

Raquel Meister Ko. Freitag  
Silvana Silva de Farias Araújo  
Valter de Carvalho Dias  
organizadores

# DESAFIOS PARA A PESQUISA EM SOCIOLINGUÍSTICA



**Blucher** Open Access



DESAFIOS PARA A PESQUISA EM  
SOCIOLINGUÍSTICA

*CONSELHO EDITORIAL*

André Costa e Silva

Cecilia Consolo

Dijon de Moraes

Jarbas Vargas Nascimento

Luis Barbosa Cortez

Marco Aurélio Cremasco

Rogério Lerner



Apoio do Edital CAPES/FAPITEC/PROMOB 10/2016

**Blucher** Open Access

Raquel Meister Ko. Freitag  
Silvana Silva de Farias Araújo  
Valter de Carvalho Dias  
(organizadores)

DESAFIOS PARA A PESQUISA EM  
SOCIOLINGUÍSTICA

2022

*Desafios para a pesquisa em Sociolinguística*

© 2022 Raquel Meister Ko. Freitag, Silvana Silva de Farias Araújo e Valter de Carvalho Dias  
Editora Edgard Blücher Ltda.

*Publisher* Edgard Blücher

*Editor* Eduardo Blücher

*Coordenação editorial* Jonas Eliakim

*Produção editorial* Kedma Marques

*Diagramação e capa* Laércio Flenic

*Revisão de texto* Samira Panini

*Imagem da capa* iStockphoto

---

# Blucher

Rua Pedroso Alvarenga, 1245, 4º andar  
04531-934 – São Paulo – SP – Brasil  
Tel 55 11 3078-5366  
contato@blucher.com.br  
www.blucher.com.br

Segundo Novo Acordo Ortográfico, conforme 5. ed.  
do Vocabulário Ortográfico da Língua Portuguesa,  
Academia Brasileira de Letras, março de 2009.

É proibida a reprodução total ou parcial por quaisquer  
meios, sem autorização escrita da Editora.

---

Todos os direitos reservados pela Editora  
Edgard Blücher Ltda.

Dados Internacionais de Catalogação na Publicação (CIP)  
Angélica Ilacqua CRB-8/7057

---

Desafios para pesquisa em sociolinguística /  
organizado por Raquel Meister Ko. Freitag, Silvana Silva  
de Farias Araújo, Valter de Carvalho Dias. - São Paulo :  
Blucher, 2022.

120 p.

Bibliografia

ISBN 978-65-5550-153-7 (impresso)

ISBN 978-65-5550-154-4 (eletrônico)

Open Access

1. Sociolinguística I. Freitag, Raquel Meister Ko. II.  
Araújo, Silvana Silva de III. Dias, Valter de Carvalho

21-5433

CDD 306.44

---

Índices para catálogo sistemático:  
1. Sociolinguística

---

# APRESENTAÇÃO

Anyone who begins to study language in its social context immediately encounters the classic methodological problem: the means used to gather the data interfere with the data to be gathered. The primary means of obtaining a large body of reliable data on the speech of one person is the individual tape-recorded interview.

(LABOV, 1972)

A entrevista sociolinguística é a forma mais consolidada para a obtenção de dados linguísticos suficientes para empreender análise de natureza linguística a partir do enfoque social. Desse modo, a Sociolinguística, como uma área autônoma dos estudos que objetivam a investigação da linguagem, vem se dedicando a compreender os processos de variação linguística que podem levar ou não à mudança linguística a partir de métodos próprios de coletas de dados e análise baseada em informações quantitativas e descritivas.

Ao longo de pouco mais de um ano, com o advento da pandemia de COVID-19, a realização das entrevistas sociolinguísticas in loco passaram a ser um grande problema no desenvolvimento das pesquisas nessa área. Os sociolinguistas precisaram se reinventar e buscar outras alternativas para continuar desenvolvendo suas pesquisas, dando o retorno acadêmico-científico que a comunidade espera.

Nessa perspectiva, o X Encontro de Sociolinguística realizado no início de dezembro de 2020, um evento já consolidado na área, reuniu pesquisadores e estudantes em torno do tema Desafios do isolamento para a pesquisa em Sociolinguística e interfaces, para debater sobre os problemas encontrados para a coleta de dados de uma forma geral, como também em um contexto especial imposto pela pandemia. Além disso, apresentaram-se também outras estratégias para a obtenção de corpora sociolinguísticos que extrapolam a clássica entrevista, perpassando o uso de textos em ambientes virtuais que se caracterizam por uma aproximação aos contextos de fala, além de ferramentas computacionais e interações virtuais para aproximação dos falantes.

Nesta obra, realizada através do apoio do Edital CAPES/FAPITEC/PROMOB 10/2016, reuniram-se alguns dos trabalhos que foram apresentados ao longo da programação do X Encontro de Sociolinguística. Desejamos que eles sejam inspiração para outras pesquisas que estão em desenvolvimento e que se deparam com os diversos problemas da atualidade relacionados com a coleta de dados, independentemente da crise sanitária imposta ao mundo. Os espaços virtuais e as tecnologias podem e devem ser usados para permitir aos cientistas da linguagem dar as respostas que a comunidade científica almeja.

*Os organizadores*

# SUMÁRIO

<b>1 DESAFIOS DA CONSTITUIÇÃO DE CORPORA LINGÜÍSTICOS .....</b>	<b>11</b>
<i>Silvia Figueiredo Brandão</i>	
1 INTRODUÇÃO.....	11
2 A NOÇÃO DE CORPUS .....	14
3 CONQUISTAS E DESAFIOS .....	14
3.1 QUESTÕES DE ORDEM TEÓRICO-METODOLÓGICA.....	15
3.1.1 NO ÂMBITO DE COMUNIDADES DE FALA.....	15
3.1.2 NO ÂMBITO DE REDES SOCIAIS E COMUNIDADES DE PRÁTICAS.....	22
3.1.3 INFORMATIZAÇÃO, DISPONIBILIZAÇÃO, CENTRALIZAÇÃO E COMPARTILHAMENTO DE CORPORA.....	24
4 BREVES CONSIDERAÇÕES FINAIS .....	29
REFERÊNCIAS .....	30
<b>2 SOCIOLINGÜÍSTICA VIRTUAL Y EL TRATAMIENTO DE LA VARIACIÓN Y CAMBIO LINGÜÍSTICOS DESDE LAS INNOVACIONES METODOLÓGICAS PARA OBTENCIÓN DE DATOS .....</b>	<b>35</b>
<i>Juan Manuel Hernández-Campoy, Belén Zapata-Barrero e Tamara García-Vidal</i>	
1 INTRODUCCIÓN: SOCIOLINGÜÍSTICA E INNOVACIÓN METODOLÓGICA.....	35
2 FUENTES RADIOFÓNICAS .....	36
3 FUENTES DE CORPUS HISTÓRICOS.....	42
4 CONCLUSIÓN.....	46
REFERENCIAS .....	46

**3 PORTVIX: ESTRATÉGIAS DE BANCOS DE DADOS PARA A CONTINUIDADE DA DOCUMENTAÇÃO EM PESQUISAS SOCIOLINGUÍSTICAS ..... 51**

*Leila Maria Tesch e Lilian Coutinho Yacovenco*

1 INTRODUÇÃO.....	51
2 PORTVIX: AMOSTRAS.....	52
2.1 O BANCO DE DADOS DO PORTVIX.....	52
2.2 O BANCO DE DADOS DE SANTA LEOPOLDINA.....	57
2.3 OS BANCOS DE DADOS DE OUTRAS COMUNIDADES DE FALA – CARAVELAS/BA.....	59
2.4 O BANCO DE DADOS DO JORNAL A GAZETA.....	60
2.5 O BANCO DE DADOS DE CARTAS, CARTÕES POSTAIS E JORNAIS ANTIGOS.....	62
2.6 O BANCO DE DADOS DE REVISTAS EM QUADRINHOS.....	64
2.7 O BANCO DE DADOS DE TELEJORNAIS CAPIXABAS.....	64
2.8 PESQUISAS COM OUTRAS ABORDAGENS.....	65
CONSIDERAÇÕES FINAIS.....	67
REFERÊNCIAS.....	67

**4 EFEITOS DAS MÁSCARAS FACIAIS NA INTERAÇÃO E A COMPENSAÇÃO NA FALA ..... 71**

*Raquel Meister Ko, Freitag e Julian Tejada*

1 INTRODUÇÃO.....	71
2 MÁSCARAS FACIAIS E A MUDANÇA NA LÍNGUA.....	72
3 COMPENSAÇÃO DA FALA.....	75
4 PARADIGMAS PARA O ESTUDO DA COMPENSAÇÃO DA FALA.....	77
5 FUTURO DA INTERAÇÃO COM MÁSCARA.....	78
REFERÊNCIAS.....	79

**5 UMA EXPERIÊNCIA DE APLICAÇÃO REMOTA DO DISCOURSE COMPLETION TASK: A ENTOAÇÃO DO PORTUGUÊS EM CONTATO EM TIMOR LESTE..... 83**

