

RELATO DE EXPERIÊNCIA

PARA A DESCRIÇÃO E ANÁLISE DE **VARIÉDADES DO PORTUGUÊS**: O DESAFIO DA CONSTITUIÇÃO DE BANCO DE DADOS

Silvia Figueiredo BRANDÃO  

Universidade Federal do Rio de Janeiro (UFRJ)

Silvia Rodrigues VIEIRA  

Universidade Federal do Rio de Janeiro (UFRJ)

 OPEN ACCESS

EDITORES

- Miguel Oliveira Jr. (UFAL)
- René Almeida (UFS)

AVALIADORES

- Elyne Vitória (UFAL)
- Alessandra Mara (UFTM)

SOBRE OS AUTORES

- Sílvia Figueiredo Brandão
Conceptualização, Metodologia,
Escrita – Rascunho Original, Escrita –
Análise e Edição.
- Sílvia Rodrigues Vieira
Conceptualização, Metodologia,
Escrita – Rascunho Original, Escrita –
Análise e Edição.

DATAS

- Recebido: 18/08/2021
- Aceito: 30/08/2021
- Publicado: 04/11/2021

COMO CITAR

BRANDÃO, S. F.; VIEIRA, S. R. (2021).
Para a descrição e análise de
variedades do português: o desafio da
constituição de banco de dados.
Cadernos de Linguística, v. 2 n. 4, e537.

RESUMO

Neste texto, apresentam-se experiências concernentes à organização de *corpora* no âmbito de alguns projetos desenvolvidos na UFRJ nas linhas geo-sociolinguística. Além de indicar suas características bem como as motivações e os critérios que presidiram à formação dos bancos de dados, focalizam-se suas contribuições científicas, no sentido de dar conta: (i) da caracterização de variedades do Português do Brasil, não somente urbanas, mas também rurais; (ii) da constituição e comparação de variedades continentais do Português; e (iii) das feições contemporâneas do Português, não só no *continuum* rural-urbano, mas também no que se refere à variação estilística e à modalidade falada/escrita. Sintetizam-se, assim, algumas das contribuições desses projetos para o conhecimento de diferentes variedades do Português e, por fim, discutem-se algumas dificuldades inerentes à constituição de *corpora* linguísticos, em especial no que se refere à sua necessária disponibilização para a comunidade em geral em virtude de sua importância sociolinguística e histórica.

ABSTRACT

In this text, experiences concerning the organization of corpora within the scope of some projects developed at UFRJ in the geo-sociolinguistic approaches are presented. In addition to indicating its characteristics as well as the motivations and criteria that guided the construction of the databases, its scientific contributions are focused, in order to account for: (i) the characterization of Brazilian Portuguese varieties, not only urban, but also rural; (ii) the constitution and comparison of Portuguese continental varieties; and (iii) the contemporary features of Portuguese, not only in the rural-urban continuum, but also in terms of stylistic variation and the spoken/written modality. Thus, some of the contributions of these projects are summarized for the knowledge of different varieties of Portuguese and, finally, some difficulties inherent to the constitution of linguistic corpora are discussed, especially with regard to its necessary availability for the community in general because of its sociolinguistic and historical importance.

PALAVRAS-CHAVE

Corpus linguístico; metodologia; geolinguística; sociolinguística; variedades do português.

KEYWORDS

Linguistic *corpus*; methodology; geolinguistics; sociolinguistics; portuguese varieties.

INTRODUÇÃO

O desafio da constituição de bancos de dados para análise linguística já foi enfrentado por diversos pesquisadores brasileiros, perseguindo ideais igualmente diversos do ponto de vista teórico-metodológico. Passado pelo menos meio século da organização do que pode ser considerado o primeiro banco de dados orais no Brasil, que permitiu a confecção do Atlas Prévio dos Falares Baianos (APFB), publicado em 1964, já se pode não só avaliar os limites desse tipo de empreitada no que tange à organização e à divulgação desses materiais, mas também aquilatar as contribuições científicas obtidas com algumas das iniciativas de elaboração e exploração de *corpora* no país, consideradas as perguntas que cada banco de dados pretende responder.

Até meados do século XX, pode-se dizer que o registro de variedades do Português era feito essencialmente no âmbito da Dialetoлогия, que tinha na Geografia Linguística seu principal método de descrição e análise. A publicação do APFB, além de se pautar por esse método e de inaugurar a prática de pesquisa de campo no país, ensejou também a formação de um *corpus*, com base em metodologia bem delimitada, embora sem os recursos que, hoje, a informática a cada dia renova. Na década de 1970, dois projetos – Norma Urbana Oral Culta e Competências Básicas do Português – em outra perspectiva, a da Sociolinguística Variacionista, desencadearam o interesse em organizar *corpora* que pudessem servir de base ao conhecimento das variedades do Português do Brasil e, ainda, ao debate sobre suas origens. Esta comunicação dá continuidade a essa história, focalizando uma parcela das iniciativas desenvolvidas na UFRJ¹ e partilhadas por diferentes gerações de pesquisadores que objetivaram e objetivam prover resultados científicos empiricamente fundamentados em bancos de dados não só do Português do Brasil, mas também de outras variedades continentais.

Tem-se, portanto, por objetivo apresentar as motivações e as características dos referidos projetos que se organizam e se apoiam em *corpora*, de modo a ampliar e redirecionar as pesquisas, sempre no intuito de melhor caracterizar as variedades do Português e contribuir para a resolução de questões investigativas. Nesse sentido, o texto, além desta introdução, desenvolve-se em três seções destinadas a focalizar as contribuições científicas dos bancos de dados em questão, no sentido de dar conta: (i) no item 2, da caracterização das variedades fluminenses não somente urbanas e, assim, permitir descrever as supostas origens das chamadas normas populares, como fez o Projeto do Atlas Etnolinguístico dos Pescadores do Estado do Rio de Janeiro (APERJ); (ii)

1 Trata-se, aqui, apenas das experiências das autoras do texto e equipe de orientandos, no que se refere à constituição, análise e divulgação de *corpora* linguísticos com dados fluminenses, no caso do Português do Brasil, e/ou dados portugueses e africanos (Moçambique e São Tomé e Príncipe).

no item 3, da constituição e comparação de variedades continentais do Português, a fim de dar conta da complexidade típica de línguas pluricêntricas, bem como da influência de variáveis extralinguísticas na formação dessas variedades, como fizeram os Projetos Análise contrastiva de variedades do Português (VARPORT) e Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias; e (iii) das feições contemporâneas do Português empregado hoje, 40/50 anos depois do primeiro banco de dados no Rio de Janeiro, não só no *continuum* rural-urbano, mas também no que se refere à variação estilística e à modalidade (fala x escrita), como propõem os projetos a serem descritos no item 4.

