

E ditar, produzir e fazer circular livros que possam colaborar com a melhoria do ensino no Brasil, estabelecer uma ponte entre a produção do conhecimento e a sociedade. Promover a circulação do saber, enfim. Esta tem sido, desde o início, a preocupação da Editora Contexto.

Boa leitura!

Siga-nos:



www.editoracontexto.com.br



SUMÁRIO

APRESENTAÇÃO	7
<i>Ataliba T. de Castilho</i>	
A PLATAFORMA DE CORPORA DO PHPB: UMA APRESENTAÇÃO AD INFINITUM.....	16
<i>Afranio Gonçalves Barbosa</i>	
EXPERIÊNCIAS METODOLÓGICAS EM CONSTITUIÇÃO DE CORPORA: PISTAS PARA UM PESQUISADOR INICIANTE	68
<i>Alexandre Xavier Lima, Leonardo Lennertz Marcotulio e Márcia Cristina de Brito Rumeu</i>	
TRATAMENTO METODOLÓGICO DAS MÃOS INÁBEIS EM CORPORA DIACRÔNICOS	92
<i>Huda da Silva Santiago e Zenaide Novais Carneiro</i>	
REFLEXÕES SOBRE A CONSTITUIÇÃO DE AMOSTRAS DERIVADAS A PARTIR DE UM CORPUS.....	120
<i>Nei Rocha e Leonardo Lennertz Marcotulio</i>	
O CONTROLE INDIRETO DE PERFIS SÓCIO-HISTÓRICOS EM CORPORA HISTÓRICO-DIACRÔNICOS: A IDENTIFICAÇÃO DE GRAUS DE LETRAMENTO PELA GRAFIA ETIMOLÓGICA DO SÉCULO XIX.....	168
<i>Afranio Gonçalves Barbosa e Alexandre Xavier Lima</i>	

O CORPUS DO PROJETO PARA A HISTÓRIA DO PORTUGUÊS BRASILEIRO: A CONSTITUIÇÃO DE CORPORA HISTÓRICOS BASEADA EM CRITÉRIOS DE TRADIÇÕES DISCURSIVAS.....	206
<i>Verena Kewitz e José da Silva Simões</i>	
CARACTERÍSTICAS E POTENCIALIDADES DOS CORPORA DO PORTUGUÊS PAULISTA.....	244
<i>Verena Kewitz e José da Silva Simões</i>	
REFLEXÕES SOBRE A CONSTITUIÇÃO E ANÁLISE DE CORPORA LINGÜÍSTICOS HISTÓRICOS E SOBRE A IDENTIFICAÇÃO DE PERFIS SOCIAIS DE REDADORES DO PASSADO.....	290
<i>Tânia Lobo e Zenaide Novais Carneiro</i>	
REFERÊNCIAS BIBLIOGRÁFICAS.....	313
APÊNDICE.....	350
OS AUTORES.....	363

APRESENTAÇÃO

Ataliba T. de Castilho

Este volume 2 da série *História do Português Brasileiro* é dedicado à organização do *corpus* diacrônico desse projeto. Ele compreende oito capítulos.

Os trabalhos reunidos nesta parte se devem, maiormente, à liderança do prof. Afranio Gonçalves Barbosa, da Universidade Federal do Rio de Janeiro, que coordenou a equipe de *corpus* do projeto em seu todo. Para bem desempenhar essas atividades, o prof. Afranio visitou várias equipes regionais, preparando os pesquisadores locais para a edição filológica dos textos.

Ao mesmo tempo, as equipes regionais se dedicaram ao árduo trabalho de identificação e edição dos documentos de interesse para o projeto Para a História do Português Brasileiro. Devem ser lembrados aqui, entre outros pesquisadores, Marcelo Módolo, que escreveu o primeiro trabalho de grau sobre o tema: Módolo (1998).

Para garantir a uniformidade dos materiais, Afranio Barbosa e Marcelo Módolo, entre outros pesquisadores, fixaram as seguintes *Normas de Transcrição de Documentos Manuscritos e Impressos – Edição Semidiplomática*:

1. A transcrição será conservadora.
2. As abreviaturas serão desenvolvidas, marcando-se – em itálico – as letras omitidas e observando-se os seguintes casos:
 - a. A norma não se aplica às abreviaturas hoje em uso corrente ou fixadas em dicionários. Exemplos: “etc.”, “Sr.”, “Sr^a”, “Itda.”, “Cia.”, “V. Ex.” e “D.” permanecem inalteradas.
 - b. Respeitar, sempre que possível, a grafia do documento, ainda que manifeste idiosincrasias ortográficas do escriba, como no caso da ocorrência “munto”, que leva a abreviatura “m.^{to}” a ser transcrita “munto”.

- c. No caso de variação no próprio documento ou em coetâneos, a opção será para a forma atual ou mais próxima da atual, como no caso de ocorrências “Deos” e “Deus”, que levam a abreviatura “D^s.” a ser transcrita “Deus”.
3. Não será estabelecida fronteira de palavras que venham escritas juntas, nem se introduzirá hífen ou apóstrofo onde não houver. Exemplos: “epor” “ser”; “aellas”; “daPiedade”; “omninino”; “dosertaõ”; “mostrandoselhe”; “achandose”; “sesegue”.
4. A pontuação original será mantida. No caso de espaço maior intercalar deixado pelo escriba será marcado [espaço]. Exemplo: “que podem perjudicar. [espaço] Osdias passaõ eninguemcomparece”. Serão observados dois casos especiais:
 - a. Em relação a trechos que demandem maior esforço para decodificação, seja pela ausência de sinais de pontuação, seja por estarem sob sistema diverso, o editor incluirá, em nota de rodapé, uma possível interpretação. Exemplo: Na edição teríamos, “Tenho uma criada que | dice que sabia fazer tudo | que eu mandace ella | fazer emtaõ perguntei | e Paõ doce voce sabe fazer | sei emtaõ mando todos | os sabados fazer.” Em nota teríamos, “Nota 1: Tenho uma criada que dice que sabia fazer tudo que eu mandace ella fazer. Emtaõ perguntei: E Paõ doce, voce sabe fazer? ‘Sei’. Emtaõ mando todos os sabados fazer”.
 - b. A sinalização [espaço] não se aplica aos espaços em cabeçalhos, títulos e/ou rótulos de seções de periódicos, fórmulas de saudação/encerramento ou na reprodução de diálogos, devendo o editor estabelecer o intervalo conforme o original.
5. A acentuação original será rigorosamente mantida, não se permitindo qualquer alteração. Exemplos: “aRepublica”; “docomercio”; “edemarcando tambem lugar”; “Rey D. Jose”; “oRio Pirahý”; “oexercicio”; “que hé munto conveniente”; “fomos a ele”; “fomos á ele”; “fomos à ele”. Os sinais de separação de sílaba ou de linha, usados pelos autores dos diversos documentos, serão mantidos como no original. Exemplos: “atira- | mos” e “atira= | mos”.
6. Será respeitado o emprego de maiúsculas e minúsculas como se apresentam no original. No caso de alguma variação física dos sinais gráficos resultar de fatores cursivos, não será considerada relevante. Assim, a comparação do traçado da mesma letra deve propiciar a melhor solução.

7. No caso dos impressos, eventuais erros de composição serão marcados com (*sic*) logo após o vocábulo e remetidos para nota de rodapé, onde se deixará registrada a lição por sua respectiva correção. Exemplo: “Nota 1: Acawado por acamado”. Se for inevitável, por limites do editor de textos, o erro será descrito. Exemplo, “Nota 2: A letra <a> inicial de ‘affirma’ está invertida.”
8. No caso dos manuscritos, eventuais grafias diferenciadas serão remetidas para nota de rodapé, onde se registrará(rão) sua(s) variante(s) mais comum(ns) e, quando possível, considerações sobre a variação em si. Exemplos: “Nota 1: ‘que eu afamado livro’ provavelmente ‘que meu afamado livro’. Talvez a escrita de eu por meu possa indicar lapso de escritura ou erro de cópia”; “Nota 2: Pirassocunda possível variante de Pirassununga: talvez a oscilação de grafia indique instabilidade para a escrita de termos Tupi”.
9. Inserções do escriba ou do copista, para não conferir à mancha gráfica um aspecto demasiado denso, obedecem aos seguintes critérios:
 - a. Se na entrelinha do documento original, entram na edição em alinhamento normal e entre os sinais: < >; <↑>, se na entrelinha superior; <↓>, se na entrelinha inferior. Por exemplo: “em dezembro recebi <↑todos> os senadores em casa”. Se houver palavra(s) riscada(s) abaixo da inserção, deverá haver menção ou, conforme sua legibilidade, transcrição em nota de rodapé. Exemplos: “Nota 1: abaixo de <↑todos> há palavra suprimida”; “Nota 2: abaixo de <↑todos> foi riscado ‘dentre’”.
 - b. Se nas margens superior, laterais ou inferior, entram na edição entre os sinais < >, na localização indicada. Exemplo: <fica definido que olugar convencionalado é acasa dePedro nolargo damatriz>. Caso seja necessário, ficará em nota de rodapé a devida descrição da direção de escritura ou quaisquer outras especificidades. Exemplo: “Nota 1: Escrito verticalmente de cima para baixo”.
10. Supressões feitas pelo escriba ou pelo copista no original serão tachadas. Exemplos: “todos ~~ninguem~~ dos presentes assignarom; sahiram ~~sahiram~~ aspressas para oadro”. No caso de repetição que o escriba ou copista não suprimiu, passa a ser suprimida pelo editor que a coloca entre colchetes duplos. Exemplo: “fugi[[gi]]ram correndo [[correndo]] emdireção opaço”.
11. Intervenções de terceiros no documento original devem aparecer em nota de rodapé informando-se a localização. Exemplos: “Nota 1: À

- direita do título encontra-se escrito por outro punho: ‘copiado’; “Nota 2: Na margem inferior encontra-se escrito por outro punho: ‘página 18’”; “Nota 3: Na margem superior encontra-se o carimbado ‘Arquivo Nacional’”.
12. Intervenções do editor hão de ser raríssimas, permitindo-se apenas em caso de extrema necessidade, desde que elucidativas a ponto de não deixarem margem à dúvida. Quando ocorrerem, devem vir entre colchetes. Exemplo: “não deixe passar neste [registro] de Areas”. Quando houver dúvida sobre a decifração de alguma letra, parte de ou vocábulo inteiro, o elemento em questão será posto entre colchetes e em itálico. Exemplos: “ent[*re*]gue o [*rapaz*]”; “faça venda a duas b[*arric*]as de vinho”.
 13. Letra ou palavra(s) não legíveis por deterioração ou rasura justificam intervenção do editor com a indicação entre colchetes conforme o caso: [.] para letras, [ilegível] para vocábulos e [ilegível. + n linhas] para a extensão de trechos maiores. Exemplos: “É assim pe[.]r.”; “É assim [ilegível.] em Java”; “É assim [ilegível + 2 linhas] em Havana.” Caso suponha ser extremamente necessário, o editor indica em nota a causa da elegibilidade: corroído, furo, borrão, rasura, etc.
 14. Letra ou palavra(s) simplesmente não decifradas, sem deterioração do suporte, justificam intervenção do editor com a indicação entre colchetes conforme o caso: [?] para letras, [*inint.*] para vocábulos e [*inint.* + número de linhas] para a extensão de trechos maiores. Exemplos: “É assim pe[?]r.”; “É assim [*inint.*] em Java”; “É assim [*inint.* + 2 linhas] em Havana.”
 15. A divisão das linhas do documento original será preservada, ao longo do texto, na edição, pela marca de uma barra vertical entre as linhas. A mudança de parágrafo será indicada pela marca de duas barras verticais. Exemplo: “Es- | taes pois muito atrazado, ponde-vos na | pira meu ignorantão. || Seria bonito que todas as.”
 16. A mudança de fólio ou página receberá a marcação entre colchetes conforme o caso:
 - a. Se em documentos manuscritos, com o respectivo número e indicação de frente ou verso. Exemplos: [fol. 1r]; [fol. 1v]; [fol. 2r]; [fol. 2v]; [fol. 3r]; [fol. 3v]; [fol. 16r].
 - b. Se em documentos impressos, indicação de página. Exemplos: [p. 1]; [p. 2]; [p. 3]; [p. 19].
 17. Na edição, as linhas serão numeradas de cinco em cinco a partir da quinta. Essa numeração será encontrada à margem direita da mancha, à esquerda do leitor. Será feita de maneira contínua por documento.

18. Os sinais públicos, diferentemente das assinaturas e rubricas simples, serão sublinhados e indicados entre colchetes. Exemplos: assinatura simples, Bernardo Jose de Lorena; sinal público, [Bernardo Jose de Lorena].
19. Informações que o editor julgar significativas sobre a diagramação e *layout* do texto em impressos devem aparecer em nota de rodapé.

Essas Normas reproduzem o texto “Normas para transcrição de documentos manuscritos para a história do português do Brasil”, publicado em Mattos e Silva (Org. 2001a, tomo 2, pp. 553-555), tendo-se agregado pequenas alterações, resultantes da prática de edição.

