

# O *CORPUS* HISTÓRICO DO PORTUGUÊS PAULISTA: CARACTERÍSTICAS E POTENCIALIDADES<sup>1</sup>

José da Silva SIMÕES

Universidade de São Paulo (USP)

Verena KEWITZ

Universidade de São Paulo (USP)

## RESUMO

*Este artigo tem por objetivo apresentar as estratégias de coleta de materiais interessantes para descrição da difusão da língua portuguesa em São Paulo, como forma de oferecer aos pesquisadores da história do Português Brasileiro em São Paulo uma gama de textos os mais representativos possíveis das várias normas em convivência em cada uma das sincronias a serem estudadas. O material aqui apresentado compõe o Corpus Histórico do Português Paulista. O trabalho toma como escopo os pressupostos teóricos do modelo de Tradições Discursivas, da linha coseriana da Linguística de Texto.*

## ABSTRACT

*The aim of this article is to present the strategies for collecting documents that are relevant to the description of Portuguese diffusion in São Paulo, so as to offer researchers in Historical Linguistics a range of representative texts for the History of Brazilian Portuguese in São Paulo. Specifically, those texts can be representative in terms of different linguistic norms found in the same period of time under attention. The documents we present here are part of the Historical Corpus of Paulista Portuguese, which is based on the Discourse Traditions Model in the Coserian Text Linguistics Perspective.*

---

<sup>1</sup> Este artigo reúne resultados obtidos pela equipe do subprojeto de pesquisa Formação de Corpora do Português Paulista do Projeto História do Português Paulista com apoio da FAPESP (FAPESP, Processo Nº 11/51787-5). Em sua versão ampliada, este trabalho será publicado no vol. 1 da Série História do Português Brasileiro, editada por Ataliba Teixeira de CASTILHO (no prelo).

## PALAVRAS-CHAVE

*Corpus histórico do Português Paulista. Português Brasileiro. Tradições Discursivas. Linguística do Texto.*

## KEYWORDS

*Historical Corpus of Paulista Portuguese. Brazilian Portuguese. Discourse Traditions. Text Linguistics.*

## Introdução

O presente artigo tem como objetivo apresentar o processo de construção de um *corpus* diacrônico que reúne textos de variada tipologia, como forma de documentar as variedades do português escrito em São Paulo. A análise de manuscritos produzidos no Brasil ao longo dos vários séculos indica a convivência de normas distintas: algumas decorrentes de renovados insumos do Português Europeu no território do Brasil Colônia e outras que já indicavam algum tipo de autonomia em relação às primeiras, fruto de interações com outras línguas faladas no território brasileiro, nomeadamente línguas indígenas autóctones e línguas africanas. Estudar a história da presença da língua portuguesa no Brasil implica necessariamente em escolher exemplares de textos que forneçam pistas linguísticas mais fidedignas dessas influências, no sentido de construir uma espécie de semiótica social, mediante a análise de estratégias de retextualização de práticas sociais ligadas à comunicação política, administrativa, jurídica, econômica, filosófica, teológica, religiosa, literária, técnica, artesanal, militar, familiar, privada etc. (cf. OESTERREICHER 1997, 2008 e 2011). O *Projeto de História do Português Paulista* (PHPP) conta um *corpus* histórico constituído de edições filológicas de documentos manuscritos, impressos e orais coletados de acordo com critérios baseados nos pressupostos teóricos do modelo

de Tradições Discursivas (COSERIU, 1994 [1981]; KOCH, 1988, 1997 e 2008; OESTERREICHER, op. cit.; KOCH e OESTERREICHER 2011 [1990]; KABATEK, 2005, 2006, 2008 entre outros). Neste artigo, apresentamos (i) as estratégias de coleta de materiais interessantes para a descrição da história da difusão da língua portuguesa no território de São Paulo, como forma de oferecer aos pesquisadores da história do Português Brasileiro (PB) em São Paulo uma gama de textos os mais representativos possíveis das várias normas em convivência em cada uma das sincronias a serem estudadas, e (ii) dados de alguns fenômenos linguísticos encontrados em diversos documentos paulistas de variados gêneros textuais, considerados, em sua maioria, portadores de norma culta: *inventários e testamentos, cartas pessoais, oficiais e de administração privada, memórias históricas* etc.

Este artigo divide-se na descrição dos três grandes conjuntos de documentos editados e disponibilizados na página de *corpus* do PHPP ([www.phpp.fflch.usp.br/corpus](http://www.phpp.fflch.usp.br/corpus)), a saber: *corpus mínimo manuscrito*, *corpus mínimo impresso* e *corpus diferencial*. Em cada subseção apresentamos a definição de cada um e os seus textos mais representativos. Por questões de espaço, apresentamos dados de alguns fenômenos linguísticos constantes em apenas alguns desses documentos, o que demonstra sua potencialidade para os estudos de mudança linguística.