*Marco Barone e Davi Borges de Albuquerque*

1 A FONOLOGIA DA ENTOAÇÃO.....	83
--------------------------------	----

<b>2 O CONTATO LINGÜÍSTICO EM TIMOR-LESTE.....</b>	<b>86</b>
2.1 FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA: O MODELO MÉTRICO AUTOSSEGMENTAL, O SISTEMA DE ANOTAÇÃO TOBI E A METODOLOGIA DO DISCOURSE COMPLETION TASK.....	<b>88</b>
<b>3 RESULTADOS E DISCUSSÃO: DECLARATIVA NEUTRA.....</b>	<b>90</b>
<b>4 RESULTADOS E DISCUSSÃO: INTERROGATIVA POLAR NEUTRA.....</b>	<b>92</b>
<b>5 RESULTADOS E DISCUSSÃO: PERGUNTA DISJUNTIVA.....</b>	<b>94</b>
<b>6 RESULTADOS E DISCUSSÃO: PERGUNTAS PARCIAIS.....</b>	<b>95</b>
<b>7 RESULTADOS E DISCUSSÃO: DECLARATIVAS ÓBVIAS E PERGUNTAS POLARES COM SURPRESA.....</b>	<b>97</b>
<b>8 CONCLUSÕES.....</b>	<b>98</b>
<b>REFERÊNCIAS.....</b>	<b>99</b>
<b>6 DESVIOS ORTOGRÁFICOS: A SEMIVOCALIZAÇÃO DA LATERAL /L/ NO CONTEXTO DO ENSINO PRIVADO.....</b>	<b>101</b>
<i>André Pedro da Silva e Yasmin Maria Macedo Torres Galindo</i>	
<b>1 INTRODUÇÃO.....</b>	<b>101</b>
1.1 PANORAMA SOCIOCONSTRUTIVISTA.....	<b>104</b>
1.2 PANORAMA FONÉTICO: A SEMIVOCALIZAÇÃO DA LATERAL /L/.....	<b>106</b>
<b>2 RESULTADOS E ANÁLISES.....</b>	<b>108</b>
2.1 FATORES FONOLÓGICOS.....	<b>108</b>
2.2 FATORES SOCIAIS.....	<b>114</b>
2.2.1 VARIÁVEL ESCOLARIDADE.....	<b>114</b>
2.2.2 CONTEXTO DE APLICAÇÃO.....	<b>115</b>
<b>3 CONSIDERAÇÕES FINAIS.....</b>	<b>116</b>
<b>REFERÊNCIAS.....</b>	<b>118</b>



## DESAFIOS DA CONSTITUIÇÃO DE CORPORA LINGUÍSTICOS

*Silvia Figueiredo Brandão*  
Universidade Federal do Rio de Janeiro/CNPq

### 1 INTRODUÇÃO

Embora já haja significativo número de bancos de dados linguísticos, que têm concorrido para a caracterização das modalidades oral e escrita do Português do Brasil (PB) dos pontos de vista sincrônico e diacrônico, muitos são, ainda, os desafios quando se trata de constituir ou ampliar corpora, como têm constatado todos os que já passaram por essa experiência e como vêm demonstrando pesquisadores que se vinculam à área da Linguística de Corpus.

Deixando à parte, entre outros, os dados recolhidos em obras precursoras, como *O dialeto caipira*, de Amadeu Amaral (1920) e *O falar carioca*, de Antenor Nascentes (1922), destaca-se o projeto do Atlas Prévio dos Falares Baianos (APFB) (ROSSI et al., 1963). Sua realização remonta a meados da década de 1950 – que iniciou, no Brasil, a prática de pesquisa de campo e de organização de corpora com base em coleta rigorosa e metodologia bem definida, numa perspectiva essencialmente diatópica e sem os recursos hoje disponíveis para registro dos dados. Mais tarde, na década de 1970, tem início a formação de corpora de perfil sociolinguístico, embora com motivações e objetivos bem diferentes.

O corpus organizado pelos pesquisadores vinculados ao NURC, surgiu no âmbito de um projeto internacional que tinha por meta observar a fala de indivíduos de nível superior de instrução no intuito de retratar a modalidade oral culta, no caso brasileiro, nas cinco cidades que, à época, tinham mais de um milhão de habitantes: Rio de Janeiro, São Paulo, Recife e Salvador, fundadas no século XVI, e Porto Alegre, no século XVIII. O projeto nasceu, em 1968, da sugestão de Nelson Rossi de estender ao Brasil o *Proyecto de Estudio Coordinado de la Norma Lingüística Culta de las Principales Ciudades de Iberoamérica y de la Península Ibérica*, por serem “tão evidentes e tão relevantes os pontos comuns à problemática do espanhol nas Américas e do português no Brasil” (ROSSI apud <https://nurcrj.letras.ufrj.br/>).

Para a constituição do corpus, que se iniciou em 1972, foram estabelecidos três tipos de gravações: aulas e conferências (Eloquções formais/EF), diálogos informais (Diálogos entre dois locutores/D2) e entrevistas (Diálogos entre locutor e documentador/DID), com indivíduos naturais das cidades-alvo, de ambos os sexos, distribuídos por três faixas etárias (de 25 a 35 anos, de 36 a 55 e 56 anos em diante). Com a implementação do Projeto Gramática do Português falado, em finais da década de 1990, 18 entrevistas de cada cidade (três por célula social) serviriam de base para as análises da equipe que o constituía.

O projeto Competências básicas do Português, coordenado por Miriam Lemle e Anthony Naro foi o ponto de partida do que mais tarde seria denominado de Programa de Estudos sobre o Uso da Língua (PEUL), cujos integrantes muito colaboraram com orientações para a formação de outros corpora, como os do VARSUL e do VAL-PB.

O projeto inseria-se no âmbito do Movimento Brasileiro de Alfabetização (MOBRAL) e tinha por objetivo principal “verificar pontos de diferenciação entre a variedade de língua portuguesa falada por esse grupo social [os alfabetizandos] e as variedades de língua escrita nas quais almejam eles adquirir competência”, de modo a contribuir para uma “gradação adequada do material didático” a eles dirigido (LEMLE; NARO, 1977, p. 2).

Para viabilizar a pesquisa, que contou também com uma amostra de língua escrita (de histórias em quadrinhos, fotonovelas, jornais classes A e B e literatura nacional) foram realizadas entrevistas com 20 alunos do MOBRAL, naturais do Grande Rio, 9 mulheres e 11 homens, 6 deles com mais de 40 anos e 14 com menos de 30. Como havia a intenção de testar o fator estilístico de grau de formalidade, as entrevistas com cada informante eram realizadas em várias etapas: no local das aulas (as 1ª, 2ª e 5ª), na residência de diferentes entrevistadores (as

3<sup>a</sup> e 4<sup>a</sup>), no local das aulas ou na residência do entrevistado, com um trecho de gravação à revelia (a 6<sup>a</sup>), no local das aulas ou na residência do entrevistado, em companhia de algum de seus amigos (a 7<sup>a</sup>). Os informantes tinham liberdade para falarem sobre temas de sua preferência, embora lhes tenha sido apresentada uma lista de tópicos de “possível interesse antropológico”, sugerida por Roberto da Matta (LEMLE; NARO, 1977, p. 5-7).

Ao corpus Censo, como foi denominada a amostra inicial, incorporaram-se outras do português falado, com informantes de ambos os sexos distribuídos por faixa etária e pelos níveis fundamental e médio de escolaridade, e, em virtude de novas questões teóricas que se iam impondo, amostras de língua escrita antiga e contemporânea, e, ainda, de Português de contato.

Com o passar do tempo, tais iniciativas, impulsionadas pela rápida difusão da Sociolinguística de inspiração laboviana, desencadeariam em diversas áreas do país, o interesse pela organização de bancos de dados e permitiriam testar hipóteses sobre a difusão e as possíveis implicações sociais de fenômenos linguísticos variáveis.

Para tratar dos desafios na constituição de corpora, sem esquecer exemplos do muito que se conquistou com base naqueles de que já se dispõe, parte-se de algumas questões presentes em Sardinha (2000), em experiências pessoais e em depoimentos e textos de colegas que já organizaram corpora. Acredita-se que, apesar dos avanços na área dos estudos linguísticos e dos recursos da informática, algumas das dificuldades inerentes à composição e divulgação de bancos de dados sejam uma constante preocupação de sociolinguistas e de dialetólogos, bem como de outros pesquisadores que desenvolvem estudos de base empírica. Tais dificuldades, no que respeita a bancos de dados de perfil sociolinguístico, decorrem, por vezes, não só de questões teórico-metodológicas advindas de novas tendências quer da Dialetologia, quer da Sociolinguística, como aponta Freitag (2015), mas também derivam de questões de ordem prática, tais como, a escolha de ferramentas para disponibilização e compartilhamento de corpora.

Nesse sentido, este texto tem continuidade em duas outras seções – a primeira, dedicada à noção de corpus (item 1), a segunda, às conquistas e desafios relacionados à constituição de bancos de dados (item 2) – a que se somam as considerações finais (item 3).

## 2 A NOÇÃO DE CORPUS

Aluísio e Almeida (2006) procuram estabelecer uma diferença entre a concepção de corpus para linguistas em geral e para pesquisadores que se vinculam à área da Linguística de Corpus. Para caracterizar a primeira delas, citam, entre outros, Ducrot e Todorov, para os quais, corpus seria “um conjunto, tão variado quanto possível, de enunciados efetivamente emitidos por usuários da referida língua, em determinada época” (DUCROT; TODOROV, 2001 apud ALUÍSIO; ALMEIDA, p. 157). Já para aqueles que se vinculam à área da Linguística de Corpus, as amostras devem estar sempre em formato eletrônico.

Como, atualmente, a tendência é viabilizar o uso de ferramentas computacionais para tornar mais fácil a busca de informações em bancos de dados, parte-se da definição de Sanchez, um linguista de corpus (1995 apud SARDINHA, 2000, p. 8-9), que os caracteriza como

um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para descrição e análise (SANCHEZ, 1995 apud SARDINHA, 2000, p. 8-9).

Nesse conceito de corpus, representatividade parece ser a palavra-chave, por estar, segundo Sardinha (2000, p. 342-348), intimamente ligada à sua extensão, quer em virtude de a linguagem ser um sistema probabilístico (pois há traços mais e menos frequentes), quer pelo fato de um corpus ser uma amostra de um todo (uma população linguística, como ele denomina), cuja real dimensão não se conhece. Desse modo, a sua representatividade seria sempre relativa, o que dependeria não só da resposta à pergunta representativa de quê, mas ainda representativa para quem, isto é, os usuários de um corpus teriam o ônus de “demonstrar a representatividade da amostra e de serem cuidadosos em relação à generalização dos seus achados no que toca ao todo (um gênero ou a língua inteira, por exemplo)”. A representatividade do corpus determinaria, segundo Biber, Conrad e Reppen (1998), os tipos de questões que podem ser formuladas e a generalização dos resultados advindos da pesquisa.