Tais Normas vieram encimadas pelas seguintes informações:

A Comissão de estabelecimento de Normas para transcrição de documentos manuscritos para a História do Português do Brasil, formada pelos pesquisadores César Nardelli Cambraia (USP), Gilvan Muller de Oliveira (UFSC), Heitor Megale (USP), Marcelo Módolo (Mestrando – USP), Permínio Souza Ferreira (UFBA), Sílvio de Almeida Toledo (USP), Tânia C. Freire Lobo (UFBA), Valdemir Klamt (UFSC), após apresentação e discussão dos subsídios colhidos nos trabalhos individuais, levou a plenário um elenco de normas que, submetidas a ampla discussão, tiveram como resultado a aprovação das seguintes *Normas para transcrição de documentos manuscritos para a história do português do Brasil* (texto acima).

Posteriormente, realizou-se em João Pessoa, por ocasião do VIII Seminário Nacional do PHPB uma reunião da Comissão de *Corpus*. A ata respectiva foi transcrita no “Apêndice” deste volume.

*

Seguem-se os resumos dos capítulos.

No primeiro capítulo, “A Plataforma de *corpora* do PHPB: uma apresentação *ad infinitum*”, Afranio Gonçalves Barbosa oferece a memória da construção do *corpus* do Projeto Nacional Para a História do português brasileiro (PHPB) de modo a servir como referência de etapas, questões e operacionalidade para outros empreendimentos de construção de

corpora histórico-linguísticos com grande cobertura diatópica. São apresentadas as opções metodológicas de base que definiram os controles externos na constituição do *corpus mínimo comum* do projeto e do *corpus diferencial*, material contrastivo inserido na *Plataforma de corpora do PHPB* na internet, que cataloga o conjunto total de edições organicamente realizadas sob a chancela do PHPB.

No segundo capítulo, “Experiências metodológicas em constituição de *corpora*: pistas para um pesquisador iniciante”, Alexandre Xavier Lima, Leonardo Lennertz Marcotulio e Márcia Cristina de Brito Rumeu trazem várias reflexões sobre a árdua tarefa da constituição de *corpora* diacrônicos. Com base na experiência em trabalhos anteriores (Lima, 2010; 2014; Marcotulio, 2010; Rumeu, 2004; 2013), os autores pretendem contribuir para a otimização do trabalho de quem inicia uma pesquisa em arquivos, constituindo um *corpus* para a sua investigação.

No capítulo seguinte, “Tratamento metodológico das *mãos inábeis* em *corpora* diacrônicos”, escrito por Huda da Silva Santiago e Zenaide Novais Carneiro, são tratadas questões tais como a identificação da inabilidade em escrita alfabética nos planos supragráfico e paleográfico, da segmentação gráfica, da escriptualidade, da escrita fonética e da repetição lexical. A seguir, são examinadas as cartas de sertanejos no século xx.

Em “Reflexões sobre a constituição de amostras derivadas a partir de um *corpus*”, os autores Nei Rocha e Leonardo Lennertz Marcotulio centram-se no *corpus histórico do português*, propondo uma reflexão sobre a constituição de amostras derivadas a partir de uma amostra base, tanto do ponto de vista da disponibilidade e comparabilidade de materiais quanto do ponto de vista estatístico, com o intuito de discutir a possibilidade de obtenção de amostras mínimas significativas que conservem o potencial de representatividade do *corpus*.

O capítulo “O controle indireto de perfis sócio-históricos em *corpora* histórico-diacrônicos: a identificação de graus de letramento pela grafia etimológica do século xix”, escrito por Afranio Gonçalves Barbosa e Alexandre Xavier Lima, oferece caminhos metodológicos para controle indireto de perfis sociais de redatores de manuscritos. Para o século xix, adotou-se

como parâmetro o controle da grafia etimológica em redatores metonímicos (reconhecidos como eruditos) e em textos modelares (jornais e gramáticas). O estabelecimento das *taxas de uso, de acerto, de abertura à variação*, às *faixas de aplicabilidade* e aos *grupos de consenso entre gramáticos* representa um avanço da reflexão metodológica, na tentativa de se reconhecer, com maior exatidão, cada ponto do contínuo de letramento, podendo, assim, servir como referencial na identificação do perfil de um redator, na ausência de informações biográficas.

No sexto capítulo, “O *corpus* do projeto Para a História do Português Brasileiro: a constituição de *corpora* históricos baseada em critérios de Tradições Discursivas”, Verena Kewitz e José da Silva Simões associam o modelo das Tradições Discursivas à Linguística de *Corpus*, mostrando a relevância dessa associação para o PHPB. Refletindo sobre as atividades das equipes regionais na organização do *corpus*, os autores mostram que há uma preocupação geral em documentar as normas popular, culta, cotidiana, entre outras, tendo em vista um trabalho descritivo e histórico que fuja ao enviesamento trazido por materiais pouco representativos da complexidade linguística brasileira. Os autores defendem os seguintes critérios para a caracterização dos documentos: (i) *grau de publicidade do documento* (particular, semiparticular ou público), (ii) *grau de simetria entre emissor e receptor* (ascendente, horizontal ou descendente), (iii) *fixidez temática* (sem fixidez, núcleo temático fixo ou tema altamente fixo), (iv) *dimensões da ação comunicativa ou objetivos do texto* (narrar, exigir, repreender, reclamar, confirmar, agradecer, atacar etc.), (v) *grau de planejamento do texto* (livre, semicontrolado ou altamente controlado). Complementam esses critérios alguns aspectos sócio-históricos, com foco no estado de São Paulo, para colocar melhor a questão da história dos textos paralelamente à história da língua portuguesa no Brasil. Esses critérios e aspectos foram apresentados no VIII Seminário do PHPB, em reunião da Comissão de *Corpus*, tendo-se decidido organizar dois elencos de documentos: o *corpus* mínimo, a ser utilizado por todos os pesquisadores, e o *corpus* diferencial, muitas vezes exclusivo de uma determinada região e/ou época. Em seguida, como forma de ilustrar a aplicação prática dos critérios sugeridos, os autores examinam o desenvolvimento histórico

do gênero *carta* sob a óptica da função comunicativa que o gênero exerceu desde suas origens e as transformações decorrentes da necessidade de uso desse gênero em diversos contextos. Esse exercício demonstra a relevância de se levar em conta os entornos comunicativos ligados às questões da história social da língua, dos contatos linguísticos, das normas linguísticas em competição nas diversas sincronias e da difusão de textos escritos (manuscritos e impressos) ao longo da história do Brasil. Mais adiante, os autores apresentam as potencialidades da *carta* e seus subgêneros para estudos de fenômenos linguísticos variados e concluem “*que a análise da história do gênero vai além dos limites da língua particular e que é necessário investigar a gênese dos textos e sua relação com outras línguas e tradições culturais*”.

Em “Características e potencialidades dos *corpora* do português paulista”, Verena Kewitz e José da Silva Simões descrevem os documentos que compõem os *corpora* do português paulista (editados até 2016) quanto à tipologia textual, às tradições discursivas e às potencialidades de cada conjunto com exemplos de vários fenômenos linguísticos. Os *corpora* dividem-se em (i) mínimo comum manuscrito, composto de cartas e seus subgêneros, atas de câmara, processo-crime e testamentos; (ii) mínimo comum impresso, com anúncios de jornais e revistas, cartas de leitor e de redator de jornais e revistas e notícias de jornais; e (iii) *corpus* diferencial, contendo vários tipos de texto (manuscritos, impressos e orais), com destaque para as memórias históricas, diários de viagem e descrições, além de alguns documentos portugueses, que se justificam pela linha de pesquisa sobre as origens do português brasileiro.

No oitavo e último capítulo, “Reflexões sobre a constituição e análise de *corpora* linguísticos históricos e sobre a identificação de perfis sociais de redatores do passado”, Tânia Lobo e Zenaide Novais Carneiro refletem sobre o percurso da equipe baiana do projeto Para a História do Português Brasileiro na organização de *corpora* linguísticos históricos, com destaque para um aspecto: a pesquisa para identificação do perfil social de *quem* escreveu em sincronias passadas da história da língua. Dada a condição tão multiétnica e, no passado, generalizadamente multilíngue da sociedade brasileira, tal questão ganha relevância, indicando a necessidade de os

corpora distinguiam textos escritos em português como L1 (quer o português europeu que se transplantava em sucessivas vagas, quer o português brasileiro que se constituía) e textos, obviamente mais raros, escritos em português como L2 por índios, africanos e, já na fase pós-colonial, por imigrantes europeus e asiáticos de procedência diversa. Adicionalmente, também se reflete sobre o fato de que, ainda que se atingisse o ideal de responder, para qualquer texto escrito no passado, ao seguinte conjunto mínimo de questões – *o que* foi escrito, *quando* foi escrito, *onde* foi escrito, *quem* escreveu, *para quem* escreveu e *por que* o fez –, um dos problemas centrais da Sintaxe Diacrônica, relativo a *como apreender as mudanças nas gramáticas através da análise dos textos*, ainda exigiria enfrentamento teórico e metodológico por parte do linguista.

Por fim, há um “Apêndice” que traz uma relação dos *corpora* apurados até 2015. Ele foi organizado com base nas informações enviadas pelos coordenadores regionais do PHPB.

Boa leitura!



A PLATAFORMA DE CORPORA DO PHPB: UMA APRESENTAÇÃO *AD INFINITUM*

Afranio Gonçalves Barbosa





SUMÁRIO

APRESENTAÇÃO	18
A PLATAFORMA DE CORPORA, O CORPUS MÍNIMO COMUM E O CORPUS DIFERENCIAL	18
Estrutura da plataforma de <i>corpora</i> do PHPB: as categorias de entrada.....	23
<i>Corpus</i> diferencial no PHPB.....	31
TRADIÇÕES DISCURSIVAS E RÓTULOS FUNCIONAIS PRETÉRITOS EM CORPORA GERAIS	36
O TRABALHO CONTÍNUO DE AMPLIAÇÃO DE CORPORA HISTÓRICOS: A EXPERIÊNCIA NO PHPB.....	42
CONSIDERAÇÕES FINAIS.....	51
ANEXO	53



APRESENTAÇÃO

O presente volume da série *História do Português Brasileiro*, intitulado *Corpus diacrônico do português brasileiro*, apresenta aos seus leitores uma fotografia em *close* de uma panorâmica de ideias e materiais construídos desde 1997, e que continuam em andamento. Portanto, falar do estado do *corpus* do Projeto é falar de uma ponta de um *iceberg* que segue crescendo *ad infinitum*, até onde for a necessidade de a Linguística Histórica contar com materiais diversificados e controlados sócio-historicamente. Sem esse controle, os estudos científicos sobre objetos sintáticos, morfológicos, fonéticos, discursivos serão diacrônicos, mas não histórico-diacrônicos. Nossa tarefa, neste capítulo, além de sistematizar experiências de trabalho do Projeto até 2013, é expor critérios de nossa construção e um quadro geral do que denominamos Plataforma de *corpora* do projeto Para a História do Português Brasileiro (PHPB), da qual uma parcela foi disponibilizada na rede mundial de computadores sob rótulo “Prévia do *corpus* mínimo comum do PHPB”. Trata-se de um texto dentro do *continuum* sem fim que olha para o feito com vistas ao fazer futuro de usuários e construtores de novos *corpora* históricos. Concentremo-nos em esclarecer qualitativamente certos critérios de construção do *corpus* com vistas a colaborar na preparação de novos pesquisadores.

A PLATAFORMA DE CORPORA, O CORPUS MÍNIMO COMUM E O CORPUS DIFERENCIAL

O *corpus* do PHPB encontra-se catalogado na denominada Plataforma de *corpora*, uma sistematização de **todos** os materiais editados pelos membros do Projeto até o *Censo de corpora do PHPB* em junho de 2010. Por meio de suas categorias de entrada, os pesquisadores poderão ter noção da massa impressionante de edições diplomático-interpretativas lançadas após anos de trabalhos de pesquisadores de várias universidades brasileiras, produzidos tanto sob a chancela do PHPB, como também por colaboradores externos. Façamos uma breve descrição das partes dessa *massa* para, logo em seguida, aprofundarmos nossas questões.

No estágio atual de trabalho, o PHPB disponibiliza – por meio da página <https://sites.google.com/site/corporaphpb/> – duas parcelas do material in-

dexado na Plataforma: O *CORPUS* MÍNIMO COMUM e o *CORPUS* DIFERENCIAL. O chamado *corpus mínimo comum* reúne, para controle diatópico, materiais de mesma natureza já levantados pelas 12 equipes. O *corpus diferencial*, por sua vez, compreende *corpora* complementares para controle contrastivo ao *corpus mínimo comum*: textos de portugueses, literários, gêneros textuais diversos etc. Nessa página, poderão ser visualizados e baixados os documentos em formatos .doc ou .pdf, bem como verificada a versão mais atual da Plataforma e a segunda versão das “Normas de Edição do Projeto PHPB”.

