meta para a ordem, contamos com, pelo menos, mais um par de repetições da desordem, para, titubeando-a, aproximá-la do eixo:³ na eventualidade de uma futura publicação, a sação para novas revisões.

Mesmo em face de possíveis errâncias, Cambraia (1999, p. 19) solta um peremptório aviso: “editar é preciso”. A primeira razão que embasa essa sua assertiva é o fato de o linguista nem sempre ter acesso direto aos manuscritos, seja por causa das dificuldades impostas pelos arquivos, seja pela distância geográfica desses centros, sendo-lhe de interesse, portanto, uma edição. A segunda diz respeito ao cariz que o manuscrito expõe: abreviaturas, legibilidade, a escrita, enfim, de outrora exige conhecimento para que possa ser lida ou paciência a fim de se obter a competência necessária para tal. Por fim, Cambraia (1999, p. 14) faz uma ressalva aos fac-símiles, fotografias e cópias xerográficas: “nem mesmo esses recursos são capazes de reproduzir com absoluta fidelidade as características de um original”. Daí por que, mesmo oferecendo-se um edição fac-símile – de grande relevância por permitir que o leitor faça, sempre que queira ou careça, um cotejo da transcrição com o documento original –, é importante que a edição, ainda que de forma superficial, descreva alguns feítios materiais dos documentos apresentados.

Este trabalho desenvolve-se precisamente sob esses auspícios: o da repetição/revisão e o da edição. Inicialmente, apresentaremos as *cartas oficiais norte-rio-grandenses*, um *corpus* diacrônico editado (2012), e reeditado (2018). Tratando-se de uma temática pouco explorada – porque irrelevante ou (quase) inexistente?! –, este artigo, após esta

² “é impossível, nas árduas tarefas de transcrição e edição, não cometer erros; a única maneira de reduzir sua aparição é a revisão” (tradução nossa, doravante TN).

³ As remissões são fatais, tanto pelo mote quanto pela gravidade borgiana (cada barra vertical simples exprime mudança de verso; a dupla, de estrofe ou espaço entre título e texto): “Errância || Só porque | erro | encontro | o que não se procura || só porque | erro | invento | o labirinto || a busca | a coisa | a causa da | procura || só porque | erro | acerto: me | construo. || Margem de | erro: margem | de liberdade” (FONTELA, 2006, p. 202).

introdução (seção 1), concentra praticamente metade de seu texto à compreensão do que são essas cartas (seção 2), do que ensajou e de como se operacionalizou sua reedição (seção 3). A segunda metade (seção 4), mais conceitual (histórica e metodologicamente, embora amiúde entrecortada por exemplos ora das *cartas* ora da pesquisa grafemática que, com base nelas, se realizou), percorremos as confluências entre a Linguística Histórica e a Linguística de *Corpus* por duas sendas mui intimamente conectadas ao próprio (re)fazimento das *cartas oficiais*: a virada na Linguística Histórica brasileira que se engendrou a partir do final dos anos 90; e algumas consequências da integração dos *corpora* diacrônicos ao mundo digital, maiormente as referentes à divulgação dos dados e aos novos instrumentais tecnológicos.

2 As cartas oficiais norte-rio-grandenses (2012-2018)

As *cartas oficiais norte-rio-grandenses* são um *corpus* diacrônico nascido no mestrado de Moraes de Melo (2012), constituindo um agrupamento de 107 documentos (da C1 até a C107)⁴ que circularam, entre 1713 e 1931, no meio público com temáticas atinentes ao que hoje é o Rio Grande do Norte. As principais razões que motivaram o autor a levar a cabo essa empresa foram seu interesse, desde a época da licenciatura, pelos fenômenos da variação e mudança linguísticas e, conjunturalmente, a sua participação, de 2010 a 2012, na equipe do Para a História do Português Brasileiro no Rio Grande do Norte (PHPB-RN), projeto local cujos objetivos centrais eram estabelecer um *corpus* mínimo comum essencial para as pesquisas em Linguística Histórica desenvolvidas no Rio Grande do Norte e contribuir na construção de um banco de dados para o programa nacional Para a História do Português Brasileiro (PHPB). Como consequência, um trabalho dessa sorte legaria material empírico para investigadores dos estudos da linguagem, auxiliando sobretudo as pesquisas diacrônicas acerca do português brasileiro. O autor, que tinha como plano de mestrado inicial efetuar uma análise de ofícios novecentistas potiguares, terminou, após a entrada no PHPB-RN, expandindo seu escopo e responsabilizando-se por uma das

⁴ C1 deve ser lido “carta um”; C107, “carta cento e sete” e assim sucessivamente. Apresentamos, ainda neste tópico 2, uma lista com os intervalos de cartas que fazem parte de cada quarto de século, desde a primeira metade do XVIII até a segunda metade do XX.

categorias, as cartas oficiais, do “*corpus* mínimo comum – manuscrito” estabelecido pelo grupo nacional.

Ao lado da edição – que constou como apêndice da dissertação e foi incorporada ao conjunto de *corpora* reunidos no site do PHPB e disponibilizados para consulta de todos os interessados,⁵ estava o corpo da pesquisa: uma caracterização das cartas, organizadas do ponto de vista da Paleografia, da Diplomática e mais pormenorizadamente das Tradições Discursivas (TD). Através desta última vertente, foram analisadas macroestruturas (no caso em tela, os gêneros textuais) e as microestruturas do tipo fórmula textual presentes no *corpus*.⁶

Na sua tese de doutorado, Morais de Melo (2018) procedeu a uma análise em que o texto cedesse seu protagonismo a outro ator, os usos (orto)gráficos sem transcendência fônica, passando as cartas a funcionar como palco para essa nova averiguação. Para tanto, deliberou revisar o labor editorial de 2012 para dirimir eventuais erros, conforme a lógica supracitada de Almeida Cabrejas (2013, p. 15), bem como excluir quaisquer textos não lavrados no Rio Grande do Norte a fim de situar o resultado final o máximo possível nas trilhas de uma dialetologia histórica mais controlada. As duas ações continham, em seu horizonte, a expectativa de engendrar um *corpus* que superasse o primeiro em rigor, fidedignidade e transparência de modo a atender melhor aos estudos científicos, mormente no campo da Linguística Histórica voltada para o português brasileiro. No tópico seguinte, “Por que reeditar?”, serão esmiuçadas melhor essas duas decisões e as demais que alteraram o primeiro *corpus*, gerando as *cartas oficiais norte-rio-grandenses* de 2018, um conjunto ordenado de 129 cartas (da C1até a C129) escritas no Rio Grande do Norte e trocadas por pelo menos um representante da administração pública no intervalo de 1713 a 1950.

O qualificativo “oficiais” é empregado para nomear o *corpus*⁷ em consonância com a seguinte definição retirada do *Manual de redação da Presidência da República* (BRASIL, 2002, p. 4):

⁵ Disponível em: <https://sites.google.com/site/corporaphpb/home/corpora-manuscritos/manuscritos-rio-grande-do-norte>. Acesso em: 12 jan. 2020.

⁶ Um recorte dessas análises pode ser lido em Morais de Melo e Lima (2016a, 2016b) e Morais de Melo (2017).

⁷ Uma explicação mais esmiuçada do porquê das expressões “cartas” e “oficiais” para batizar o *corpus* encontra-se em Morais de Melo (2012, p. 90-92).

As comunicações oficiais são necessariamente uniformes, pois há sempre um único comunicador (o Serviço Público) e o receptor dessas comunicações ou é o Serviço Público (no caso de expedientes dirigidos por um órgão a outro) – ou o conjunto dos cidadãos ou instituições tratadas de forma homogênea.

Contudo, esse paradigma é relativizado por Barbosa (2002, p. 428), para quem os “documentos da administração colonial em circulação pública” podem ser de duas naturezas: os de *caráter deliberativo oficial*, respeitando aos documentos emanados de órgãos ou autoridades do governo, do Serviço Público, em concordância com os conceitos focalizados acima; e os de *requerimento pessoal*, que são endereçados a representantes ou entidades do serviço público, mas não são deste necessariamente provenientes, podendo terem sido escritos por cidadãos externos à administração pública. É precisamente essa definição que nos brinda Barbosa (2002) a que melhor convém às *cartas oficiais*.

Inclui-se nesse círculo de correspondências públicas a esfera jurídica, que não participava do primeiro *corpus*, de 2012, uma vez que, para a reedição de 2018, grande parte das cartas obtidas no Arquivo Público Estadual do Rio Grande do Norte provinha de transações judiciais. Embora estranho à conformação original, essa esfera indiscutivelmente integra a administração pública, o que se comprova por qualquer busca na internet em que se encontra, a modo de ilustração, que o judiciário é um dos três poderes do Estado (público, portanto, não privado).

A grande maioria dos documentos de nosso *corpus*⁸ (102 dos 129, isto é, 79,06%) tem como missivistas pessoas que representam o Governo, sejam capitães-mores (C1, C2 e C3, por exemplo), provedores (C5 e C6, v.g.), escrivães da Fazenda Real (C8 e C31, e.g.), presidente de uma província (C62, C67), secretários do governo (C72, C80) ou juizes (C108, C109). As demais 27 cartas (20,93%) são enviadas para (mas não por) membros da administração pública, sendo a maioria delas (20 documentos) requerimentos, como a C32, de Francisco de Paulo Moreira, solicitando confirmação de seu posto de cirurgião; ou a C119, de um 3º sargento ao presidente do Supremo Tribunal de Justiça.

⁸ Neste artigo, o vocábulo *corpus* empregado de modo absoluto dirá respeito sempre às *cartas oficiais norte-rio-grandenses* reeditadas em 2018. A referência à primeira edição sempre virá acompanhada de algum especificador do tipo “o *corpus* de 2012”, “o primeiro *corpus*” ou soluções afins.

Somente em duas dessas 27 cartas não é nenhum dos interlocutores um membro estrito da administração pública, na C42 e na C111: a primeira é uma cópia de um ofício enviado por José Barbosa e João Luís Pereira a membros do clero com o intuito de que estes intercedessem à Rainha para a obtenção de verba em prol da construção de um hospício de Santo Antônio na cidade de Natal; e a segunda, um ofício enviado por Arthur Napoleão Soares de Macedo ao coronel José Soares Filgueira Sobrinho questionando o militar sobre a compra de terras. De toda forma, além de, na C42, haver uma remissão final à Rainha, aparecem, como uma das pontas da correspondência, em ambos os casos, componentes que, a seu modo, se incluíam no *modus operandi* da administração brasileira. Afinal, não é gratuita a inclusão, na obra *A administração no Brasil Colonial*, organizada por Salgado (1985), dos quatro seguintes eixos organizacionais: “Estrutura judicial”, “Administração fazendária”, “Organização militar” (a exemplo da C111, escrita já fora dos lindes coloniais, mas num contexto em que essa organização exercia papel de destaque na administração pública) e “Administração eclesiástica” (no caso da C42). Sobre essa última instância, de modo especial, comprovando seu papel integrante na dinâmica administrativa (pelo menos na colonial, da qual a C42 faz parte), lemos:

A expressão ‘funcionários eclesiásticos’ dá bem a medida de como a Igreja nascente nas terras americanas dependia do Estado português, situação que se prolongou por todo o período colonial e durante o Império, cujas raízes, conforme já dissemos, estavam no direito do padroado. Além da integração político-religiosa, a Coroa se beneficiou, e muito, da sua condição de administradora dos dízimos eclesiásticos, em muitas regiões talvez a principal fonte da renda colonial (SALGADO, 1985, p. 115).

Afora todas essas justificativas, encetadas pela concepção de Barbosa (2002) sobre documentos da administração pública e aliadas às de Salgado (1985), essas cartas que não são lavradas por pessoas que assumem, no sentido mais estreito, função na máquina pública alinham-se às demais por uma nota que não se pode perder de vista: todos os 129 textos que integralizam as *cartas* representam uma minoria letrada e com razoável ou alto nível de instrução numa sociedade em que rareavam as condições de oferta e acesso à educação.

Além dessas questões por trás de “oficiais”, o uso de “cartas oficiais” é também uma declarada evocação às cartas oficiais da Paraíba

organizadas por Fonseca (2003, p. 120-121) e à decisão nomenclatória da autora:

Dos documentos selecionados, encontram-se neste *corpus*, de acordo com a classificação feita por Martinheira: *carta de lei*, *carta régia*, *aviso dos secretários*, *ofícios*. Dentre esses, os *ofícios* serão focalizados com a denominação geral de cartas por serem maioria absoluta e por apresentarem maior heterogeneidade.