## 1. O corpus mínimo manuscrito do PHPP

Desde o início do *Projeto Para a História do Português Brasileiro* (PHPB) em 1997, *cartas* vem sendo coletadas e usadas como *corpus* ideal para encontrar pistas de oralidade conceptual. Em função disso, decidiu-se constituir um *corpus* mínimo de manuscritos que fossem comuns a várias regiões brasileiras.

Para a formação de *corpus* do Português Paulista, as cartas sempre estiveram na agenda de coleta e edição, com especial atenção às cartas

pessoais. No entanto, em arquivos públicos nem sempre se encontram cartas pessoais, ou ainda cartas que contenham as pistas de oralidade ou textos em norma popular, até mesmo em função de ser delicado definir o que era norma popular em séculos anteriores ao XX.

As cartas pessoais encontradas em arquivos públicos provêm geralmente dos acervos pessoais de alguma pessoa ilustre, como políticos, escritores, jornalistas etc., cuja doação é mais recente (do século XIX para cá). Acervos pessoais anteriores ao século XIX são raros ou inexistentes. Encontram-se, em grande parte, documentos relacionados à esfera jurídica e administrativa, dentre os quais cartas oficiais e de administração privada.

A carta oficial pode ser caracterizada prototipicamente por ser um documento público, em que remetente e destinatário têm relação assimétrica (ascendente ou descendente), o tema é altamente fixo, as dimensões da ação comunicativa limitam-se a *requerer*, *ordenar*, *recriminar*, *obedecer*, *relatar*, e o texto é altamente controlado, com fórmulas que seguem uma determinada tradição discursiva (cf. SIMÕES; KEWITZ 2009b: 488).

Nem todas as cartas rotuladas como *oficiais* apresentam as condições e estratégias prototípicas desse subgênero, pois algumas tratam de mais de um tema, e o texto pode apresentar, por vezes, grau médio de controle.

Conforme descrevemos em Simões; Kewitz (2009b: 490), uma *carta de administração privada* (Barbosa, 2000) prototípica apresenta grau de publicidade semiparticular; relação assimétrica entre remetente e destinatário; núcleo temático fixo; dimensões da ação comunicativa mais ampla (que da carta oficial), como *requerer*, *relatar*, *recriminar*, *reclamar*, *oporse*, *ordenar*, *obedecer* etc.; planejamento semicontrolado do texto, ainda que apresente domínio de fórmulas de abertura e de fechamento.

Cartas pessoais prototipicamente têm caráter particular, os interlocutores têm relação simétrica, o tema é variado e livre, as dimensões da ação comunicativa são ainda mais amplas (*narrar*, *agradecer*,

*reclamar, desejar, pedir, opinar, desabafar* etc.), e o grau de planejamento é, em geral, semicontrolado ou livre (cf. Simões; Kewitz, 2009b: 492).

Em vários conjuntos de cartas paulistas é possível observar a sobreposição de traços do que se caracteriza prototipicamente como carta pessoal e carta de administração privada. Exemplificam essa perspectiva os seguintes conjuntos de cartas: (i) Cartas em torno da figura de José Bonifácio (1ª metade do XIX), de vários remetentes (CH-PHPP: SIMÕES & KEWITZ 2006)<sup>2</sup>; (ii) *Cartas Andradinas (1820-1828)*, escritas por José Bonifácio e Martim Francisco destinadas a Antônio de Menezes Vasconcellos de Drummond e José Joaquim da Rocha (CH-PHPP: Simões; Oliveira; Lopes 2015), (iii) Cartas de Fernando Prestes e Julio Prestes a Washington Luís, 1ª metade do XX, editadas por Albuquerque, Ferreira e Kewitz (CH-PHPP: 2015) e (iv) Cartas de um comerciante do Vale do Ribeira (1ª e 2ª metades do XX), editadas por Araújo; Kewitz (CH-PHPP: 2015).

Além das *cartas*, os materiais que compõem o *corpus* mínimo manuscrito contêm *atas*, particularmente atas de câmara do século XVII, *processos-crime* e documentos judiciais afins e *testamentos*.