## 3 CONQUISTAS E DESAFIOS

Passa-se, agora, a tratar, com ênfase na perspectiva sociolinguística, inicialmente, de desafios de ordem teórico-metodológica, com base na noção de

representatividade e comparabilidade (2.1), para, em seguida, focalizar desafios que dizem respeito a disponibilização, centralização e compartilhamento, nem sempre consensual, de corpora linguísticos (2.2), com base nas primeiras iniciativas de debate sobre esses temas no âmbito da ABRALIN, bem como de questões propostas por Freitag em recente *live* por ela mediada.

### 3.1 Questões de ordem teórico-metodológica

#### 3.1.1 No âmbito de comunidades de fala

Não resta dúvida de que, nos últimos 70 anos, tomando como ponto de partida o APFB, a organização de corpora redundou em inúmeras pesquisas sobre diferentes comunidades de fala. Além dos atlas já publicados e de outros desenvolvidos, como teses e dissertações, sobretudo a partir de 1996, quando se instalou o comitê encarregado de elaborar o ALiB, projetos, como o NURC, o PEUL, o VARSUL, o ALIP, o VALPB, o FALA-NATAL, o SP2010, o APERJ, o COMPARAPORT, entre vários outros, vêm fornecendo valiosas descrições de variedades cultas e populares do PB.

Devem-se destacar, ainda, corpora representativos de variedades não brasileiras: o Corpus Concordância, reúne, além de entrevistas relativas ao PB (realizadas em Copacabana e em Nova Iguaçu, na região Metropolitana do Rio de Janeiro) as gravadas em duas localidades da região Metropolitana de Lisboa (Oeiras e Cacém) bem como, as realizadas por Tjerk Hagemeijer em São Tomé e Príncipe. O Corpus Moçambique-Port, recolhido por Vieira e Pissurno (2016) em Maputo conta com 35 entrevistas a que se juntarão as que estão sendo realizadas remotamente.<sup>1</sup> Por sua vez, o corpus do Projeto Em busca das raízes do português brasileiro compõe-se de entrevistas realizadas em Luanda.

Como observa Brandão (2013, p. 3):

Os conhecimentos já existentes sobre a estrutura e a dinâmica das línguas em diversas perspectivas teóricas bem como as descrições sócio e geolinguísticas de que já se dispõe se, de um lado, podem servir de parâmetros para a realização de novas amostras, de outro, requerem que o pesquisador não só tenha clareza quanto à possibilidade de encontrar bem representadas as estruturas/variáveis que objetiva analisar, mas também que considere as características demográficas, socioeconômicas e culturais da área focalizada. Cabe, portanto, decidir, no caso de um corpus oral, sobre a metodologia de coleta, o número de informantes e os parâmetros que presidirão à sua seleção (BRANDÃO, 2013, p. 3).

---

<sup>1</sup> As amostras relativas ao Português do Brasil, ao Português Europeu e ao Português de Moçambique podem ser acessadas em <https://corporaport.letas.ufrj.br>.

A grande dificuldade em retratar adequadamente a realidade linguística brasileira advém da complexidade e da heterogeneidade sociocultural do país, devida, de um lado, ao maciço contato multilinguístico e multiétnico que se verifica desde a época da colonização, de outro, a um processo de deslocamento de grande parte da população rural para áreas urbanas.

No Censo Geral do Império, de 1872, o primeiro realizado no país, o Brasil contava com 9.930.478 habitantes, dos quais apenas 10,41% nas capitais provinciais e no Município Neutro,<sup>2</sup> sendo que, deste percentual, 48% se concentravam nas cidades do Rio de Janeiro, Salvador e Recife.<sup>3</sup> A transformação do “vasto país rural” nas palavras de Celso Cunha, em vasto país urbano fica mais evidente a partir dos anos 1940 do século XX.

A mudança de perfil foi abrupta em algumas regiões, como no Centro-Oeste (21,5% para 88,8%) e gradativa no Norte e Nordeste. No Sudeste, onde sempre se encontraram as taxas mais altas de urbanização, o aumento, em 70 anos, foi de cerca de 53%, devendo-se, ainda, levar em conta que tais deslocamentos não se deram apenas no âmbito intrarregional, tendo havido, ainda,

[...] deslocamentos inter-regionais, determinados pela busca de melhores oportunidades de trabalho. Na Região Sudeste, em que se localizam as duas maiores regiões metropolitanas do país, a de São Paulo, com 20.935.204 habitantes, e a do Rio de Janeiro, com 11.973.505, é onde a concentração de população urbana atinge o maior índice: 99,3%. Grande parte desse contingente é oriundo de outras regiões do país, sobretudo do Nordeste e de Minas Gerais, o que torna esses espaços extremamente complexos não só do ponto de vista social, mas também linguístico, uma vez que neles se configura um intenso contato interdialetoal (BRANDÃO, 2015, p. 201).

Nas pesquisas tradicionais que visam à caracterização de uma determinada área, os informantes são naturais da localidade pesquisada bem como seus pais, o que, em zonas urbanas, por vezes, torna difícil a seleção de informantes. Na amostra do Atlas Fonético do Entorno da Baía de Guanabara (AFEGB) (LIMA, 2006), que contempla quatro comunidades da região Metropolitana do Rio de Janeiro, foi necessário flexibilizar esse critério tendo em vista que a grande maioria da população das localidades-alvo era natural de outras cidades fluminenses ou de outros pontos do país, sobretudo do Nordeste, de Minas Gerais e do Espírito Santo. Assim, convencionou-se que o informante poderia ser oriundo de outra

---

<sup>2</sup> O Município Neutro, unidade administrativa criada em 1834, corresponde ao atual Município do Rio de Janeiro, que passou a ser denominado de Distrito Federal de 1891 até 1960, quando Brasília se tornou a capital do Brasil e, consequentemente, Distrito Federal.

<sup>3</sup> [https://pt.wikipedia.org/wiki/Censo\\_demogr%C3%A1fico\\_do\\_Brasil\\_de\\_1872](https://pt.wikipedia.org/wiki/Censo_demogr%C3%A1fico_do_Brasil_de_1872). Acesso em: 3 nov. 2020.

localidade, desde que tivesse ido morar na comunidade-alvo com até cinco anos de idade, desconsiderando-se a exigência inicial quanto à naturalidade dos pais. Dos 24 informantes, 7 procediam de outras localidades. Quanto às mães e pais dos informantes, respectivamente, 15 (62,5%) e 17 (70,8%) não tinham nascido nas localidades-alvo.

No site do Projeto SP2010, embora os critérios gerais para seleção de informantes tenham levado em conta sexo/gênero, faixa etária, escolaridade, região e zona da cidade, anotaram-se dados sobre a naturalidade dos pais e avós dos entrevistados.

Borttonni-Ricardo (2004) indica como um dos eixos fundamentais para o estudo da variação no PB, o que ela denominou de *continuum* de urbanização que tem como polos variedades rurais isoladas e variedades urbanas padronizadas intermediadas pela zona rurbana.

Entre os corpora que contemplaram comunidades rurais encontram-se os que compõem o projeto Vertentes do Português Popular da Bahia<sup>4</sup> que conta com entrevistas em quatro comunidades afrodescendentes, dois municípios do interior, e, ainda, amostras de fala de quatro bairros populares de Salvador e de um município de sua região Metropolitana. Outro projeto, A Língua Portuguesa no Semiárido Baiano, nas suas duas primeiras fases, voltou-se para a constituição de corpora representativos de comunidades rurais (1996-2001) e, a partir de 2007, a Feira de Santana, havendo, ainda, a intenção de estudar o português falado em áreas indígenas especiais.<sup>5</sup>

Como observa Souza (2009, p. 181), entre os espaços rurais e urbanos, “as fronteiras esmaecem, seus contornos, outrora nítidos, borram-se, tornam-se imprecisos; dilatam-se e esfacelam-se em inúmeras situações intermediárias”, constituindo o que se convencionou denominar de espaços rurbanos, por sua vez, também difíceis de caracterizar.

Espaços rurbanos poderiam ser a periferia das cidades e as favelas, por vezes incrustadas nas chamadas áreas nobres de grandes centros urbanos, como a Rocinha, em São Conrado, Cantagalo-Pavão-Pavãozinho, na fronteira entre Ipanema e Copacabana, no Rio de Janeiro, ou Paraisópolis, na região do Morumbi na capital paulista. Nessas comunidades, que, em geral, abrigam indivíduos vindos de outras áreas do país, há, em maior ou menor grau, a manutenção de seus dialetos de origem e a acomodação ao dialeto local a depender de pertencerem

---

<sup>4</sup> Cf. <http://www.vertentes.ufba.br/>.

<sup>5</sup> Cf. [http://www2.uefs.br/nelp/fases\\_subprojetos.htm](http://www2.uefs.br/nelp/fases_subprojetos.htm).

a redes sociais mais ou menos densas. Além disso, em algumas comunidades há um nítido sentimento de pertença a um grupo específico, o que leva à adoção de determinadas estruturas linguísticas.