Embora na nova edição os *ofícios* não representem a maioria absoluta (no antigo, ocupavam 70% do *corpus*), ainda é o gênero mais expressivo, uma vez que das 129 cartas (denominação geral), 63 (48,83%) são *ofícios* que formam, somados às 25 cartas,⁹ 68,2% do total. A maior diferença entre as primeiras *cartas* e a reedição de 2018 foi o crescimento relevante de requerimentos, que constituem agora o segundo gênero mais recorrente, com 26 exemplares (20,15%). Os três juntos – *ofícios*, *requerimentos* e *cartas* – expressam 88,35% das 129 *cartas oficiais norte-rio-grandenses*.

Outra elucidação é sobre o porquê da marcação do termo com itálico. Decidimos fazê-lo para ressaltar a identidade do *corpus* organizado e editado nesta tese, e evitar, pelo registro de cartas oficiais norte-rio-grandenses, assim, sem marcas, a abertura para se ler, nas menções que fazemos da expressão, a ideia, lassa e equívoca para este estudo, de estarmos tratando de qualquer carta escrita ou no Rio Grande do Norte ou por norte-rio-grandenses e em qualquer período, a despeito do perímetro bem delimitado com o qual trabalhamos e no qual as análises aqui promovidas e suas conclusões estão estritamente circunscritas.

⁹ O *ofício* é um gênero de estrutura bastante semelhante à da carta e da carta régia, diferenciando-se, contudo, por uma razão de ordem sociofuncional: são cartas oficiais trocadas, em sua maioria, entre membros da administração pública, sob a condição de nenhum dos coenunciadores, o remetente ou o destinatário, ser o Rei. A tradição da carta é, formalmente, muito símile à do *ofício*, mas é ativada pela necessidade de comunicar algum assunto relativo à administrações pública ao Rei, daí todas elas trazerem como *inscriptio* apenas o pronome de tratamento “Senhor” que, conforme explica Fonseca (2003, p. 150), é exclusivo para autoridade real. Devido à fixidez de seu endereço, o Rei, a carta goza sempre de um caráter ascendente, isto é, é escrita por alguém hierarquicamente inferior à pessoa a quem o documento é dirigido. *Inscripção* é um termo da diplomática o qual podemos encontrar na obra de Belloto (2002) e indica uma das partes que pode ocorrer no protocolo inicial (equivalente à seção pré-textual) de um documento.

Do ponto de vista metodológico, decidimos fixar um mínimo de 2.500 palavras por quarto de século. Embora o trabalho de 2018 já não estivesse atrelado a nenhuma ação do PHPB – ao contrário do empreendimento de 2012, que esteve formalmente envolvido com o projeto –, resolvemos seguir, conquanto de forma adaptada,¹⁰ a orientação dada pelo programa nacional de obter 5.000 palavras para cada metade de centúria. Vejamos abaixo as cartas de cada período seguidas por uma descrição geral sobre os documentos de cada século.

- XVIII.1.1¹¹ – 13 cartas, da C1 (1713) até a C13 (1725);
- XVIII.1.2¹² – 10 cartas, da C14 (1726*)¹³ até a C23 (1747);
- XVIII.2.1 – 12 cartas, da C24 (1756) até a C35 (1772*);
- XVIII.2.2 – 10 cartas, da C36 (1777) até a C45 (1798);
- XIX.1.1 – 14 cartas, da C46 (1806) até a C59 (1822*);
- XIX.1.2 – 16 cartas, da C60 (1833) até a C75 (1849);
- XIX.2.1 – 13 cartas, da C76 (1852) até a C88 (1874);
- XIX.2.2 – 16 cartas, da C89 (1876) até a C104 (1891);
- XX.1.1 – 14 cartas, da C105 (1913) até a C118 (1919);
- XX.1.2 – 11 cartas, da C119 (1931) até a C129 (1950).

¹⁰ A adaptação concerne a trabalharmos com quarteis, e não com metades de século, como estabelece o PHPB e como foi acatado em Moraes de Melo (2012). Essa alteração deu-se durante a pesquisa doutoral para que se pudesse controlar melhor eventuais mudanças nos usos gráficos – objeto de análise de Moraes de Melo (2018) – no eixo diacrônico investigado (1713-1950).

¹¹ O primeiro algarismo arábico após a indicação do século alude a alguma metade do século (1 representa a primeira e 2, a segunda) e a segunda cifra arábica indica os primeiros ou os últimos 25 anos dessa metade. Assim, de cada centenário correspondente, 1.1 expressa o intervalo que vai do ano 01 ao ano 25; 1.2, de 26 até 50; 2.1, de 51 a 75; e 2.2, de 76 a 00. Nessa linha, XVIII.2.1 engloba um hiato que vai de 1751 a 1775. Nas *cartas oficiais* de 2018, a primeira carta desse período é a C24, de 1756, e a última é a C35, escrita provavelmente em 1772.

¹² Agradecemos à professora Carmen Alveal, do Departamento de História da UFRN, e a seus bolsistas, por terem cedido, em 2011, a transcrição de quase 2.000 palavras da primeira metade do século XVIII para nosso *corpus*.

¹³ O asterisco informa que a data não é exata (não consta no documento), mas sugerida por Lopes (2000).

Textos do século XVIII: somando 45 cartas oficiais (19 cartas, 10 requerimentos, 6 ofícios, 3 cartas de registro, 3 certidões, 1 provisão, 1 carta de sesmaria, 1 carta régia e 1 carta patente), os documentos setecentistas de nosso *corpus* foram retirados de um conjunto de textos do Arquivo Ultramarino de Lisboa (AHU) relacionados ao Rio Grande do Norte. Esse conjunto foi catalogado por Lopes (2000), professora aposentada do departamento de História da UFRN, como parte do projeto Resgate Barão do Rio Branco, e disponibilizado pela equipe de História da referida universidade. Esses documentos do AHU inventariados por Lopes (2000) abarcam dois séculos de cartas oficiais, datadas as mais antigas de 1623 até as de 1823. Interessaram-nos apenas aquelas escritas a partir de 1701, conforme nosso recorte cronológico. Mesmo comungando da afirmação de Acioli (1994, p. 62), para quem “a dificuldade de leitura não reside obrigatoriamente no retrocesso cronológico”, porquanto houve, de fato, cartas setecentistas fáceis de serem transcritas, vimos nessa fonte os textos mais difíceis para serem transcritos, principalmente pela ilegibilidade de alguns e pela considerável presença de documentos com trechos em escrita que julgamos ser processada,¹⁴ ambas evitadas em nossa seleção.

Textos do século XIX: contendo 59 cartas oficiais (46 ofícios, 6 cartas, 5 requerimentos e 2 relatórios), os textos oitocentistas mais antigos das *cartas oficiais* ainda foram obtidos dos documentos do AHU catalogados por Lopes (2000). Mas cerca de 50% da primeira metade e todos os da segunda metade da centúria em pauta são ofícios localizados no Instituto Histórico e Geográfico do Rio Grande do Norte (IHGRN), preservados em oito caixas de arquivo, havendo em cada uma delas subpastas organizadas por data. O ofício mais antigo oriundo do Instituto de que dispusemos em nosso *corpus* data de 1812, e o mais recente, de 1891. As imagens desses documentos não estavam captadas, como os do século XVIII, exigindo um trabalho de campo para conhecer o acervo, selecionar as cartas e fotografá-las a fim de transcrevê-las.

¹⁴ Segundo Acioli (1994, p. 42), foi uma escrita que promovia “uma rapidez de traçado extraordinária”, tornando-se “a preferida dos escrivães”. É uma escrita cursiva e de traços deturpados (degeneração da Cortesã, consoante a autora), caracterizada pela separação irregular das palavras, confusão no traçado das letras e grande quantidade de traços supérfluos.

Textos do século XX: com 25 cartas oficiais (11 ofícios, 11 requerimentos, 1 auto, 1 termo e 1 certidão) – número reduzido, quando comparado aos dos séculos anteriores, por nosso *corpus* cobrir apenas a primeira metade dos Novecentos –, os documentos foram conseguidos graças a visitas ao Arquivo Público Estadual do Rio Grande do Norte para seleção, fotografação e posterior transcrição para a edição. Esse intervalo de 50 anos é marcado pela presença majoritária de documentos do circuito jurídico, com cartas expedidas, *v. g.*, por juízes, promotores públicos e vários pedidos/ordens de habeas-corpus (categorizados como requerimentos), o que não se verificava nas *cartas oficiais* de 2012.

Duas últimas considerações sobre as *cartas* de um modo geral, antes de entrarmos nas razões que motivaram sua reedição. A primeira tange à sua ordenação. A numeração que vemos, de 1 a 129, atende quase absolutamente à cronologia das cartas, com exceção da C111 e C112. Devido à desatenção, só percebemos, em fase bastante avançada da tese, que a posição deveria ser invertida, posto que a C111 data de 26 e 27 de agosto de 1918, ao passo que a C112 foi escrita em 10 de agosto do mesmo ano, devendo, por conseguinte, ser a C111. Devido aos vários desdobramentos que essa mudança implicaria e diante do pouco tempo que nos restava quando avistamos o erro, preferimos deixá-lo.

O segundo ponto tem a ver exatamente com essa dupla datação da C111. São duas datas porque, na realidade, existem dois ofícios escritos, não obstante no mesmo fôlio. O primeiro é remetido por Arthur Napoleão Soares de Macedo ao coronel José Soares Filgueira Sobrinho, perguntando ao militar se ele comprara quinhentos braços de terra à firma Severo & Irmão. Ao pé deste documento, conforme solicitado por Arthur Soares (<Preciso que VS., abem da verdade, me | responda ao pé desta>), há um segundo ofício com a resposta do coronel. Se optamos por evitar a correção na numeração das C111 e C112, pois, ainda que se tratasse de apenas uma permuta, reclamaria uma cadeia de pequenos ajustes, imagine-se a inclusão de uma outra carta: afetaria a numeração das 19 cartas que sucedem a C111, gerando a necessidade de um sem-fim de reparos. Mantivemos, assim, uma carta dupla, que, desmembrada, faria com que esta nova reedição contabilizasse não 129, mas 130 documentos.

3 Por que reeditar?

Nesta seção, apresentaremos as razões que subjazem a decisão de revisitar e reeditar as *cartas oficiais norte-rio-grandenses* (MORAIS DE MELO, 2012) através de cinco eixos: a exclusão de cartas, a revisão das transcrições, a procura de novas cartas, a transcrição desses novos documentos e a nova apresentação desse *corpus* diacrônico.

3.1 Exclusão de cartas

A primeira edição das *cartas* constitui um conjunto de 107 missivas escritas na administração pública no intervalo de 1713 a 1931 em que questões vinculadas ao Rio Grande do Norte são abordadas.

Quantitativamente, 82, 24% do *corpus* de 2012 é redigido no Rio Grande do Norte, o que representa 88 das 107 cartas, ao que se somam 5 (4,6% do total) escritas no Rio de Janeiro, 1 (0,9%) em Sergipe, 4 (3,7%) em Lisboa e outras 9 sem identificação tópica precisa. Daí pode-se afirmar que 87,8% (a soma das cartas escritas no Rio Grande do Norte, no Rio de Janeiro e em Sergipe) representam o que Barbosa (2002) chama de português *no* Brasil, isto é, textos lavrados em solo brasileiro, sem terem sido emanados, necessariamente, de um punho brasileiro (entenda-se uma pessoa que, mesmo não tendo nascido no Brasil, tenha vivido a maior parte de sua vida, principalmente seus anos de formação, em terras brasileiras).

Embora nem todas as cartas tivessem sido escritas no Rio Grande do Norte, decidimos, à época, cunhá-las *cartas oficiais norte-rio-grandenses* no rasto de Acioli (1994, p. 55, grifo nosso), para quem, diante da influência ibérica impressa patentemente nos manuscritos brasileiros, avisa:

Entendeu-se que brasileiros seriam todos os manuscritos relacionados com o nosso país, quer oriundos do Brasil, quer de Portugal. Assim sendo, tanto cartas remetidas da colônia quanto documentos régios ou consultas do Conselho Ultramarino, despachados da metrópole, foram considerados brasileiros quando o assunto em questão descreve problemas desta possessão portuguesa na América.