Na Seção 5, tecem-se algumas considerações finais, entre as quais não só as que dizem respeito às contribuições da constituição de *corpora*, mas também as que se relacionam à sua disponibilização para a comunidade científica em geral. Trata-se de empreendimento mais do que necessário, sobretudo por sua importância sociolinguística e histórica, uma vez que, em cada um deles, estão registrados os perfis de comunidades que se intercomunicam em Língua Portuguesa e por meio dela constroem suas identidades e difundem seus bens culturais.

1. PARA A DESCRIÇÃO DAS NORMAS FLUMINENSES

Primeiramente, trata-se, aqui, da contribuição de *corpora* no que se refere ao problema da representação das comunidades de fala. Os bancos de dados inicialmente planejados e elaborados no Rio de Janeiro priorizaram comunidades em centros urbanos, seja para o registro da fala de indivíduos pouco ou nada escolarizados, seja para o registro das variedades cultas. Assim, iniciativa pioneira no cenário fluminense foi a de constituir um banco de dados que permitisse descrever como falam os brasileiros sem instrução formal em regiões menos urbanas/mais rurais do que as da região metropolitana do Rio de Janeiro.

Nesse contexto, a ideia de desenvolver o projeto do Atlas Etnolinguístico dos Pescadores do Estado do Rio de Janeiro (APERJ), inicialmente coordenado pelo filólogo e professor Celso Cunha, surgiu, em finais da década de 1970, por sugestão do dialetólogo Emmanuel Companyns, um dos colaboradores do Atlas Linguistique et Ethnographique de la Gascogne. Para ele, o projeto se justificava, entre outros motivos, pelo fato de o Estado constituir uma das áreas brasileiras de povoamento mais antigo, sobretudo no litoral, o léxico da pesca ser o repositório de um saber popular, difundido de geração a geração e as comunidades pesqueiras poderem, por isso mesmo, resguardar formas linguísticas mais conservadoras (BRANDÃO, 2005, p. 359).

Essa ideia, no entanto, só se concretizou em meados da década de 1980, quando se formou uma equipe constituída pelo coordenador e cinco professores em fase de

doutoramento: Cilene da Cunha Pereira, Maria Emília Barcelos da Silva, Sílvia Figueiredo Brandão, Edila Vianna da Silva e Laerte Carpena de Amorim², que, com a morte de Celso Cunha em 1989, passou a coordenar o projeto até a sua aposentadoria em 1991. Para viabilizarem a elaboração do Atlas, parte da equipe – mais especificamente as quatro referidas professoras – desenvolveu teses na linha dialetológica, com base na fala, respectivamente, de comunidades pesqueiras de Araruama, da Região dos Lagos, do Município de Campos e do então Município de São João da Barra e que constituiriam alguns dos futuros pontos de inquérito do atlas. Essas teses – que podem ser caracterizadas como monografias dialetais – obedeciam, de certa forma, a um modelo que incluía a gravação de entrevistas guiadas por um questionário, com base nas quais se analisaram algumas variáveis linguísticas e se elaboraram glossários que contemplavam aspectos da atividade pesqueira. As quatro teses serviram, portanto, de base para a definição dos procedimentos metodológicos que norteariam a pesquisa.

Inicialmente, definiram-se 49 pontos de inquérito que abrangiam comunidades pesqueiras litorâneas e interioranas em todos as regiões do estado: 13, na Norte e Noroeste; 12, na das Baixadas Litorâneas; 14, na Metropolitana do RJ e 10, na Sul. Durante dois anos, a equipe dedicou-se à formulação do questionário, ao treinamento para a sua aplicação, à definição das normas relativas à transcrição fonética das respostas, bem como à transcrição grafemática das entrevistas e à realização de pesquisas-piloto nas regiões Sul e Metropolitana da cidade do Rio de Janeiro, para testar o questionário e para eventuais reformulações e acréscimos.

Ocorre que alguns dos integrantes da equipe passaram também a realizar estudos na linha sociolinguística, o que, aos poucos, foi contribuindo para que se reformulasse a metodologia do atlas, desde 1992 coordenado por Sílvia Figueiredo Brandão.

Em função das novas decisões, foi-se chegando à conclusão de que o atlas, em virtude da sua nova dimensão, deveria ser elaborado por regiões, decidindo-se, então, iniciar as investigações pelas regiões Norte e Noroeste, o que implicava dilatar o prazo para divulgação do Atlas.

Os 13 pontos de inquérito das regiões Norte e Noroeste, cujas entrevistas serão disponibilizadas no site, podem ser visualizados na Figura 1: (a) Barra de Itabapoana (BIT), Guaxindiba (GUA), Gargaú (GAR), Atafona (ATA), Farol de São Tomé (FST), Macaé (MAC); (b) São João da Barra (SJB), São Fidélis (SFI), Cambuci (CAM), Itaocara (ITO), Itaperuna (ITA); (c) Ponta Grossa dos Fidalgos (PGF) e São Benedito (SBE), comunidades em que a atividade pesqueira se desenvolve, respectivamente, no a) mar; b) nos rios Paraíba do Sul (as quatro primeiras) e Muriaé; c) nas lagoas Feia e de Cima, respectivamente.