Num primeiro momento, *atas de câmara* podem parecer bastante formulaicas e repletas de repetições, em função da própria estrutura e do grau de publicidade do texto. Ainda assim, não se podem desconsiderar os textos que mais se aproximem no polo da escrituralidade, como os documentos jurídicos e administrativos de forma geral. Os primeiros documentos paulistas dessa tipologia, editados pelo PHPP, datam do século XVII e foram produzidos nas vilas de Jundiá (CH-PHPP: MORAIS, 2014) e de Mogi das Cruzes (CH-PHPP: FERREIRA, 2007), duas das mais antigas da Capitania de São Paulo. Embora sejam textos portadores de tradições discursivas que se repetem através de fórmulas,

<sup>2</sup> A partir deste ponto, sempre que a referência bibliográfica se referir a um conjunto de documentos específico do *Corpus Histórico do Projeto História do Português Paulista*, a chamada de autores/editores será antecedida da notação [CH-PHPP]. Esses materiais estão disponíveis sob o rótulo de seus autores/editores na página [www.phpp.fflch.usp.br/corpus](http://www.phpp.fflch.usp.br/corpus) e por esse motivo não se encontra como referência desdobrada ao final deste artigo.

é possível encontrar trechos de escrita menos controlada, como em *Estes vinte Equatro mil Reis sam **dos que Esta** | **va** Em depozição namão do dito jozephbealves* (CH-PHPP: MORAIS, 2014).

Ao considerar *processos* como textos interessantes aos estudos linguísticos, uma questão que se coloca geralmente é se todo o processo deve ser editado e analisado ou partes dele e, assim, quais partes. Essa pergunta deve ser respondida do ponto de vista do analista: se se quer estudar determinado item linguístico, talvez nem todo o processo seja relevante. Mas se a análise envolver questões discursivas, semânticas e argumentativas, talvez o processo em sua totalidade deva ser considerado. Com base nessa reflexão, Dias (CH-PHPP: 2017) editou um processo-crime do século XIX e identificou as partes mais e menos formulaicas e suas tradições discursivas, o que pode auxiliar na identificação de diferentes normas em cada uma, até mesmo por conta de quem produziu cada um desses documentos. A título de exemplo, na parte em que testemunhas são ouvidas e têm sua fala transcrita, categoricamente, neste processo se faz uso da fórmula “Perguntado se...”, independentemente de ser homem ou mulher.

Semelhantemente às atas de câmara do século XVII, os *testamentos* inserem-se na tipologia de documentos jurídicos, considerados os mais conservadores quanto às partes que os compõem e às tradições discursivas (cf. KABATEK, 2006). O PHPP conta com 37 documentos de Taubaté da 2ª metade do XVII, editados por Monteiro (CH-PHPP: 2014).

Analisando testamentos paulistas e norte-rio-grandenses, Castilho da Costa (2015) identificou algumas mudanças nesse gênero em documentos dos séculos XVII a XX quanto à macroestrutura e à estrutura do texto. Segundo ela, deixam de aparecer os atos de fala “encomendar a alma” e “fazer pedidos pios”, “encomendar missas” e “ordenar a forma de sepultamento”, cristalizados na tradição romana jurídico-religiosa e normatizados pelos manuais de bem morrer (2015: 294).

Para cada uma das macroestruturas dos testamentos analisados, CASTILHO DA COSTA (2015: 304-214) observou a repetição de formas linguísticas, como preposições complexas (*em nome de, no ano de, na forma seguinte, a rogo de* etc.), uso de verbos *dicendi* (*mandar, declarar*), reduzidas de gerúndio para justificar a produção do testamento (*temendo, desejando, estando* doente etc.), redobramento (*todos estes filhos e filhas casamos em sua vida della dita mulher*), entre outras estruturas.

## 2. O corpus mínimo impresso do PHPP

A formação de um *corpus* para o PHPB teve início justamente com documentos impressos, nomeadamente de jornais brasileiros do século XIX. O interesse nesses textos se deu, primeiramente, pelo fato de não haver cortes e edição nos jornais do século XIX, como ocorre em jornais atuais. Dado que várias capitais e cidades brasileiras produziram jornais locais ao menos a partir da 2ª metade do XIX, passou-se a formar o *corpus* mínimo impresso, composto por *cartas de leitores e de redatores e anúncios*.

Num estudo amplo das mudanças de alguns gêneros jornalísticos, Castilho da Costa (2010) analisou edições completas de dois jornais paulistanos de 1854, 1875 e 1901, *O Correio Paulistano* e a *Província de São Paulo*, nome original de *O Estado de S. Paulo* (CH-PHPP: CASTILHO DA COSTA, 2010). Aos três gêneros acima mencionados, acrescentou-se a *notícia*, configurando entre os tipos de texto do *corpus* mínimo impresso do PHPP. A autora identificou ainda a presença de vários gêneros jurídicos administrativos nos jornais do século XIX, como despacho, acórdão, sentença, declaração, ata de assembleia provincial etc., ausentes nos jornais atuais. Segundo ela, *a linguagem jurídica exerceu influência na constituição dos gêneros textuais como a notícia e o anúncio publicitário* (2010: 270), demonstrando uma intersecção entre os domínios do Direito, Jornalismo e Publicidade, que gradualmente vão se desvinculando nos jornais do século XX.