Após vários diálogos no período da tese, especialmente com os orientadores, decidimos excluir, no processo de reedição, todos os documentos do primeiro *corpus* que não tivessem sido escritos no

Estado. Muito embora o contorno analítico levado a cabo em Morais de Melo (2018) – o exame dos usos gráficos sem transcendência fônica, um plaino, portanto, sumamente gráfico – não estivesse a princípio atrelado a variáveis diatópicas, a escrita, em sua latência, pode revelar liames com o geográfico, máxime no que toca aos virtuais rastros da oralidade deixados nos traços visuais.¹⁵ E como aspirávamos a que o novo *corpus* não estivesse a serviço apenas da fatia investigativa executada dentro das balizas doutorais, mas que se constituísse em um *corpus* útil à mais ampla gama de finalidades dentro das lindas que oferece, resolvemos proceder a essa talha.

¹⁵ Numa disciplina sobre metodologia de pesquisa ofertada por José Luis Ramírez Luengo aos alunos de mestrado e doutorado da Universidad Autónoma de Querétaro (México), em 2015, a que o autor deste artigo assistiu na condição de ouvinte, comentou-se mais de uma vez sobre a existência, entre os historiadores da língua do domínio hispânico que se interessam por questões grafemáticas, de certo entendimento segundo o qual a variável geográfica não acarreta influxo nas realizações gráficas sem transcendência fônica. Eximiam-se, porém, dessa aparente incolumidade as possíveis marcas da oralidade decalcadas na escrita e alguns usos gráficos relativos ao período de formação das escritas romances, na Idade Média. No que tange a esse último ponto, Sánchez-Prieto Borja (2006), por exemplo, mostra como as tradições de escrita medievais também eram criadas no seio de povoados, geralmente pela fonte irradiadora de mosteiros, chancelarias e câmaras régias, como se pode depreender de trechos como: “Claro que la valoración de las [*sic*] usos díspares que muestran los documentos también puede hacerse en clave sociolingüística, teniendo en cuenta posibles diferencias entre un núcleo de población en el que los notarios públicos adoptaban usos que si no habían logrado una estandarización sí estaban extendidos más allá del primitivo reino de Castilla, mientras que la tradición monástica de Vega de Espinaredo se revela más arcaizante, o quizá simplemente más acorde con la lengua hablada” [“Claro que a valoração dos usos díspares que os documentos mostram também pode se tornar chave sociolingüística, levando em conta possíveis diferenças entre um núcleo de povoação no qual os notários públicos adotavam usos que, se não haviam alcançado uma standardização, estavam, sim, estendidos para além do reino de Castela, enquanto a tradição monástica de Vega de Espinaredo se revela mais arcaizante, ou talvez simplesmente mais de acordo com a língua falada” (TN)] (SÁNCHEZ-PRIETO BORJA, 2006, p. 240) e “Desde años me ha llamado la atención las marcadas diferencias dentro del llamado ‘castellano alfonsí’ entre los grandes códices de la Cámara Régia y los documentos salidos de la Cancillería” [“Desde muitos anos tem me chamado a atenção as diferenças marcadas dentro do chamado ‘castelhano alfonsino’ entre os grandes códices da Câmara Real e os documentos saídos da Chancelaria” (TN)] (SÁNCHEZ-PRIETO BORJA, 2006, p. 241).

Além dessas 19 cartas eliminadas por critérios diatópicos, também ficou decidida a eliminação de cópias que não fossem de época. Destarte, apenas mantivemos, das antigas cartas, a C23 (atual C27), C31 (atual C42) e C66 (atual C85). Por outro lado, todas as 21 cartas oficiais que compunham a primeira metade do século XX foram descartadas por se tratar de códices diplomáticos em registros,¹⁶ todos transcritos por uma mesma mão (registradas talvez num mesmo dia) com a intenção de arquivar em livro os ofícios, não tendo sido efetivamente, as referidas cópias, enviadas. Ao final desse processo de filtro, 35 das 107 cartas de 2012 tinham sido rejeitadas para inteirar o *corpus* de 2018.

3.2 Revisão das transcrições

Silva Neto (1956, p. 27) discorre sobre uma lista de erros “mais comumente observados na leitura de manuscritos medievais”. A despeito de ser obra escrita na década de 50 e estar pensada para documentos medievais, o que nos relata o grande filólogo é um rol de alertas bastante atuais e de extrema importância para qualquer pesquisador que, no seu afã investigativo, tenha de praticar o exercício de transcrição textual, independentemente da época do documento a ser copiado.¹⁷ São eles os erros devidos à má compreensão de letras, a desconhecimentos de fatos linguísticos, à ignorância de siglas e abreviaturas, os saltos-borrões¹⁸ e os erros devidos à má separação das sílabas. Muito embora, na nossa labuta do mestrado, tenhamos efetuado uma transcrição cuidadosa, atenta sobremaneira a esses quesitos, e que foi revisada pelo autor, pelos orientadores e talvez pela equipe do PHPB nacional, uma vez que foi

¹⁶ Os códices diplomáticos estão agrupados em livro. Quando esses códices são do tipo registro, as cópias dos documentos exarados são organizadas por sequência cronológica.

¹⁷ Uma prova explícita da validade desse elenco que, na década de 50 fichou Silva Neto (1956), podemos encontrá-la em Rumeu, Barbosa e Callou (2002). Os autores escolhem alguns exemplos de textos coloniais transcritos, a maioria deles ligados à Biblioteca Nacional (anais, acervos), em que houve ou equívoco na transcrição ou na interpretação feita da transcrição por causa de um dos cinco erros (todos os cinco são ordeira e ilustrativamente contemplados no artigo dos três pesquisadores) de modo a evidenciar as achegas do grande filólogo brasileiro.

¹⁸ Assim inicia o tópico destinado ao problema dos saltos-bordões: “Muitas vezes, por falta de atenção, o copista saltava uma ou várias linhas” (SILVA NETO, 1956, p. 31).

disponibilizada no site de *corpora* do projeto,¹⁹ achamos por bem repassar cada uma das 72 cartas mantidas daquele *corpus* de 2012. Descansados pelo distanciamento dos anos, os olhos – quiçá sazoados, pelo repouso, para o labor transcripcional – lograram decifrar muitas palavras (e até frases) que tinham sido assinaladas com [*inint.*] (ininteligíveis), bem como localizar alguns equívocos (efeitos ora de uma, ora de outra causa de deslize alistada por SILVA NETO, 1956), que foram devidamente consertados.

Algumas cartas, no tempo do mestrado, foram transcritas no próprio IHGRN, tendo sido suas fotos tiradas, para o trabalho de edição, dos documentos arquivados nas caixas de ofícios que nos foram disponibilizadas.

IMAGEM 1 – Fotos com caixas de ofícios disponibilizadas pelo IHGRN entre 2010 e 2012



Fonte: Elaboração própria.

Algumas das fotos tiradas na época, culpa de nosso descuido, não ficaram tão nítidas como deveriam. Durante o processo de revisão, vimos a necessidade de sanar algumas dúvidas voltando ao Instituto. Infelizmente, porém, ao longo de todo o período do doutorado (2014-2018), o mais importante arquivo público do Estado permaneceu com suas portas fechadas para pesquisa devido a reformas que acabaram sendo embargadas num processo que parecia sem fim. Em mais de uma ocasião, explicando a responsáveis pela casa nossa necessidade de consumir apenas algumas verificações para um trabalho que gozou de todo o apoio da própria equipe do IHGRN entre 2010 e 2012, tentamos, sem êxito, ter acesso às caixas de ofícios (mostradas nas fotos acima)

¹⁹ <https://sites.google.com/site/corporaphpb/home/corpora-manuscritos/manuscritos-rio-grande-do-norte>.

nas quais estavam guardadas essas cartas oficiais. Em março de 2017, o Instituto reabriu suas portas, mas apenas de modo parcial: para exposições e eventos. Continua fechado para pesquisa, impedindo a consecução dessa nova fotografiação.²⁰

No final das contas, mesmo diante de persistentes percalços, Morais de Melo (2018) oferece, com absoluta certeza, um *corpus* com transcrições mais atiladas, com menos lacunas e com menos erros. Não perfeito, porém. Tão pertinente e irrepreensível é o comentário de Almeida Cabrejas (2013) acerca do trabalho de transcrição e tão bem resume o nosso sentimento ao final da reedição que reiteramos, malgrado já possa ser lido na introdução deste artigo, uma parte dele: “incluso donde no hay dudas, es imposible en las arduas tareas de transcripción y edición no cometer errores; la única manera de reducir su aparición es la revisión”²¹ (ALMEIDA CABREJAS, 2013, p. 15). Sem dúvida, caso venhamos a publicar em livro as *cartas oficiais norte-rio-grandenses* (e esperamos que, para este então, já tenha retomado seu pleno funcionamento o IHGRN, de modo que possamos rever alguns documentos e precisar algumas outras informações), procederemos a mais uma revisão com o intuito de identificar e solucionar o máximo possível de problemas remanescentes, aproximando, assim, nosso ofício transcricional de um estado ótimo.

3.3 Procura de novas cartas

Quando elaboramos a edição de 2012, tínhamos como meta suprir as demandas do PHPB, isto é, obter 5.000 palavras para cada metade de século da primeira metade do século XVIII até a segunda do XX. Como estávamos responsáveis por textos manuscritos, abandonamos o intervalo dos últimos 50 anos (1950-2000) por não encontrarmos, nas caixas de documentos a que tivemos acesso, textos escritos à mão após 1950. Todos já estavam tipografados. Após a supressão dos 35 documentos desenredada no tópico acima, terminamos com uma quantidade bastante irregular de palavras para cada corte cronológico. Ainda que nossa tese não estivesse atrelada a nenhum projeto que assentasse uma quantidade

²⁰ Em outubro de 2020, lê-se “Temporariamente, não está aberto para pesquisa em seus acervos”, na página principal de seu *site*: <http://ihgrn.org.br/instituicao/instituto>.

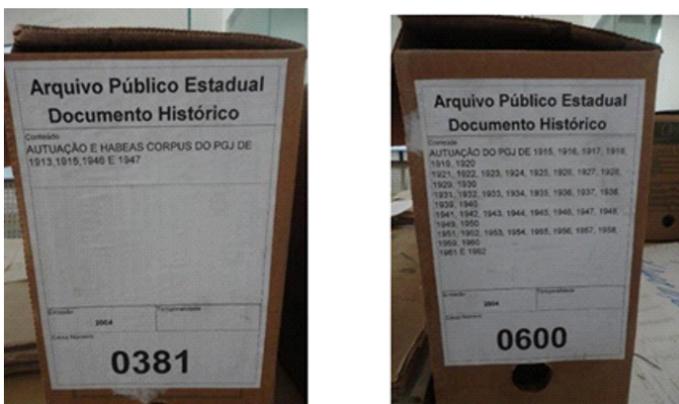
²¹ “Inclusive onde não há dúvida, é impossível nas árduas tarefas de transcrição e edição não cometer erros; a única maneira de reduzir sua aparição é a revisão” (TN).

mínima de palavras e mesmo sabendo da possibilidade de se trabalhar com quantidades desiguais (efetuando-se os cálculos de forma proporcional ao total de cada lapso de tempo delimitado, neutralizando-se, dessarte, por meio de um equilíbrio relativo, as somas conflitantes de palavras), optamos por reestabelecer um equilíbrio absoluto entre os intervalos temporais.

Para se conseguir preencher os vãos abertos após a exclusão das 35 cartas, foi preciso voltar à lide primeira: ir a arquivos, procurar textos, selecioná-los, trazer à luz velhos espíritos em escrita sobrevividos. Parafraseando o efusivo mote que nos presenteia Tarallo (1990, p. 175), *picaretas em punho*, voltamos a *cavar*. Para todo o século XVIII até 1823, ainda podíamos contar com a farta base de dados do Arquivo Ultramarino de Lisboa catalogada por Lopes (2000) e a cujo conjunto integral de fotos temos acesso. O problema maior era recauchutar as baixas, que não foram muitas, do século XIX após 1823 e, de modo crítico, todo a primeira metade do século XX. Malogrados os intentos de conseguir aceder ao IHGRN, vimos no Arquivo Público do Estado nossa nova manancial de documentos.

Revivemos todas os estágios por que passamos entre 2010 e 2012: fizemos algumas visitas ao Arquivo, quando, guiados pelo crivo do olho para reconhecer os textos que fossem aparentemente mais legíveis e assegurados de que tinham sido escritos em Natal e circulados na administração pública, fotografamos uma série de cartas para posterior sondagem e eventual transcrição e inclusão como parcela documental do *corpus*.

IMAGEM 2 – Algumas das caixas cedidas para pesquisa pelo Arquivo Público



Fonte: Elaboração própria.