2 O tema da tese de Laerte Carpena de Amorim foram as salinas de Cabo Frio-RJ.

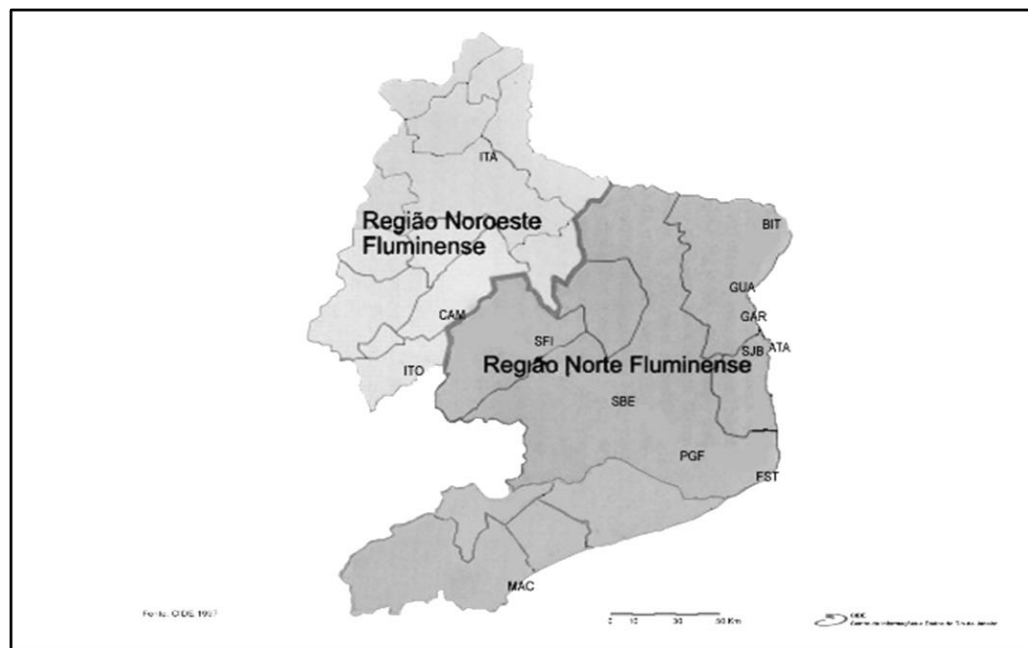


Figura 1. Pontos de inquérito do Projeto APERJ no N e NO fluminenses.
Fonte: Brandão; Vieira (1998, p. 243), com alterações.

Assim, cada uma das 13 localidades, de perfil rural ou urbano, à exceção de Macaé, contou com 6 informantes principais, todos do sexo masculino, distribuídos por 3 faixas etárias: A-18-35; B-36-55 e C-56 anos em diante, analfabetos ou escolarizados até a quarta série do ensino fundamental, num total de 78. A seleção, inteiramente aleatória quanto a esta variável, acabou por refletir, de modo bastante aproximado, a realidade não só dessas comunidades, em que os indivíduos raramente frequentavam a escola por mais de cinco anos, mas também da região como um todo, em que o índice de analfabetismo, a depender da unidade geopolítica, era da ordem de 40% a 60% entre os segmentos mais velhos da população e de 10% a 30% entre os mais jovens, o que correspondia, respectivamente, a um índice de 17%, 50% e 66% de analfabetos por grupo etário de informantes (BRANDÃO; VIEIRA, 1998, p. 228-229).

As entrevistas resultam da conjugação de dois tipos de inquérito: (a) aplicação de questionário composto de 422 itens sobre a atividade pesqueira, abrangendo três grandes campos semânticos: meios físico, biótico e antrópico e (b) DIDs (diálogos entre informante e documentador), em que o pescador era estimulado a falar de temas genéricos e de episódios de seu cotidiano.

O atlas deveria conter: (1) cartas linguísticas convencionais, em que se registrariam itens lexicais provenientes da aplicação do questionário transcritos foneticamente; (2) cartas sintéticas (fonéticas, léxicas e morfossintáticas) que apresentariam os resultados decorrentes da aplicação aos dados da metodologia variacionista, isto é, cartas em que,

por meio de gráficos, se demonstrariam os índices de frequência de ocorrência de variantes. Para esse tipo de cartas, elegeram-se as seguintes variáveis: no plano fonético fonológico, (a) as vogais pretônicas /e/, /o/ e, ainda, /a/, que, durante as sondagens preliminares efetuadas no Norte do Estado, demonstrou ser suscetível a realizar-se como [ɛ], fato que não tinha sido registrado em outras variedades do PB; (b) o /R/ e o /S/ em coda silábica, cujas diferentes realizações parecem servir de parâmetro para a caracterização das áreas linguísticas brasileiras; (c) /t/ e /d/ diante de [i] que apresentam diferentes graus de palatalização, chegando a concretizar-se como africadas pós-alveolares; /r/ e /r̃/ que, em alguns falares, podem ser despalatalizadas ou vocalizadas; (2) no plano morfossintático, (a) a colocação dos pronomes possessivos; (b) a concordância verbal de terceira pessoa do plural e a de número em sintagmas nominais; (c) o uso de artigo diante de possessivos; e (d) o emprego de nós/a gente; (3) no plano lexical: o registro de variantes do léxico da pesca artesanal.

Deve-se salientar que, com base no *Corpus* APERJ, essas variáveis, focalizadas pelos membros da equipe, a que se somaram seus orientandos, foram objeto de muitos trabalhos de iniciação científica, mestrado e doutorado, bem como de publicações e comunicações a congressos. Dentre esses trabalhos³, que ainda hoje constituem os únicos registros e análises sobre a variedade popular das localidades do Norte e Noroeste fluminenses, encontram-se teses e dissertações, como as de Vieira (1995), sobre concordância verbal, Silveira (1996), sobre o uso de artigo diante de possessivo, Almeida (1997), sobre concordância nominal, Rodrigues (2001) sobre o S em coda, Quandt (2004) sobre /l/ em coda e em ataque complexo, Carvalho (2002), sobre /t d/ diante de [i]. Há, ainda, estudos sobre os róticos (BRANDÃO, 1995, 1997, entre outros) e aquele em que houve a divulgação de 11 cartas fonéticas, 7 sobre /R/ e 4 sobre o /S/ em coda silábica (BRANDÃO, 1994), elaboradas segundo os critérios acima descritos.

Em relação ao armazenamento dos dados, as entrevistas do *Corpus* APERJ (feitas originalmente em gravadores de fitas cassetes) encontram-se digitalizadas; somam 178 horas de gravação, 78 delas também transcritas grafematicamente e disponíveis na versão impressa. Embora essas gravações se encontrem armazenadas em dispositivos móveis ou em computadores e ainda não contem com um *site* para divulgação, duas integrantes da antiga equipe estão providenciando sua disponibilização na *web*, por acreditarem na importância do acervo, não só do ponto de vista sociolinguístico, visto que retrata comunidades de perfil rural/rurbano em áreas pouco exploradas do Estado, mas também por seu valor etnográfico e histórico.

3 Nesta e nas demais seções, são destacados, à guisa de ilustração, apenas alguns dos produtos orientados ou elaborados pelas autoras e que tiveram por base os *corpora* de que trata o texto. Não seria possível apresentar uma listagem completa, tendo em vista o número de pesquisadores, vinculados ou não aos projetos aqui focalizados, que a eles vêm recorrendo.

2. PARA A DESCRIÇÃO E A COMPREENSÃO DE VARIEDADES CONTINENTAIS DO PORTUGUÊS

A constituição de bancos de dados que contemple a língua em diversas expressões e espaços geográficos é fundamental não só para a descrição das variedades em questão, mas também para a compreensão das motivações para a variação e a mudança. Observar o que é comum e o que difere em variedades de uma língua permite observar e interpretar as motivações, sejam estas de ordem sócio-histórica, sejam de natureza estrutural ou linguística. Nesse sentido, duas iniciativas, que envolveram pesquisadores brasileiros e portugueses em cooperação, permitiram constituir bancos de dados fundamentais à análise comparativa de variedades do Português: Análise Contrastiva de Variedade do Português (Projeto VARPORT) e o Projeto Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias.