O *corpus mínimo comum* conta com redatores brasileiros ou de naturalidade não identificada, mas com produção em redes de escrita no Brasil. Os identificados como portugueses passam ao *corpus diferencial*. Sua organização distribui: equipe regional por equipe regional, três rótulos textuais de impressos – cartas de leitores, cartas de editores e anúncios – e dois outros para manuscritos: cartas particulares e cartas de leitores. Cada um desses rótulos constitui uma célula comparativa com o número mínimo de 5 mil palavras para cada metade de século. Todo o potencial comparativo entre esses gêneros de materiais em termos regionais e temporais serão explorados à frente em capítulo específico sobre a questão do tamanho da amostra do PHPB e o tamanho mínimo de amostras para diferentes tipos de pesquisas conforme tipos de fenômenos linguísticos que delas retirem dados.

Se a configuração do *corpus mínimo comum* do PHPB não cobre a totalidade de materiais na dimensão geográfica de todas as equipes regionais, importantes vias de contrastes podem ser feitas cruzando um mínimo de três equipes regionais. São eles:

1. O contraste entre periódicos IMPRESSOS e MANUSCRITOS pessoais na diacronia interna do século XIX.
2. O contraste entre documentos manuscritos OFICIAIS e documentos manuscritos PESSOAIS na diacronia interna do século XVIII.
3. O contraste diacrônico de MANUSCRITOS entre os séculos XVIII, XIX e XX.

Essas três vias de contraste agrupam materiais de forma semelhante ao que se chamou de *corpus compartilhado* entre equipes regionais do famoso Projeto Nacional Norma Urbana Culta (Nurc) na década de 1990: um conjunto menor de fontes tirado do conjunto maior produzido pelo Projeto de modo a disponibilizá-las de modo equilibrado **por local**. No caso do Nurc,

cinco capitais compunham a dimensão espacial dos estudos sobre a norma culta do português brasileiro; no caso do *corpus mínimo comum*, 12 equipes regionais recolhem, em acervos locais, fontes primárias para a pesquisa. O controle diatópico acaba sendo o *fio de prumo* dessas fontes em ambos os Projetos. Esse deve ser o primeiro controle a ser discutido.

Os critérios de recorte para a dimensão espacial nos *corpora* do PHPB coincidem com a distribuição geográfica de suas equipes regionais, independentemente de estarem em capitais de seus estados ou não. É o limite do bom senso. Vale, contudo, lembrar dos limites desse controle externo em *corpora escritos* em geral.

A rotulação *Rio de Janeiro* ou *Salvador*, a título de ilustração, para controle extralinguístico de dados levantados em *corpora orais* implica uma metonímia entre o espaço geográfico-administrativo usado como categoria nominal de entrada no *corpus* escolhido (aqui, nesses exemplos, em nível municipal) e o grupo humano sob clara identidade cultural reunido no espaço físico de isoglossas e no espaço abstrato do imaginário social (aqui, nesses exemplos, cariocas e soteropolitanos). Trata-se de identidades culturais conhecidas por toda sociedade e que se traduzem, inclusive, em estereótipos populares bem antes de se escolherem informantes a serem gravados para um *corpus*. Trata-se de identidades culturais que estão na base do pressuposto de haver traços linguísticos comuns entre os membros dessas comunidades de fala em tal ordem de normalidade, que se tornam capturáveis em amostras tratadas sob rigorosos critérios estatísticos. Como já dito em Barbosa (2010: 14):

No plano da linguagem, de uma identidade reconhecida em outros planos culturais. Sabe-se, de antemão, das identidades baiana e carioca. Sabe-se haver traços linguísticos constantes que se somam e se sobrepõem às variações diafásicas e de redes sociais em que os falantes estão inseridos. Esse conhecimento das identidades é partilhado previamente aos estudos linguísticos.

O contraponto a ser feito é não existir essa tranquila objetividade em relação aos *corpora orais* quando aplicamos o controle extralinguístico diatópico em codificações de dados levantados de *corpora escritos* de sincronias passadas. Que conhecimento prévio podemos ter sobre as identidades culturais no passado para estabelecermos a referida metonímia entre conjunto de textos escritos de um local e os grupos humanos desse local?

Se um dado *corpus* ideal somente contar com naturalidades plenamente identificadas, seus recortes geográficos evidentemente representarão metonímia de normas vernáculas filtradas pela imagem difusa da escrita. Suas representações serão filtradas pelas normas não vernáculas das diferentes redes de escrita em que se inseriam os autores, por tradições discursivas em linguagem especial, dentre outros elementos. As isoglossas pretéritas estão perdidas, os estereótipos vivos no passado, esquecidos na bruma do tempo. É um limite da Linguística Histórica. Como os *corpora* ideais (contando somente com autores plenamente identificados) são exceções, é preciso entender o alcance de nossas afirmações calcadas no controle diatópico proporcionado pelos recortes em *corpora* histórico-diacrônicos reais, com diversos autores não identificados.

Veja-se que, no caso dos *impressos*, cuja maioria dos textos de jornais nem vem assinados, não se sabe se os *anúncios*, *carta de leitores* e *editoriais* publicados em periódicos do Rio de Janeiro foram escritos por cariocas. O jornal é carioca, por isso está no *corpus*, mas quem escreve pode não ser da cidade. A presença do jornal no *corpus* carioca proporciona a pesquisa da linguagem dos textos escritos modelares a segmentos sociais participantes de suas redes de escritura no e a partir do Rio de Janeiro. Proporciona, também, recuperar o principal veículo de gêneros textuais promovidos e promotores da cultura letrada em diferentes fases da história social da escrita no Brasil. Quando se descobre quem eram os autores que assinaram materiais, abre-se a possibilidade de se estudar materiais escritos para um grau máximo de publicidade e, por isso mesmo, escritos sob diferente estado de vigília de seus autores: eles escrevem sem saber quem serão seus leitores. No caso de parte dos *manuscritos* nos *corpora* do PHPB é semelhante a situação. Quando não se sabem dados biográficos de quem assina o documento, a localização *Rio de Janeiro* posta na *carta particular* ou no *documento oficial* é o elemento que a faz ser selecionada e editada pela equipe carioca, mas, na verdade, não se sabe se seu autor era carioca. São limites normais dos *corpora* históricos que não impedem generalizações dos resultados de cada pesquisa, mas exigem maior cuidado em assumi-las com a mesma tranquilidade de quando se trabalha com *corpora orais*.

Como se vê, os resultados estatísticos, em uma dada pesquisa histórico-linguística, do cruzamento de controles internos e o controle diatópico estão, a princípio, mais ligados a explorar o conjunto de papéis em si do que simplesmente assumir o local de escritura expresso nos textos por seus

autores. A contextualização histórica é tarefa dos membros do Projeto que editam os textos, bem como dos organizadores dos *corpora*. Mas, a bem da verdade, em grande parte, passa a ser tarefa de investigação a ser realizada pelos usuários dos *corpora*.

Evidente que essa realidade não impede, necessariamente, que os resultados do cruzamento possam ser representativos de costumes linguísticos locais, mas não se pode fazer essa relação diretamente, sem considerar a forte possibilidade de estarem, na verdade, espelhando, por exemplo, as normas linguísticas de grupos sociais/profissionais/de interesse da rede de escrita-leitura dos autores. Além de tantas ressalvas para se evitarem falsas generalizações nas investigações com dados levantados nos *corpora* do PHPB, o que mais o controle externo diatópico permite afirmar? O mesmo que outros *corpora* escritos de mesma natureza:

QUE O CONJUNTO DE TEXTOS REUNIDOS SOB ETIQUETA DE CADA EQUIPE REGIONAL DO PHPB, A CADA 50 ANOS, PODE EXPLICITAR DIFERENTES DISTRIBUIÇÕES DE DADOS LINGUÍSTICOS CAPTURÁVEIS NA AMOSTRA REPRESENTATIVA DE DIFERENTES GRUPOS DE TEXTOS SEPARADOS POR RÓTULOS SOCIOPRAGMÁTICOS GENÉRICOS.

A partir da captura estatística de diferenças patentes na distribuição de variantes linguísticas, a generalização desses resultados é tarefa qualitativa de cada usuário-pesquisador à medida que se possa cruzá-los com resultados de outros trabalhos histórico-sincrônicos sobre os mesmos lugares. São as constantes encontradas pelos pesquisadores nesses contrastes com outros materiais que abrirão, um pouco mais, a fresta por que olhamos a língua usada no passado.

Ao fim e ao cabo, na prática, aqui, são constantes a serem encontradas a partir das três hipóteses de controle resumidas em (1), (2) e (3), pelos os usuários do *corpus mínimo* disponibilizado na internet e, se quiserem, encontradas nas novas possibilidades de controle com quaisquer materiais presentes na Plataforma de *corpora*, solicitados aos coordenadores das equipes regionais via endereços eletrônicos. Por exemplo, pode-se criar uma nova célula comparativa entre manuscritos pessoais do Rio de Janeiro da primeira metade do século XX e dados de mãos inábeis da Bahia da mesma época verificando, assim, um *continuum* mais completo de níveis de escrita alfabética, desde diferentes níveis de inabilidade, passando por redatores hábeis em nível elementar e mediano, chegando a graus maiores de domínio de escrita

em redatores letrados. Para isso, bastaria baixar os manuscritos escritos do Rio de Janeiro no *corpus mínimo comum* e solicitar, via e-mail, manuscritos de mãos inábeis de redatores baianos catalogados na Plataforma de *corpora* anexada ao fim deste capítulo.

Estrutura da plataforma de corpora do PHPB: as categorias de entrada

DUAS METADES DOS SÉCULOS XVII, XVIII, XIX E XX

A distribuição de células de materiais editados dentro do PHPB, ou seja, as categorias de entrada dispostas em tabelas, revela opções metodológicas prévias à constituição de *corpora*. Quais seriam elas?

O primeiro recorte de qualquer *corpus* linguístico é a dimensão temporal. Sem quaisquer agrupamentos cronológicos a partir das datações expressas nos documentos levantados, não há como caracterizar historicamente o estado sincrônico dos fenômenos em tempos pretéritos. É bem verdade que *corpora* linguísticos podem, por conta de diferentes pontos de vista teóricos e/ou metodológicos, ter outros referenciais de datação, como (i) as datas de nascimento dos redatores ou (ii) eventos histórico-sociais.

No primeiro caso, agrupam-se os textos no *corpus* em células conforme a época de aquisição da linguagem dos autores. No segundo caso, os textos são agrupados em função de mudanças consideradas significativas na estrutura da sociedade que, por hipótese, teriam deixado reflexos na relação social com a escrita e, por conseguinte, na linguagem transparecida nos textos de então. No entanto, são parâmetros a serviço de objetivos e questões de pontos de vista específicos que dificultam a comparabilidade entre *corpora* construídos por diferentes projetos. Se não se perde de vista que a datação e etapas de aquisição de escrita não equivalem às da aquisição de fala, não se perde a noção de que a produção escrita pretérita não permite discernir com clareza que elementos ficaram da fase de aquisição de linguagem em usos mais vernáculos e que elementos nos textos são incorporações de sistemas mais periféricos, ou mesmo quase que uma Língua 2 advinda dos sistemas de uso não vernáculos das Tradições Discursivas (TDs) modelares para os redatores.

Sem perder de vista essas questões, um *corpus* com recortes externos de datação específicos como esse tem comparabilidade imediata com outros de igual base. A recuperação incidental dos usuários para reagrupar os do-

cumentos pela data de escritura será *a posteriori*. A experiência nos mostra que essa necessidade nem sempre é percebida por usuários iniciantes no levantamento de dados em *corpora* histórico-diacrônicos. Esse é o caso do Projeto Varport, que optou por contrastar a produção escrita no século xx no Brasil e em Portugal sem paralelo exato no recorte temporal em suas células nominais das tabelas do *corpus* disponibilizados. Enquanto o século xx é dividido em quatro fases no Brasil (dentre as quais a Fase 3 com recorte entre 1950-1974 e a Fase 4 entre 1975-2000), ele vem dividido em cinco fases em Portugal. Em função das mudanças profundas entre o Portugal sob a ditadura política e ainda colonizador em África e o Portugal democrático e subsequente às guerras coloniais, optou-se por já dispor a entrada temporal compreendendo a Fase 3 entre 1950-1960, uma quarta fase entre 1961-1974 e, ainda, uma Fase 5 entre 1975-2000. O usuário pode reuni-las conforme as Fases 3 e 4 pensadas para o Brasil. É uma ação incidente ao *corpus* posterior e conforme os objetivos comparativos dos usuários com outros *corpora*. É semelhante ao que terá de ser feito com quem compara as duas metades do PHPB com outros conjuntos. A questão aqui é de caráter conceitual: um *corpus* construído para atendimento geral, sem focar objetivos específicos deveria deixar o maior número de ações incidentais de reagrupamento para os usuários. A opção de partir das categorias nominais mais externas, como a divisão temporal, com menor neutralidade sempre acaba confundindo usuários mais incipientes e gerando confusões.

Por essa razão, *corpora* que optam pelo caráter aberto, geral, acabam buscando a neutralidade da simples linha cronológica segmentada em décadas, séculos etc. e deixam a recuperação da data de nascimento dos redatores e as informações sobre a sociedade à época como matéria para os cabeçalhos postos em cada fonte neles editada. Dessa forma, são os usuários, se lhes for importante, que organizam a documentação ali disponibilizada conforme suas questões específicas. Esse é o caso dos *corpora* histórico-diacrônicos do PHPB.