A maioria das missivas preservadas nas caixas que nos foram cedidas era relativa a trâmites da burocracia jurídica. Não nos agradou, a princípio, incluirmos sob o rótulo de “cartas oficiais” textos de ordem jurídica. Todavia, uma vez que quase todos os textos se afinavam com os gêneros com que já vínhamos trabalhando (ofício e requerimento,²² majoritariamente) e que o poder judiciário se incorpora à administração pública, sofreamos a inquietude e aceitamo-los como “cartas oficiais”.²³ A seguir, elencaremos, para cada quarto de século, a quantidade e tipo de documentos na configuração final após o trabalho de revisão das cartas remanescentes, seleção e transcrição das novas. Entre parênteses, temos a quantidade de textos inéditos (num total de 57), isto é, que não faziam parte de Moraes de Melo (2012). Assim, “7 cartas (2)”, na primeira linha, significa que há 7 cartas, duas das quais foram obtidas durante os afazeres doutorais; e “8 requerimentos (8)”, na última linha, revela que dos 8 requerimentos, todos são “novos”.

²² Cinco dos documentos que classificamos como requerimentos a rigor são pedidos/ordem de *habeas corpus*. Analisando, não obstante, sua estrutura com base no estudo prévio que realizamos sobre as tradições discursivas dos gêneros textuais/espécies documentais que compunham o *corpus* de 2012 (MORAIS DE MELO, 2012), verificamos que se tratava da mesma macroestrutura, contendo, ainda por cima, algumas das mesmas expressões formulaicas que mapeamos para esse gênero no referido estudo. Além disso, após conversa com advogados a quem mostramos esses cinco documentos, eles afirmaram que, antes de serem pedidos/ordens de *habeas corpus*, todos são com efeito requerimentos.

²³ Inclusive se olharmos a tabela que pautou as agendas de trabalho do PHPB disponível em Hora e Silva (2010, p. 423), para a macrocategoria “corpus mínimo comum - manuscritos”, notaremos que, dentre as seis tipologias definidas (a saber: testamento, processos-crime, atas da câmara, cartas particulares, cartas da administração privada e cartas oficiais), é na de cartas oficiais que esses textos que circularam no meio jurídico melhor se inserem.

QUADRO 1 – Relação com intervalo de cartas, quantia de cada gênero e, entre parênteses, número de exemplares inéditos para cada quartel de século

SÉCULO XVIII	XVIII.1.1	C1 a C13 – 7 cartas (2), 3 cartas de registro, 1 requerimento (1), 1 certidão e 1 ofício. TOTAL: 13 cartas, sendo 3 inéditas.
	XVIII.1.2	C14 a C23 – 8 cartas (5), 1 carta de sesmaria (1) e 1 requerimento. TOTAL: 10 cartas, sendo 6 inéditas.
	XVIII.2.1	C24 a C35 – 4 cartas (1), 2 certidões (1), 1 carta régia, 1 carta patente (1), 3 requerimentos e 1 ofício. TOTAL: 12 cartas, sendo 3 inéditas.
	XVIII.2.2	C36 a C45 – 1 provisão, 5 requerimentos (4) e 4 ofícios. TOTAL: 10 cartas, sendo 4 inéditas.
SÉCULO XIX	XIX.1.1	C46 a C59 – 6 cartas (6), 4 requerimentos (4), 4 ofícios (3). TOTAL: 14 cartas, sendo 13 inéditas.
	XIX.1.2	C60 a C75 – 15 ofícios e 1 requerimento. TOTAL: 16 cartas.
	XIX.2.1	C76 a C88 – 7 ofícios, 2 relatórios (2) e 4 ofícios (1). TOTAL: 13 cartas, sendo 3 inéditas.
	XIX.2.2	C89 a C104 – 16 ofícios. TOTAL: 16 ofícios.
SÉCULO XX	XX.1.1	C105 a C118 – 11 ofícios (11) e 3 requerimentos (3). TOTAL: 14 cartas, todas inéditas.
	XX.1.2	C119 a C129 – 8 requerimentos (8), 1 auto (1), 1 certidão (1) e 1 termo (1). TOTAL: 11 cartas, todas inéditas.

Fonte: Elaboração própria.

3.4 A transcrição das cartas

Para a transcrição, continuamos seguindo grande parte das normas propostas pelo PHPB. O modelo de transcrição foi publicado por Mattos e Silva (2001), intitulado “Normas para transcrição de documentos manuscritos para a história do português brasileiro”. Em 2010, elas foram repassadas para todas as equipes regionais do projeto, dentre as quais para o PHPB-RN, com algumas atualizações, chamando-se “Normas de transcrição de documentos manuscritos e impressos”. O modelo direciona uma edição semidiplomática dos documentos, que Spina (1977, p. 79) também chama de diplomático-interpretativa, uma vez que “vai mais longe na interpretação do original, pois já representa uma tentativa de melhoramento do texto, com a divisão das palavras, o desdobramento das abreviaturas (trazendo as letras, que não figuram no original, colocadas entre parênteses) e às vezes até com pontuação”.

Cambráia indigita esse tipo de transcrição como presumivelmente a melhor para uma edição que tenha em mente os linguistas:

Quando se tem em mente como principal público-alvo (mas não o único) lingüistas, o tipo mais adequado parece ser a **edição semidiplomática**, pois esse tipo de edição, em uma versão um pouco mais conservadora do que como definida por Spina (1994), tem como vantagem respeitar ao máximo as características do original, fazendo-se, no entanto, pequenas intervenções (sempre assinaladas!) com o objetivo de viabilizar a leitura de seu público. Embora voltada para um público em especial, isto não significa que estudiosos de outras áreas não possam também utilizá-la: este tipo de edição serve também, por exemplo, a pesquisadores de literatura ou historiadores que, com um pequeno esforço inicial para se habituarem ao sistema de transcrição adotado, certamente não encontrarão maiores dificuldades na leitura do texto (CAMBRAIA, 1999, p. 16).

No caso dos critérios propostos pelo PHPB, não há inclusão de sinais de pontuação e as palavras que estão unidas por razões lingüísticas, como a natureza clítica de algum termo, são mantidas, sendo separadas, contudo, aquelas unidas devido à morfologia das letras do original manuscrito, que, por apresentarem geralmente as extremidades alongadas, alcançam a palavra seguinte e a ela se ligam por seus ornatos. Numa transcrição diplomática, por outro lado, ocorre uma “reprodução tipográfica do original manuscrito, como se fosse completa e perfeita cópia do mesmo” (SPINA, 1977, p. 78).

3.5 Lição justalinear e facsímile

Desde a época do mestrado, não julgamos o arranjo dos documentos transcritos proposto pelo PHPB como o melhor no sentido de promover uma justa correspondência visual entre a transcrição e o documento original. A funcionalidade do formato se via ainda mais comprometida à medida que pensávamos numa edição facsimilar, pois invalidava parcialmente o potencial de localização e comparação que o facsímile compreende. Trazemos, à continuação, um recorte da C61 seguido pela transcrição antiga acompanhada pela atual. Através delas, podemos contrastar seus efeitos, perceptíveis sobremodo quando vistos em cotejo com a foto.

Aproveitemos, antes, para, a partir desse microexemplo que virá, detalhar alguns poucos parâmetros que, embora talvez já sejam dados por certos numa lição justalinear, julgamos merecedores de menção.

Um dos vários trabalhos que efetuam esse tipo de registro nos moldes que seguimos, conquanto não reserve nenhuma parte do livro para dar, como o fazemos (inda que assaz sumariamente), um esclarecimento sobre os critérios seguidos, é Megale, Toledo Neto e Fachin (2009, p. 29-61). Entendemos que a bússola básica desse tipo de apresentação dita: preservar o máximo possível da disposição visual que caracteriza o documento original. Dessa maneira, um texto que esteja mais à direita será transcrito mais à direita (vejamos, no quadro abaixo, na coluna reservada à lição justalinear, a localização da assinatura <Bazilio Quaresma Torreão>, colacionando-a com a foto), as proporções (veja que <Bazilio> está alinhado verticalmente a <Natal> de forma parelha ao original), os espaços (note-se a sangria em <Deus Guarde>) tentarão ser mantidas etc.

Esse rigor, sem embargo, não é total, máxime no que respeita aos espaçamentos. Ao atentarmos para o fragmento de foto, divisamos que o espaço entre a linha que termina em <1833> e a assinatura de Basílio Torreão é maior do que o que se emprega entre a assinatura e a linha seguinte (que começa com <Ill.^{mos} Senhores>). Na transcrição, contudo, as duas foram registradas como um espaço simples (uma linha em branco apenas). Há outros casos, como o da C1, em que a assinatura, no manuscrito, se dá no final da página, ao passo que o texto imediatamente anterior (o final da seção textual) se encontra na parte de cima. Seriam necessários, caso se quisesse ater, na transcrição, para esse aspecto, pelo menos 20 ou 25 linhas em branco para se emular o que figura o original. Deixamos, entretanto, uma apenas.

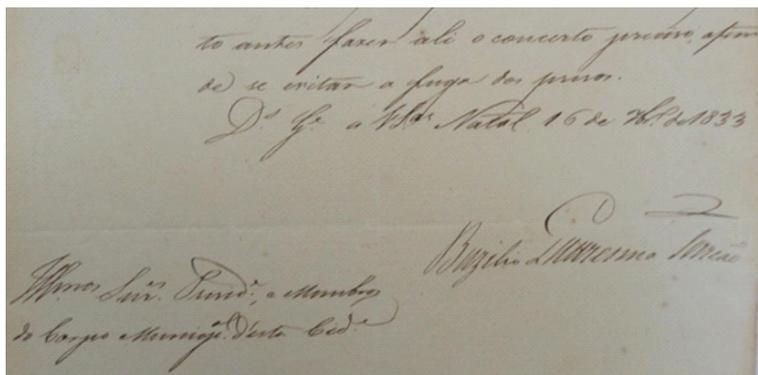
Duas últimas convenções devem vir à luz antes de irmos à mostra. A primeira diz respeito a uma situação em que a transcrição não honrará a biunivocidade linha a linha ansiada pela lição justalinear. Haverá casos esporádicos em que a transcrição de uma linha do documento original, devido à quantidade de abreviaturas desenvolvidas, não pôde ser comportada numa única linha do Word.

Nesses eventos casuais, lançaremos mão do colchete aberto ([) para indicar que aquela linha introduzida por esse sinal não corresponde a uma linha do original, mas ao final da linha anterior. Por fim, uma decisão que diz respeito à disposição da edição nas páginas da tese e que foi tomada em vista de se tentar reduzir um pouco o volume do trabalho. Sempre que a carta for suficientemente pequena de modo que caibam, numa única página, tanto a foto do fólio quanto sua transcrição,

elas serão justapostas nessa mesma folha. Do contrário (a maioria dos casos), surgirá, de cada fôlio e nesta ordem, a transcrição e a imagem cada qual numa página.

Casos explanados, aduzimos o exemplo.

IMAGEM 3 – Fragmento da C61



Fonte: Elaboração própria.

QUADRO 2 – Comparação entre a disposição da transcrição seguida em Morais de Melo (2012) e Morais de Melo (2018)

Morais de Melo (2012) seguindo o modelo do PHPB	Lição justalinear de Morais de Melo (2018)
<p>to antes fazer ali o concerto preciso, afim de se evitar a fuga dos presos. Deus Guarde a VS.ªs Natal 16 de abril de 1833 Bazilio Quaresma Torreão Ill.ªs Senhores Presidente, e Membros do Corpo Municipal d'esta Cidade</p>	<p>to antes fazer ali o concerto preciso, afim de se evitar a fuga dos presos. Deus Guarde a VS.ªs Natal 16 de abril de 1833 Bazilio Quaresma Torreão Ill.ªs Senhores Presidente, e Membros do Corpo Municipal d'esta Cidade</p>

Fonte: Elaboração própria.

4 Linguística de *Corpus* e Linguística Histórica

Depois de apresentada a reedição das *cartas oficiais norte-rio-grandenses*, desenvolvemos, como última parte deste artigo, uma

discussão mais geral acerca da relação entre a Linguística de *Corpus* e a Linguística Histórica nos estudos da linguagem. Essa conexão possui um escopo assaz vasto e poderia, portanto, ser abordada segundo uma miríade de perspectivas diversas. A que trazemos nas páginas seguintes é bastante parcial, na medida em que trata de um recorte espacial (a guinada dos estudos históricos no Brasil pós-década de 90) e temático (a disponibilização de *corpora* diacrônicos na rede mundial de computadores) que estão vinculadas diretamente a reflexões, inquietações, vivências e roteiros que formaram a travessia do autor deste artigo pelas *cartas oficiais* ao longo de 8 anos (2010, início do mestrado – 2018, final do doutorado), período no qual foi sujeito-(agente e paciente)-pesquisador sob a constelação desse entrelaçamento.