O Projeto Análise Contrastiva de Variedades do Português (VARPORT), de cooperação internacional CAPES-ICCTI, coordenado por Silvia Figueiredo Brandão (UFRJ) e Maria Antónia Mota, do Centro de Linguística da Universidade de Lisboa (CLUL), foi um dos pioneiros nesse tipo de empreendimento⁴. Tendo início em 2000, contou com uma equipe formada por 14 docentes e 11 doutorandos das duas instituições. Os estudos conjuntos foram realizados segundo diferentes perspectivas teóricas, com o objetivo de fornecer contribuições para fazer frente a duas questões básicas: (a) se haveria uma única ou mais de uma gramática do Português e (b) se a diversidade atingiria, com a mesma intensidade, todos os níveis da língua nas diferentes variedades focalizadas.

Entre os tópicos abordados, em cada uma das variedades nacionais, encontram-se, nos planos (a) fonético-fonológico: padrões prosódicos, atuação de processos de enfraquecimento vocálico e consonantal, nasalidade vocálica; (b) morfossintático: ordem dos constituintes no nível da oração e da frase, sujeito preenchido e sujeito nulo, topicalização, verbos leves e plenos, regência nominal e verbal, dêiticos, processos morfológicos flexionais e derivacionais; (c) morfológico, morfofonológico e morfossintático: padrões de flexão verbal, concordância sujeito-verbo, formas nominais do verbo; (d) léxico-semântico: a frequência e a distribuição de itens lexicais.

Para viabilizar o estudo conjunto, que poderia se concentrar nas modalidades oral e/ou na escrita, criou-se um *corpus* compartilhado, que foi disponibilizado no site do Projeto (www.varport.letras.ufrj.br) e que contou, para a sua formação, com a expertise de Maria Fernanda Bacelar do Nascimento, a principal responsável pela

4 Na ocasião, além do VARPORT, só existia outro projeto de cooperação internacional Brasil-Portugal (Português-Europeu-Português Brasileiro: Unidade e Diversidade no Século XXI), coordenado por Mary Kato e João Peres.

organização do *Corpus* do Português Fundamental e do *Corpus* de Referência do Português Contemporâneo – CRPC, no âmbito do Centro Linguístico da Universidade de Lisboa (CLUL).

Os materiais que compõem o *Corpus* Compartilhado do VARPORT advêm de recortes de *corpora* já existentes, como os acima indicados, no caso do PE oral, e dos Projetos NURC e APERJ, para representação, respectivamente, das variedades orais culta e popular do PB. No que se refere ao *corpus* escrito, foram feitos levantamentos, nas Bibliotecas Nacionais sediadas no Rio de Janeiro e em Lisboa, de excertos de notícias, anúncios e editoriais de jornais brasileiros e portugueses que abarcam o período de 1808 a 2000. No quadro 1, tem-se uma síntese do formato do *corpus*, cujos textos escritos, gravações e respectivas transcrições grafemáticas estão disponíveis no referido site.

PB E PE – MODALIDADE ORAL			
CULTO		POPULAR	
Década de 70	3 faixas etárias: 18-35 anos 36-55 anos 56 anos em diante	Décadas de 1980-1990	3 faixas etárias: 18-35 anos 36-55 anos 56 anos em diante
Década de 90	Sexo/gênero: masculino feminino		Sexo/gênero: masculino
Recontato			
PB E PE – MODALIDADE ESCRITA			
Gêneros textuais	Períodos		
	Século XIX	Século XX	
Anúncios	Fase 1 (De 1808-1840)	Fase 1 (De 1901-1924)	
Editoriais	Fase 2 (De 1841-1870)	Fase 2 (De 1925-1949)	
Notícias	Fase 3 (De 1871-1900)	Fase 3 (De 1950-1974)	
		Fase 4 (De 1975-2000)	

Quadro 1. Formato do *Corpus* Compartilhado VARPORT

No site do Projeto, está elencada a produção científica da equipe no período de 2000 a 2003 e que abarca teses, livros e capítulos de livros, artigos, conferências, palestras, comunicações a eventos científicos e os cursos e seminários realizados. Pretende-se incluir no site a publicação com que se inaugurou o projeto e a que marcou o seu término, ambas organizadas pelas coordenadoras do projeto (BRANDÃO; MOTA, 2000; 2003).

Dando continuidade às pesquisas compartilhadas entre o Programa de Pós-Graduação em Letras Vernáculas/UFRJ, mais concretamente à linha de pesquisa Língua e Sociedade: Variação e Mudança, e o Grupo Fontes, Conhecimento e Modelização em Linguística, do Centro de Linguística da Universidade de Lisboa (CLUL), Maria Antónia R. C. da Mota e Sílvia Rodrigues Vieira desenvolveram o segundo projeto de cooperação internacional (CAPES / GRICES - 2008-2011), intitulado Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias do Português.

O referido projeto, ao buscar a comparação entre variedades do Português usadas em três continentes (África, Brasil e Europa), no que diz respeito a uma questão central das gramáticas, a concordância morfossintática, buscou estabelecer padrões variáveis, associando-os a variedades de ordem geográfica e/ou social, o que, em última instância, vem contribuindo para determinar diferentes normas em coexistência e em concorrência, em cada espaço geográfico onde se fala português.

Para cumprir o referido propósito, as pesquisas do projeto precisavam levar em conta dados contemporâneos de variedades do Português em diferentes países, razão pela qual constituíram, no período sobretudo de 2008 a 2011, um *corpus* de entrevistas sociolinguísticas realizadas no Brasil (Rio de Janeiro) e em Portugal continental (Grande Lisboa) e insular (Funchal), estratificado segundo os mesmos critérios, que ficou conhecido como *Corpus Concordância*⁵. No âmbito português, de todas as gravações feitas, são consideradas, para a amostra básica, 18 gravações em Lisboa/Oeiras, 18 gravações em Cacém, cidade-dormitório vizinha de Lisboa, além de 18 gravações em Funchal, na Ilha da Madeira. No Brasil, do mesmo modo, são consideradas 18 gravações na Zona Sul da capital, em Copacabana e adjacências, e 18 gravações na cidade de Nova Iguaçu, também vizinha da capital, na área metropolitana do Rio de Janeiro. Além desses informantes que se enquadram mais apropriadamente no perfil desejado, outras entrevistas foram realizadas e constam do referido *corpus* como integrantes de uma amostra complementar.