Mais recentemente, foram coletados e editados os editoriais do jornal *A Província de São Paulo* (séc. XIX) / *O Estado de S. Paulo* (séc. XX) com o objetivo de verificar mudanças nos processos constitutivos do texto (CH-PHPP: LOPES-DAMÁSIO; JUBRAN *et al.* Orgs. 2010), cartas do editor/editorial de uma revista feminina da 2ª metade do XIX (CH-PHPP: MEDEIROS; ARAÚJO; GONÇALVES SEGUNDO 2015a), anúncios de revistas paulistanas da 1ª metade do XX (CH-PHPP: MEDEIROS; ARAÚJO; GONÇALVES SEGUNDO 2015b) e anúncios de emprego em jornais paulistanos do XIX (CH-PHPP: OLIVEIRA, 2012).

### 3. O *corpus* diferencial do PHPP

Neste conjunto inserem-se tipologias textuais diversas, tanto impressas quanto manuscritas, além de inquéritos orais; ora são exclusivas de uma região, ora começam a aparecer em determinada época, por isso denomina-se *diferencial*.

Dentre os textos impressos estão os editoriais de jornais de bairros paulistanos (de 1973 a 1991), editados por GONÇALVES SEGUNDO (CH-PHPP: 2011), e o folhetim *A Justiça dos Bohemios*, de Ponson du Terrail, cuja tradução para o português foi publicada sequencialmente em várias edições de um jornal paulistano de 1884 (CH-PHPP: MÓDOLO; SILVA 2010). Duas peças teatrais de autores paulistas foram editadas para compor parte do *corpus* diferencial: *Sangue Limpo*, de Paulo Eiró (1863) e *As Noivas*, do santista Paulo Gonçalves (década de 1920/30), editadas ambas por MÓDOLO; SANTOS (CH-PHPP: 2010).

O material manuscrito compõe a maior parte do conjunto de textos do *corpus* diferencial: *diários de viagem*, *memórias históricas*, *descrições de municípios paulistas* e *documentos portugueses*. A edição desses textos está relacionada à pesquisa específica de alguns autores, como resultado de teses e dissertações, e à pesquisa de alguns subprojetos do PHPP em



sua segunda fase, como é o caso dos documentos portugueses, descritos em 3.1. Os *diários de viagem* e as *memórias históricas* são examinados em conjunto dadas as tradições discursivas que os entrelaçam, seguidos das *descrições dos municípios paulistas*. Por fim, apontamos os inquéritos orais, completando assim os materiais do *corpus* diferencial do PHPP.

### 3.1 Documentos portugueses

No conjunto de *corpus* diferencial do PHPP foram inseridos documentos portugueses por duas razões: (a) compõem objeto de edição e estudo de pesquisadores dos subprojetos do PHPP e (b) servem de base para os estudos do Português Médio, levando em conta as hipóteses levantadas por Cardeira (2005, 2013) e Castro (2012).

A partir das sugestões de Castro (2012), foi proposta uma nova linha de pesquisa de fenômenos sintáticos, tomando como ponto de partida o século XV pelo fato de os primeiros navegadores que aqui chegaram terem adquirido sua gramática antes de 1500 (KEWITZ; PAIXÃO DE SOUSA, 2011). Por isso, os documentos portugueses foram incluídos no *corpus* do PHPP para complementar o que já está disponível no *corpus* do projeto *Thyco-Brabe* e no *Corpus Informatizado do Português Medieval*. Até 2016, contamos com dois documentos: um contrato datado de 1506, (CH-PHPP: FACHIN; MONTE, 2015) e uma crônica anônima do século XVI (CH-PHPP: LOMBARDO, 2015).

### 3.2 Diários de viagem, memórias históricas e descrições de municípios paulistas

A ideia de incorporar *diários de viagem*, *memórias históricas* e *descrições de municípios* aos *corpora* históricos do PHPB vem do entendimento de que se deve levar em conta textos cuja orientação principal seja a descrição e a narratividade.

Simões (2007) defende a hipótese de que tradições discursivas ligadas ao universo discursivo da História (SCHLIEBEN-LANGE,

1983) contribuíram para a circulação e *habitualização* (KOCH, 2008) de rotinas textuais, modelos estruturais da arquitetura dos textos históricos que circulavam tanto em língua espanhola como em língua portuguesa e serviram como modelos dessas mesmas rotinas para a construção das *memórias históricas*.