Observe-se que, aqui, não se emprega a expressão geral e consagrada *corpus diacrônico* como equivalente a *corpus* linguístico histórico-diacrônico, pois um *corpus* histórico é sempre preparado para entradas sincrônicas, que possibilitam aos usuários estudar propriedades funcionais e/ou fatores favorecedores de variáveis dependentes em um dado momento da história. Uma tese com materiais do século XVIII não é, a rigor, uma tese diacrônica, é uma tese histórica, com a história linguística da sincronia século XVIII. O *corpus* linguístico montado com vários recortes sincrônicos torna-se um *corpus* his-

tórico-diacrônico na abordagem dos dados empreendida pelos usuários. Se ele toma os dados do século XVIII como um todo, em contagem somando todos os dados, sua abordagem é histórico-sincrônica; se ele toma os dados do século XVIII em contagem que separe os mesmos dados entre primeira e segunda metades do século XVIII, sua abordagem é diacrônica, ou seja, compara duas sincronias para avaliar estabilidades ou mudanças de um ponto a outro no tempo. Nesse exemplo, o *ponto no tempo* reúne 50 anos.

Seria melhor confrontar fenômenos linguísticos por meio de entradas temporais de dez em dez anos ao invés de separar a documentação em metades de séculos? Evidente que sim. As linhas que unem os pontos nos gráficos resultantes dos fenômenos cruzados com o eixo temporal apresentam traçados mais fiéis aos movimentos desse mesmo fenômeno na língua em uso quando as categorias de entrada no *corpus* estão mais segmentadas. Mas o que se escolhe como *ponto no tempo* em cada *corpus* depende das possibilidades de avanço orgânico, em parêntese, comum das equipes regionais. No caso do PHPB, apesar de se pretender, inicialmente, que cada equipe levantasse sua documentação impressa e manuscrita em três pontos dentro de cada século (sem três pontos não há curva ascendente ou descendente feita dentro do século), só foi possível organizar internamente os séculos em dois pontos, em duas metades. As curvas diacrônicas ficaram para o cotejo de três metades, quando se passa de um século a outro. Na Plataforma, isso poderá ser feito com materiais levantados e editados do século XVII ao século XX, ficando o *núcleo ótimo comparativo*, contudo, entre quatro pontos no tempo: a segunda metade do século XVIII, primeira metade do século XIX, a segunda metade do século XIX e a primeira metade do século XX.

Não por acaso, é a partir do século XIX que se torna maior a concentração de documentação sobrevivente nos arquivos públicos e privados dentro e fora do PHPB. O avanço das instituições que vão estruturando a nação, em especial desde o Império até a República, reflete-se na maior condição de se guardarem textos escritos nos âmbitos administrativos e nas redes de escrita, que se vão tornando mais complexas. Por exemplo, sem as iniciativas de criação de institutos históricos e geográficos em várias cidades brasileiras (no Rio de Janeiro, desde 1838), não teríamos disponíveis hoje, muitas já digitalizadas na internet, as revistas criadas por esses institutos, que trazem à luz textos sob Tradições Discursivas típicas de círculos de leitura e escritura entre acadêmicos. Da mesma forma, sem o avanço dos Correios e Telégrafos por terra, ao longo século XIX, os jornais não teriam tido acesso ao serviço de

distribuição de seus exemplares pelo interior e entre capitais para assinantes de diferentes províncias. Tal infraestrutura criou redes de escrita e leitura com provocações e respostas entre leitores de jornais paulistas, mineiros e cariocas. Essas redes aumentam nossa oportunidade, por conta do maior número de exemplares em circulação, de hoje encontrarmos documentação produzida numa cidade e guardada em acervos de outra. Os pesquisadores, em todo o país, podem, portanto, recuperar materiais dessas redes de escrita e leitura mais complexas, formar *corpora* especiais linearizados em sincronias de 50 em 50 anos e cotejá-los aos do PHPB como *corpora diferenciais*. A neutralidade da distribuição de sincronias na simples linha do tempo permitiria essa relação mais direta, ao menos, entre a segunda metade do século XIX e a primeira metade do século XX, três pontos dentro do chamado *núcleo ótimo comparativo* do PHPB.

IMPRESSOS E MANUSCRITOS

Aos olhos de um usuário incipiente de *corpora* linguísticos histórico-diacrônicos, na recolha de dados para sua dissertação de mestrado, por exemplo, talvez pareça óbvio que se deva contrapor *manuscritos* a *impressos*. Se sua inserção no uso social da escrita é anterior aos anos 2010 – quando a popularização das mídias eletrônicas em redes sociais subverte a relação público/privado da escrita na internet –, talvez sua experiência com a oposição manuscrito/impresso lhe sugira sempre algo mais pessoal ou circunscrito a pequenos círculos com manuscritos e algo sempre mais voltado ao público em impressos. Teria sido assim na escrita de cartas, bilhetes, cartões postais, redações e cadernos escolares, por um lado, e na escrita de santinhos de propaganda de rua, jornaizinhos na escola, jornais e livros, por outro. No entanto, na relação histórica entre produtos escritos pelas diferentes sociedades ao longo do tempo, essas fronteiras que se estabelecerem no século XX, agora desfeitas com o quase fim dos manuscritos no século XXI, apresentam diferentes relações nos séculos XVII, XVIII e XIX.

Até o início do século XIX, a sociedade brasileira era uma sociedade manuscrita. Era proibida, por leis coloniais, a impressão de livros, panfletos e jornais no Brasil. A despeito de uma ou outra situação de exceção, somente com a vinda da família real para o Rio de Janeiro é que a difusão de prensas se espalha rapidamente. Em três décadas, multiplicam-se jornais do norte ao sul de modo quase “exponencial”. Em 1830, a revista *O Beija-flor*, no

Rio de Janeiro, afirmava que, desde 1827, as publicações periódicas haviam quadruplicado no Brasil, passando de 12 ou 13 para 54. Antes, manuscruver não era, em si, um ato voltado para círculos fechados ou destinatário único. Na colônia, por exemplo, a prática de serem concedidas licenças para livros manuscritos perdura por todos os setecentos. Mesmo na administração pública, a impessoalidade era total na esmagadora maioria de redes de relacionamento entre remetente e destinatário dos documentos. As cartas à rainha, ou a qualquer autoridade portuguesa enviada do Brasil no século XVIII, não seriam lidas pelo destinatário evocado no texto. Eram milhares de mercês que hoje lotam as caixas do Arquivo Histórico Ultramarino de Lisboa. A mudança na história social da escrita no Brasil que se opera no século XIX cria uma nova relação com impressos que os torna lugar da maior publicidade que antes não opunha manuscritos.

Em função dessa mudança na história social da escrita no Brasil, para além de incluir, como categoria de entrada, *impressos versus manuscritos* no século XIX, o recorte externo do *corpus* que controla a maior/menor publicidade teve de opor *manuscritos de circulação na administração pública* versus *manuscritos de circulação privada* para o século XVIII. Os livros impressos que circulavam na colônia poderiam ser postos no século XVIII em oposição a manuscritos tal qual foi feito no *corpus* para o século XIX? Pelo critério genérico de oposição *manuscritos versus impressos* poderia. Mas haveria, nesse caso, uma desconexão com o impacto e valor dos manuscritos em cada século. Os raros impressos no Brasil colônia, objetos merecedores de figurar em inventários, não serviam para tornar públicas as matérias escritas, tampouco valiam de modelos de escrita para seus leitores-redatores da mesma forma que os impressos periódicos o foram ao longo do século XIX. Além disso, a circulação de documentos públicos continua nos séculos XIX e XX, de forma impressa, mas perde para os jornais o caráter que apresentava até o século XVIII: o de textos modelares para os manuscritos pessoais. Esse é o critério de base para as oposições ao longo do tempo na Plataforma e no corpus *mínimo comum* do PHPB: a relação entre textos modeladores e textos modelados em cada sociedade. *Grosso modo* falando, até o avanço, no século XIX, dos impressos, os manuscritos pessoais refletem mais a pluriortografia presente nos documentos burocráticos do que a pluriortografia dos raros impressos disponíveis aos letrados. Com os jornais, a leitura dos impressos aparece refletida nos manuscritos de redatores letrados, que, à semelhança do que

ocorre nos impressos, reduzem ou mesmo somem com as grafias mais fonéticas e exacerbam as grafias latinizadas. Se nos impressos e manuscritos até o século XVIII havia uma relação pendular entre grafias fonéticas e latinizadas – consequência, conforme Barbosa (2008: 195), prática da “*pluriortografia: havia diferentes obras ortográficas predicando formas de escrever diferentes*” –, isso se define para o lado etimológico latinizante no XIX. Essa mudança modelar no nível gráfico, no nível do epifenômeno, também deve se fazer sentir no nível dos fenômenos linguísticos. Se, em nível de modelos simbólicos de heróis, analfabetos que ouviam folhetins nas rodas de leitura popularizaram um dos tipos de herói do Romantismo, como não seria no nível de usos linguísticos modelados em variantes mais normais em periódicos impressos? Certamente a atitude discursiva dos redatores de manuscritos imprimia um grau maior de vigília sobre o próprio texto se fosse um manuscrito feito para circular na esfera pública do que se o fosse para uma esfera privada. Da mesma forma, o grau de vigília aumenta no século XIX, mas agora a oposição geral se traduz entre manuscritos e impressos. As tabelas do PHPB traduzem essa tensão historicamente diferenciada nos meios sem desprezar, no âmbito de sua Plataforma, um modo de entrada que permita detalhamentos. Um deles é a proposta de inserir, entre a divisão consagrada *documentação pública/privada*, a documentação produzida fora desses dois escopos, mas igualmente numerosa e variada na história: a administração privada. Por vezes, ainda que com fórmulas e evocações textuais especiais, são espaços de recolha de documentação de variados vernáculos.

Parte-se, portanto, de relações externas amplas, mas com implicações nas atitudes pragmáticas do redator para com seu próprio texto, ficando, no *corpus* construído sem objetivos de estudo ou controles específicos, os rótulos de entrada bem mais reveladores da prática de cada época do que nosso olhar teórico contemporâneo. Essa tarefa é do usuário que recebe material controlado para não enviesar suas generalizações linguísticas por misturar diferentes aproximações a modelos de escrita. Veja-se que, por exemplo, o rótulo *carta* não mistura *carta* no uso sociopragmático de correspondência pessoal com *carta* no sentido neutro de *documento* (em geral mais formulai-cos, portanto, menos vernáculos) ainda vivo hoje em, por exemplo, Carta Magna (a Constituição Brasileira) ou *carta de motorista* (em alguns lugares do Brasil equivalente à *carteira de motorista*). O rótulo *carta pessoal* também não mistura as diferentes sintonias de escritura de cartas de fato pessoais,

particulares, contudo produzidas em diferentes modos de circulação dentro das esferas pública e privada.

Um ponto importante é compreender os traços internos e externos para reunirmos missivas, correspondências pessoais, sob o rótulo *cartas particulares no corpus mínimo comum* ou distingui-las em grupos na Plataforma de corpora. O gênero *carta* como correspondência pessoal, *grosso modo*, envolve estrutura geral simples:

Destinatário e fórmulas de saudação, evocação ou atração de benevolência no início (protocolo); o corpo do texto pouco ou não formulaico onde se desenvolve escrita mais livre de fórmulas, ainda que não isenta, e com temática de interesse no eixo remetente-destinatário. Remetente, fórmulas de despedida e de autenticação no fim (escatocolo), tais como datação, localização (por vezes no Protocolo), assinatura, marcas pessoais e *post scriptum* assinalado. Diversas rotulações dessas cartas têm sido moldadas a partir de seus conteúdos temáticos (cartas políticas, cartas de amor etc.), suas relações pessoais imediatas (cartas familiares, cartas de negócios), suas particularidades sociogeográficas (cartas do sertão) ou de circunstâncias de escrituras (cartas da prisão), enfim, múltiplas possibilidades tão numerosas quanto são numerosos os pontos de vista teórico-metodológicos sobre aspectos de forma, de linguagem ou de suas funções sociopragmáticas. Todos válidos, pois nenhum gênero é natural e absoluto, nenhum gênero *brota do chão*; são sempre frutos dos múltiplos olhares da pesquisa acadêmica. Contudo, são rótulos oriundos de questões e interesses muito específicos, espelhos de estudos específicos. Se fôssemos capazes de reunir conjuntos de cartas pessoais conforme cada um desses olhares, escritas do século XVII ao XX, em de cada região brasileira, por homens e mulheres oriundos de pelo menos três níveis socioculturais, então poderíamos ter uma ótima recuperação da diacronia de vários fenômenos linguísticos. Mas não somos capazes disso. Para o *corpus* geral, valerá, como já dito aqui, o caráter geral *grosso modo* descrito anteriormente. Esse é o critério para o *corpus comum mínimo*. Na Plataforma, que visa catalogar uma produção mais ampla do Projeto, mantemos como caráter categórico o modo de circulação dessa correspondência para tentar reunir sob esse critério as marcas da sintonia de escritura evocadas ao redator em função da esfera por onde circulará seu texto e, por conseguinte, por quem poderá ser lido e onde ele poderá ficar guardado. É o critério geral do *corpus* assim distribuído:

I) **Categorias de entrada**

- a. Séculos: XVII, XVIII, XIX e XX;
- b. Impressos e manuscritos;
- c. Esferas institucional pública, institucional privada e particular;
- d. Textos manuscritos mais formulaicos e textos menos formulaicos (correspondências);
- e. Rótulos de época usados nos periódicos impressos e livros.