Os critérios empregados para a constituição da amostra respeitam três variáveis extralinguísticas – faixa etária, nível de instrução e sexo/gênero do informante –, conforme sintetizado no quadro a seguir:

5 Embora a fase inicial do referido projeto tenha se encerrado após o triênio previsto no convênio de cooperação internacional, a análise comparativa de variedades do Português com base no *Corpus Concordância* e em outros materiais teve continuidade no âmbito do Projeto 21 da Associação de Linguística e Filologia da América Latina (ALFAL), sob a coordenação de Sílvia Rodrigues Vieira e Sílvia Figueiredo Brandão. No período de 2011 a 2018, o Projeto 21 foi intitulado “Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias do Português” – período sintetizado nos artigos publicados em número específico dos Cadernos da ALFAL (VIEIRA; BRANDÃO, 2015). Tendo acolhido outros pesquisadores, o grupo, a partir de 2019, passou a contemplar quaisquer fenômenos morfossintáticos ou fonético-fonológicos em variedades do Português, motivo pelo qual hoje é identificado como “Projeto 21: Estudo comparativo de variedades africanas, brasileiras e europeias do Português (COMPARAPORT)”.

Faixa etária	Nível de escolaridade	Gênero/gênero
A = 18 a 35 anos	1 = Fundamental – 2º segmento (6º ao 9º ano)	H = Masculino
B = 36 a 55 anos	2 = Médio (10º ao 12º ano)	M = Feminino
C = 56 a 75 anos	3 = Superior	
3 células x 3 células x 2 células = 18 informantes (naturais das localidades em estudo)		
COP = PB-COPACABANA; NIG = PB-NOVA IGUAÇU; OEI = PE-LISBOA-OEIRAS; CAC = PE-CACÉM; FNC = PE-FUNCHAL		

Quadro 2. Formato do *Corpus* CONCORDÂNCIA

Diversos alunos brasileiros de graduação e de pós-graduação da UFRJ (além de uma pós-graduanda portuguesa, assessorada pela Professora Maria Antónia Mota) e alguns professores participaram do longo período que envolveu a feitura de gravações e transcrições das entrevistas sociolinguísticas. O procedimento-padrão prevê que cada um desses colaboradores esteja identificado na ficha do informante referente às entrevistas. No caso da amostra de Funchal, na Ilha da Madeira, a constituição do *corpus* ficou a cargo da Professora Aline Bazenga, que realizou todo o trabalho com o auxílio de alunos da Universidade da Madeira.

Participaram, ainda, do trabalho de revisão das transcrições coletadas no Brasil, diversos estudantes de graduação e pós-graduação, de modo a preparar todo o material que ficou inicialmente hospedado, para divulgação, na UFRJ, via Núcleo de Computação Eletrônica - NCE. Atualmente, após diversas etapas de tratamento de dados e problemas com a segurança e manutenção do espaço, todo o material está, além de armazenado em dispositivos móveis e computadores, hospedado no *site* www.corporaport.lettras.ufrj.br (em fase de atualização e revisão), que agrega a divulgação de informações relacionadas às pesquisas de Sílvia Figueiredo Brandão e de Sílvia Rodrigues Vieira.

No que se refere a variedades africanas do Português, a equipe do Projeto valeu-se, inicialmente, da amostra do Português de São Tomé, organizada, em 2009, por Tjerk Hagemeyer, disponibilizada no chamado *Corpus* VAPOR – Variedades Africanas do

Português, que, assim como o *Corpus África*⁶, está disponível em www.clul.ul.pt, site do Centro de Linguística da Universidade de Lisboa, em que são também elencados estudos realizados por membros da equipe brasileira com base nessas fontes de dados. No site da Cátedra do Português: língua segunda e estrangeira (<https://www.catedraportugues.uem.mz/variedades-nao-europeias>), da Universidade Eduardo Mondlane, de Moçambique, podem-se encontrar, na seção Bibliografia, na versão completa, trabalhos realizados por membros da equipe do *Corpus Concordância* com amostras relativas a São Tomé.

Ao longo dos dez anos de disponibilização desses dados, muitas análises contrastivas dos padrões de concordância, além de outros fenômenos fonético-fonológicos e morfossintáticos, foram desenvolvidas, de modo a permitir não só o conhecimento de variedades ainda não descritas segundo os métodos sociolinguísticos, como também a comparação dos resultados com os obtidos em *corpora* da segunda metade do século XX. Considerando apenas a experiência dos membros da equipe, pode-se destacar, primeiramente, a produtiva divulgação de resultados relacionados à concordância verbal e nominal em diversos artigos em periódicos e livros. Merece destaque, nesse sentido, um volume da Revista *Journal of Portuguese Linguistics* (MOTA; VIEIRA, 2013), integralmente dedicado às contribuições do Projeto em sua fase inicial, e o livro *A concordância verbal em variedades do Português: a interface Fonética-Morfossintaxe* (VIEIRA, 2015). Essas referências valem-se do conjunto de amostras comparáveis providas pelo *Corpus Concordância* – duas comunidades de fala no Brasil, na área metropolitana do Rio de Janeiro, duas em Portugal peninsular – Lisboa e Cacém – e, ainda, uma na Ilha da Madeira –, além dos dados da Ilha de São Tomé, referentes ao *Corpus VAPOR*. Conforme testemunha o professor Gregory Guy, que prefaciou o referido livro,

Esse conjunto de dados permite novas análises da concordância verbal variável, atacando empiricamente não só as questões de condicionamento da regra, mas também as questões sobre a evolução da Língua Portuguesa e as origens das características distintivas do PB. Os resultados mostram claramente os contrastes entre os vários dialetos examinados, e, no Brasil, além de sinalizar as diferenças entre zonas rurais ou urbanas, evidenciam aquelas relativas aos diferentes níveis de escolarização. (GUY, 2015, p. 9)

Tendo os pesquisadores do Projeto percebido a urgente necessidade de investigação de outras variedades africanas do Português, a constituição de bancos de dados teve continuidade por meio de projetos investigativos específicos. Nesse sentido, em 2016, Sílvia Rodrigues Vieira e Karen Cristina da Silva Pissurno realizam, em viagem a Maputo, capital de Moçambique, a gravação de uma amostra nas mesmas condições do *Corpus Concordância*, com entrevistas de 35 indivíduos, falantes de Português como L1 ou L2,

6 O *Corpus África* é constituído de dados de cinco amostras das variedades de Angola, Cabo Verde, Guiné-Bissau, Moçambique e São Tomé e Príncipe

material que foi identificado como *Corpus* Moçambique-Port. Tendo em vista as dificuldades de encontrar perfis totalmente compatíveis com os das amostras brasileira, europeia e são-tomense dos *corpora* anteriormente descritos, foram selecionados 18 informantes, de acordo com o perfil básico dos sujeitos entrevistados quanto aos critérios de idade, sexo e escolaridade.