Alguns paulistas dedicaram-se à tarefa de relatar viagens, sobretudo encomendadas por algum superior em Portugal durante o período colonial. Entre eles destaca-se Francisco José de Lacerda e Almeida (1753-1798) que teve a incumbência de demarcar as fronteiras entre o Brasil e a América espanhola. No *corpus* histórico do PHPP encontra-se o “*Diario de viagem, que de Villa Bella deMatto-Grosso fis para a Cidade deSãoPaulo pelas ordinarias derrotas de terra, eRios que delle constar no anno de 1788*” (a partir do manuscrito da Biblioteca Nacional de Lisboa, PBA 642), editado por Simões; Manoel; Morais (CH-PHPP: 2013a). Além deste documento há outras edições relevantes para o estudo da produção de memórias do séc. XVIII e XIX: as memórias de Frei Gaspar da Madre de Deus de 1780 (CH-PHPP: SIMÕES; MANOEL; MORAIS 2013a) e de Manoel Cardoso de Abreu de 1796 (CH-PHPP: COSTA 2007), e a *Dissertação* do português Marcelino Pereira Cleto de 1788 (CH-PHPP: SIMÕES; MANOEL; MORAIS 2013a).

### 3.2.1 Descrição de municípios paulistas

Na Biblioteca Nacional do Rio de Janeiro, encontram-se descrições de municípios brasileiros datadas sobretudo do ano de 1881. Trata-se de uma encomenda feita às câmaras municipais brasileiras pelo então Diretor B.F. Ramiz Galvão, com o objetivo de reunir dados para a Exposição *História do Brasil* em homenagem a D. Pedro II. Patriótico, Ramiz Galvão preocupava-se em construir o passado do Brasil: CALDEIRA (2010: 65).

Segundo os dados do catálogo *online* da BNRJ, foram enviadas cerca de 140 descrições, acompanhadas de ofício de encaminhamento a Ramiz

Galvão. Poucos estão digitalizados e a maioria conta com 5 a 10 páginas em média. Boa parte dos Estados estão representados com a descrição de pelo menos um município; outros Estados contêm maior número de descrições, a exemplo de Minas Gerais, com 19 municípios, assim como São Paulo.

Dessa relação, digitalizamos e editamos filologicamente as Descrições dos Municípios de Santo Antônio do Apiaí e de Xiririca (atual Eldorado), ambas no Vale do Ribeira, uma das regiões mais antigas do Estado de São Paulo. Ambos escritos em 1881, são, de longe, os mais extensos e contêm informações ricas para diversas áreas do conhecimento, como Botânica, Zoologia, História, Geografia, Geologia etc. No ofício de encaminhamento, a Câmara Municipal de Apiaí esclarece que a descrição *é producto do trabalho de muitos redigindo alguns artigos completos e muitos outros respondendo as perguntas que se lhes fazião* (CH-PHPP: KEWITZ; ARAÚJO; BERTO 2016a). Nos exemplos abaixo, apresentamos alguns fenômenos linguísticos identificados na *Descrição do Município de Apiaí*, para ilustrar sua potencialidade:

- (1) Concordância zero:  
**E’ desconhecida** nesta villa, as febres intermitentes, bixigas, dysenterias, e tetanos. (linhas 196-198)
- (2) Concordância por reanálise:  
O rheumatismo **saõ motivados** das muitas chuvas (...) (linhas 222-223)
- (3) Uso categórico das formas *seje* e *esteje*:
  - (a) Na Capella da Ribeira ou **seje** no terreno ribeirino (...) (linhas 229-230)
  - (b) Não deve ser este o elemento da prospe\_ | ridade do Apiahÿ pois embora **esteje** | provado que este ouro acha\_se muito | abundante... (linhas 1953-57)

- (4) Expressão do Objeto Direto:  
Deffcil enpresa é escrever a historia par\_ | ticular do Apiahÿ pela grande deffcul\_ | dade de reunir os dados precisos e co\_ | ordenar **elles** de uma forma regular. (linhas 873-876)
- (5) Falsos começos (sentenças sem verbo):  
Não cons\_ | ta que fossem agredidos por bugres, | pois existiam poucos e estes errantes **Ø** | como attesta a falta de utencilios acha\_ | dos nestes logares. Taõ somente o Senhor (...) (linhas 971-5)
- (6) Orações relativas: abundante uso do pronome *cujó*, mas há também casos de relativa não padrão:  
Destrebuia | receitas **que** cobrava 2000 **por cada uma** (...) (linhas 1507-8).