A Linguística Histórica é o grande campo dos estudos da linguagem que trata de estudar as mudanças por que passa, com o transcurso do tempo, a língua em seus mais diversos níveis de análise. Mattos e Silva (2008) distingue uma linguística histórica no sentido *lato* e outra no *stricto sensu*. A primeira corresponde a qualquer tipo de linguística que trabalha com *corpora* datados e localizados; a segunda, apenas com dados provenientes de sincronias passadas. Esta última, a linguística histórica *stricto sensu*, por sua vez, ainda se subdivide, para a autora, em uma linguística histórica de caráter sócio-histórico, por considerar fatores extralinguísticos na análise de fenômenos da língua; e uma linguística diacrônica, que possui uma natureza associal, considerando, sobretudo, os fatores imanentes à estrutura. Como não efetivamos qualquer tipo de controle sobre a biografia dos escreventes (uma tarefa, por si só, já bastante complicada, o que se complexifica ainda mais quando se busca desvendar a real mão que escreveu a carta),²⁴ a análise promovida por Morais de Melo (2018) sobre os usos gráficos nas *cartas oficiais* 8 não teve qualquer cacife para, com propriedade, imiscuir-se por variáveis extralinguísticas, fazendo, portanto, para usar dos termos lançados por Mattos e Silva (2008), uma linguística diacrônica de caráter grafemático.

²⁴ Sobre as dificuldades implicadas na identificação da mão que lavra um documento escrito em sincronias pretéritas, *vide* Gonçalves e Ferreira (2001, p. 485-486). Muito valioso igualmente é o testemunho de Mattos e Silva (2002) sobre os óbices enfrentados por vários trabalhos ligados ao PHPB quanto à constituição de *corpora* diacrônicos.

Os estudos históricos da língua,²⁵ que hoje se albergam preponderantemente sob o título destrinchado pela autora baiana, dimanam de uma tradição filológica.²⁶ Concentrando nossa atenção em modelos clássicos de obras que faziam filologia e mesmo História da Língua no século XX, notamos que a fonte primária para os manuais e artigos/ensaios investigativos eram, de modo semelhante ao que se faz hoje, manuscritos ou, mais comumente, edições feitas desses manuscritos... desde que se tratasse do período medieval. Para os demais séculos, a fonte básica eram geralmente os textos literários ou, às vezes, obras metalinguísticas da época a ser descrita. Podemos confirmar este panorama novecentista através de três marcos consagrados: para o início do século, as *Lições de Filologia portuguesa*, escritas nos primeiros anos da década de 10 por Vasconcelos (s/d); para o meio, os *Ensaio de Filologia portuguesa*, de Silva Neto (1954 [primeira edição]); e o *I seminário de Filologia e língua portuguesa*, ocorrido em 1997 (RODRIGUES; ALVES; GOLDSTEIN, 1999). Um passar de olhos sobre os índices ou uma folheada nessas obras é suficiente para que se comprove o diagnóstico feito acima. Isso justifica o porquê de, em 1996, Castro escrever:

Em relação ao português clássico, em especial dos séculos XVII e XVIII (para não falar do desconhecidíssimo XIX), quem o quiser estudar tem de se resignar a fazer de cabouqueiro, desenterrando penosamente os seus documentos, peneirando os

²⁵ Aqui estamos falando do contexto lusófono, e mais especificamente do brasileiro, porque é sobre o qual temos maior conhecimento, muito embora a mudança no *zeitgeist* metodológico marcada na década de 90 que será comentada na sequência do parágrafo do qual esta nota deriva, pelas discussões que tivemos na disciplina dada por José Luis Ramírez Luengo já anteriormente referida em nota de rodapé, também tiveram lugar no quadro hispânico, inclusive um pouco antes do que se dá no Brasil, na década de 80 pelo menos. Vejam-se, por exemplo, as alusões que Fontanella de Weinberg (1998, p. 95) faz a seus trabalhos anteriores.

²⁶ Sobre esse sulco filológico, recomendamos a bela – e mais extensiva, haja vista trazer alguns exemplos que extrapolam os muros da realidade brasileira, dentro dos quais se situam nossos apontamentos – discussão que promove Maia (2012). Principiando seu artigo com a contextualização da guinada (cujo expoente nas terras brasileiras foi o PHPB) que inseriu, no final do século XX, a Linguística Histórica na agenda dos estudos da linguagem, concentra-se, então, na relação entre Filologia e Linguística Histórica, enfocando, primeiramente, o peso daquela nesta e, logo, a importância desta (e da Linguística moderna como um todo) para aquela.

dados, organizando uma taxionomia inexistente e, se ainda tiver coragem e tempo de vida, formulando hipóteses interpretativas que ficarão à espera de um debate crítico só possível se outros investigadores se transviarem pelos mesmos terrenos (CASTRO, 1996, p. 137).

O registro de uma transição radical vemo-lo um ano após a citação de Castro e no mesmo ano em que sucede o supra aludido Seminário de Filologia: em 1997, quando ocorreu o I Seminário para a História do Português Brasileiro, cujos textos frutos do evento foram publicados, em livro, no ano seguinte (CASTILHO, 1998). A grande mudança que as discussões levantadas nesse seminário traziam, no bojo dos compromissos a que se propunha o PHPB em formação, era – ao lado das preocupações com a sintaxe histórica e a história social, que despontam desde a publicação de Castilho (1998) e às quais se unem, em obras futuras do projeto, outras, como as TD – a constituição de um *corpus* diacrônico do português brasileiro. Reflexo disso é uma série de artigos em que se vão relatando depoimentos acerca de processos de formulação de *corpora*, alguns dos quais mencionamos: Carneiro e Almeida (1998) para documentos do século XVIII ao XX de Feira de Santa, na Bahia; uma primeira sondagem dos textos disponíveis na Biblioteca Nacional, por Barbosa (1998); os primeiros passos para a formação de um banco de documentos paranaenses, por Cyrino, Barrichello e Paula (2002); um *corpus* quase concluído com cartas de homens “ilustres” na Bahia, por Carneiro e Almeida (2002); dois *corpora* (um do século XVIII e o outro, do XIX) com cartas públicas e privadas paranaenses (o último texto da série de livros *Para a história do português brasileiro* cujo cerne está no testemunho acerca da confecção de *corpora*), por Rumeu (2006).

Os filólogos e linguistas interessados pela História da Língua viam-se envolvidos numa maratona que demandava idas a arquivos, manuseio de textos (manuscritos em seu grosso), seleção, transcrição (nos próprios arquivos ou, caso permitida fotografiação ou microfilmagem, em outros ambientes) e edição de documentos. Numa aplicação cabal do que delinea Berber Sardinha (2000, p. 325) sobre um de seus campos de interesse – “A Linguística de *Corpus* ocupa-se da coleta e exploração de *corpora*, ou conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística” –, os novos linguistas históricos estavam

fazendo linguística de *corpus*, e em completude: não só explorando os conjuntos ordenados de dados, mas coletando-os, conformando-os.²⁷

Muito rapidamente essa caminhada de labor arquivístico viu-se, pelas premissas da modernidade, interpelada pelo mundo digital, o que fazia jus à própria esteira constitutiva da Linguística de *Corpus*, cuja história, segundo Berber Sardinha (2000, p. 329), está “intimamente ligada à disponibilidade de corpora eletrônicos”. Se desde o primeiro da série *Para a história do português brasileiro* viamos, a cada livro, uma seção devotada a “estudos sobre a organização do *corpus* diacrônico” (ALKMIN, 2002, p. 7, por exemplo), somos surpreendidos, no quarto volume da série (DUARTE; CALLOU, 2002), com uma produção congruente e judiciosa em que é cogitada, por prismas diversos (o ético, inclusive), a vinculação desses *corpora* à rede mundial de computadores: trata-se do artigo de Barbosa, Lopes e Callou (2002).

Os autores, após definirem o termo *corpus diacrônico* como um “conjunto de materiais de diferentes sincronias passadas” (BARBOSA; LOPES; CALLOU, 2002, p. 30), alertam que a consolidação desse agrupamento de dados organizado não reside apenas nas precisões tecnológicas das ferramentas de busca, mas na “formação do *corpus* de textos sobre o qual os instrumentos incidem”. Daí o necessário esmero que tivemos na confecção das *cartas oficiais norte-rio-grandenses*, e em sua reedição, no que toca à observância de fatores como o recorte temporal, local de produção e natureza do documento. Dada essa definição de *corpus diacrônico*, os autores anunciam:

A equipe carioca [do PHPB], que tem suas raízes no Projeto NURC-Rio, deixa disponível, desde já, na *internet*, seu *corpus* diacrônico: transcrições de impressos do século XIX e edições diplomático-interpretativas de manuscritos dos séculos XVIII e XIX. *Essa ação pioneira* contribui para que a comunidade acadêmica venha a trabalhar com dados mais seguros na reconstrução da história do português brasileiro, a modalidade linguística de mais de 160 milhões de pessoas (BARBOSA; LOPES; CALLOU, 2002, p. 30, grifo em “Essa ação pioneira” nosso).

²⁷ Bawarshi e Reiff (2013, p. 55-59) dedicam um item de sua obra para refletir sobre as contribuições que a “linguística histórica/de *corpus*” (assim, em quase todas as aparições, grafada) tem para dar “com a pesquisa e ensino de gêneros ao dar conta da natureza das tipologias e da mudança linguística”.

Esse passo de integração de uma, poderíamos cunhá-la deste modo, Linguística de *Corpus* Histórica à rede virtual, realizada, neste novo cenário do fazer diacronia linguística surgido na década de 90 em terras *brasilis*, pela primeira vez (ao menos segundo a autotaxonomia de pioneira lida na citação acima) pela equipe do PHPB-RJ viria a se tornar a recomendação repassada pelo PHPB (a meta era a constituição desses *corpora*, havendo a orientação de que fossem, após concluídos, disponibilizados em rede) a todas as equipes locais. Efeito disso foi a criação da plataforma <https://sites.google.com/site/corporaphpb/>. Tornou-se, assim, acessível ao grande público uma série de *corpora*, ordenados em três grandes grupos: *corpora* impressos, *corpora* manuscritos e *corpus* diferencial. Em cada um deles, tem-se via para materiais produzidos pelas equipes regionais de diversos estados e resultantes, como afirma a página principal, “de teses, dissertações, livros, CDs, relatórios, enfim, não apenas diferentes resultados acadêmicos de trabalhos concebidos e produzidos sob a chancela do PHPB, mas também de colaboradores externos”. Aí estão as *cartas oficiais* fruto de Morais de Melo (2012).

Amplamente promissora, portanto, a investida de Barbosa, Lopes e Callou (2002). O artigo segue, no entanto, com algumas vigilâncias a que esse tipo de associação – *a priori*, tão prolífica, salutar e, como muito bem asseveram os autores, legítima/necessária, mormente quando esse trabalho de coleta e formação de *corpora* contém, em seus passos, financiamento público²⁸ – deve atentar. Alguns deles estão relacionados com o crédito dado ao autor do trabalho. Os autores exibem um cabeçalho de sua edição eletrônica para fazer-nos ver que, a despeito de haver um trabalho conjunto, deve ficar discriminado o autor exato que se responsabilizou por cada documento. Com isso, prosseguem, “os indivíduos não verão sua produção acadêmica dispersa e escamoteada no trabalho coletivo, mas, pelo contrário, ela será contabilizada na edição eletrônica a ser citada pelas pesquisas de outrem” (BARBOSA; LOPES; CALLOU, 2002, p. 31). Isso atenua, mas não extingue, os riscos de um uso – quiçá oriundo de uma

²⁸ “Nossa política de divulgação assume que o acesso virtual deva ser gratuito, uma vez que o ônus da execução foi, direta ou indiretamente, do dinheiro público. Somente faz sentido cobrar pela confecção de versões em CDs em função dos gastos materiais e de prestação de serviço posteriores à publicação da página da rede. A cobrança por *softwares* deveria seguir a mesma lógica, caso não tenham sido financiadas pelo setor privado” (BARBOSA; LOPES; CALLOU, 2002, p. 31).

compreensão (insonte ou não) de que a disponibilização de um material em linha, ao permitir o “domínio” público (escorrito seria dizer “acesso” público), é de todos, é da rede e, por conseguinte, não tem autoria – do acervo aberto sem as devidas referências.²⁹

Outro cuidado (e com ele fechamos nosso bosquejo sobre esse artigo que aporta questões extremamente relevantes sobre o fazer linguística de *corpus* histórica nos tempos de hoje, questões muitas vezes dadas por certas e, por isso, talvez, não trazidas à baila, não ditas) ao qual se deve atender afeta ao ineditismo que se queira preservado por parte do autor: “É primordial que se resguardem edições que sejam objeto de teses, dissertações ou de trabalhos que os editores julguem de maior relevo antes de serem incluídos no acervo de cada equipe” (BARBOSA; LOPES; CALLOU, 2002, p. 32).