Dos entrevistados selecionados, 61% afirmam que têm o Português como língua materna, enquanto o restante informa ter aprendido o Português mais tarde, especialmente na escola, utilizando-a como uma L2. Todos são naturais de Moçambique, sendo 11 de Maputo e os demais, embora morem na capital, são migrantes de outras províncias. No que se refere às línguas por eles utilizadas, aqueles que afirmam ter aprendido o Português como L2 indicam, principalmente, Changana e Rhonga como L1, caso dos informantes de Maputo, e outras línguas, como, por exemplo, Makhwa, Chuwabu, Moniga, Bitonga, caso dos informantes das outras localidades.

Com base no *Corpus* Moçambique-Port, foram ou estão sendo investigados diversos temas morfossintáticos, como a concordância verbal de 3ª pessoa (PISSURNO, 2017), de 1ª pessoa do plural, a concordância nominal interna ao sintagma nominal, a ordem dos clíticos pronominais, ou fonético-fonológicos, como a realização variável dos róticos, do vocalismo átono, e da realização de ditongo /ei/, dentre outros. Para o conhecimento desses empreendimentos com dados não só do Português de Moçambique, mas também no de São Tomé, Brandão (2018) oferece um panorama dos resultados das investigações, na obra intitulada *Duas variedades africanas do Português: variáveis fonético-fonológicas e morfossintáticas*.

Considerando a complexa situação de multilinguismo em Moçambique e os resultados obtidos com o *Corpus* Moçambique-PORT, interessa recolher novas amostras dessa variedade do Português, seja como língua 1 ou como língua 2. Nesse sentido, encontra-se em curso a realização de entrevistas sociolinguísticas mediada por aplicativos de interação de longa distância, amostra identificada como *Corpus* Moçambique *online*. Dão andamento à constituição desse material as pós-graduandas do Programa de Pós-graduação em Letras Vernáculas Karen Cristina da Silva Pissurno e Bianca Ferreira da Costa. Está em curso a realização das gravações *online*⁷ com moradores de uma província próxima a Maputo, mas com perfil menos urbano.

7 As entrevistas são feitas via chamada de whatsapp, com gravação no Google meet ou no Zoom. Já foram realizadas, até o momento, 16 entrevistas sociolinguísticas.

3. PARA A COMPREENSÃO DE EXPRESSÕES VARIÁVEIS DO PORTUGUÊS CONTEMPORÂNEO EM CONTINUA: RURALIDADE-URBANIDADE, MONITORAÇÃO ESTILÍSTICA E FALA-ESCRITA

Embora os resultados obtidos a partir de investigações de dados dos *corpora* anteriormente descritos tenham permitido a descrição e, ainda, a compreensão de dinâmicas de formação de variedades do Português, restam questões a serem respondidas que não podem ser contempladas exclusivamente em amostras que se valem dos critérios clássicos de estratificação de amostras, como idade, sexo e escolaridade.

Nesse sentido, vale destacar iniciativas de constituição de *corpora*, sobretudo em pesquisas de pós-graduação, que têm por objetivo não só obter dados atualizados com base em entrevistas sociolinguísticas prototípicas, mas também de outras coletas de expressões da fala e da escrita no Estado do Rio de Janeiro, que, em alguma medida, refletem os *continua* de variação propostos por Bortoni-Ricardo (2005) como um modelo de descrição do Português do Brasil: o *continuum* rural-urbano; o *continuum* de monitoração estilística; e o *continuum* de oralidade-letramento.

A fim de dar conta do eixo ruralidade-urbanidade, algumas iniciativas merecem destaque no sentido de buscar atualizar a descrição das normas cultas e populares no complexo tecido social contemporâneo do Brasil, que conta com constante afluxo de migrantes de regiões e estratos socioeconômicos diversos.

Considerando o Estado do Rio de Janeiro, Corrêa (2019) realizou, em sua tese de doutorado, detalhada investigação sobre o uso e a avaliação das marcas de concordância verbal de terceira pessoa em comunidades urbanas (Nova Iguaçu e Rio de Janeiro, mesmas cidades contempladas na elaboração do *Corpus* Concordância) e em outras consideradas mais rurais/menos urbanas (em Guapimirim/Cachoeira de Macacu). Para tanto, constituiu o chamado *Corpus* Rural-Urbano-PORT, empreendimento que contempla entrevistas sociolinguísticas, além de experimentos/testes de avaliação.

Conforme propõe a investigadora, a pesquisa pressupõe a existência de uma escala de ruralidade/urbanização, mesmo consciente de que as pessoas, nos dias atuais, independentemente das comunidades em que habitam, estão minimamente integradas por meios de comunicação, o que determina uma nova concepção de ruralidade, não mais como área de nítido isolamento, mas com algumas características físicas, demográficas e econômicas que a distinguem. Considerando esses pressupostos, foram realizadas 45 entrevistas (dentre as quais 24 da região urbana e 21 da região rural) com

indivíduos de escolaridade fundamental ou média, homens e mulheres, distribuídos nas mesmas faixas etárias do *Corpus* Concordância.

Todas as entrevistas sociolinguísticas realizadas estão sendo transcritas e também serão disponibilizadas oportunamente no site <https://corporaport.letras.ufrj.br/>. A amostra de entrevistas rurais está sendo ampliada, aos poucos, de modo a não só completar o que já foi feito (houve dificuldade de encontrar três idosos com ensino médio na amostra inicialmente organizada, em virtude de os índices de escolarização no campo serem mais baixos do que os verificados na cidade), mas também abarcar em geral mais indivíduos de comunidades distantes da capital, com hábitos que podem ser considerados mais rurais.

Pesquisas como as de Correia (2019) bem como as realizadas com o *Corpus* Concordância têm registrado resultados diversos sobre o que representaria a variedade urbana do Português do Brasil, a partir da fala fluminense. Ocorre que a maioria desses resultados parece representar uma comunidade de fala urbana com maior acesso a bens culturais, típico de grandes cidades. Tendo em vista, entretanto, a complexidade de megametrópoles como o Rio de Janeiro, pesquisas em curso têm procurado coletar, então, dados mais diversos das comunidades consideradas urbanas.

Nesse sentido, cabe citar dois trabalhos: o de Monteiro (2020), monografia de fim de curso de graduação, e o de Danieli Silva Chagas, em sua tese de doutorado em andamento. O primeiro, tendo por propósito investigar a fala de cariocas na Vila Cruzeiro, favela do Rio de Janeiro, iniciou a composição de banco de dados com moradores da região, advogando em prol de um *continuum* “asfalto-morro”, que não costuma ser contemplado na maioria dos *corpora* brasileiros. O segundo objetiva compor um banco de dados de uma comunidade carioca remanescente de Quilombo do Camorim, localizada no bairro de Jacarepaguá, região oeste da cidade.