Pelos fenômenos exemplificados acima é possível observar que, de um lado, há pistas de menor grau de planejamento do texto, até mesmo talvez pela pressa em mandar a descrição ao Diretor da BN, Ramiz Galvão, em tempo para a Exposição da História do Brasil; de outro identifica-se também grau médio-alto de elaboração do texto pela presença de referências a obras técnicas estrangeiras e detalhamento na descrição de cada item do questionário, por exemplo.

### 3.3 Inquéritos Orais

Os primeiros projetos de coleta, edição e análise de inquéritos orais já estão bem documentados, a exemplo dos trabalhos do Projeto NURC e *Projeto da Gramática do Português Falado*: Castilho (2010), entre outros. A importância da oralidade para os estudos linguísticos não caberá aqui examinar, mas vale lembrar duas abordagens decorrentes dos estudos da língua falada (culta) no Brasil, nomeadamente a abordagem multissistêmica (CASTILHO, 2007, 2010) e a perspectiva textual-interativa (JUBRAN, 2006, 2015).

Outros tantos projetos se formaram, particularizando ora uma determinada região brasileira, ora certas classes sociais e/ou níveis de escolaridade, entre outros. No âmbito do PHPP, foram levantados inquéritos orais de falantes analfabetos ou com pouca escolaridade, residentes em comunidades da Grande São Paulo, reunidos em Rodrigues (CH-PHPP: 2013). Soma-se a isso a incorporação dos inquéritos orais produzidos pelo *Projeto Filologia Bandeirante* (CH-PHPP: MEGALE Org. 2000), com cerca de 16 entrevistas com falantes no interior paulista com mais de 60 anos de idade e de baixa ou nenhuma escolaridade.

Nos dois casos, as transcrições foram feitas de acordo com as normas propostas por Rodrigues; Ferreira Netto (2000), com acomodações fonéticas utilizando a ortografia convencional, permitindo entrever fenômenos fonológicos, além dos sintáticos, lexicais, semânticos e discursivos.

Da 1ª metade do século XX há poucas gravações, ora pela sua raridade, ora por depender da coleta em acervos específicos (como o *Museu da Imagem e do Som* de São Paulo, entre outros), tarefa essa ainda não empreendida. Temos, até o momento, uma gravação com a entrevista de Monteiro Lobato, realizada na Rádio Record em 1948. A gravação está disponível na *internet* e foi transcrita por Kewitz (CH-PHPP 2013). Nessa gravação especificamente, é possível observar o recorrente uso de parentetização pelo escritor, o que pode ser explicado, talvez, pelo fato de ter sido preso duas vezes antes da entrevista e, assim, recear tocar em assuntos delicados para a época, como o petróleo, a política da ditadura do governo Gaspar Dutra etc.

## Considerações finais

Neste artigo, apresentamos um panorama de textos que compõem os três *corpora* do português paulista – *mínimo manuscrito*, *mínimo impresso* e *diferencial* – quanto à diversidade de gêneros textuais: suas tradições

discursivas e suas potencialidades para os estudos linguísticos. Esse olhar sobre a constituição de um *corpus* representa bem o que disse Rodolfo Ilari (c.p.), a propósito das atividades do *Projeto de Gramática do Português Falado*: *estamos construindo o barco e navegando ao mesmo tempo*.

A caracterização das potencialidades dos documentos, especificamente, ainda que de forma breve, demonstra a relevância do exame tanto das macroestruturas dos textos quanto dos fenômenos que podem ser ou já foram analisados mais detalhadamente, como a concordância, uso de pronomes pessoais e de tratamento, estratégias de junção, expressão do objeto direto e do sujeito, entre outros fenômenos.

Embora bastante diversificado, o que apresentamos a partir dos documentos constantes na página do PHPP na internet ([www.phpp.fllch.usp.br/corpus](http://www.phpp.fllch.usp.br/corpus)) corresponde ao que MATTOS E SILVA (2006) bem coloca sobre o labor da Linguística Histórica, de que documentar a diacronia de uma língua é sempre um trabalho fragmentado, “porque fragmentário é o espólio de que dispõe o pesquisador” (2006: 34), respeitando sempre a máxima laboviana de que é preciso empreender “a arte de fazer o melhor uso de maus dados”. O que mostramos neste artigo, no entanto, demonstra que é possível tentar fazer um bom uso dos *bons* dados.

## Referências

BARBOSA, Afrânio G. **Análise linguística do gerúndio**. Cartas de comércio do séc. XVIII. Tese (Doutorado em Língua Portuguesa) Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.

CALDEIRA, Ana Paula Sampaio. **A Biblioteca Nacional nos tempos de Ramiz Galvão (1879-1882)**. Anais da Biblioteca Nacional, 2010, vol. 130, p. 9-109.