Exemplo:

1. **Impressos século XIX**

- 1.1 Cartas de leitores
- 1.2 Cartas de redatores/editoriais
- 1.3 Notícias
- 1.4 Anúncios
- 1.5 Textos literários: folhetins e crônicas

2. **Manuscritos século XVIII**

- 2.1 Manuscritos na esfera privada
 - 2.1.1 Correspondência particular XVIII – **cartas pessoais** em circulação restrita – família, amigos, colegas etc.;
 - 2.1.2 Correspondência institucional privada XVIII – **cartas pessoais** em circulação na esfera administrativa de instituições privadas – casas de comércio, pequenas lojas, ordens religiosas fora da máquina pública, irmandades laicas, hospitais, clubes;
 - 2.1.3 Demais documentação privada – documentação geral da esfera institucional privada – casas de comércio, pequenas lojas, ordens religiosas fora da máquina pública, irmandades laicas, hospitais, clubes.

3. **Manuscritos na esfera pública**

- 3.1 Correspondência oficial XVIII – **cartas pessoais** em circulação na esfera administrativa de instituições públicas.
- 3.2 Demais documentação oficial – documentação geral da esfera pública – notariais, jurídico, administrativos, relatórios etc.

4. Manuscritos na esfera literária

Dessa forma, na Plataforma, as cartas pessoais figuram distribuídas em *circulação restrita*, *circulação administração institucional privada*, *circulação administração institucional pública*. Um caráter categórico de base externa às cartas em si talvez mais objetivamente reunindo em um fio condutor o que se multiplica nos olhares da investigação científica.

Corpus diferencial no PHPB

Vários materiais encontram-se, hoje, no *corpus diferencial* em função de não contarmos com exemplares em todas as equipes, e não por uma questão de sua natureza. Dessa forma, permanecem complementares para efeito comparativo, mas passarão ao *corpus mínimo comum* à medida que sejam feitos cronogramas de ampliação dos *corpora* de cada equipe local do PHPB. É o caso, por exemplo, das *notícias* em jornais. No *corpus diferencial*, *notícias* em jornais do século XX do interior do Pará, de Santarém, e do interior de Minas Gerais, de Ouro Preto cobrem nascimentos locais da mesma forma que em outros periódicos nas capitais um século antes.

- (1) [20,2 NJOR PA] *Nascimentos: || Os sr Felipe correia de Castro e sua | digna esposa d. Paqueta Teixeira de | Castro tiveram a gentileza de nos par- | ticipar o nascimento de sua filhinha Cléa | Damazia. ocorrido no dia 21 do corrente.|| - O sr. Raimundo [...].*
- (2) [20,2 NJOR MG] *NASCIMENTOS || Dia 28 do p. p.,¹ no Rio del Janeiro, nasceu e menino Már-| cio, gracioso primogênito do|Dr. Albino Sartori Júnior e del sua Exma. Esposa.|| O lar do Dr.| Artur Guimarães Drumond e del sua consorte, D. Maria da Trin-|dade Fortes Drumond foi, dial 30 do passado, enriquecido com| o aparecimento de uma graciosa| menina. Aos recém-nascidos, mill votos de felicidades, e parabéns,| que mandamos aos respectivos pais.*

Notícias como essas são um retrato da função pragmático-social dos impressos em comunidades relativamente pequenas no século XX. Ao fazerem correr a notícia de nascimentos ao alcance de sua pequena circulação pelas cercanias rurais em torno do pequeno núcleo urbano, jornais e outros periódicos conferem à família conexão aos círculos socialmente bem

colocados. O *status* de estar na rede social escrita à época era forma de mostrar *status* e presença na rede das chamadas *boas famílias*.

Não que inexista esse tipo de cobertura nas capitais no século XX, mas aparecem bem mais em seções onde esse rótulo *notícia* não mais se confunde com o rótulo *anúncios*, e bem menos em seções que intercalem *anúncios* e *notícias*, por tradições que avancem desde o XIX. “Notícia” e “anúncio” são vocábulos que remetem à mesma ação básica contida em *contar*, *relatar*, *anunciar*, *reportar*, *noticiar*, dentre outros. A nominalização dessas ações básicas presentes desde a circulação de manuscritos como em cartas pessoais ou administrativo-oficiais quando feitas nos impressos periódicos partem de uma fase de equivalência nas estruturas mais simples no início de cada jornal e chegam a se referir a especializações de seções geograficamente delimitadas nas edições. Dessa forma, as nominalizações *notícias* e *anúncios* de escravos fugidos, vapores e mercadorias que chegam ao porto e de eventos na Europa ficam misturadas numa seção única em fases iniciais de *layout* mais simples dos periódicos diários ou semanais (com uma ou duas colunas de texto na página) e, conforme crescem formato e alcance dos jornais via expansão dos Correios, essas mesmas nominalizações vão se especializando. Se, por um lado, *notícias* vão se configurando como uma reportagem de fatos de interesse e impacto mais gerais e assumem lugar claramente setorizado no exemplar dos periódicos, *anúncios* vão se especializando em anúncios sociais e anúncios comerciais, materiais classificados postos, com o tempo, no século XX, numa seção de *classificados*. Esses caminhos evolutivos não são unitemporais e uniformes, mas seguem uma tendência geral nos jornais das maiores cidades já na segunda metade do século XIX. Nesse sentido, a presença de uma *notícia* como reportagem internacional na primeira metade do século XIX não nega a tendência, apenas exemplifica uma das funções ainda amalgamadas numa mesma seção e ainda dependentes das redes manuscritas de circulação como as cartas. Observe-se que, no meio da notícia, são cartas que trazem as notícias para a *notícia* difundida pelo periódico impresso em Pernambuco.

- (3) [19,1 NJORPE] NOTÍCIAS DE PORTUGAL, || MARÇO 29. || BADAJOZ. — *As notícias desta praça | da nossa fronteira são em data de 17 | do corrente: e são de pessoa de | cretido. || As cartas de Madrio do dia 9 davaõ | como certa a invasão de Portugal, e | tanto assim, que haviaõ sahido da Ca- | pital para reforçar a divisaõ do Tejo e | 3. Brigada composta*

*de 3.561 homens, | inclusive Artilheria, chegando com ef- | [f]eito esta
força a Zebol[l]a junto a Ta-|lavera; o que alegrou muito os após- |
tolicos desta praça, que sem pejo fize-| raõ saudes a invasão de Portu-
gal, e | marcarão já o dia 3 de Abril em que | se devia começar. Porém
as cartas | posteriores, do dia 13, dizem que se | passaraõ ordens para
suspender a mar- | cha, [...].*

A esses três materiais com notícias já na página do *corpus diferencial*, somam-se outros da Plataforma de *corpora*, como notícias do Rio de Janeiro, da segunda metade do século XIX em jornais de grande circulação. São materiais contrastivos com o *corpus mínimo comum* por complementariedade. Já os textos escritos por portugueses sempre serão *diferencial*, por natureza, contrastivos devido ao caráter intrínseco para a formação do português brasileiro.

Veja-se o material para o século XVIII com manuscritos pessoais de mercadores portugueses que viviam no Brasil.² Trata-se de conjunto de missivas enviadas do Brasil para Antônio Esteves Costa, um abastado comerciante, dono de uma *casa de comércio* que, de Lisboa, por cartas, controlava seus negócios por comissionários espalhados pela Europa e colônias portuguesas. É material editado em Barbosa (1999) e disponibilizado no *corpus diferencial*. Esse conjunto compõe-se de cartas oriundas de várias cidades ao longo da costa brasileira, distribuídas, no acervo de origem, pelas rubricas *particular e comercial*. No entanto, no trabalho de edição dos originais, percebe-se não haver diferenças, nem de conteúdo, nem de forma, entre as ditas *particulares e comerciais* do século XVIII. Veja-se que, descoberto o fato de esse comerciante espalhar patrícios pelo mundo e, deles, recebia cartas em Lisboa, passamos a contar com 83 cartas de portugueses identificados que viviam, sendo a maior parte, do nordeste brasileiro: 4 cartas escritas no Rio de Janeiro, 13 na Bahia, 31 em Pernambuco, 28 no Maranhão e 7 no Pará. Como afirma Barbosa (1999: 132):

[...] o conjunto de cartas, guardado na caixa de manuscritos número 224 da seção de reservados da Biblioteca Nacional de Lisboa, reúne um precioso acervo da prosa não-literária em Língua Portuguesa Clássica, construído sob condições particularmente favoráveis à formação de um *corpus*, se não homogêneo em relação às circunstâncias de produção, e fidedigno à escrita distensa de seus autores, pelo menos o mais próximo possível dessas qualidades imprescindíveis para o estudo comparativo entre modalidades dentro da modalidade escrita de uma dada língua.

São redatores em grau mediano de habilidade, ou com marcas de inabilidade de escrita alfabética, que deixam transparecer traços grafofonéticos e fatos do dia a dia do mercado nos portos dessas cidades brasileiras. São redatores que estiveram em contato cotidiano com colonos nascidos no Brasil. É uma fonte de dados do português europeu no Brasil em contato não palaciano com o português do Brasil. Diferencial por natureza. Os usuários de *corpora* históricos podem seguir a trilha de materiais aqui traçada no *diferencial* e buscar outros conjuntos de materiais semelhantes fora do PHPB, comuns no século XVIII como peças avulsas, mas também encontrados em raras coleções. É o caso de material de idêntica natureza administrativo-privada de outra casa de comércio, pertencente a outro megacomerciante sediado em Lisboa com portugueses comissionados em todo o mundo e escrito cinquenta anos antes, na primeira metade do século XVIII. São cartas de mercadores portugueses editadas em Lisanti Filho (1973). Em diálogos de materiais *intra* e *extracorpora* do PHPB, portanto, potencializam-se novos recortes diacrônicos, como esse, dentro do século XVIII.

O *corpus diferencial* vale, nesse e noutros casos, como um mapeamento de filões textuais fora do eixo propositalmente mais restrito do *corpus mínimo comum*, que restringe o leque de materiais a bem do cumprimento do avanço coordenado da construção de *corpora* simultaneamente feito por equipes em diferentes regiões do país. Foi uma decisão metodológica que muito provavelmente será enfrentada por outros projetos nacionais que pretendam comparar dados de diferentes regiões: maior diversidade de gêneros textuais ou maior cobertura na dimensão espacial de um número mais restrito de gêneros. O ideal de unir as duas coisas se faz com gerações de pesquisadores. Para, em cinco anos, se equipararem, em termos de quantidade e qualidade de documentação editada, equipes fundadoras e equipes menos antigas do Projeto que completem o quadro diatópico nacional, é preciso optar. Sempre haverá equipes, fundadoras ou novas, em estados mais avançados de quantidade e variedade de materiais. É uma realidade que depende do número e da experiência (prévia ou inaugural) dos membros locais das equipes para o trabalho em acervos públicos e privados de documentação histórica. É uma proposta tentadora programar a ampliação diatópica do *corpus mínimo comum* para o conjunto de equipes conforme o elenco de Tradições Discursivas (TDs) já editadas pelas equipes mais antigas e/ou pelas equipes com mais pós-graduandos editando

materiais para seus trabalhos de conclusão de curso. Mas é uma tentação que se opõe ao trabalho conjunto que faça o *corpus mínimo comum* seguir crescendo organicamente na cobertura do espaço. Se a escolha de anúncios, editoriais, cartas de leitores, manuscritos oficiais e pessoais já deixou lacunas criadas pelas realidades locais, que se impuseram em cada etapa coordenada de ampliação, a opção de propor ao coletivo nacional o estado mais avançado pontual deixaria o *corpus* fadado ao descumprimento e à desproporção irreversível.

O *corpus diferencial*, da mesma forma como no caso dos materiais contrastivos do português europeu potencializados com as cartas de mercadores portugueses do século XVIII, assume o caráter de planejamento de expansões futuras e recortes complementares imediatos a serem feitos com essas parcelas postas na página prévia de *corpora* do PHPB. Essa instrumentalização da configuração atual, em gêneros textuais, do *corpus diferencial* torna a aparente colcha de retalhos de materiais numa sistematização da riqueza documental aberta a novos recortes comparativos a serem feitos pelos usuários. Aos construtores de *corpora* cabe dispor a otimização comparativa nessa separação metodológica de uma parcela da Plataforma de *corpora* em duas realidades, o avanço conjunto do *corpus mínimo comum* e o avanço mapeado da riqueza de materiais e de relações *intra* e *extra*-Projeto no *corpus diferencial*. Neste ponto da história do PHPB, a riqueza do *corpus diferencial* conta com cartas de portugueses identificados, textos teatrais, testamentos, processos-crime, atas de câmaras, memórias, relatos históricos, diários, históricos de viagem, inventários, folhetim, crônica, notícia, demais materiais já editados no PHPB, assim como vínculos eletrônicos para páginas do Projeto Nurc que disponibilizem gravações e transcrições de entrevistas das décadas de 1970 e 1980. Tudo isso pronto para ser baixado.