Com iniciativas como essas, entende-se ser possível apresentar resultados que sejam mais representativos da realidade complexa do Português, em meios tanto urbanos, como rurais.

Considerando o *continuum* de monitoração estilística, Rosa (2019), em sua dissertação de Mestrado, compôs amostra que foi intitulada *Corpus* Registro-PORT. Tendo percebido os altos índices de concordância padrão verificados nas entrevistas do Rio de Janeiro do *Corpus* Concordância, a pesquisadora buscou responder à seguinte questão: estaria um possível grau de atenção prestada à fala em situação semi-spontânea, como a da entrevista sociolinguística, afetando os resultados obtidos?

Desse modo, a pesquisadora realizou a gravação de 12 falantes (um homem e 11 mulheres) na cidade do Rio de Janeiro, em duas situações dialógico-interacionais. Primeiramente, foram gravadas interações com os participantes de forma secreta; posteriormente, foram feitas entrevistas sociolinguísticas com os mesmos indivíduos. Ao

fim das duas etapas, os informantes autorizaram as duas recolhas de dados. Essas interações também estão em fase de transcrição, para que oportunamente sejam disponibilizadas no recém-referido *site*.

Com base nessas entrevistas, Rosa (2019) pôde confirmar sua hipótese, demonstrando que os índices de concordância internos ao sintagma nominal são afetados pelo grau de atenção prestado à fala, sendo menores nas gravações secretas.

Por fim, no que se refere ao *continuum* de oralidade-letramento, foi realizada, no âmbito de uma disciplina de Pós-graduação em Letras Vernáculas da UFRJ, ministrada no ano de 2018, uma primeira experiência de constituição de banco de textos de fala e escrita cultas, com base em um *continuum* de gêneros textuais dos domínios jornalístico e acadêmico. Distribuídos supostamente do mais oral ao mais escrito, esses textos revelaram, padrões variáveis de uso sistemático que evidenciaram a relevância de tal *continuum*.

Em Vieira e Lima (2019), apresenta-se parte do levantamento de dados – feito com os estudantes da referida disciplina – concernentes a oito fenômenos variáveis: preenchimento do objeto direto/acusativo de 3ª pessoa; preenchimento do objeto indireto/dativo de 3ª pessoa; construções existenciais; orações relativas; colocação pronominal; estratégias de indeterminação do sujeito; e expressão de futuro. Nesse empreendimento, foram selecionados textos de dez gêneros textuais distintos, quais sejam: entrevistas sociolinguísticas; tirinhas; anúncios; entrevistas impressas (em jornais ou revistas); cartas de leitor; crônicas jornalísticas; notícias; editoriais; teses/dissertações; artigos em revistas científicas – esses três últimos também pertencentes à área da Comunicação Social.

Dando continuidade a esse empreendimento, Monique Débora Alves Lima, em sua tese de doutorado, está construindo, sob as mesmas condições de coleta, o banco de textos escritos no domínio jornalístico, enquanto Ana Carolina Alves Caetano, em sua Dissertação de Mestrado, está construindo o banco de textos referentes ao domínio acadêmico, que também vai comparar áreas distintas de formação (Letras, Economia e Comunicação Social).⁸ Com base em tais trabalhos, a pesquisa sociolinguística também contará com um banco de textos da modalidade escrita, com concepções diversas de oralidade/letramento, o que permitirá amplo mapeamento de fenômenos morfossintáticos variáveis diversos.

8 Iniciativa recente, desenvolvida por Sílvia Rodrigues Vieira, Monique Débora Lima e pesquisadores de outras instituições brasileiras, com a consultoria de Carlos Alberto Faraco, busca ampliar essa experiência considerando também as capitais de São Paulo, Porto Alegre, Recife e Salvador. O objetivo é constituir um banco de dados que permita descrever a variabilidade intrínseca à escrita culta, o que, em última instância, permitirá formulações de padronização linguística fundamentada sobretudo na descrição dos usos.

4. CONSIDERAÇÕES FINAIS

O relato dessas diversas experiências permite constatar que, ao longo de cerca de quase sessenta anos, a contar da publicação do primeiro atlas brasileiro, foi grande o avanço das pesquisas, sobretudo as desenvolvidas com base na metodologia da Sociolinguística Variacionista e da Dialetoлогия, no sentido de constituir bancos de dados que permitam conhecer os fenômenos variáveis que atuam no Português do Rio de Janeiro, no âmbito brasileiro, e em outras variedades do Português. Com base nessas experiências, pode-se certamente afirmar que a observação de *corpus* continua a ser uma das mais relevantes ferramentas dos pesquisadores, não só para a descrição de dados, mas também para a explicação do funcionamento das línguas do mundo, em termos de suas motivações estruturais e sócio-históricas.

Não obstante as comprovadas contribuições dos bancos de dados para a descrição das variedades do Português do Brasil, as experiências já permitiram formular as perguntas que não puderam ser respondidas com base nos limites da estratificação sociolinguística dos *corpora*. Desse modo, fica evidente a necessidade de construção de novas amostras que possibilitem não só atualizar o conhecimento acerca da língua em uso, em suas diversas fases, mas também avaliar outras influências sobre as variedades que extrapolam a esfera geo-sociolinguística, como, por exemplo, as relativas à atuação dos *continua* de ruralidade-urbanidade, monitoração estilística, fala-escrita, ou, ainda, as que venham a resolver o problema laboviano da avaliação de variantes.

Além dos aspectos atinentes à organização de bancos de dados, outras questões precisam receber investimento da Linguística Brasileira, no sentido de prover a divulgação e a transparência das atividades científicas para toda a sociedade brasileira, o que se coaduna com o importante movimento de Ciência Aberta no país.

O campo da divulgação dos bancos de dados linguísticos exige, entretanto, requisitos que extrapolam os fundamentos teórico-metodológicos da investigação científica em si e esbarra em necessidades materiais, financeiras, para o armazenamento e disponibilização dos *corpora* em redes nacionais/internacionais, além do compartilhamento dos materiais (fundamental para garantir a replicabilidade das pesquisas) e dos resultados dos projetos investigativos. Essas tarefas implicam contar não só com um espaço virtual para o armazenamento dos dados, mas também com uma equipe preparada para acompanhamento permanente das atividades envolvidas na tarefa (da coleta, transcrição, revisão e informatização dos dados).