CARDEIRA, Esperança. **Entre o português antigo e o português clássico**. Lisboa: Imprensa Nacional-Casa da Moeda, 2005.

\_\_\_\_\_. **Do português médio ao clássico: o Cancioneiro Geral de Garcia de Resende**. In: E. C. Herrero; C. C. Rigual (eds.). **Actas del XXVI Congreso Internacional de Linguística y de Filología Románicas** (Valencia 2010). Vol. I. De Gruyter. p.543-554, 2013.

CASTILHO, Ataliba T. **Proposta funcionalista de mudança linguística**. Os processos de lexicalização, semanticização, discursivização na constituição das línguas. In: LOBO, Tânia; RIBEIRO, Ilza; CARNEIRO, Zenaide; ALMEIDA, Norma. (Orgs.). **Para a história do português brasileiro: novos dados, novas análises**. Salvador: EDUFBA, 2006, p. 223-296.

\_\_\_\_\_. **Abordagem da língua como um sistema complexo. Contribuições para uma nova Linguística Histórica**. Em: A.T. de Castilho; M.A.T. Moraes; R.E.V. Lopes / S.M.L. Cyrino (Orgs. 2007). **Descrição, História e Aquisição do Português Brasileiro**. Homenagem a Mary A. Kato. Campinas: Pontes /Fapesp, p. 329-360, 2007.

\_\_\_\_\_. **Nova Gramática do Português Brasileiro**. São Paulo: Editora Contexto, 2010.

\_\_\_\_\_. **Historiando o Português Brasileiro**. Vol. 1 – *Corpus* Diacrônico do Português Brasileiro. São Paulo, Editora Contexto, no prelo.

CASTILHO DA COSTA, Alessandra F. **Tradições discursivas da mídia impressa no português brasileiro: gênese, formação e transformação de gêneros discursivos em uma perspectiva diacrônica**. Relatório final de pesquisa de Pós-Doutorado, USP/ FAPESP, 2010.

\_\_\_\_\_. **Por quanto esta he minha ultima vontade do modo que tenho dito:** Tradições discursivas, textuais e linguísticas em testamentos norte-rio-grandenses dos séculos XVIII a XX. In: WINTER-FROEMEL, Esme; LÓPEZ SERENA, Araceli; TOLEDO Y HUERTA, Álvaro Octavio de; FRANK-JOB, Barbara. (Org.). **Diskurstraditionelles und Einzelsprachliches im Sprachwandel.** Tübingen: Narr, 2015, v. 1, p. 285-316.

CASTRO, Ivo. **Vésperas Brasilianas.** In: SANTIAGO-ALMEIDA, M.M.; Lima-Hernandes, M.C. (Orgs.) **História do Português Paulista**, vol. III, Série Estudos. Campinas: IEL Publicações/FAPESP, 2012, p.45-72.

COSERIU, Eugenio. **Textlinguistik.** Eine Einführung. Tübingen: Gunter Narr, 1994 [1981].

DIAS, Carla Regiane. **E morreo curado por pózes e raízes: edição semidiplomática e estudo de um processo-crime de feitiçaria e homicídio no Brasil Império (século XIX).** Dissertação (Mestrado em Filologia e Língua Portuguesa) Faculdade de Filosofia Letras e Ciências Humanas da Universidade de São Paulo, São Paulo, 2017.

KABATEK, Johannes. **Sobre a historicidade de textos.** Tradução de José da Silva Simões. **Linha d'água.** 17. São Paulo: Associação de Professores de Língua e Literatura / FFLCH / USP / DLCV, 2005: 159-172.

\_\_\_\_\_. **Tradições discursivas e mudança linguística.** In: LOBO, Tânia; RIBEIRO, Ilza; CARNEIRO, Zenaide; ALMEIDA, Norma. (Orgs.). **Para a história do português brasileiro: novos dados, novas análises.** Vol. VI. Salvador: EDUFBA, 2006: 505-530.

\_\_\_\_\_. (org). **Sintaxis histórica del español y cambio lingüístico: nuevas perspectivas desde las Tradiciones Discursivas.** Madrid: Iberoamericana, 2008.



KEWITZ, Verena; PAIXÃO DE SOUSA, Maria Clara. **Vésperas Brasilianas: uma agenda para os estudos sintáticos do Português Brasileiro nos primeiros séculos. Revista Portuguesa de Humanidades – Estudos Linguísticos**, v. 15, p. 65-90, 2011.

KOCH, Peter. **Norm und Sprache**. In: ALBRECHT, Jörn; LÜDTKE, Jens; THUN, Harald (eds.). **Energie und Ergon. Sprachliche Variation – Sprachgeschichte – Sprachtypologie**. Studia in honorem Eugenio Coseriu. 3 vol. Tübingen. (Tübinger Beiträge zur Linguistik, 300), 1988: 327-354.