O PHPB não tem um *corpus* fechado não apenas porque as equipes continuarão editando materiais, mas também porque nele se tem as combinações em aberto à mercê das demandas dos usuários. Se não interessarem contrastes diacrônicos, tem-se a oportunidade de juntar todos os textos da Plataforma de *corpora* (recebendo via internet), ou baixar todos os materiais do *corpus mínimo comparativo* e do *corpus diferencial* para um trabalho em pancronia com uma massa invejável de dados. São decisões *ad hoc* que dependem do objeto, questão e objetivo de cada estudo.

TRADIÇÕES DISCURSIVAS E RÓTULOS FUNCIONAIS PRETÉRITOS EM CORPORA GERAIS

Há uma relação de pertinência e não de equivalência entre Tradições Discursivas (doravante TDs) e gêneros textuais: todo gênero textual é uma tradição discursiva, mas nem toda tradição discursiva é um gênero textual. O conceito de Tradição Discursiva prevê que toda habitualização cria uma expectativa de continuidade de modo que o elemento linguístico habitualizado torna-se significável em si. A quebra da expectativa cria estranhamento. A habitualização das marcas que criam nossa percepção que um dado texto se enquadra em um dado gênero textual é a mesma que torna uma fórmula de despedida em cartas pessoais uma tradição discursiva. Portanto, Tradições Discursivas não teoriza sobre gêneros textuais, mas é um conceito para explicar, entre outras questões da linguagem humana, porque em textos reunidos ao longo do tempo podem fornecer dados tanto da diacronia da linguagem usada pela sociedade em cada ponto do tempo coberto pelo conjunto reunido em um *corpus*, como também da linguagem conservada e mudada em diferente compasso nas TDs também ali reunidas. A história da linguagem dos textos não é necessariamente a história da linguagem da sociedade. Isso é demonstrado largamente na bibliografia sobre TDs. Aplicar o conceito de TD aos *corpora* discerne esses dois níveis por exclusão: quando se identifica e se descreve a TD, mapeia-se no mesmo material o que seria, nos textos, reflexo da linguagem dentro de alguma norma da época, fosse vernácula, fosse culta.

Como as TDs criam abstrações das habitualizações, até mesmo as questões grafológicas entre impressos e manuscritos, abordadas no capítulo “O *corpus* do projeto Para a História do Português Brasileiro: a constituição de *corpora* históricos baseada em critérios de tradições discursivas” deste volume, seriam abarcadas nas TDs. Contudo, como os gêneros textuais sempre serão TDs, do ponto de vista metodológico da construção de *corpora* gerais – montados para usuários explorarem toda sorte de fenômenos e não apenas aqueles ligados a TD em si – o uso de um desses rótulos para criar variáveis independentes com TDs/gêneros textuais no controle de uma variável dependente em pesquisa quantitativa não faz diferença epistemológica. De fato, de acordo com o objeto da pesquisa, por vezes a construção de um grupo de fatores de uma variável independente vale-se mais de uma boa teoria de gêneros. Sem se saber quais são os elementos habitualizados em cartas de leitores, os editoriais ou em cartas pessoais por um tempo razoável, a forma de discernir

fatores de controle desses rótulos do ponto no tempo coberto pela amostra será mais bem fundada por recortes teóricos genéticos. No âmbito de um *corpus* geral, não direcionado em sua construção para esta ou aquela questão sobre os estudos da própria tipologia da amostra de escrita pretérita, os rótulos *cartas pessoais*, *carta de leitores* ou *carta de editor/ editoriais* são rótulos que reproduzem a nomeação dada aos materiais por seus contemporâneos. São rótulos que, para eles, significavam algum papel funcional sociopragmático para lidar com os textos. Os gêneros ali observados sempre dependerão do olhar teórico de gênero que se lhes aplique; a Tradição Discursiva descrita sempre dependerá de um recorte temporal mais amplo para detectar um elemento habitualizado. Com dez exemplares de um jornal não é muito seguro afirmar que uma marca linguística ali constante tenha se tornado significável em si de modo a ser, para os homens *daquela época*, uma Tradição Discursiva. Um suposto mestrando que venha a usar o *corpus* geral do PHPB poderá operacionalizar em seu grupo de fatores de controle baseado na tipologia textual para ver o quanto as variantes sob estudo são favorecidas por aquelas tipos de material sociopragmáticamente vistos como editoriais, cartas de leitores ou cartas pessoais, mas entender as tradições discursivas ou gêneros ou cogêneros ali presentes é um objeto de estudo talvez fora de seus objetivos.

Diante do nosso imenso desconhecimento de quais eram as relações pragmático-discursivas, por exemplo, com os jornais em cada região do Brasil no XIX, fica difícil saber o que, nos valores dados aos rótulos textuais em cada jornal, já era um hibridismo trazido dos jornais ingleses, franceses, portugueses (e desses podem vir, quem sabe, estabilidades genéticas desde seus jornais manuscritos) e o que é hibridismo que vai, na diacronia dentro da própria história dos jornais no Brasil, estabelecendo transmutações ao lado de estabilizações de formas e práticas discursivas que hoje entendemos por/percebemos como/temos expectativa de serem artigos, editoriais etc.

Caso optássemos por tentar transformar em categoria de entrada no nosso *corpus* comum os gêneros distintos montados pela observação de cada diferença mais saliente (seja ela estrutural, comunicativa, de conteúdo, estilo, e tudo mais que nossa inteligência alcançar), nossas tabelas seriam *intermináveis* e nosso trabalho *impraticável*. Mais que isso, perderíamos o registro do que pode haver, a cada época, em cada lugar, às vezes em cada jornal, os significados de cada rótulo apenso aos gêneros ali encaixados. Exemplo emblemático é o trecho do jornal oitocentista *Diário do Rio de Janeiro* na edição de 12 de fevereiro de 1841:

A palavra folhetim, adoptada pelo jornal | do commercio para dar ideia dos artigos de | recreio que os francezes chamao feuilleton, | está geralmente recebida: nós, para não | contrariar o uso, substituímos o nosso appen- | dice pelo folhetim. || Publicamos hoje algumas fabulas e uma ode, composição d'um nosso compatriota | o Sr. Doutor J. J. T.: o publico apreciará | seu merecimento [...]. (apud Lima, 2010: 72)

Em seguida a esse trecho do *Diário do Rio de Janeiro*, Lima (2010: 73) comenta que “a palavra *Folhetim*, através da repetição, tornou-se essencial para o sucesso do texto a ela submetida. Da repetição, surgiu a ritualização, caracterizada pela localização no jornal (rodapé) e a sua presença, o que legitimou a sua identidade”.

Um jornal na primeira metade dos oitocentos dizendo que passaria a usar *folhetim* a espelho do que fazia o *Jornal do Commercio*, deixando o *appendice* para reunir *fábulas* e *ode*. Rótulo ou TD em transformação? Foi por isso que, na Plataforma de *corpora* do PHPB preferimos sempre indicar o rótulo usado pelo meio de circulação dos textos ao lado de uma visão de gênero do organizador dos *corpora*. Por exemplo, o usuário lê o rótulo *folhetim*, que é o que está no jornal levantado, e ao lado lê, entre parêntesis, ou *romance em folhetim*, ou *romance-folhetim*, ou *crônica*, enfim, que é o olhar contemporâneo que vê nos textos sob aquele rótulo características bem distintas. A Plataforma de *corpora* é a organização **do todo** produzido no âmbito do PHPB sob uma estrutura organizacional baseada em aspectos mais externos exatamente para fugir um pouco da intensa profusão de olhares teóricos possíveis sem deixar de lado a recuperabilidade dessa profusão a partir das categorias nominais de entrada baseadas em práticas e marcas de época. Assim, mesmo com coisas tão distintas, se os editores de um jornal usam um mesmo rótulo funcional, deve ser porque aproveite um traço em comum que, por mais tênue que seja, deve ser, além de mais prático para eles, favorecedor de algum tipo de forma de linguagem. Nem que seja o que vem desde o sentido primeiro de folhetim: “textos de entretenimento localizados no rodapé dos jornais” (o *feuilleton* dos jornais franceses) – um sentido ligado apenas ao *layout* (ou, como gostamos de precisar, a *geografia da página*). Sentido que não se apaga, nem deixa de conviver com os outros pelo menos até 1880. É o que comenta Lima (2014: 99):

Nós verificamos que não havia linearidade evolutiva nas acepções de folhetim, no Brasil (Lima, 2010, p. 69). No mesmo periódico que encontrávamos romances no folhetim, também se publicavam romances-

folhetim e folhetim histórico ou de costume, que, aos nossos olhos contemporâneos são crônicas. No periódico *Gazeta de Notícias*, era comum a publicação de uma crônica folhetinesca na primeira página e a publicação de um romance-folhetim na segunda página, traduzida e já consagrada pelo público europeu.

Essa prática social de atribuir rótulos operacionais parece um pouco com os rótulos que locadoras e emissoras de televisão usam para classificar filmes em contextos de leitura rápida dos usuários: todos sabemos que *drama*, *aventura*, *ação*, dentre outros, não dão, nem de longe, conta dos entrecruzamentos de gêneros dentro de cada rótulo desses, mas sabemos que são, hoje, rótulos pragmáticos que operam em eficiência suficiente para nossas primeiras escolhas. Quer dizer, são formas vigentes de práticas discursivo-sociais hoje – se não perdemos a rotulação operacional do passado, talvez vejamos um pouco das passadas.

É bem sabido que a Plataforma não é um *corpus* e sim uma *base para corpora*. Assim, o formato de entrada nominal de células (que passarão a grupo de controle para os futuros usuários) deve ser calcado nos traços comuns que nos façam perder menos informações, já que qualquer opção implica perdas. Sempre viveremos uma “escolha de Sofia”. Não pensamos em manter o rótulo de época como entrada nominal absoluta, mesmo que fosse mais prático para quem quisesse ver transmutação, hibridismo e conservações a partir dentro de cada rótulo (que é o que geralmente se faz com os rótulos que encontramos para os manuscritos coloniais nas fichas catalográficas: chamamos de carta, ofício, requerimento ou representação alguns tipos de correspondências da máquina colonial porque assim estão catalogados, ainda que percebamos igualdades ou diferenças outras). Se fosse assim, se por um lado veríamos o que vai sendo incluído e o que vai saindo dentro rótulo *anúncio* ao longo do XIX (desde o rótulo inicial *anúncios variados*), por outro, perderíamos seus pontos de contato com *avisos*, *notícias*, *comunicado* etc. quando, a partir de um traço comum preestabelecido para o levantamento de textos, tomamos qualquer um desses rótulos como o recorte de anúncio que queremos estudar numa linha do tempo até o que temos hoje, ponto de vista inevitável de quem já vive as tradições textuais que nos alcançaram. Por isso, o traço “comercial” é um fio condutor que acaba se impondo, ainda que haja outros, para vermos as estabilidades e transmutações dentro desse recorte que denominamos “anúncio” como célula de entrada para gênero no *corpus* mínimo.

É claro que, na “escolha de Sofia” que se nos é imposta pela realidade, perdemos vários tipos de anúncios que não tenham nenhum fundo comercial, de negócios, de economia (por isso valem as aproximações por conjectura que temos feito em anúncios de busca de escravos, ou até em avisos/notícias de partidas de navios e inaugurações de hotéis), mas na verdade, para não perdermos teríamos que ter por objetivo bienal montar *corpora* só de anúncios/avisos/notícias em vários jornais de vários lugares (nesse caso até caberia usar o rótulo como guia absoluto). Esse é um ponto ótimo para todo um projeto de estudo mais aprofundado que deverá ser feito um dia para este e outros grupos textuais impressos e manuscritos hoje mencionados em nossos *corpora*. Serão *corpora* que avançarão confirmando e revendo o que hoje dissermos a partir do que é possível cobrir. Não podemos nos cobrar uma coerência absoluta a cada olhar teórico que a partir de diferenças/igualdades vislumbrem gêneros diferentes dentro de nossos recortes porque assim também o seria em levantamentos exaustivos só de *anúncios* e *notícias*. O importante é manter a coerência ao(s) traço(s) comuns escolhidos para guiar nossos levantamentos e usar o bom senso para adaptações. Assim, se tenho para uma época bem inicial da imprensa no Brasil – geralmente sem muita fartura de alguns gêneros – ou uma época de difícil circulação de informações, um jornal usando uma carta pessoal para noticiar, em sua seção de notícias, esse texto entra sem problema na célula nominal *notícias* na primeira metade do século. Como a opção operacional foi a divisão por metades dos séculos, temos de ter cuidado para não inflacionarmos de materiais das décadas de 1840 e 1850 e deixarmos de cobrir as décadas iniciais.

Pelo que dissemos antes, mesma coisa para *folhetim*: se o traço fosse (porque folhetim não está no *corpus* comum) algo que traduzisse o que fica no eixo de discussão romance-novela-conto da prosa ficcional (estruturada com narrador, personagens, diálogos, vários/poucos núcleos temáticos...), então as crônicas e os artigos que aparecem com frequência sob o rótulo *folhetim* não entrariam na célula nominal *folhetim* em nossas tabelas, e sim em *crônicas* e *artigos*. Em relação ao exemplo dado na Plataforma, podemos inverter e colocar, ao invés de rótulo e gênero entre parêntesis FOLHETIM (romance em folhetim), FOLHETIM (romance-folhetim) e FOLHETIM (crônica), podemos colocar gênero e rótulo, ou destacar o rótulo de época na ficha que encabeça cada texto em nossos *corpora*: o importante é não perder a informação.