Entre os trabalhos nessa linha, há o de Mollica et al. (2008), que discute, tendo como referência a Favela da Maré, no Rio de Janeiro, “aspectos teórico-metodológicos [...] para analisar os processos migratórios no Brasil [...] e suas repercussões, [...] os modelos através dos quais o contato linguístico é analisado mais adequadamente, em se tratando de comunidades rurbanas localizadas nas periferias das grandes cidades brasileiras” (p. 64). Há, ainda, o trabalho de Rodrigues (2004, p. 120), que tratou da concordância verbal em sua tese com base em 40 entrevistas realizadas em favelas da zona Oeste de São Paulo. Ela comenta:

[...] instalam-se as dicotomias rural/urbano e culto/popular quando se considera a realidade social da capital paulistana. Nela se verifica um fenômeno especial de variação sociolinguística resultante desse fenômeno de migração interna: a variedade linguística que utilizam os migrantes em seus estados de origem, deixa de representar, simbolizar sua região; tal variedade, regional na origem, torna-se variedade social, símbolo de uma posição social inferior. Os migrantes vão constituir, com a população dessas cidades e de regiões próximas a elas, pertencentes ao mesmo estrato populacional, um extenso grupo de usuários de uma variedade popular ou não-padrão, estigmatizada, que se torna ela mesma um indicador da classe socioeconômica a que pertencem. É lícito esperar que as novas gerações formadas por filhos de migrantes tendam a abandonar os hábitos linguísticos de seus pais, adotando uma variedade de língua que vai, então, refletir a estratificação social urbana e as atitudes sociais que servem para sustentá-la (RODRIGUES, 2004, p. 120).

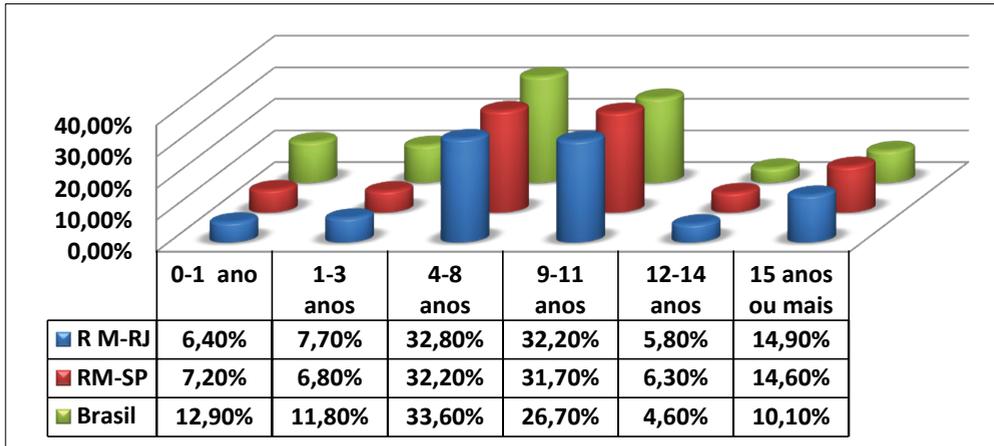
Estão por conhecer também muitas comunidades de médio porte, interiores. No site do Projeto ALIP (Amostra Linguística do Interior Paulista), afirma-se que uma das motivações para a consecução do Corpus IBORUNA foi a necessidade de “representar o dialeto falado no interior paulista, em razão de este ser ainda pouco conhecido, em bases científicas, por seus usuários, e pelos próprios linguistas”.

O *continuum* de urbanização, por sua vez, está associado a nível de escolaridade, parâmetro muito utilizado em estudos sociolinguísticos, em que podem estar embutidas outras variáveis, como o nível socioeconômico, o maior ou menor acesso a bens culturais, maior ou menor contato com indivíduos de diferentes origens geográficas e estratos sociais. Em alguns casos, parece haver sobreposição desse contínuo com o que Brandão (2013a, p. 77) denominou de *continuum* de nível de escolaridade. Em áreas rurais e rurbanas, predominam, via de regra, indivíduos analfabetos ou com baixo índice de escolaridade (0 a 4

ou 5 a 8 anos), enquanto nas grandes cidades há mais probabilidade de se encontrarem indivíduos de nível médio e superior.

Na Figura 1, expõem-se índices percentuais correspondentes a anos de escolarização no Brasil e nas regiões Metropolitanas do Rio de Janeiro (RM-RJ) e de São Paulo (RM-SP).

Figura 1 – Índices percentuais correspondentes a anos de escolarização no Brasil e nas regiões Metropolitanas do Rio de Janeiro (RM-RJ) e de São Paulo (RM-SP).



Fonte: Brandão (2015, p. 203), com base em dados fornecidos pelo IBGE.

Comparando-se as duas regiões metropolitanas, não se observam discrepâncias significativas: os índices percentuais são bastante semelhantes, em todos os níveis. As diferenças se mostram mais claramente quando se comparam os percentuais referentes aos que frequentaram a escola por até 8 anos nas regiões metropolitanas aos do Brasil como um todo. Observe-se, ainda, que os índices correspondentes aos extremos da cadeia de escolarização no Brasil são semelhantes (12,90% que congregam analfabetos e indivíduos que foram à escola por até um ano e 10,10% de indivíduos com nível superior), enquanto nas regiões metropolitanas os percentuais referentes a 15 ou mais anos de escolaridade correspondem ao dobro dos de menor nível de instrução.

Nos grandes bancos de dados, os padrões de distribuição etária dos informantes são variados. Sem mencionar o NURC, cuja repartição por faixas é compatível com o segmento social de falantes com nível superior, no VARSUL, inicialmente, na amostragem urbana, consideraram-se, além de sexo e de três níveis de escolaridade (fundamental I - de 1 a 4 anos, fundamental II - de 5 a 8 anos e nível médio - de 9 a 11), duas faixas etárias: de 25 até 50 e acima de 50 anos. No ALiB, também foram consideradas apenas duas faixas etárias (18 a

30 anos e 50 a 65 anos), opção metodológica que se deveu ao grande número de pontos de inquérito (250) e de informantes, cujo número passaria dos 1.100 considerados para cerca de 1.600, caso se definisse uma faixa etária intermediária.

No projeto FalaPOA, cujo corpus foi organizado entre 2015 e 2019 sob a coordenação de Elisa Batistti, levaram-se em conta indivíduos distribuídos por sexo, três faixas etárias (20-39 anos, 40-59 anos, 60 ou mais anos), três níveis de escolaridade – fundamental, médio, superior –, dois bairros (por renda média mensal em salários mínimos) em cada uma das 4 zonas da cidade (Centro, Norte, Sul, Leste), com base em indicadores sociais e econômicos do site OBSERVA POA (Observatório da Cidade de Porto Alegre).<sup>6</sup>

Sexo/gênero, nível de escolaridade e faixa etária são os parâmetros mais usuais na estratificação dos informantes, embora se venha buscando retratar outras dimensões da variação, diante da já aludida complexidade que caracteriza qualquer comunidade de fala. Dentre elas, estão a variação diafásica ou de registro, contemplada em alguns corpora, como no NURC, com seus três já citados tipos de entrevistas ou no IBORUNA, que, além da Amostra Comunidade (Censo) tem uma amostra de Interação Dialógica, constituída de gravações secretas, na tentativa de fugir ao paradoxo do observador (LABOV, 1972) e chegar ao nível ideal de espontaneidade.

Em alguns bancos de dados de grande porte, por exemplo, as diferentes formas de tratamento são pouco usuais, em função sobretudo do perfil das entrevistas, que, em geral, retratam a interação entre um documentador e um entrevistado. A apreensão das restrições que presidem ao uso de uma ou outra forma, depende não só de diferentes situações intercomunicativas, mas também das relações simétricas/assimétricas que se estabelecem entre os envolvidos na interação, determinadas quer por diferenças relativas à faixa etária, escolaridade e gênero, quer pelos distintos papéis sociais por eles desempenhados. Nesse sentido, seria mais adequado contar com gravações secretas, o que é sempre mais trabalhoso e implica ajustes nos critérios de definição do perfil dos informantes. Para contornar essa lacuna, Célia Regina Lopes, por exemplo, que trabalha com o sistema pronominal, tem recorrido não só a gravações secretas na rua, mas também a filmes.

Novos bancos de dados (e mesmo os antigos, ampliando seus objetivos) deveriam contar com gravações secretas e buscar formas alternativas de gravação de entrevistas por WhatsApp, ou por outros aplicativos, como o Meet ou o Zoom, como têm feito orientados de Silvia Rodrigues Vieira no sentido de alargar o

---

<sup>6</sup> Cf. [www.observapoa.com.br](http://www.observapoa.com.br).

corpus Moçambique-Port. Outro exemplo, é o de Labov, que, em suas recolhas para o *Atlas of North American English* (LABOV; ASH; BOBERG, 2006), contactou os informantes por telefone.

Também a Geolinguística, na perspectiva pluridimensional (RADKE; THUN, 1996) tem buscado integrar à diatopia outras dimensões, entre as quais a diageracional, a diastrática, a diazonal (rural/urbana), a diafásica, a diarreferencial, como indica a proposta de Margotti (2008, p. 2),<sup>7</sup> que, no texto, discute as dificuldades inerentes à apresentação dos resultados que advêm do emprego de métodos da sociolinguística.

Quanto à variação diamésica (variação fala/escrita), o PEUL e o Grupo de Estudos Discurso e Gramática, ambos sediados na UFRJ, dispõem de amostras de fala e escrita. Este último, também atuante na UFF e na UFRN, tem um corpus recolhido em cinco cidades e que obedeceu aos seguintes critérios: cada informante produziu cinco tipos distintos de textos orais e, a partir deles, cinco textos escritos (narrativa de experiência pessoal, narrativa recontada, descrição de local, relato de procedimento, relato de opinião), no intuito de garantir a comparabilidade entre as modalidades falada e escrita. Os informantes, todos alunos, foram distribuídos por níveis de escolaridade/faixa etária, abrangendo desde a alfabetização (de 5 a 8 anos) até o último ano do Ensino Superior – acima de 23 anos.