Conforme afirmou Brandão (2020), “desafios não faltam, mas também não falta empenho por parte dos pesquisadores no sentido de aprimorar métodos de análise e, por

parte da ABRALIN, no sentido de incentivar o compartilhamento de *corpora*, que é também uma forma de incentivar novas pesquisas.”⁹

Sabe-se que a resolução dessas dificuldades constitui necessidade antiga e afeta igualmente diversos projetos nacionais. Contar com os servidores de instituições públicas – como fazem as autoras do presente texto – torna o trabalho dependente das condições materiais e técnicas, que são muito diversas e, por vezes, instáveis no contexto brasileiro. Sem dúvida, seria de alta relevância para a pesquisa contar com espaço nacional que pudesse hospedar os materiais para investigação linguística.

REFERÊNCIAS

ALMEIDA, E. M. *A variação da concordância nominal num dialeto rural*. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 1997.

BORTONI-RICARDO, S. M. *Nós chegemu na escola, e agora? Sociolinguística e educação*. São Paulo: Parábola Editorial, 2005.

BRANDÃO, S. F. Algumas cartas fonéticas do Atlas Etnolinguístico dos Pescadores do Estado do Rio de Janeiro. Comunicação apresentada ao Congresso Internacional sobre o Português. Associação Portuguesa de Linguística/Universidade de Lisboa. Lisboa, abril de 1994.

BRANDÃO, S. F. O /R/ Implosivo no Norte do Estado do Rio de Janeiro. In: PEREIRA, C. C.; PEREIRA, P. R. D. (org.) *Miscelânea de estudos linguísticos, filológicos e literários in memoriam Celso Cunha*. Rio de Janeiro: Nova Fronteira. 1995. p. 49-58.

BRANDÃO, S. F. Aspectos de um dialeto rural. In: HORA, D. *Diversidade Linguística no Brasil*. João Pessoa: Idéia, 1997. p. 61-69.

BRANDÃO, S. F. A história e as contribuições de um projeto na linha geolinguística. In: AGUILERA, V. (org.) *A geolinguística no Brasil: caminhos e perspectivas*. Londrina: Editora da Universidade Estadual de Londrina. 2005. p. 357-370.

BRANDÃO, S. F. (org.) *Dois variedades africanas do Português: variáveis fonético-fonológicas e morfossintáticas*. São Paulo: Blucher, 2018.

BRANDÃO, S. F.; MOTA, M. A. (org.) *Análise contrastiva de variedades do português: projeto luso-brasileiro em curso*. Rio de Janeiro: Faculdade de Letras da UFRJ, 2000.

BRANDÃO, S. F.; MOTA, M. A. (org.) *Análise contrastiva de variedades do Português: primeiros estudos*. Rio de Janeiro: In-Fólio, 2003.

BRANDÃO, S. F.; VIEIRA, S. R. Aspectos morfossintáticos da fala de comunidades pesqueiras: um estudo variacionista. In: GROBE, Sybille; ZIMMERMANN, Klaus (ed.) *Substandard e mudança no Português do Brasil*. Frankfurt am Main: TFM, 1998. p. 227-254.

9 Conferência de abertura do X Encontro de Sociolinguística (remoto), promovido por instituições universitárias dos Estados da Bahia e de Sergipe, em via de publicação.

CARVALHO, S. D. M. *A palatalização das plosivas dentais na fala de pescadores do Norte e Noroeste do Estado do Rio de Janeiro*. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2002.

CORRÊA, C. M. M. de L. *Concordância verbal de terceira pessoa do plural em comunidades rurais e urbanas do estado do Rio de Janeiro: avaliação e produção*. Dissertação (Doutorado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2019.

GUY, G. Prefácio. In: VIEIRA, S. R. (org.). *A concordância verbal em variedades do Português: a interface Fonética-Morfossintaxe*. Rio de Janeiro: Vermelho Marinho, 2015. p. 7-9.

MONTEIRO, L. de S. *"Eles fala umas fanfarronagem que nós não entende": A concordância verbal de primeira pessoa do plural em favela do Rio de Janeiro*. Monografia (Licenciatura em Português-Espanhol) - Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2020.

MOTA, M. A.; VIEIRA, S. R. (org.). *Journal of Portuguese Linguistics*, v. 12, n. 2, 2013.

PISSURNO, Karen Cristina da Silva. *A concordância verbal de terceira pessoa do plural na variedade moçambicana do Português: uma abordagem sociolinguística*. 2017. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Universidade Federal do Rio de Janeiro, 2017.

QUANDT, V. O. *O comportamento da lateral anterior na fala do Norte-Noroeste fluminense*. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2004.

RODRIGUES, S. H. Arouca. *O S pós-vocálico na fala da Região Norte-Noroeste do Estado do Rio de Janeiro*. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2001.

ROSA, F. F. S. da. *A influência da monitoração estilística na concordância nominal de número interna ao SN*. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2019.

SILVEIRA, E. F. B. *O uso de artigo diante de pronomes possessivos na fala de comunidades rurais do Estado do Rio de Janeiro*. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 1996.

VIEIRA, S. R. *Concordância verbal: variação em dialetos populares do Norte fluminense*. Dissertação (Mestrado em Letras Vernáculas) - Programa de Pós-graduação em Letras Vernáculas, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 1995.

VIEIRA, S. R. (org.). *A concordância verbal em variedades do Português: a interface Fonética-Morfossintaxe*. Rio de Janeiro: Vermelho Marinho, 2015.

VIEIRA, S. R.; BRANDÃO, S. F. (org.) *Cadernos da ALFAL*, n. 7, 2015.

VIEIRA, S. R.; BRANDÃO, S. F. *CORPORAPORT: Variedades do Português em análise*. Rio de Janeiro: Faculdade de Letras-UFRJ. Disponível em: <www.corporaport.letras.ufrj.br>

VIEIRA, S. R.; LIMA, M. D. A. de O. (org.). *Variação, gêneros textuais e ensino de Português: da norma culta à norma-padrão*. Rio de Janeiro: Letras UFRJ, 2019.

VIEIRA, S. R.; MOTA, M. A. C. da (org.). *Corpus Concordância*. Rio de Janeiro: Faculdade de Letras/UFRJ. Disponível em: VIEIRA, S. R.; BRANDÃO, S. F. *CORPORAPORT: Variedades do Português em análise*. Rio de Janeiro: Faculdade de Letras-UFRJ. www.corporaport.letras.ufrj.br.

VIEIRA, S. R.; PISSURNO, K. C. da S. (org.). *Corpus Moçambique-PORT*. Rio de Janeiro: Faculdade de Letras/UFRJ. Disponível em VIEIRA, S. R.; BRANDÃO, S. F. *CORPORAPORT: Variedades do Português em análise*. Rio de Janeiro: Faculdade de Letras-UFRJ. www.corporaport.letras.ufrj.br.