Mesma coisa para editorial: se definirmos como traço para ser um editorial como gênero de entrada em nosso *corpus* ser “a voz, a posição do jornal”, não importando se dos próprios editores ou de vozes assumidas de outrem, então, inevitavelmente misturamos editoriais (geralmente um lugar regular na geografia dos exemplares, com ou sem o rótulo expresso) com cartas de editores (a expressão direta do que pensam os editores), que parece ter sido a opção na proposta inicial. Se nosso traço de controle passasse a fazer a distinção, então ou teríamos duas entradas nominais ou, por motivos práticos, passaríamos cartas de editores para o diferencial. Enfim, a solução dada é mais operacional para o *corpus* comum e não impede que quem se dedicar mais de perto à diferença entre editoriais e cartas de editores possa constituir um *corpus* diferencial paralelo para testar as hipóteses genéticas. Construir *corpus* geral para projetos coletivos optando por incluir, desde o início, com um *corpus* diferencial é tão importante quanto os *corpora* comuns, pois são a possibilidade de refino dialético de afirmações que se façam com o trato contrastivo: esmiuçar várias possibilidades de recortes e eixos de diferenças funcionais estruturais, temáticas ou de estilo; marcas salientes nos materiais. Estamos trabalhando do ponto de vista de quem tenta construir *corpora* coerentes e reveladores de realidades em sincronias passadas dentro dos limites de *corpora* comuns mínimos. O esmiuçar de outros recortes fica para o ponto de vista dos futuros usuários, que trabalharão contrastando nossos recortes com outros *corpora* a serem construídos dentro ou fora do PHPB. Para tal, é fundamental haver tanto coerência no recorte de traços comuns claramente estabelecidos para a escolha dos gêneros de entrada reunidos sob os rótulos funcionais das tabelas no *corpus*, quanto garantir a recuperabilidade de traços singularizadores dos materiais do *corpus* mínimo, informações dispostas na ficha-cabeçalho de cada documento.

O TRABALHO CONTÍNUO DE AMPLIAÇÃO DE CORPORA HISTÓRICOS: A EXPERIÊNCIA NO PHPB

Ampliar organicamente a cobertura diatópica de um *corpus* é a base da representatividade dos dados dele levantados. Não há comparabilidade garantida quando os esforços desde a definição do que se busca nos arquivos históricos até o caráter das edições da documentação não seguem coordenados. Por ocasião da ampliação do *corpus mínimo comum* (tarefa coletiva empreendia por equipes locais tais como Rio Grande do Norte, Paraíba, Pernambuco, Ceará, Bahia, Alagoas, Rio de Janeiro, São Paulo, Minas Gerais, Paraná, Santa Catarina), seria necessário estabelecer prioridades, prazos e homogeneizar formulários. Contudo, a primeira etapa de avanço precisou ser um verdadeiro censo de materiais organizados desde 2007 sob a chancela do PHPB Nacional. Para avançar, para estabelecer prioridades comuns a todos, era preciso saber de todos os materiais prontos, mas não partilhados nacionalmente. A Plataforma de corpora estava defasada em 2010. Seguindo os Formulários 1 e 2, as equipes puderam informar, de modo padronizado, os resultados dos esforços de produção locais.

I) Formulário 1: guia de perguntas/orientações para *censo de corpora* (para coordenadores regionais)

CENSO DE EDIÇÕES DE FONTES PRIMÁRIAS PARA O *CORPUS* GERAL AMPLIAÇÃO DO *CORPUS* MÍNIMO 1ª FASE (2010)

- a) Existem materiais editados que ainda não estão disponíveis (por força de ineditismo de tese, por exemplo), mas poderiam ser noticiados, de forma genérica, sem precisar arquivo ou localização, no *corpus* geral do PHPB? Como cada editor quer citá-las? Por favor, liste-os a seguir e ponha o e-mail do editor para ficar disponível ao contato de prováveis interessados.
- b) Existem materiais publicados em livros, mas que, em função de contrato, não poderão ser disponibilizados eletronicamente? Como devemos citá-los no *corpus* geral do PHPB? Como poderão ser adquiridos? Há e-mail para pedidos?
- c) Existem materiais publicados em livros cujos arquivos originais em doc ou pdf poderão ser disponibilizados eletronicamente no *corpus* geral do PHPB? Como deverão ser citadas suas publicações em papel?
- d) Existem materiais publicados eletronicamente que já podem ser distribuídos ou disponibilizados no *corpus* geral/*corpus* compartilhado/*corpus* do VIII PHPB? Qual o título do CD ou o endereço eletrônico para distribuição? Se já estiverem em uma página na internet, há preferência de apenas incluir um *link* na página do PHPB?

COORDENADOR REGIONAL: se possível, preencha as informações a seguir para agilizar a organização do material para que todos recebam um MAPA GERAL do que me for enviado. Isso ajudará no trânsito de textos para nossas pesquisas.

Equipe Regional PHPB:

Coordenador:

Editor do material:

E-mail do editor:

Telefone do editor:

Título da edição:

Descrição do material editado (manuscrito autógrafo, 1ª edição impressa, publicado em jornal de grande circulação, gênero textual, modo de circulação, especificidades sócio-históricas etc.):

Como citar o material em trabalhos acadêmicos?

- e) Só haverá ampliação do *corpus* diferencial SE E SOMENTE SE terminado o *corpus* mínimo da 1ª fase e DO QUE FOR POSSÍVEL de qualquer um dos seguintes materiais (*corpus* diferencial):

II) Formulário 2: tabela a ser preenchida mapeando os materiais já editados para o *corpus* diferencial

Equipe regional:						
CORPUS DIFERENCIAL MANUSCRITO – PHPB	Séc. XVIII 1701-1750	Séc. XVIII 1751-1800	Séc. XIX 1801-1850	Séc. XIX 1851-1900	Séc. XX 1901-1950	Séc. XX 1951-2000
Cartas de portugueses identificados						
Textos teatrais						
Língua falada: Nurc						
Testamentos						
Processos-crime						
Atas de câmaras						
Memórias/relatos históricos/diários históricos de viagens						
Inventários						
Folhetins						
Crônicas						
Demais materiais prontos						
Novos gêneros						
COORDENADOR REGIONAL: uma vez pronta a primeira fase, garantindo células comparativas por tipo de material e por três séculos, passamos à segunda etapa de ampliação. É a fase da macrocategoria administração privada em todas as equipes, nos três séculos.						

Em nome da maior comparabilidade promovida por *corpora* gerais, a ampliação sempre deverá ser pensada em etapas. No caso do PHPB, ainda que não cumprida por todos, o que se conseguiu reunir manteve a garantia de comparabilidade futura porque houve etapas coletivas coordenadas. Mesmo no cumprimento do censo de *corpora* para a Plataforma, já se embutia a reorganização do *corpus mínimo comum*, estabelecendo metas

comuns na busca das equipes nos arquivos em suas localidades. É o caso da orientação priorizar o mapeamento de materiais antes de se envidar esforços locais na direção de ampliar do *corpus mínimo*: “Uma vez pronta a primeira fase, garantindo células comparativas por tipo de material e por três séculos, passamos à segunda etapa de ampliação. É a fase da macrocategoria administração privada em todas as equipes, nos três séculos.” Mesma coisa em relação ao *corpus* diferencial: “Só haverá ampliação do *corpus diferencial* SE E SOMENTE SE terminado o *corpus mínimo* da 1ª fase e DO QUE FOR POSSÍVEL de qualquer um dos seguintes materiais (*corpus diferencial*)”. Essa etapa preliminar à ampliação, o *Censo de edições de fontes primárias para o corpus geral do PHPB*, será etapa de outros projetos com semelhantes ambições. Observar a formatação e condução aqui promovidas pode colaborar como instrumentalização de execução. Veja-se que, somente em relação às fases de ampliação do *corpus mínimo comum*, é preciso observar um detalhe de estratégia: propor a procura de novos tipos de manuscritos somente após o esforço de tentar equalizar o quadro de impressos entre as equipes locais e aumentar a *latitude* comparativa do *corpus* para os usuários. Isso foi mais importante do que partir para a *administração privada*, deixada para uma segunda fase. Começar por impressos foi o caminho natural na fase de construção dos *corpora* do PHPB empreendido pelas equipes originais. Assinalamos o mesmo a projetos que se valham de nossa experiência. E por quê? Porque, para equipes novas, sempre é melhor começar a decodificar textos para transcrição conservadora que não exijam treinamento em paleografia de leitura. Já são dificuldades suficientes para iniciantes enfrentar questões de crítica textual de impressos oitocentistas; enfrentar, de pronto, os manuscritos da administração privada, ou quaisquer outros, atrapalharia o avanço conjunto das equipes. Em geral, o avanço de materiais manuscritos, especialmente do século XVIII para trás, depende do número de trabalhos de conclusão de curso em nível de pós-graduação que envolvam construção de *corpora*. O número de pessoas com semelhante perfil varia de equipe para equipe, de modo a não ser equiparável a capacidade de cada local em cumprir os prazos previstos pelo trabalho coordenado.

O preenchimento periódico de quadros e a resposta a perguntas como essas servem, também, para fazer o *corpus mínimo comum* avançar. A segunda fase, se executada de acordo com nossa experiência, aumentará o quilate de seus materiais. A chamada *administração privada* é espaço privilegiado de

diversificação pragmática dos contextos de escritura. Delineada em Barbosa (1999) em relação ao século XVIII, chama a atenção para o fato de o modo de circulação dos documentos abrir um terceiro filão de produção escrita que desfaz o tradicional binômio entre a documentação oficial e particular: entende que a documentação particular, em oposição à pública, na verdade, circulava numa esfera *privada* de duas maneiras: ou dentro de uma administração privada, ou dentro círculos realmente particulares. Assim, diários e cartas pessoais escritos por um dado indivíduo viviam sob diferente sintonia de escritura (grau de vigília sobre a normatização linguística e atendimento a marcas e ritos genético-textuais) em relação ao mesmo indivíduo escrevendo para ter seu texto circulando na esfera administrativa, porém não na pública, e sim na esfera administrativa privada de instituições como ordens religiosas, casas de comércio, hospitais, irmandades laicas, lojas, enfim, uma terceira situação de atendimento a modelos e padrões por vezes mais abertos a usos de uma escrita cotidiana, por vezes mais próximos de usos especiais diferentes dos da documentação oficial (executiva, legislativa, judiciária e cartorial). A produção no âmbito do PHPB acolheu ótimos exemplos desse filão, como as edições de Klebson Oliveira (2006) de escritos produzidos por uma irmandade de negros na Bahia oitocentista, que em muito ampliaram nosso olhar para a complexidade da sociedade de então. A segunda fase de ampliação, no caso do PHPB, não promoveu avanço orgânico entre as equipes em relação à administração privada por conta da priorização de se completar o avanço orgânico nos impressos, deixando-a como já estava, um ponto qualitativo do *corpus diferencial*. No caso de projetos futuros, pode figurar na programação de partida das equipes coordenadas para a cobertura de manuscritos. O que no Formulário 3 do PHPB consta como 2ª fase seria tabela organizacional de início dos trabalhos coordenados. A fase terceira de ampliação, Formulário 4 no PHPB, será sempre uma fase em aberto na medida em que é reaplicada periodicamente na evolução de materiais que passam do *corpus diferencial* para o *corpus comum* entre todas as equipes. Como dito, o Formulário 4 “organiza o avanço de todas as equipes por sobre os materiais até então disponíveis no *corpus diferencial*”.

III) Formulário 3: ampliação do *corpus* mínimo 2ª fase – cartas da administração privada (até junho de 2011)

CORPUS MÍNIMO COMUM: MANUSCRITOS						
CORPUS MÍNIMO MANUSCRITO – PHPB	Séc. XVIII 1701-1750	Séc. XVIII 1751-1800	Séc. XIX 1801-1850	Séc. XIX 1851-1900	Séc. XX 1901-1950	Séc. XX 1951-2000
Cartas particulares	Pronto	Pronto	Pronto	Pronto	Pronto	Pronto
Cartas oficiais	Pronto	Pronto	Pronto	Pronto	Pronto	Pronto
Cartas da administração privada	2ª fase	2ª fase	2ª fase	2ª fase	2ª fase	2ª fase
Equipe regional:						
CORPUS MÍNIMO COMUM: IMPRESSOS						
CORPUS MÍNIMO IMPRESSO – PHPB	Séc. XVIII 1701-1750	Séc. XVIII 1751-1800	Séc. XIX 1801-1850	Séc. XIX 1851-1900	Séc. XX 1901-1950	Séc. XX 1951-2000
Cartas de leitores	Pronto	Pronto	Pronto	Pronto	Pronto	Pronto
Cartas de redatores/ editoriais	Pronto	Pronto	Pronto	Pronto	Pronto	Pronto
Anúncios	Pronto	Pronto	Pronto	Pronto	Pronto	Pronto
Cartas da administração privada	Pronto	Pronto	Pronto	Pronto	Pronto	Pronto
Equipe regional:						
COORDENADOR REGIONAL: uma terceira fase de complementação cobre a inevitável falta de paralelo secular entre impressos e organiza o avanço de todas as equipes por sobre os materiais até então disponíveis no <i>corpus</i> diferencial. Essa fase não tem final ainda definido em relação ao <i>corpus</i> mínimo. Como não atingiremos o ideal de todos os tipos de materiais estarem cobertos por todas as equipes e nos três séculos, será necessário, na terceira fase, discutir prioridades no levantamento coletivo.						