Outro aspecto relacionado à representatividade retoma a discussão sobre como determinar o número de informantes no corpus como um todo e quantos deles por célula social em análises sociolinguísticas sobre comunidades de fala. Como se sabe, convencionou-se que, em análises variacionistas, o mínimo requerido seriam cinco informantes por célula, o que, se levadas em conta apenas as variáveis clássicas – sexo, faixa etária (3), nível de escolaridade (3) – as análises poderiam requerer, no mínimo, 90 informantes. No entanto, por uma série de motivos, elas, em geral, baseiam-se em dois informantes por célula (por vezes apenas um), o que, apesar disso, tem apresentado resultados consistentes. Os trabalhos realizados com o Corpus Compartilhado do NURC, por exemplo, levam em conta 18 informantes por cidade: 3 homens e 3 mulheres em cada uma das 3 faixas etárias.

Freitag, Martins e Tavares (2012, p. 928), com base no Censo 2010 do IBGE e no índice de 0,5% de representatividade da população estabelecido por Labov em seu estudo sobre Martha's Vineyard, afirmam haver inconsistências nos

---

<sup>7</sup> Disponível em: [http://www.leffa.pro.br/tela4/Textos/Textos/Anais/CELSUL\\_VIII/geolinguistica\\_pluridimensional.pdf](http://www.leffa.pro.br/tela4/Textos/Textos/Anais/CELSUL_VIII/geolinguistica_pluridimensional.pdf).

recortes em geral estabelecidos, uma vez que a distribuição por faixas etárias é diferente de área para área e cidades densamente povoadas são representadas pelo mesmo número de informantes que cidades de pequeno e médio portes. Eles exemplificam, no primeiro caso, com dados de Sergipe, em que o maior contingente da população está na faixa de até 24 anos (46,3%) em contraste com a de 40 a 64 (22,9%) e, no segundo, com o Banco IBORUNA, que estipulou o número de informantes segundo a densidade populacional de cada cidade (GONÇALVES, 2008 apud FREITAG; MARTINS; TAVARES, 2012, p. 930).

Uma comparação, por exemplo, com base no Censo de 2010, entre os estados de Sergipe, com 2.068.017 habitantes, e do Rio de Janeiro, com 15.989.929, demonstra que a situação é ligeiramente diferente nas faixas extremas, mas não no que respeita à faixa entre 25 e 39 anos, em que os percentuais são idênticos (24,5%). Sem dúvida, em cada município desses Estados, a situação pode ser diferente, no que é fundamental observar os indicadores sociais ao se definirem critérios de seleção de informantes.

### 3.1.2 No âmbito de redes sociais e comunidades de práticas

Tendências recentes da Sociolinguística podem também determinar nova dinâmica na constituição de corpora. Freitag, Martins e Tavares (2012) focalizam bancos de dados sociolinguísticos do Português do Brasil, avaliando as suas potencialidades e limitações frente aos estudos da chamada terceira onda. Num quadro (a seguir), eles apresentam, resumidamente, as diferenças entre estudos voltados para a caracterização de comunidades de fala e de comunidades de práticas.

Quadro 1 – Comparação entre abordagens sociolinguísticas de comunidades de fala e de comunidades de práticas.

Abordagem de comunidade de fala	Abordagem de comunidade de práticas
- estratificação baseada em fatores sociodemográficos amplos	- estratificação baseada em valores localmente estabelecidos
- distribuição homogênea, tanto quanto ao tamanho quanto às categorias controladas	- distribuição variável, definida caso a caso
- categorias definidas a priori	- categorias definidas a posteriori
- permissão para captar tendências amplas da comunidade	- permissão para captar valores sociais localmente estabelecidos nas relações
- coleta padronizada (entrevista sociolinguística)	- coleta etnográfica (observação participante, interações entre grupos)
- constituição da amostra em curto prazo	- constituição da amostra em longo prazo

Fonte: Freitag; Martins; Tavares (2012, p. 931).

Como se deduz do Quadro 1, bancos de dados de grande porte, organizados até a década de 2000 (PEUL, VARSUL, NURC, VALPB, entre outros) não servem de base para a análise de comunidades de práticas e de redes de sociais, embora os resultados das análises que Eckert (2012) inclui na primeira onda da Sociolinguística continuem a ser fundamentais para a realização de estudos que se enquadram na segunda e na terceira. Tais estudos, que também operam com dados estatísticos, requerem um trabalho de natureza mais propriamente etnográfica e o convívio do pesquisador com a comunidade, de modo a, detectando como se dá a dinâmica sociolinguística entre os membros de um determinado grupo, determinar como se instaura o significado social da variação.

Há significativas diferenças no que respeita ao perfil do corpus, à metodologia empregada, aos procedimentos analíticos. Para contemplar a nova abordagem em corpora de amplo espectro, os autores sugerem que os bancos de dados incorporem aspectos metodológicos de terceira onda, tais como os que foram implementados no Banco de Fala Culta de Itabaiana - SE, no Banco Falares Sergipanos e no Banco Fala Natal.

O Banco Falares Sergipanos, criado sem “abrir mão da comparabilidade com os bancos de dados já constituídos” (p. 934) “segue duas linhas de coleta – a de estratificação homogeneizada e a de comunidades de prática” (p. 935). Em Freitag (2013, p. 160-162), estão sintetizadas algumas das diretrizes que concorreram para a sua definição: (a) a pesquisa abrange 6 cidades com 40 entrevistas por cidade, prevendo-se 18 comunidades de práticas (religiosas, recreativas e escolares); (b) a seleção de informantes (2 por célula, a princípio) segue o método “bola de neve”, isto é, os indivíduos indicam novos participantes da sua rede de amigos e conhecidos para participarem da pesquisa; (c) a estratificação dos informantes segue a maior/menor distribuição percentual da população do Estado de Sergipe por cinco faixas etárias com base no IBGE (até 14 anos; de 15 a 24; de 25 a 39; de 40 a 64; mais de 65 anos) e, de início, não contempla escolaridade; (d) os informantes passam por uma fase de (i) pré-seleção, checagem para levantamento do seu perfil sociocultural e (ii) entrevista sociolinguística com roteiro prévio, prevendo questões/tópicos que induzam tipos textuais quer narrativos, quer argumentativos/explanativos e, ainda, questões sobre a comunidade e sobre avaliação da fala; (e) a amostra de comunidades de práticas, em cada uma das seis cidades, pretende ser representativa de grupos de indivíduos observados em ação nos espaços escolar, de trabalho, religioso, recreativo, por meio de “gravações de longo termo, em intervalo semanal, por um período de seis meses, a fim

de captar nuances de estilo e adequação de papéis sociopessoais dos participantes, por meio de coletas longitudinais”.

Amostras que servem de base para estudos sobre comunidades de práticas em bancos de dados nos moldes do Falares Sergipanos constitui, sem dúvida, um avanço quanto à observação de fenômenos variáveis, podendo, inclusive, esclarecer aspectos analisados em comunidades de fala, como os relacionados à variável sexo/gênero, por exemplo.

### **3.1.3 Informatização, disponibilização, centralização e compartilhamento de corpora**

Além de aspectos mais propriamente metodológicos, há, ainda, a considerar aspectos práticos, que tornam oneroso, em termos quer financeiros, quer cronológicos, a constituição de corpora, que requer, a depender de sua extensão, uma equipe determinada a realizar entrevistas no mais curto prazo e treinada de forma a evitar vieses que ponham em risco os resultados das análises. Nesse sentido, é de fundamental importância que os integrantes das equipes, embora trabalhando de forma complementar, pela inerente especificidade de cada uma das tarefas próprias da organização e informatização de um corpus, participem ativamente de todas as etapas de trabalho ou, quando se trata da utilização de um corpus já constituído, fiquem inteirados dos parâmetros que o fundamentaram. Assim, parece, a cada dia, mais natural compartilhar corpora por conta, seja do interesse em contrastar diferentes dialetos, seja pelo próprio objeto em análise requerer um campo de observação mais amplo.

A discussão sobre informatização, disponibilização, centralização e compartilhamento de corpora vem de longa data. Em 2007, no II Congresso da Associação Internacional de Linguística Portuguesa (AILP), realizado no Rio de Janeiro, foi organizada a mesa redonda *Corpora do Português: formação e políticas de disponibilização*, que contou com a participação de Maria Fernanda Bacelar do Nascimento (Universidade de Lisboa), Paulino Vandresen (UFSC), Izete Coelho (UFSC) e Sílvia Figueiredo Brandão (UFRJ). Brandão, embora tenha focalizado, em especial, os corpora linguísticos existentes, naquele momento, na Faculdade de Letras da UFRJ, tentou esboçar uma cronologia desses debates, a maior parte deles registrada em *Boletins da ABRALIN*. O tema, inclusive, já tinha sido debatido, em 2001, em Lisboa, numa outra mesa redonda<sup>8</sup> também no âmbito da AILP, por ocasião do primeiro congresso dessa associação.

---

<sup>8</sup> Da mesa participaram Fernanda Bacelar do Nascimento (Universidade de Lisboa), Ataliba de Castilho (Universidade de São Paulo) e Perpétua Gonçalves (Universidade Eduardo Mondlane).

Brandão observou que, a partir da década de 1980, algum tempo após a criação e início de desenvolvimento, na década de 1970, de projetos de pesquisa linguística que dependiam da formação de bancos de dados de maior porte, vem, intermitentemente, à discussão a necessidade de se criarem parâmetros e políticas que permitam o compartilhamento de dados e, conseqüentemente, o incentivo à realização, entre outras, de pesquisas de caráter comparativo.

No histórico apresentado e publicado em 2008 (p. 65-67), ela demonstra que o tema vem sendo discutido há, pelo menos, cerca de 36 anos, sem que se tenha chegado a um consenso sobre alguns tópicos. Os debates se deram pela primeira vez, ao que tudo indica, em 1984, na 36ª Reunião da SBPC, na mesa redonda Problemas de Sociolinguística.<sup>9</sup> Organizada pela ABRALIN, girou em torno de três eixos: (a) reflexões de natureza ética, (b) sugestões no sentido de tornar comparáveis os dados recolhidos por diferentes grupos e delimitar temas de interesse de cada um deles; (c) propostas concretas de levantamento de projetos que contassem com amostras, sendo inclusive apresentado por Sebastião Vôtre um esboço de ficha de dados que permitisse traçar o perfil desses corpora.