Dando prosseguimento a essa pauta e, ao mesmo tempo, cosendo o remate deste item 4, que, por sua vez, faz despontar o fecho do próprio artigo, seguimos a trilha desenhada por alguns dos artigos que compõem o livro organizado por Kabatek (2016a). Abrindo o livro, deparamo-nos com um estudo do próprio organizador em que, partindo da constatação de que existem diferentes “disciplinas” contidas na denominação Linguística de *Corpus*, o autor distingue três vertentes: uma primeira, destinada à elaboração de *corpus*, englobando suas etapas, que vão desde a coleta até a apresentação dos dados organizados; uma segunda, à primeira estreitamente interligada, que respeita ao tratamento dos dados por meio de etiquetagem, contagens e estatísticas; e uma terceira – que o linguista considera a mais presente nos estudos reunidos na obra – que faz um linguística “com *corpus*”, em que se focalizam fenômenos específicos da história da língua com base nos dados que comporta determinado *corpus*

²⁹ Um exemplo que ilustra esse risco é o artigo de Braga e Bispo (2016, p. 3) em que se recorre às *cartas* que editamos em 2012 e que foram divulgadas na plataforma do PHPB (onde há, não obstante, a discriminação dos organizadores: o autor e seus dois orientadores de mestrado), referenciando-as apenas como dados “extraídos das amostras de 86 cartas oficiais, provenientes da plataforma de *corpora* do projeto Para a História do Português Brasileiro (PHPB), que circularam no estado do Rio Grande do Norte entre os séculos XVIII e XIX”. Uma citação que traz um nítido detrimento autoral em favor de um agente usurpador: a plataforma praticamente “se passa” (!) por autor. O escamoteamento de que falam Barbosa, Lopes e Callou (2002) se dá aqui não (apenas) em nome de um trabalho coletivo, mas (também) em nome do suporte que veicula a produção acadêmica.

(cf. KABATEK, 2016b, cap. 1, par. 4).³⁰ O exercício praticado para a confecção das *cartas oficiais* mobilizou todas as instâncias implicadas por essas três vertentes, inclusive porque, como assume o autor, “entre las tres vertentes hay, obviamente, una estrecha relación y, en tanto que un corpus no es un fin en sí, sino que se crea *para algo*, se necesita un intercambio continuo del creador del corpus con los usuarios que

³⁰ Alguns dos livros utilizados nesta tese só foram obtidos em sua versão eletrônica, os *e-books* (não confundir com livros em pdf), disponibilizados, após compra, em plataformas de leitura, como o Kindle e o Kobo. Não há consenso – às vezes, na realidade, sequer orientação – sobre como se devem fixar as citações dessas obras, já que, na maioria das vezes (ao menos com relação aos livros de que dispomos através desses leitores digitais), não está disponível a informação da página correspondente na edição impressa junto aos números das localizações (*location numbers*) que aparecem nos aplicativos. Na falta de uma resolução, decidimos seguir, ainda que não ortodoxamente, as orientações dadas pelas Bibliotecas Bodleian (Bodleian Libraries) da Universidade de Oxford (Oxford University), por julgarmos que elas atendem efetivamente à meta de orientar o leitor a encontrar (seja na edição impressa, seja na virtual) a citação em apreço. Essas recomendações podem ser lidas no link <https://libguides.bodleian.ox.ac.uk/e-books/citing>. Uma vez que, como salienta o próprio site, os números de localização (os *location numbers*), encontrados ao menos em todos os e-books compatíveis com Kindle ou Kobo, não fornecem uma saída estável, haja vista se alterarem conforme tipo de fonte ou tamanho (zoom) da tela de leitura, as Bodleian Libraries sugerem, em caso de não haver a informação da página (correlata à edição impressa), que se indique primeiramente o capítulo (que abreviaremos “cap.”) seguido pelo parágrafo (que transformaremos em “par.”) no qual aparece o excerto. Mesmo cômicos de que em alguns livros essa solução pode gerar caminhos bastante árduos para o autor que dela se valer (imaginemos um livro sem capítulos), é, na ausência de modelos pré-estabelecidos institucionalmente, o rumo que melhor nos convenceu. Talvez não fosse necessário esclarecer – fazemo-lo, contudo, para evitar desencontros quando de um possível confronto em busca de determinada passagem que nos chegou via e-book e também por serem os livros eletrônicos uma realidade recente (que o diga o quase desconcerto dos profissionais especializados, os bibliotecários, a que recorremos ao serem questionados sobre possíveis vias a seguir) – que citações deslocadas, por sua extensão (por conterem mais de 3 linhas, por exemplo, para evocar a ABNT), à semelhança de um parágrafo, não serão contabilizadas como tal, contando apenas os parágrafos no sentido mais consabido do termo, ainda que, por razões editoriais ou de apresentação cibernética, alguns (os primeiros de um capítulo, e.g.) não apresentem sangria. Essa contagem será feita a partir do início do capítulo (desconsiderando-se conteúdos supratextuais, como a epígrafe e as notas de rodapé, e organizacionais, como o título e os subtítulos). O resultado de todos esses passos pode-se ver em “(cf. KABATEK, 2016b, cap. 1, par. 4)”, reproduzido aqui em bis ilustrativo.

lo utilizan para un estudio concreto”³¹ Em nosso caso, neste primeiro momento, criador e usuário coincidiram numa mesma pessoa.

O capítulo escrito por Enrique-Arias (2016), por sua vez, abre com uma certificação – a qual, em que pese os problemas levantados por Barbosa, Lopes e Callou (2002), é aplicável perfeitamente ao que se vê nos últimos anos, quando *corpora*, a exemplo dos criados, vêm sendo lançados na internet no Brasil – acerca dos efeitos e/ou impactos da divulgação de *corpora* na rede mundial de computadores: “La investigación en diacronía del español se ha visto beneficiada en los tiempos recientes por la disponibilidad de grandes bases de datos textuales de uso libre en la red”³² (ENRIQUE-ARIAS, 2016, cap. 2, par. 1).

Entendemos que uma das grandes vantagens que introduz o acesso a esses *corpora* em linha é efetivamente a possibilidade de utilizar ferramentas de busca – sejam as mais simples, como o Ctrl L do Microsoft Word ou Ctrl F do Adobe PDF, até programas desenvolvidos para agilizar ou até determinar a eficácia no trabalho de coleta – sobre o conjunto ordenado de dados. No caso das *cartas oficiais* reeditadas, para o controle dos usos de grafemas sem transcendência fônica, valemo-nos de um *software* criado para a identificação de palavras contendo os 23 ambientes gráficos definidos para análise em Moraes de Melo (2018). Recorremos, contudo, no momento de coletar os dados a serem escrutinados, também a esses mecanismos já dados, como o Ctrl L, para contornar algumas vicissitudes que eventualmente escapassem às malhas do programa.

Dando continuidade à sua reflexão sobre as vantagens e os desafios que o *corpus* eletrônico ou disponível digitalmente carrega, Enrique-Arias (2016, cap. 2, par. 9) aponta que numa leitura linear, isto é, numa leitura feita em material impresso na sua versão física, o leitor tem de processar o texto em sua ordem de aparição ao passo que “en el texto electrónico es mucho más habitual acceder al texto a través de una concordancia generada por una máquina de búsqueda”³³ De fato, ao se

³¹ “Entre as três vertentes há, obviamente, uma relação estreita e, posto que um corpus não é um fim em si, mas que se cria *para algo*, é necessário um intercâmbio contínuo do criador do corpus com os usuários que o utilizam para um estudo concreto” (TN).

³² “A pesquisa em diacronia do espanhol tem-se visto beneficiada nos tempos recentes pela disponibilidade de grandes bases de dados textuais de uso livre na rede” (TN).

³³ “no texto eletrônico, é muito mais habitual acessar o texto através de um padrão gerado por uma ferramenta de busca” (TN).

ter acesso a um *corpus* impresso apenas em versão física, sem sequer um redobro digitalizado ou disponível virtualmente, o pesquisador está atado ao dever de averiguar o material em sua inteireza, linha por linha e palavra por palavra, e ir rascunhando e agrupando os elementos que lhe sejam de interesse. Essa labuta é consideravelmente mitigada quando se pode ir direto – um direto passível de ser torto segundo os limites do dispositivo de busca – ao que se almeja examinar.

Esse contraste entre a leitura linear e a leitura de um *corpus* digital é sumária e assaz polidamente esquematizado pelo autor: “en la lectura lineal se accede a las estructuras lingüísticas en el orden *contexto* > *forma*, mientras que en el corpus informatizado se accede en el orden contrario, es decir, *forma* > *contexto*”³⁴ (ENRIQUE-ARIAS, 2016, cap. 2, par. 9). Conquanto assinala os benefícios que o *corpus* digital contém, o pesquisador não deixa de sublinhar um possível complicador. Ao se lidar com textos de sincronias passadas – em que há uma manifesta variação, com relação ao que se tornou padrão hoje, na forma das palavras, além de em várias outras camadas: estratégias retóricas, encadeamento textual etc. –, o investigador vê-se obrigado, para poder contar com as benesses do texto informatizado, a “conocer de antemano, a partir de gramáticas históricas, diccionarios o estudios previos, cuáles son las formas utilizadas para expresar la función que el investigador se propone reastrear en el corpus”³⁵ (par. 9). Daí, progride o autor, ser crucial o modo de acesso aos dados, a que se pode conferir uma parcela dos eventuais limites da Linguística de *Corpus* (cf. ENRIQUE-ARIAS, 2016, cap. 2, par. 9).

Destaca-se, nessa enseada e dimensionando a questão para nossa realidade, o papel da tabela matriz, uma das estratégias desenvolvidas por Morais de Melo (2018), a fim de operacionalizar a análise dos usos de grafemas sem transcendência fônica nas *cartas oficiais*. Para cada palavra não abreviada identificada nas *cartas* – a qual ficava disposta, com as demais variantes gráficas que porventura surgissem, em uma linha –, registrava-se, numa coluna criada à extrema esquerda da tabela, o vocábulo

³⁴ “na leitura linear, acessa-se às estruturas linguísticas na ordem *contexto* > *forma*, ao passo que, no *corpus* informatizado, o acesso se dá na ordem contrária, ou seja, *forma* > *contexto*” (TN).