\_\_\_\_\_. **Diskurstraditionen: zu ihrem sprachtheoretischen Status und zu ihrer Dynamik**. In: FRANCK, Barbara; HAYE, Thomas; TOPHINKE, Doris (Orgs.). **Gattungen mittelalterlicher Schriftlichkeit**. Tübingen: Narr, (ScriptOralia; 99), 1997: 43-79.

\_\_\_\_\_. **Tradiciones discursivas y cambio lingüístico: el ejemplo del tratamiento *vuestra merced* en español**. In: KABATEK, Johannes (org.). **Sintaxis histórica del español y cambio lingüístico: nuevas perspectivas desde las Tradiciones Discursivas**. Madrid: Iberoamericana, 2008: 53-88.

\_\_\_\_\_.; OESTERREICHER, Wulf. **Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch**. Tübingen: Niemeyer. (Romanistische Arbeitshefte, 31), 2011 [1990].

MATTOS E SILVA, Rosa Virgínia. **O Português Arcaico. Fonologia, morfologia e sintaxe**. São Paulo: Contexto, 2006.

JUBRAN, Clélia Cândida S. **A perspectiva textual-interativa**. In Jubran & Koch (2006 Orgs.) **Gramática do Português Culto Falado no Brasil: A construção do texto falado**. Campinas: Editora da Unicamp, 2006, p.27-36.

\_\_\_\_\_. (Org.). **A construção do texto falado**. São Paulo: Editora Contexto, 2015.

OESTERREICHER, Wulf. **Zur Fundierung von Diskurstraditionen**. In: FRANCK, Barbara; HAYE, Thomas; TOPHINKE, Doris (Orgs.). **Gattungen mittelalterlicher Schriftlichkeit**. Tübingen: Narr, (ScriptOralia; 99), 1997: 19-41.

\_\_\_\_\_. **Dinámica de estructuras actanciales em los Siglos de Oro: el ejemplo del verbo *encabalar***. In: KABATEK, Johannes (org.). **Sintaxis histórica del español y cambio lingüístico: nuevas perspectivas desde las Tradiciones Discursivas**. Madrid: Iberoamericana, 2008: 225-248.

\_\_\_\_\_. **Conquistas metodológicas en la lingüística diacrónica actual**. La historicidad del lenguaje: lenguas, variedades y tradiciones discursivas en el marco de una semiótica social. In: CASTILLO LLUCH, Mónica e PONS RODRIGUES, Lola (orgs.). **Así se van las lenguas variando**. Nuevas tendencias en la investigación del cambio lingüístico en español. Berna: Peter Lang (Colección Fondo Hispánico de Lingüística y Filología), 2011: 305-334.

RODRIGUES, Ângela Cecília de Souza; FERREIRA NETTO, Waldemar. **Transcrição de inqueritos: problemas e sugestões**. In: MEGALE, H. (org.). **Filologia Bandeirante**. São Paulo: Humanitas / FAPESP, 2000, p. 171-193.

SCHLIEBEN-LANGE, Brigitte. **Tradition des Sprechens**. Elemente einer pragmatischen Sprachgeschichtsschreibung. Stuttgart / Berlin / Köln / Mainz: Kohlhammer, 1983.

SIMÕES, José da Silva. **Sintaticização, discursivização, e semanticização das orações de gerúndio no português brasileiro**. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.

\_\_\_\_\_; KEWITZ, Verena. **Traços linguístico-discursivos em *corpora* do português brasileiro.** In: **Estudos Linguísticos XXXV**, 2006b: 1018-1027.

\_\_\_\_\_; KEWITZ, Verena (orgs.). **Cartas paulistas dos séculos XVIII e XIX:** uma contribuição para os corpora do PHPB. São Paulo: Humanitas / FFLCH / USP, 2006a (CD-Rom).

\_\_\_\_\_; KEWITZ, Verena. **Tradições discursivas e organização de *corpora*.** In: AGUILERA, Vanderci (org.). **Para a História do Português Brasileiro.** Vol. VII. Londrina: UEL, 2009a: 467-558.

\_\_\_\_\_; KEWITZ, Verena. **Normas linguísticas, história social, contatos linguísticos e tradições discursivas:** transformando encruzilhadas em novos caminhos para a constituição de *corpora* diacrônicos. In: CASTILHO, Ataliba Teixeira de (org.). **História do Português Paulista.** Série Estudos, vol. 1. Campinas: Instituto de Estudos da Linguagem / Universidade Estadual de Campinas, 2009b: 699-720.

Recebido em 10/10/2016 e aceito em 06/12/2016