Os debates seriam retomados quase 10 anos depois: (a) em 1993, em Campinas, no Seminário sobre a Informatização de Acervos de Língua Portuguesa; (b) em março de 1994, na “Oficina de trabalho sobre Programas de Análise e Tratamento de Textos”, em que se fez a demonstração de softwares para tratamento de dados linguísticos<sup>10</sup> (CASTILHO, A.; SILVA, G. M. O.; LUCCHESI, D., 1995, p. 147-148); (c) em julho de 1994, na 46ª Reunião Anual da SBPC, na Universidade Federal do Espírito Santo, no Encontro sobre “Informatização de acervos de língua portuguesa”, em que Ataliba T. de Castilho, Giselle Machline de Oliveira e Silva e Dante Lucchesi, após sintetizarem os principais tópicos da referida mesa-redonda e mencionarem algumas iniciativas no sentido de viabilizar o compartilhamento e informatização das amostras, apresentaram um levantamento, ainda que parcial, dos corpora existentes, feito com base em ficha preparada por Rodolfo Ilari e enviada aos grupos de pesquisa pela ABRALIN (CASTILHO, A.; SILVA, G. M. O.; LUCCHESI, D., 1995, p. 148-152).

Voltando à reunião de 1984, no que parece constituir o primeiro debate público sobre corpora linguísticos, o primeiro eixo de discussão dizia respeito a questões de natureza ética.

---

<sup>9</sup> Participaram da mesa Miriam Lemle (UFRJ), Sebastião Vôtre (UFRJ), Claiz Passos (UFBA) e Fernando Tarallo (UNICAMP).

<sup>10</sup> Entre eles, Notebuilder e Wordcruncher; Microconcord e Wordlist; Folio Views; Varbrul, STABLEX. Tact; The Ethnograph.

Segundo Pereira e Cardoso (2013, p. 72), no Brasil, remonta a 1988 “a primeira regulamentação referente a pesquisas com seres humanos embasada em documentos e discussões internacionais e focada principalmente na questão biomédica e em pesquisas ligadas a saúde”. Segundo as autoras, tal regulamentação teve pouca adesão e a questão só viria a ser retomada em 1995, redundando na Resolução 196/96 do Conselho Nacional de Saúde, que estabelecia diretrizes relativas a pesquisas com seres humanos e criava a Comissão Nacional de Ética em Pesquisa – CONEP. Esse documento “passou por complementações relacionadas a assuntos específicos, ligados a populações indígenas, genética, reprodução humana, dentre outras” (p. 72). Até o final de 2011, os projetos eram enviados a um comitê de ética, vinculado ou não à instituição a que o pesquisador pertencia, e, a partir de janeiro de 2012, encaminhados pela Plataforma Brasil.

Em 1988, ano em que se começa a discutir a questão em nosso país, já se havia iniciado a constituição de bancos de dados de alguns dos grandes projetos que ainda hoje têm continuidade, como o do NURC, em 1973, e a Amostra Censo da Variação Linguística no Rio de Janeiro, de 1976-1977.

Naquela época, sem que houvesse legislação específica sobre procedimentos éticos, as entrevistas eram realizadas com o consentimento oral dos participantes, que tinham seus dados econômico-sociais registrados nas então chamadas fichas dos informantes, cujos dados eram devidamente resguardados, sendo de domínio apenas dos pesquisadores, que atribuíam um código (ou um nome fictício) a cada um deles.

Ao se organizar, por exemplo, na década de 1980, o corpus do Projeto APERJ, continha também elocuições livres que serviram de base a várias análises sociolinguísticas com foco em áreas rurais do Norte e Noroeste fluminenses. O procedimento ético implicava explicar, previamente, que se tinha como objetivo principal recolher a linguagem da pesca e dos pescadores e, ao final da entrevista, perguntar ao informante se queria escutar o seu depoimento, como outra forma de concordância ou não de participação na pesquisa. Era, ainda, praxe, em cada área, visitar as sedes das colônias de pesca e entrar em contato com os líderes comunitários, que acabavam por indicar os pescadores mais representativos, isto é, os indivíduos que se dedicavam precipuamente à pesca.

É fundamental, quando se adentra uma pequena comunidade, ser sensível às normas que regem o convívio entre seus membros e destes com a de forasteiros. Um acontecimento, um detalhe, mesmo involuntário, pode trazer problemas no momento da recolha de dados. Um caso que serve de exemplo foram as entrevistas feitas pela autora deste texto em São Benedito da Lagoa de Cima, no

Município de Campos. Nas duas primeiras vezes que lá foi, tudo correu muito bem, mas, na terceira, nenhum pescador, a princípio, queria dar entrevistas. Só descobriu o motivo, por acaso, quando um dos informantes lhe disse que dois dos entrevistados por ela (inclusive um dos mais jovens e aparentemente muito saudável) tinha morrido de repente, o que a tornou uma pessoa, naquele momento, indesejável. Levou certo tempo, mas readquiriu a confiança do grupo. Dinah Callou, em conversas informais, mencionou que alguns dos potenciais primeiros informantes do NURC, por vezes hesitavam em dar entrevistas, em função do momento político atravessado pelo país nos anos 1970.

Outro dos tópicos de discussão, incluído na rubrica ética no encontro de 1984, dizia respeito à relação entre pesquisadores em diferentes estágios de formação num grupo de pesquisa. É premissa básica que todos os discentes envolvidos na organização de corpora partilhem dos mesmos princípios éticos, sendo fundamental que os coordenadores de projetos promovam atividades com mestrandos, doutorandos ou graduandos de Iniciação Científica de modo a não só lhes transmitir procedimentos metodológicos e éticos, mas também aproveitar suas sugestões. Nos idos da década de 1980, ainda no âmbito do APERJ, os discentes recebiam treinamento para a pesquisa de campo, iam para as áreas de pesquisa, participavam da discussão das normas de transcrição grafemática das entrevistas, normalmente realizada por eles e revista pelos pesquisadores responsáveis pelo projeto.

Hoje, parece já ser consenso não só solicitar ao informante a assinatura de um documento em que explicita sua concordância em participar da pesquisa, mas também, previamente, submeter a uma Comissão de Ética, via Plataforma Brasil, os parâmetros que a nortearão. É essencial, no entanto, que haja em todas as universidades comitês de ética específicos para a área da Linguística.

Certamente, outras questões éticas se impoem no que concerne à centralização e ao compartilhamento de corpora, como as questões formuladas por Raquel Freitag aos participantes do Simpósio Gestão de dados linguísticos.<sup>11</sup> Foram cinco os questionamentos: (a) como atender aos princípios de ciência aberta quanto ao armazenamento, reuso e autoria do conjunto de dados linguísticos? (b) como lidar com a questão entre transparência na ciência e o sigilo dos participantes? (c) quais as ferramentas mais adequadas para a vitalidade dos conjuntos de dados linguísticos? (d) que ferramentas permitem melhor

---

<sup>11</sup> Trata-se do simpósio *Gestão de dados linguísticos*, que teve lugar em 21 de julho de 2020 como parte do evento Abralín ao vivo - *Linguists online*. Disponível em: <https://aovivo.abralin.org/lives/gestao-de-dados-linguisticos/>.

armazenamento e um sistema de interface para consulta e pesquisa? (e) como ficam os grupos minoritários e variedades sub-representadas?

Não há dúvida de que bancos de dados linguísticos, sobretudo os organizados em instituições públicas, devem ser disponibilizados não só para a comunidade científica da área, mas também para a sociedade em geral, pelo seu valor intrínseco e pelo aporte de recursos financeiros que os viabilizam. Além disso, um corpus linguístico pode ser útil a historiadores, antropólogos, sociólogos, historiadores, geógrafos, entre outros.

Em 1993 e em 1994, nos mencionados Seminário e Oficina, já se discutia a necessidade de proceder ao tratamento dos dados por meio dos softwares então existentes. De lá para cá, os recursos de informática avançaram exponencialmente, havendo uma série de opções, algumas, inclusive, aparentemente de fácil manejo, como é o caso dos aplicativos EXMARaLDA e do ELAN, que possibilitam o alinhamento das transcrições com os arquivos de som, permitindo, ainda, que áudio, transcrições e metadados sejam pesquisáveis. Oushiro (2014), inclusive, procura mostrar o caráter amigável do ELAN.

Exemplo bem-sucedido de corpus eletrônico anotado, é o Corpus Histórico do Português Tycho Brahe,<sup>12</sup> idealizado por Charlotte Galves e composto de 76 textos em português escritos por autores nascidos entre 1380 e 1881, 44 deles com anotação morfológica e 27, com anotação sintática.

Alguns dos corpora mais longevos já disponibilizam seus dados online, embora ainda de forma convencional, como é o caso do NURC-RJ e do VAL-PB, este com entrevistas realizadas em 1993, em 2015 e 2018. No entanto, como informa Oliveira Jr. (2012), em 2012, por meio de Chamada Universal do CNPq, obtiveram-se os recursos financeiros que permitiram a implantação do NURC-Recife Digital,<sup>13</sup> com corpus anotado e que serviria de modelo para as demais quatro cidades.

Apesar de todos os esforços, a manutenção de sites é um problema recorrente não só pela falta de recursos financeiros regulares, mas também por não se contar com a assessoria permanente de uma equipe (informatas, transcritores de entrevistas, revisores, anotadores de corpora) para as necessárias atualização e manutenção. Grande parte dos bancos de dados está alocada em servidores de instituições públicas e, portanto, na dependência de condições técnicas muito variáveis. No site CORPORAPORT, que visa a divulgar

---

<sup>12</sup> Cf. <http://www.tycho.iel.unicamp.br/corpus/>.