³⁵ “conhecer de antemão, a partir de gramáticas históricas, dicionários ou estudos prévios, quais são as formas utilizadas para expressar a função que o pesquisador se propõe a rastrear no *corpus*” (TN).

conforme sua forma ortográfica atual. Estavam as realizações, assim, lematizadas.³⁶ Foi graças a essa estaca referencial que se puderam efetuar com sucesso, por meio do *software*, a busca e a apreensão das palavras segundo os padrões pré-estabelecidos e nos moldes que idealizamos. A palavra “aceitação”, por exemplo, foi materializada em nosso *corpus* apenas como <aseitação> na C42 e <acceitação> na C74; nenhuma em forma compatível com a da ortografia hodierna. Se não houvesse essa entrada estândar, o lema, complicar-se-ia em altíssima potência o trabalho de obtenção das palavras através da rotação do *software* sobre as 23 notações (cada uma corresponde a um contexto grafemático de análise) instituídas, inviabilizando-o até, na verdade. É evidente que estamos relatando aqui um processamento que não recai de modo direto nas *cartas oficiais*, mas nessa captura que delas fizemos, a tabela matriz, a qual, sem embargo, pode ser perfeitamente compreendida como uma extensão de acesso do próprio *corpus* linguístico. É igualmente óbvio, por sua vez, que esse acesso é um acesso parcial ao *corpus*: a tabela contém todas as palavras não abreviadas que puderam ser transcritas das *cartas*, o que

³⁶ Vaamonde (2015) aponta a falta de lematização como uma das limitações dos grandes *corpora* do espanhol, nas figuras do *Corpus Diacrónico del Español*, o CORDE, concebido pela RAE, e do *Corpus del Español*, o CdE, criado por Mark Davis. Os outros limites são: 1º o uso de textos escritos como fonte de dados, uma “limitación forzosa y obvia de todo corpus diacrónico” [“limitação forçosa e óbvia de todo *corpus* diacrônico” (TN)] (VAAMONDE, 2015, p. 2), diante da inexistência de fontes orais dos períodos pretéritos da língua; 2º níveis de acesso ao documento, pois o autor defende que, para que possa o *corpus* atender às mais plurais expectativas, deveria ser apresentado em distintas edições: semidiplomática, crítica e fac-símiles; 3º anotação linguística detalhada, com, para além da lematização, as etiquetas morfológicas e sintáticas; e 4º gerenciamento dos fatores extratextuais que vão além do cronológico e do local de produção, como a caracterização social do autor, sua procedência dialetal e propósito comunicativo do texto (cf. p. 4). Saltando o primeiro ponto, por razões evidentes, ao encarar nosso próprio labor e premeditando a possibilidade de uma futura publicação (linear e informatizada) das *cartas*, gostaríamos de oferecer, ao lado das atuais edições semidiplomática e fac-similar (no caso desta, com a revisão de algumas fotos, logo que o IHGRN reabra suas portas para pesquisa), uma edição crítica/modernizada, que faria às vezes de um hiperlema textual, bem como de tentar uma supervisão de outros critérios extralinguísticos, como os que cita Vaamonde (2015). A princípio, por outro lado, não prevemos nenhuma investida em etiquetagens morfossintáticas. Na sequência do artigo, o autor nos traz um exemplo de um *corpus* diacrônico em que se supera grande parte dessas limitações assinaladas por ele, o Post Scriptum.

permite uma série de estudos relativos aos grafemas alfabéticos, mas, claramente, não serve para perscrutações focadas em outros aspectos grafemáticos (abreviações e pontuação, por exemplo) ou em demais estratos de organização da língua (sintático, textual, discursivo).

Outro texto compilado em Kabatek (2016a) que lança luz – pela apresentação de um competente modelo – sobre as confluências fecundantes e necessárias entre Linguística de *Corpus* e Linguística Histórica é o artigo de Carvalheiro *et al.* (2016). Os autores relatam um pouco do Post Scriptum, *corpus* linguístico desenvolvido pelo Centro de Linguística da Universidade de Lisboa que promove um acervo regulado de cartas privadas – merecendo já os devidos louros, pois são um dos gêneros mais difíceis de serem angariados, precisamente por circularem na esfera privada – escritas entre o século XVI e o início do século XIX em Portugal e na Espanha com a meta (hoje já atingida) de oferecer 2 milhões de palavras (1 milhão oriundo de cartas privadas em português e o outro milhão, em espanhol). Reportamo-nos a este capítulo pelo horizonte de expectativa que enxergamos no comprometimento e seriedade com que a equipe trouxe à luz um *corpus* médio-grande³⁷ tão esmerado em cumprir uma série de requisitos que são agudamente complicados (controle biográfico, zona dialetal, notação morfossintática) e, por isso, muitas vezes deixados de lado.³⁸ Encerram o artigo dando uma amostragem do potencial do *corpus* com uma análise do marcador discursivo “pois”, em português, e de dois mecanismos anafóricos, o “cuja” do português e os pronomes oblíquos de terceira pessoa “le/la/lo” do espanhol. Por meio dessas amostragens analíticas, os pesquisadores ilustram dois argumentos que pretendem defender:

³⁷ Consoante Berber Sardinha (2000, p. 346), para um *corpus* de abordagem histórica, existem cinco classificações no que toca à sua extensão: pequena (com menos de 80 mil palavras, sendo, portanto, a categoria na qual se incluem nossas *cartas oficiais*), pequeno-médio (entre 80 e 250 mil palavras), médio (entre 250 mil e 1 milhão de palavras), médio-grande (entre 1 e 10 milhões, onde entra o Post Scriptum) e grande (com 10 milhões ou mais de palavras).

³⁸ O exemplo mais próximo de Post Scriptum no Brasil seja, quiçá, o Projeto Tycho Brahe, conduzido por uma equipe da Unicamp, em que as anotações linguísticas, tanto morfológicas (em 44 textos, num total de 1.962.176 palavras, segundo a página oficial do projeto) quanto sintáticas (em 27 textos, atingindo 1.234.323 palavras) são operacionalizadas. O site desse *corpus* histórico é o <http://www.tycho.iel.unicamp.br/corpus/index.html>.

i) o de que o conceito de *desvio linguístico* é facilmente desmontável com a demonstração da antiguidade de processos que, enquanto <<desvios>>, costumam ser estigmatizados pelo público leigo; ii) o de que os atlas histórico-dialetais são um recurso linguístico que se pode realizar, hoje em dia, com um grau de incerteza cada vez mais insignificante (CARVALHEIRO *et al.*, 2016, cap. 9, par. 5).

O último estudo de que nos valemos, encerrando, com ele, este artigo, é o de Nieuwenhuijsen (2016). Abre seu texto a autora com um fragmento que, à guisa de ciclo, retoma – deitando-o agora em solo hodierno – a tônica com que encetamos este item 4.3 relativa à conjunção entre as duas rotas da linguística aqui consideradas: “Hoy en día es un hecho que la lingüística histórica como disciplina académica está cambiada *profunda y definitivamente*, debido al desarrollo de la lingüística de corpus y las nuevas metodologías ofrecidas y hasta impuestas por los corpus diacrónicos digitales”³⁹ (NIEUWENHUIJSEN, 2016, cap. 11, par. 1, grifo nosso). Essa mudança acarreta uma latente perda da empatia (termo que a autora toma de uma expressão cunhada por Kabatek, “lingüística empática”) com a matéria maior do *corpus*, na medida em que o pesquisador perde a obrigação de palmilhar os textos em sua plenitude; já pode transitar fácil e diretamente, sem desvios, do objetivo ao objeto. O que vemos nessa corretíssima consideração é quase uma metáfrase do debate que Enrique-Arias (2016), acima glosado, faz ao diferenciar as duas formas de acesso às estruturas linguísticas: da ordem *contexto > forma* à ordem *forma > contexto*.

A pesquisadora elenca alguns impasses que podem ameaçar a busca de palavras e, conseqüentemente, azar algumas inconsistências nas análises. Foi muito grato ver que nossa faina, mesmo com suas eivas, oferece condições para que se driblem esses fatores de embaraço. Tomando as mesmas medidas de análise de que se vale Vaamonde (2015), o Corpus Diacrónico del Español (CORDE) e o Corpus del Español (CdE), o primeiro óbice que Nieuwenhuijsen aponta são as buscas de elementos clíticos que, por muitas vezes virem aglutinados a palavras pesadas (nacionais, substantivas, v. g.), ou são de difícil obtenção pelas

³⁹ “Hoje em dia é um fato que a Linguística Histórica como disciplina acadêmica está mudada profunda e definitivamente devido ao desenvolvimento da linguística de corpus e às novas metodologias oferecidas até impostas pelos *corpora* diacrônicos digitais” (TN).

ferramentas de busca oferecidas por esses *corpora* ou são tomadas por outras (NIEUWENHUIJSEN, 2016, cap. 11, par. 3).⁴⁰ Um segundo risco com um *corpus* digital de que fala a autora é a categorização errônea de algumas palavras (ela traz o exemplo de homônimos: uma realização de “fuera” [“fora”], advérbio de lugar, coletado como “fuera” [“fosse”], pretérito imperfeito do subjuntivo de “ser”). Como a separação e organização de cada uma das cerca de 26.000 palavras da tabela matriz medraram do trabalho manual de Morais de Melo (2018), e não de uma equipe ou de um programa, acreditamos que erros – desse tipo, ao menos – não ocorram ou, na pior das hipóteses, sejam raros. Se para o primeiro tipo de problema, a linguista não enxerga um remédio, para o segundo ela aventa a revisão manual de todos os exemplos. Assim, propõe como solução o que fizemos como missão, o que não nos isenta, *a posteriori*, do dever da reinspeção.

Por fim, a autora demonstra como a aplicação de um teste de regressão logística binário pode precisar e esclarecer melhor o comportamento das formas linguísticas em análise. Embora o exame dos usos gráficos nas *cartas* tenha sido efetuado, em Morais de Melo (2018), sobre os cálculos dos valores brutos e percentuais, e não por meio de um teste logístico binário, confiamos ter obtido um quadro suficientemente eficaz para pensar e acompanhar com rigor as tendências dos usos gráficos nos documentos potiguares em tela. De mais a mais, o modo como arquitetamos a apresentação dos resultados (que chamamos de “perfis de saída” na tese) lega a transparência oportuna para que se possa, aquando das análises, transitar descomplicada e fertilmente de interpretações de base mais quantitativa às de cunho mais qualitativo. Mesmo sugerindo o teste de regressão logística, Nieuwenhuijsen (2016, cap. 11, par. 51) fecha seu artigo com um belo remate no qual abona a experiência que enarramos neste parágrafo. Com ele, destarte, fechamos o tópico:

A pesar del gran valor de los test estadísticos para la lingüística histórica, no queremos abogar aquí por la supresión de los análisis tradicionales y la sustitución completa de los análisis tradicionales

⁴⁰ Repetimos que a tabela matriz, embora desfaça esse obstáculo, se revela exígua – devido às convenções tomadas em sua criação (cf. MORAIS DE MELO, 2018, p. 148-155) – para o reconhecimento de outras formas, quais sejam a colocação do til, maiúscula e minúsculas e pontuação. Para esses casos, é necessária uma leitura linear da transcrição completa, invalidando-se, a princípio, a utilização de acessórios de busca.

por las pruebas estadísticas. El análisis cuantitativo sigue siendo imprescindible para formarse una idea global de la frecuencia y desarrollo de una forma o construcción sintáctica. Asimismo, el análisis cualitativo, es decir el detenido estudio de ejemplos específicos en su contexto, permite identificar posibles factores que hayan influido en la evolución del cambio lingüístico. El tratamiento estadístico, en cambio, constituye una herramienta complementaria muy potente, que sirve para comprobar la validez de las conclusiones sacadas en ambos tipos de análisis y para medirla posible influencia de los distintos factores identificados en el material estudiado.⁴¹

5 Considerações finais

O movimento renovador pelo qual a Linguística Histórica brasileira passou na década de 90 – num contínuo progressivo que transformou essa área de estudos em uma das mais produtivas atualmente dentro das ciências da linguagem no Brasil – fez da edição de *corpora* seu fito basilar, estando, desde então, sua consecução em perene fabrico. Se, por um lado, nesta seara dos estudos diacrônicos, editar textos goza de um protagonismo notório, o que se tem dito sobre reeditá-los? Este trabalho propôs-se, então, a apresentar as *cartas oficiais nortério-grandenses*, um conjunto de documentos oficiais organizados por primeira vez em 2012 como parte das ações locais vinculadas ao PHPB. Devido a novas orientações e a alterações nos objetivos de pesquisa, as *cartas* foram reeditadas, afã finalizado em 2018, quando veio à luz um *corpus* composto por 129 missivas consignadas entre 1713 e 1950 no estado do Rio Grande do Norte.

⁴¹ “Apesar do grande valor dos testes estatísticos para a Linguística Histórica, não queremos advogar aqui pela supressão das análises tradicionais e pela substituição completa das análises tradicionais pelas provas estatísticas. A análise quantitativa continua sendo imprescindível para se formar uma ideia global da frequência e desenvolvimento de uma forma ou construção sintática. Da mesma forma, a análise qualitativa, isto é, o estudo detido de exemplos específicos em seu contexto, permite identificar possíveis fatores que tenham influenciado na evolução da mudança linguística. O tratamento estatístico, por outro lado, constitui uma ferramenta complementar muito potente que serve para comprovar a validade das conclusões tiradas de ambos os tipos de análise e para medir a possível influência de distintos fatores identificados no material estudado” (TN).

O processo de reedição exigiu uma série de ingerências no *corpus* de 2012. O saldo final foi positivo: afinou-se o feixe ordenado de cartas a novos critérios e necessidades surgidos ao longo do percurso doutoral e realizou-se a revisão das cartas antigas, logrando-se, inclusive, a decifração de muitas passagens ininteligíveis aos olhos que no começo desta década iniciaram a missão transcricional dos documentos. Porquanto tanto a tarefa de 2012 quanto a de 2018 nascem do cruzamento entre Linguística Histórica e Linguística de *Corpus*, resolvemos dedicar a segunda metade deste artigo para refletir sobre algumas repercussões desse encontro. Destacamos dois pontos: a disponibilização de *corpora* diacrônicos na internet facilita muito o acesso dos dados aos pesquisadores interessados, mas precisa atentar para a preservação dos direitos autorais; as novas tecnologias potencializam as abordagens e os filtros aplicados aos *corpora*, mas não invalidam estratégias menos avançadas.