IV) Formulário 4: ampliação do *corpus* mínimo 3ª fase – (a partir de julho de 2011)

COMPLEMENTAÇÃO CORPUS MÍNIMO COMUM – PHPB	Séc. XVIII 1701-1750	Séc. XVIII 1751-1800	Séc. XIX 1801-1850	Séc. XIX 1851-1900	Séc. XX 1901-1950	Séc. XX 1951-2000
Textos em prosa não literária em livros impressos no século XVIII	Levantar	Levantar	Levantar	Levantar	Levantar	Levantar
Testamentos Processos-crime Atas de câmara	Completar	Completar	Completar	Completar	Completar	Completar
Prioridades definidas de novos levantamentos em todas as equipes	Levantar	Levantar	Levantar	Levantar	Levantar	Levantar
Equipe regional:						

O Formulário 5 e o Formulário 6 são os fundamentais para tudo já referido no plano organizacional com vistas aos cruzamentos de informações que proporcionarão aos usuários reorganizar o *corpus mínimo comum*, *diferencial* ou materiais da Plataforma em *corpora* específicos para seus objetivos específicos. Seja na reorganização externa da nossa linha neutra de entrada da datação do documento em metades dos séculos numa linha que siga a data de nascimento dos autores, seja por reunião dos gêneros por critérios que sirvam a questões de TDs sob foco, esses cruzamentos serão mais bem potencializados mediante o preenchimento mais bem detalhado. Quem conhece as informações circunstanciais de cada material é quem o edita. O cabeçalho deve ter sua forma padronizada para agilizar aos usuários a recuperabilidade e sistematização de informações. Ele é um bloco de “fontes secundárias” em relação à fonte primária ali disponibilizada. Quando for possível contar com aplicativos de concordância nas páginas do Projeto, essa função cairá de importância. Mas enquanto esse tempo não chega, a sistematização será mais bem executada com cabeçalhos apropriadamente preenchidos. É, portanto, fundamental chamar a atenção dos membros encarregados de construir os *corpora* para que

já registrem as informações durante o lidar inicial com a documentação. Isso porque, passada a fase de transcrição/edição dos materiais (quando cada um aprofunda a pesquisa sobre as condições externas de produção e a história social da época e das pessoas que aparecem nos textos), muito dificilmente se voltará a isso. Quem coordena a construção coletiva de equipes distanciadas no espaço não terá como voltar aos arquivos com todos os membros para recuperar informações que enriquecem os itens 7 e 9 do Formulário 6. De fato, é atividade de registro a ser moldada (Formulário 5), exemplificada (Formulário 6) e orientada durante a fase de partida para os acervos. Idealmente falando, seria fase propedêutica à ida aos arquivos. Nesse sentido, o que se escolhe como exemplificação, que no caso da ampliação do PHPB foi o sexto Formulário, terá caráter de exemplo do amplamente discutido via correio eletrônico. Entre o ideal e o real obtido, seja na experiência do PHPB, seja em novas experiências, fica o avanço qualitativo obtido muito maior do que sem semelhante cuidado.

V) Formulário 5: guia de formatação das informações sobre os documentos

MODELO DE CABEÇALHO DE IDENTIFICAÇÃO DE CADA DOCUMENTO
DENTRO DO QUADRO: VERDANA 12; CABEÇALHO EM FONTE VERDANA 10 E SEM
ESPAÇO ENTRELINHAS

Projeto Nacional PHPB – Equipe Regional XXX
Século XXX – Tipo de Manuscrito ou Impresso/Rótulo no *Corpus*
Edição: SOBRENOME, Nome

1. Modalidade: língua escrita
2. Tipo de texto:
3. Assunto:
4. Data do documento:
5. Local de origem do documento:
6. Local de depósito do documento:
7. Identificação do autor:
8. Número de palavras:
9. Informações levantadas:
10. Editor do documento:

TEXTO EDITADO: MARGENS 3CM, COM FONTE “VERDANA 9”, entrelinhas 1,5

VI) Formulário 6: forma de preenchimento e exemplos de cabeçalhos presentes no início das edições

Projeto Nacional PHPB – Equipe Regional Rio de Janeiro

Século XIX – Jornal Impresso/Notícia

Edição: LIMA, Alexandre Xavier

1. Modalidade: língua escrita.
2. Tipo de texto: (colocar o rótulo de época e, se for o caso, o gênero textual entre parêntesis a seguir, segundo a visão teórica do editor).
Só colocar o rótulo de época se constar do manuscrito – geralmente repetido na ficha catalográfica dos arquivos – ou se constar dos títulos e das seções internas dos impressos periódicos ou não periódicos. Por exemplo: Editorial (artigo-editorial) do jornal *GAZETA DE NOTÍCIAS* – Ano III – nº 225. Outro exemplo: Carta pessoal (Carta familiar). Caso o gênero coincida com o rótulo de época, não precisa repeti-lo entre parêntesis. Por exemplo: Carta de editor. Outro exemplo: Notícia.
3. Assunto: (um resumo do conteúdo do texto e, se for o caso, as considerações sobre aspectos temáticos – como seu grau de fixidez, por exemplo: Medidas sanitárias de combate à cólera).
4. Data do documento: (Colocar data de redação e, se for o caso, de publicação do documento sob forma de dia – numérico; mês – por extenso; e ano – numérico. Por exemplo: 7 de janeiro de 1862 ou 7 de janeiro de 1862/publicado em 15 de fevereiro de 1862.)
5. Local de origem do documento: (do geral para o específico. Por exemplo: Brasil – Bahia – Feira de Santana).
6. Local de depósito do documento: (identificação do arquivo público/privado. Por exemplo: Fundação Biblioteca Nacional do Rio de Janeiro – Seção de Periódicos – PRc SPR 61 (5) / C2, 4, 15; ou Biblioteca Nacional de Lisboa – Seção de Manuscritos – caixa mss. 224/133).
7. Identificação do autor: (iniciar com o nome completo – se houver, pois em anúncios, por exemplo, muitas vezes não há autoria indicada – e depois todas as informações coletadas sobre o autor do texto, tais como nome, naturalidade – se estrangeiro português ou de outra origem, se brasileiro nato suas possíveis ascendências: branca europeia portuguesa ou outra, africana, indígena. Informações sobre a região em que nasceu/vivia; sobre filiação, profissão, posição social, formação escolar etc. Por exemplo: Antonio Polycarpo Cabral/Brasileiro natural da Bahia/Nascimento: 1789 etc.).
8. Número de palavras: (total de palavras, contadas no Word, por exemplo: 469).
9. Informações levantadas: (considerações, se for o caso, sobre detalhamentos possíveis, por exemplo, de aspectos significativos do *layout*, como, por exemplo, “Os textos na seção rotulada NOTÍCIAS, neste jornal, encontram-se sempre na 2ª página”. As considerações podem ser de outras naturezas, tais como sobre o grau de planejamento textual, a simetria entre emissor/receptor, o registro em que o texto está escrito, ou quaisquer outras informações sobre o lugar do texto na história social da escrita, como, por exemplo, se modelo de norma culta, se exemplar de normas vernáculas etc.).
10. Editor do documento: (nome do membro do PHPB que assina a edição e, se possível, já sob forma de entrada bibliográfica. Por exemplo: CYRINO, Sonia M. Lazzarini; BARRICHELO, Jerusa; PAULA, Flávia Figueiredo de. Cartas não-oficiais – Curitiba, Paraná. Curitiba: Projeto PHPB/PR, 2004, CD-ROM. Cartas familiares ao Sr. José Lourenço. Carta 1.)

O Globo, interessante diário que se publica nesta | côrte, merece a maior atenção do Correio Paulistano, | ao extractar as noticias de cada mala: sobre todas as | questões e redação se colloca na altura de quem com- | prehende a missão da imprensa, e muito especialmente | na materia, de que se occupam as minbas cartas. || Na presente apontarei um exemplo do que acabo de | afirmar; mas presto um serviço ao Correio recom- | mendando-lhe em geral aquelles illustrados jorna- | listas [...].

Sempre vale destacar, com vistas à formação de membros que partirão para o trabalho em arquivos, a importância da sistematização de informações do cabeçalho para o caráter qualitativo-metodológico do *corpus geral* do qual eles participam. Por exemplo, se para a pesquisa sobre PB é preciso discriminar dados, na listagem que vai informar materiais editados, devem ser discriminados brasileiros realmente identificados e os textos de redatores não identificados. Sem isso, podem ser viesados dados com portugueses misturados ou com brasileiros de diferentes regiões que, como é praxe na história da escrita, datam e localizam seus textos onde e quando os escreveram, não onde e quando nasceram. Com essa informação preenchida por quem edita os *corpora*, os usuários poderão rodar cálculos juntando e separando *identificados* e *não identificados* e verificar se há ou não interferência dos dados dos textos de *não identificados*.

Como a entrada de materiais em *corpora* linguísticos gerais usa para a dimensão espacial as informações escritas nos documentos, cartas escritas no Rio de Janeiro são testemunhos dos círculos de escritura que incluem o Rio de Janeiro *intra* ou *extrafronteiras*, mas somente se tornam cartas *cariocas* ou cartas *fluminenses* se a naturalidade de cada redator for identificada. Quando um usuário lança mão de *corpora* criados a partir de, e para cobrir questões/objetivos mais específicos, esse problema pode ser minorado ou resolvido. Se um dado *corpus* for construído somente com cariocas identificados e reunindo três ou quatro tradições discursivas manuscritas e/ou impressas bem identificadas em habituações observadas em duzentos anos, seus usuários terão um *nirvana* linguístico-histórico. Mas, nesse caso, responderiam a perguntas e observariam diacronias de fenômenos linguísticos somente para a cidade do Rio de Janeiro. Esperar até que várias equipes regionais cheguem a esse *nirvana* para fazer comparações inter-regionais é contraproducente. Por isso, a opção por um *corpus* geral coordenado, como o do PHPB, que vá antecipando recortes mais gerais, é uma opção metodológica para minorar uma história geral do português brasileiro feita com generalizações a partir de dados só de *cariocas*, ou só de *mineiros*, ou só de *baianos*, enfim, pontos sem parelha.

CONSIDERAÇÕES FINAIS

Algumas considerações feitas neste capítulo inicial voltam aprofundadas, complementadas, e mesmo contrastadas, por outras considerações feitas nos demais capítulos. Não poderia ser diferente, pois o desejo de contar com

pontos de vista outros sobre implicações positivas, negativas ou sobre lacunas decorrentes das opções metodológicas na construção coordenada de *corpora* é desejo que já vem incutido desde o convite aos autores do volume. É fruto da consciência de que a tradução em capítulos da complexidade de áreas e de visões distintas em colaboração na construção de *corpora* valerá para o caráter de formação de quadros de pesquisadores que venham a construir ou serem usuários de *corpora* de semelhante natureza histórico-diacrônica. Aprofundar, complementar e contrastar pontos abordados neste capítulo inicial é, no fundo, reflexo do avanço das atividades dentro PHPB na fase subsequente à ampliação coordenada do *corpus* agora contemplado na Plataforma e disponível diretamente aos usuários no *corpus mínimo comum* e *corpus diferencial*. O tom bem próximo do da fase “imatura” das equipes originais do PHPB que relatava coisas como “fomos aos arquivos e fizemos assim” é o tom proposital deste capítulo. É o tom de quem se vir incipiente e enfrentar, com o mesmo idealismo, a tarefa de começar a encontrar na Linguística de *Corpus* caráter central ou complementar em sua formação de pesquisador. É o tom para quem quiser encontrar nas reflexões e opções metodológicas do passado do PHPB caminhos para iniciar suas próprias reflexões e opções que aprofundem, complementem e contrastem este capítulo, o volume 2 como um todo, os outros volumes da *História do Português Brasileiro*.

Que este capítulo cumpra seu caráter inspirado nas duas faces do deus latino *Janus*: uma voltada para trás, para o que passou, outra voltada para frente, para o que virá, marcando fins e começos, transições, passagem do tempo. Que os leitores incipientes nas questões dos *corpora* para pesquisas históricas se sintam aqui motivados, com a energia de quem, olhando o ano que passou, começa o ano novo, como se estivessem sempre no mês de *januarius*.

NOTAS

- ¹ A abreviatura *p.p.* no início da notícia deve se tratar da expressão “próximo passado”, para se referir ao mês imediatamente anterior à publicação do jornal.
- ² Rotulada na página do *corpus* diferencial como *Administração privada XVIII 2 – cartas de mercadores portugueses* conforme critério para manuscritos pessoais de Barbosa (1999).