<sup>13</sup> Cf. <https://fale.ufal.br/projeto/nurcdigital/>.

projetos de pesquisa de suas idealizadoras e disponibilizar os corpora que têm servido de base às suas pesquisas e às de seus orientandos, constantemente enfrentam-se dificuldades nesse sentido.

A centralização, em nível nacional, de bancos de dados linguísticos, sem dúvida, seria um grande passo para a preservação e difusão de um acervo linguístico-cultural de valor inestimável, mas constituiria um megaprojeto, que dependeria, entre outros fatores, da constituição de uma equipe multidisciplinar que indicasse se haveria parâmetros de digitalização e anotação comuns aos diferentes bancos de dados ou se eles seriam incorporados no seu formato original. Isso implicaria novas questões de natureza metodológica, a obtenção de recursos financeiros, decisões sobre a representatividade dos corpora, isto é, sobre que bancos de dados seriam reunidos num repositório central e quais os critérios que presidiriam a seu acesso e sua curadoria.

#### **4 BREVES CONSIDERAÇÕES FINAIS**

Do que aqui foi exposto, deduz-se que, nos últimos 70 anos, com base na metodologia da Sociolinguística Variacionista e da Dialetoлогия, muito já se realizou no sentido de conhecer os fenômenos variáveis que atuam no Português do Brasil.

Há, no entanto, muitos desafios para a ampliação desse conhecimento, entre outros: (a) criar novos bancos de dados formatados segundo a metodologia mais adequada para descrever, por exemplo, comunidades que vivem em relativo isolamento, sejam comunidades quilombolas, indígenas, rurais, periféricas aos grandes centros urbanos, para tanto definindo o que se considera isolamento nos dias atuais (geográfico, cultural, econômico, tecnológico); (b) desenvolver metodologia adequada ao estudo de redes sociais e comunidades de prática; (c) ir realimentando os bancos de dados existentes, por exemplo, com o acréscimo de gravações secretas, com entrevistas por meio de aplicativos; (d) buscar métodos que permitam correlacionar resultados obtidos em estudos sobre comunidades de fala e comunidades de prática; (e) encontrar métodos que permitam a comparabilidade quantitativa e qualitativa dos resultados obtidos nas análises, de modo a chegar a generalizações quanto a determinadas variáveis linguísticas, por meio de recursos estatísticos de meta-análise, como vem sendo proposto por Raquel Freitag, tendo em vista que os corpora em que se baseiam análises variacionistas apresentam diferenças quanto ao perfil dos informantes, ao seu número por célula, às variáveis consideradas; (f) proceder à anotação dos

corpora e divulgá-los na web, primeiro passo para a criação de um repositório central de bancos de dados, acessível a todos os cidadãos.

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.

Johanson (1991, p. 313), em texto intitulado “Times change, and so do corpora”, ao especular sobre o futuro dos corpora, faz a seguinte observação, aqui endossada:

Apesar das grandes mudanças operadas, em menos de três décadas, desde o primeiro corpus concebido para uso em computador, há um aspecto em que o papel do corpus na pesquisa linguística não mudou. O corpus continua sendo uma das ferramentas do linguista, a ser usada junto com a introspecção e técnicas de dedução. Linguistas criteriosos, assim como artífices experientes, afiam suas ferramentas e reconhecem seus usos apropriados.<sup>14</sup> (JOHANSON, 1991, p. 313).

## REFERÊNCIAS

ALMEIDA, Fabiana da Silva Campos. Micro Atlas Fonético do Estado do Rio de Janeiro: uma contribuição para o conhecimento dos falares fluminenses. 2 v. Tese (Doutorado em Letras Vernáculas) - Faculdade de Letras, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2008.

ALUÍSIO, Sandra Maria; ALMEIDA, Gladis Maria de Barcellos. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa linguística. *Calidoscópico*, v. 4, n. 3, p. 156-178, set./dez.2006.

AMARAL, Amadeu. O dialeto caipira. 3. ed. São Paulo: Hucitec; Secretaria de Cultura, Ciência e Tecnologia, 1976 [1920].

BIBER, Douglas; CONRAD, Susan; REPPEN, Randi. *Corpus linguistics – Investigating language structure and use*. Cambridge: University Press, 1998.

BORTONI-RICARDO, Stela Maris. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola Editorial, 2004.

---

<sup>14</sup> “In spite of the great changes in the less than three decades since the first computer corpus, there is one way in which the role of the corpus in linguistic research has not changed. The corpus remains one of the linguist’s tools, to be used together with introspection and elicitation techniques. Wise linguists, like experienced craftsmen, sharpen their tools and recognize their appropriate uses.” (Tradução da autora).

BRANDÃO, Silvia Figueiredo. Corpora linguísticos no Rio de Janeiro In: GONÇALVES, Carlos Alexandre Victório; ALMEIDA, Maria Lúcia Leitão de (org.). *Língua Portuguesa: identidade, difusão e variabilidade*. Rio de Janeiro: AILP/UFRJ, 2008, v. 1, p. 65-73. Disponível em: [http://www.ailp-edu.org/download\\_livro\\_I.htm](http://www.ailp-edu.org/download_livro_I.htm). Acesso em: 10 set. 2020.

BRANDÃO, Silvia. Figueiredo. Réalité sociolinguistique brésilienne et géolinguistique pluridimensionnelle. In: CARRILHO, Ernestina; MAGRO, Catarina; ÁLVAREZ, Xosé. *Current Approaches to Limits and Areas in Dialectology*. Cambridge: Cambridge Schollars Publishing, p. 3-26, 2013.

BRANDÃO, Silvia Figueiredo. Pour une approche géo-sociolinguistique de la réalité linguistique brésilienne. *Géolinguistique*, Grenoble, n. 15, p. 191-214, 2015.

CASTILHO, Ataliba T.; SILVA, Giselle Machline; LUCCHESI, Dante. Informatização de acervos da língua portuguesa. *Boletim da ABRALIN*, v. 17, p. 143-154, jul. 1995.

DUCROT, Oswald; TODOROV, Tzevetan. *Dicionário enciclopédico das ciências da linguagem*. 3. ed. São Paulo: Perspectiva, 2001.

ECKERT, Penelope. Three waves of Variation Study: the emergency of meaning in the study of Variation. *Annual Review of Antropology*, v. 41, p. 87-100, 2012.

FREITAG, Raquel Meister Ko.; MARTINS, Marco Antônio; TAVARES, Maria Alice. Banco de dados sociolinguísticos do português brasileiro e estudos da terceira onda: potencialidades e limitações. *Alfa*, São Paulo, v. 56, n. 3, p. 917-944, 2012.

FREITAG, Raquel Meister Ko. Banco de dados Falares Sergipanos: falares sergipanos database. *Working Papers in Linguística*, Florianópolis, v. 13, n. 2, p. 156-164, abr.-jul., 2013.

FREITAG, Raquel Meister Ko. Desafios teórico-metodológicos da sociolinguística variacionista. In: PARREIRA, M. C.; CAVALARI, S. M. S.; ABREU-TARDELLI, L.; NADIN, O. L.; COSTA, D. S. *Pesquisas em Linguística no século XXI: perspectivas e desafios teóricos-metodológicos*. São Paulo: Cultura Acadêmica, p. 29-43, 2015. Disponível em: <https://www.fclar.unesp.br/Home/Instituicao/Administracao/DivisaoTecnicaAcademica/ApoioaoEnsino/LaboratorioEditorial/serie-trilhas-linguisticas-n27.pdf>. Acesso em: 18 nov. 2020.

JOHANSSON, Stig. Times change, and so do corpora. *In*: AIJMER, Karin; ALTENBERG, Bengt (org.). *English corpus linguistic. Studies in Honour of Jan Svartvik*. New York: Longman, 1991. p. 305-314.

LABOV, William. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LABOV, William; ASH, Sharon; BOBERG, Charles. *Atlas of North American English: phonetics, phonology and sound change*. Berlin: Mouton de Gruyter, 2006.

LEMLE, Miriam. Texto gerador. *Boletim da ABRALIN*, v. 6, p. 5-11, mai.1984.

LEMLE, Miriam; NARO, Anthony Julius. *Competências básicas do português. Relatório final de pesquisa apresentado às instituições patrocinadoras Fundação Movimento Brasileiro de Alfabetização (MOBRAL) e Fundação Ford*. Rio de Janeiro, 1977.

LIMA, Luciana Gomes de. 2 v. *Atlas Fonético do entorno da Baía de Guanabara-AFeBG. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro*. Rio de Janeiro, 2006.

MARGOTTI, Felício. *Geolinguística pluridimensional: desafios metodológicos*. *In*: *Anais do Encontro do CELSUL*, 2008, 9 p. Disponível em: [http://www.leffa.pro.br/tela4/Textos/Textos/Anais/CELSUL\\_VIII/geolinguistica\\_pluridimensional.pdf](http://www.leffa.pro.br/tela4/Textos/Textos/Anais/CELSUL_VIII/geolinguistica_pluridimensional.pdf). Acesso em: 18 out. 2020.

MENDES, Ronald Belini. *A terceira onda da Sociolinguística*. *In*: FIORIN, José Luiz. (org.). *Novos caminhos da Linguística*. São Paulo: Contexto, 2017. p. 103-123.

MOLLICA Maria Cecilia; MELLO, Luciana; LOUREIRO, Fernando; ALÍPIO, Rodrigo. *Comunidades rurbanas e conflitos linguísticos*. Gragoatá, Niterói, n. 25, p. 63-73, 2. sem. 2008.

NASCENTES, Antenor. *O linguajar carioca*. 2. ed. Rio de Janeiro: Simões, 1953.