Esperamos que as experiências compartilhadas neste artigo e os argumentos levantados possam dar maior visibilidade a estas facetas que certamente, ao menos em alguma medida, participam da engrenagem dos estudos históricos da língua: a reedição de *corpus* linguístico, as vantagens e riscos da liberação dos *corpora* em meio virtual e os horizontes abertos pelas novas tecnologias para aproximações cada vez mais ricas dos dados.

Agradecimento

A Hozanete Lima, mais do que orientadora, uma querida amiga, por ter me incentivado a querer publicar.

Referências

ACIOLI, V. L. C. *A escrita no Brasil colônia: um guia para a leitura de documentos manuscritos*. Recife: FUNDAJ; Massangana; Ed. Universitária da UFPE, 1994.

ALKMIM, T. M. (org.). *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas, 2002. v. 3.

ALMEIDA CABREJAS, B. Escuchar los textos: el análisis de los textos en el estudio de la fonética y fonología de épocas pasadas. *Lingüística en la Red*, Alcalá de Henares, n. 11, p. 1-18, 2013.

BARBOSA, A. G. O português escrito no século XVIII: fontes reunidas na Biblioteca Nacional do Rio de Janeiro. In: CASTILHO, A. T. (org.). *Para a história do português brasileiro: primeiras idéias*. São Paulo: Humanitas, 1998. v. 1. p. 229-238.

BARBOSA, A. G. O contexto dos textos coloniais. In: ALKMIM, T. M. (org.). *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas, 2002. v. 3. p. 421-431.

BARBOSA, A.; LOPES, C. R. S.; CALLOU, D. Organização dos *corpora* diacrônicos do PHPB-RJ na rede mundial de computadores. In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ/FAPERJ, 2002. v. 4. p. 29-37.

BAWARSHI, A. S.; REIFF, M. J. *Gênero: história, teoria, pesquisa, ensino*. São Paulo: Parábola, 2013.

BELLOTO, H. L. *Como fazer análise diplomática e análise tipológica de documentos de arquivo*. São Paulo: Arquivo do Estado e Imprensa Oficial do Estado, 2002.

BERBER SARDINHA, T. Linguística de corpus: história e problemática. *Delta*, São Paulo, v. 16, n. 2, p. 323-367, 2000. DOI: <https://doi.org/10.1590/S0102-44502000000200005>.

BORGES, J. L. *Obras completas I: 1923-1949*. Buenos Aires: Emecé, 2009.

BORGES, J. L. *Obras completas*. São Paulo: Globo, 1998. v. 1.

BRAGA, A. P. A.; BISPO, E. B. Estratégias de relativização em cartas oficiais norte-rio-grandenses dos séculos XVIII e XIX. *Odisseia*, Natal, v. 1, n. 2, p. 3-16, 2016.

BRASIL. Presidência da República. *Manual de redação da Presidência da República*. Brasil: Presidência da República, 2002.

CAMBRAIA, C. N. Subsídios para uma proposta de normas de edição de textos antigos para estudos lingüísticos. In: RODRIGUES, A. C. S.; ALVES, I. M.; GOLDSTEIN, N. S. (org.). *I Seminário de Filologia e Língua Portuguesa*. São Paulo: Humanitas, 1999. p. 13-23.

CARNEIRO, Z. O. N.; ALMEIDA, N. L. F. Documentos dos séculos XVIII-XX para a constituição de um banco de dados do português. In: CASTILHO, A. T. (org.). *Para a história do português brasileiro: primeiras idéias*. v. 1. São Paulo: Humanitas, 1998. p. 197-210.

CARNEIRO, Z. O. N.; ALMEIDA, N. L. F. Informes sobre *corpus* em fase de conclusão: cartas de homens “ilustres” do século XIX (PB/Bahia). In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ/FAPERJ, 2002. v. 4. p. 61-75.

CARVALHEIRO, C. *et al.* A idade dos <<desvios>>: diacronia, variação social e lingüística de *corpus*. In: KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016. [E-book] p. 175-196. DOI: <https://doi.org/10.1515/9783110462357-009>

CASTILHO, A. T. (org.). *Para a história do português brasileiro*. São Paulo: Humanitas; FAPESP, 1998.

CASTRO, I. Para uma história do português clássico. In: CONGRESSO INTERNACIONAL SOBRE O PORTUGUÊS, 1996, Lisboa. *Actas* [...]. Lisboa: Associação Portuguesa de Linguística; Edições Colibri, 1996. p. 135-150. DOI: <https://doi.org/10.2307/3980045>

CYRINO, S. M. L.; BARRICHELLO, J.; PAULA, F. F. Formação de um banco de documentos paranaenses: primeiros resultados. In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ; FAPERJ, 2002. v. 4. p. 77-85.

DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do Português Brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ; FAPERJ, 2002. v. 4.

ENRIQUE-ARIAS, A. Sobre la noción de perspectiva en lingüística de corpus: algunas ventajas de los corpus paralelos. In: KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016. [E-book]. p. 21-39. DOI: <https://doi.org/10.1515/9783110462357-002>

FONSECA, M. C. A. P. *Caracterização lingüística de cartas oficiais da Paraíba dos séculos XVIII e XIX*. 2003. 499f. Tese (Doutorado em Linguística) – Centro de Artes e Comunicação, Universidade Federal de Pernambuco, 2003.

FONTANELLA DE WEINBERG, M. B. La variable sexo y las grafías de los hablantes bonaerenses en los siglos XVIII y XIX. In: BLECUA, J. M.; GUTIÉRREZ, J.; SALA, L. (org.). *Estudios de grafemática en el dominio hispánico*. Bogotá: Ediciones Universidad Salamanca; Instituto Caro y Cuervo, 1998. p. 83-95.

FONTELA, O. *Poesia reunida [1969-1996]*. São Paulo: Cosac Naify; Rio de Janeiro: 7 Letras, 2006.

GONÇALVES, U. S.; FERREIRA, P. S. Aventura no reino das traças: contribuindo para uma história lingüística da Bahia. In: MATTOS E SILVA, R. V. (org.). *Para a história do português brasileiro: primeiros estudos*. São Paulo: Humanitas; FAPESP, 2001. v. 2. p. 483-504.

HORA, D.; SILVA, C. R. (org.). *Para a história do português brasileiro: abordagens e perspectivas*. João Pessoa: Ideia, 2010. v. 8.

KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016a. [E-book]. DOI: <https://doi.org/10.1515/9783110462357>

KABATEK, J. Un nuevo capítulo en la lingüística histórica iberorrománica: el trabajo crítico con los corpus. Introducción a este volumen. In: _____ (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016b. [E-book]. p. 1-19. DOI: <https://doi.org/10.1515/9783110462357-001>

LOPES, F. M. (org.). *Catálogo de manuscritos avulsos da capitania do Rio Grande do Norte (1623-1823)*. Natal: EDUFRN, 2000.

MAIA, C. Lingüística Histórica e Filologia. In: LOBO, T. et al. (org.). *ROSAE: lingüística histórica, história das línguas e outras histórias*. Salvador: EDUFBA, 2012. p. 533-542.

MATTOS E SILVA, R. V. (org.). *Para a história do português brasileiro: Primeiros estudos*. São Paulo: Humanitas; FAPESP, 2001. v. 2.

MATTOS E SILVA, R. V. Reflexões e questionamentos sobre a constituição de *corpora* para o Projeto *Para a história do português brasileiro*. In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: FAPERJ, 2002. v. 4. p. 17-28.

MATTOS E SILVA, R. V. *Caminhos da Linguística Histórica: ouvir o inaudível*. São Paulo: Parábola Editorial, 2008.

MEGALE, H.; TOLEDO NETO, S. A.; FACHIN, P. R. M. (org.). *Caminhando mato dentro: documentos do ouro do século XVIII*. São Paulo: Espaço Editorial, 2009.

MORAIS DE MELO, F. *Cartas oficiais norte-riograndenses dos séculos XVIII, XIX e XX: constituição e caracterização de um corpus diacrônico*. 2012. 329f. Dissertação (Mestrado em Estudos da Linguagem) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal do Rio Grande do Norte, 2012.

MORAIS DE MELO, F. As fórmulas textuais das “Cartas oficiais norte-rio-grandenses” (1713-1931). In: NEGRO ROMERO, M.; ÁLVAREZ, R.; MOSCOSO MATO, E. (org.). *Gallaecia*. Estudos de lingüística portuguesa e galega. Santiago de Compostela: Universidade de Santiago de Compostela; Servizo de Publicacións e Intercambio Científico, 2017. p. 465-482.

MORAIS DE MELO, F. *Nas trilhas da escrita: reedição e análise grafemática das cartas oficiais norte-rio-grandenses (1713-1950)*. 2018. 961f. Tese (Doutorado em Estudos da Linguagem) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal do Rio Grande do Norte, 2018.

MORAIS DE MELO, F.; LIMA, M. H. A. Uma microanálise de cartas oficiais norte-rio-grandenses. *Alfa: revista de Linguística da UNESP, São José do Rio Preto*, v. 60, p. 61-77, 2016a. DOI: <https://doi.org/10.1590/1981-5794-1604-3>. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/7438>. Acesso em: 12 jan. 2020.

MORAIS DE MELO, F.; LIMA, M. H. A. Four Analyses on the Official Letters of Rio Grande do Norte (Brazil). *Scriptum digital*, [S.l.], n. 5, p. 25-43, 2016b. Disponível em: http://www.scriptumdigital.org/documents/2-Melo_y_Lima_Wok.pdf. Acesso em: 12 jan. 2020.

NIEUWENHUIJSEN, D. Notas sobre la aportación del análisis estadístico a la lingüística de corpus. In: KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016. [E-book]. p. 215-237. DOI: <https://doi.org/10.1515/9783110462357-011>

RODRIGUES, A. C. S.; ALVES, I. M.; GOLDSTEIN, N. S. (org.). *I Seminário de Filologia e Língua Portuguesa*. São Paulo: Humanitas, 1999.

RUMEU, M.; BARBOSA, A.; CALLOU, D. Textos coloniais na América Portuguesa e seus problemas. In: ALKMIM, T. M. (org.). *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas, 2002. v. 3. p. 433-442.

RUMEU, M. C. B. Para uma história do português no Brasil: edição de cartas setecentistas e oitocentistas. In: LOBO, T. et al. (org.). *Para a história do português brasileiro: novos dados, novas análises*. Salvador: EDUFBA, 2006. v. 6. p. 819-844.

SALGADO, G. (org.). *Fiscais e meirinhos: a administração no Brasil Colonial*. Arquivo Nacional. Rio de Janeiro: Nova Fronteira, 1985.

SÁNCHEZ-PRIETO BORJA, P. Interpretación fonemática de las grafías medievales. In: CONGRESO INTERNACIONAL DE HISTORIA DE LA LENGUA ESPAÑOLA, VI., 2003, Madrid. *Actas [...]*. Madrid: Arco; Libros, 2006. v. 1. p. 219-260. Disponível em: <https://ebuah.uah.es/dspace/bitstream/handle/10017/7223/Interpretación%20Fonemática.pdf?sequence=1>. Acesso em: 11 abr. 2016.

SILVA NETO, S. *Ensaio de Filologia Portuguesa*. São Paulo: Companhia Editora Nacional, 1954.

SILVA NETO, S. *Textos medievais portugueses e seus problemas*. Rio de Janeiro: Casa de Rui Barbosa, 1956.

SPINA, S. *Introdução à edótica*. São Paulo: Cultrix, 1977.

TARALLO, F. *Tempos lingüísticos: itinerário histórico da língua portuguesa*. São Paulo: Ática, 1990.

VAAMONDE, G. Limitaciones en el uso de corpus diacrónicos del español. Nuevas aportaciones desde el proyecto de investigación Post Scriptum. *E-AESLA*, Madrid, n. 1, p. 1-10, 2015. Disponível em: <https://cvc.cervantes.es/lengua/eaesla/pdf/01/60.pdf>. Acesso em: 03 jun. 2017.

VASCONCELOS, C. M. V. *Lições de Filologia Portuguesa: segundo as preleções feitas aos cursos de 1911/12 e de 1912/13 seguidas das Lições Práticas de Português Arcaico*. Lisboa: Martins Fontes, [s./d.].