OUSHIRO, Livia. *Transcrição de entrevistas sociolinguísticas com o ELAN*. *In*: FREITAG, R. M. K. (org.). *Metodologia de coleta em manipulação de dados em sociolinguística*. São Paulo: Blucher, 2014. p. 46-50.

PAIVA, Maria da Conceição de; SCHERRE, Maria Marta Pereira. *Retrospectiva sociolinguística: contribuições do PEUL*. D.E.L.T.A., São Paulo-SP, v. 15, ed. especial, p. 203-230, 1999.

PASSOS, Claiz. Reflexões sobre a profissão de linguista. *Boletim da ABRALIN*, v. 6, p. 17-26, mai. 1984.

PEREIRA, Lara Rodrigues; CARDOSO, Jaqueline Henrique. Comitês de ética: regulamentando a história oral? *Tempos históricos*, v. 17, p. 68-82, 2º. sem. 2013.

PEREZ, Aquilino Sanchez. Definicion e historia de los corpus. *In: PEREZ, A. S.; SARMIENTO, R.; CANTOS, Pascual; SIMÓN, J. (org.). CUMBRE. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y análisis. Madrid: SGEL, 1995.*

RADKE, Edgar; THUN, Harald. Radtke Edgar et Thun Harald, Novos caminhos da geolinguística românica. Um balanço. *In: Radke, E.; THUN, H. (org.). Neue Wege der romanischen Geolinguistik: Akten des Symposiums zur Empirischen Dialektologie (Heilderberg/Mainz, 21-24.10.1991). Kiel: Westensee-Verlag, 1996. p. 25-49.*

RODRIGUES, Ângela Cecília de Souza. Concordância verbal: sociolingüística e história do português brasileiro. *Fórum Linguístico, Florianópolis*, v. 4, n. 1, p. 115-145, jul 2004.

ROSSI, Nelson; ISENSÉE, Dinah Maria; FERREIRA, Carlota. *Atlas Prévio dos Falares Baianos. Rio de Janeiro: Instituto Nacional do Livro, 1963.*

SARDINHA, Tony Berber. Linguística de corpus, histórico e problemática. *D.E.L.T.A*, v. 16, n. 2, p. 323-367, São Paulo-SP, 2000.

SCHERRE, Maria Marta Pereira. Breve histórico do Programa de Estudos sobre o Uso da Língua. *In: SILVA, Giselle Macheline de Oliveira; SCHERRE, Maria Marta Pereira (org.). Padrões sociolingüísticos. Rio de Janeiro: Tempo Brasileiro, 1986. p. 27-50.*

SOUZA, Gisela Barcellos. Paisagens rurbanas: a tensão entre práticas rurais e valores urbanos na morfogênese dos espaços públicos de sedes de municípios rurais. Um estudo de caso. *Sociedade & Natureza, Uberlândia-MG*, v. 21, n. 2, p. 181-192, ago. 2009.

VÔTRE, Sebastião. Para uma política de banco de dados. *Boletim da ABRALIN*, v. 6, p. 12-16, mai. 1984.



# SOCIOLINGÜÍSTICA VIRTUAL Y EL TRATAMIENTO DE LA VARIACIÓN Y CAMBIO LINGÜÍSTICOS DESDE LAS INNOVACIONES METODOLÓGICAS PARA OBTENCIÓN DE DATOS

*Juan Manuel Hernández-Campoy  
Belén Zapata-Barrero  
Tamara García-Vidal  
Universidad de Murcia*

## 1 INTRODUCCIÓN: SOCIOLINGÜÍSTICA E INNOVACIÓN METODOLÓGICA

Uno de los principales logros de la Sociolingüística en sus inicios durante la década de 1960 fue demostrar que la variación lingüística no es libre, sino condicionada por factores sociales, situacionales y/o lingüísticos, y que, consiguientemente, el uso lingüístico y el cambio no están sólo condicionados regionalmente sino también social, estilística, y lingüísticamente (véanse Labov 1966, 1972; Trudgill 1974). Así, desde entonces, los variacionistas vienen explorando el funcionamiento del lenguaje con especial atención a: los patrones de comportamiento sociolingüístico, la estratificación social del lenguaje, la estructura de los sistemas lingüísticos, la naturaleza de la variación lingüística, los mecanismos del cambio lingüístico, los patrones de desarrollos y tendencias en sistemas lingüísticos, la predicción de procesos evolutivos en las lenguas, o el diagnóstico de la salud y vitalidad de las variedades lingüísticas (véase Chambers, Trudgill y Schilling-Estes 2002).

Metodológicamente, desde que Georg Wenker inició el primer estudio dialectológico moderno en Alemania en 1876, empleando el método indirecto

postal para la cumplimentación de un cuestionario, y unos años después Jules Gilliéron con su método directo de encuesta en Francia en 1896 (véanse Chambers y Trudgill 1980; Francis 1983), la obtención de datos en el trabajo de campo sociolingüístico se ha ido beneficiando de las innovaciones tecnológicas (véase Hernández-Campoy 2014). De este modo, si Wenker empleó el correo postal a finales del siglo XIX, en el siglo XXI ya se están empleando cuestionarios a través de navegadores World Wide Web como Google e interfaces y aplicaciones de redes sociales como Facebook, WhatsApp, Telegraph, o Twitter, por ejemplo (véase Yus 2021). Si Gilliéron envió a Edmond Edmont a realizar sus encuestas presenciales con método directo y anotación fonética sistemática por la geografía francesa decimonónica más vernácula, William Labov dio un paso más en dicho método etnográfico tras bautizarse con las entrevistas grabadas de los años de 1960 e ideó a principios del siglo XXI la encuesta telefónica para su Proyecto TELSUR (véase Labov, Ash y Boberg 2006).

La Sociolingüística, por tanto, en tanto que paradigma, se encuentra en un proceso continuo de reformulación teórica y redefinición metodológica en consonancia con la evolución de la epistemología, e incluso la filosofía social, así como con el desarrollo de nuevos métodos de trabajo de campo, técnicas de recolección de datos y análisis cualitativos y cuantitativos (Hernández-Campoy 2014, 2016), como demostró Eckert (2012, 2018) con sus generaciones sociolingüísticas, o ‘waves’. El objetivo del presente trabajo es mostrar dos tipos de fuentes de datos lingüísticos con sus respectivos ejemplos de estudios transversales y/o longitudinales virtuales que proporcionan evidencia comparativa de la variación y cambio sociolingüísticos: muestras de los medios de comunicación y muestras procedentes de corpus lingüísticos históricos digitalizados.

## 2 FUENTES RADIOFÓNICAS

El uso de corpus orales electrónicos, como los procedentes de muestras de habla radiofónica, son materiales excelentes para la detección y análisis cuantitativo y cualitativo de la variación estilística (Bell 1982, 1984; Coupland 1985, 2001; Cutillas-Espinosa y Hernández-Campoy 2007; Hernández-Campoy y Cutillas-Espinosa 2010, 2012; entre otros), el cambio lingüístico en curso (Van de Velde, Gerritsen y Van Hout 1996; Van de Velde, Van Hout y Gerritsen 1997; Hernández-Campoy y Jiménez-Cano 2003) o para el estudio del lenguaje de los medios de comunicación (Sclafani 2018; Soukup 2011; o Zapata-Barrero 2020).

Hernández Campoy y Jiménez Cano (2003) destacaron la utilidad del empleo de archivos sonoros radiofónicos como fuente para el estudio tanto en tiempo real como aparente del habla de una comunidad sociolingüística. Así, la combinación de diacronía y sincronía logra detectar un proceso de estandarización lingüística de una comunidad tradicionalmente no estándar. Emplearon archivos sonoros radiofónicos en su estudio sobre el proceso de estandarización del castellano estándar en Murcia en detrimento de la variedad local no estándar. Los datos lingüísticos se obtuvieron de la fonoteca de Radio Murcia (Cadena SER), que permitieron el análisis de distintos programas locales conservados, fundamentalmente informativos y tertulias de entre 1975 hasta 2001. Este período de 26 años se dividió en cinco intervalos éticos (Eckert 1997) como referentes temporales para la medición.

Dado que ese periodo estudiado de 1975 a 2000 obligaba inevitablemente a tratar de llenar el período de la transición política de la dictadura a la democracia en Murcia, las grabaciones disponibles consistían en entrevistas o declaraciones de personajes predominantemente políticos de sexo masculino (Grupo 1), quedando todavía socialmente lejos de la mujer la participación en la toma de decisiones de esta índole. Las entrevistas a personas de la sociedad eran más escasas y excepcionales en los intervalos temporales iniciales (Grupo 2), pudiendo ser ya sistemáticas a partir de la década de 1980. Dada la presencia tan irregular de los mismos informantes políticos en todos los intervalos establecidos, junto con la esporádica presencia de los no políticos, la investigación longitudinal, en tiempo real, se tuvo que realizar aplicando un estudio de tendencias en lugar de un estudio de panel (véase Eckert 1997). De este modo, con el uso de informantes distintos, la atención prestada a la manifestación del cambio en la comunidad de habla murciana a lo largo del tiempo (cambio histórico), en lugar de la de los individuos concretos a lo largo de su vida, obligaba a evitar confundir un cambio demográfico con un cambio real en la población murciana. Todas las variables lingüísticas analizadas eran rasgos prominentes en el habla de Murcia: i) la supresión de la /s/ postvocálica a final de palabra —variable (s); ii) la supresión de la /r/ postvocálica a final de palabra —variable (r); iii) la supresión de la /l/ postvocálica a final de palabra —variable (l); iv) la supresión de la /d/ intervocálica —variable (d); v) la supresión de la /r/ en posición intervocálica en la palabra ‘para’ —variable (para); vi) la asimilación consonántica regresiva; y vii) el trueque de líquidas concretamente. Las Figuras 1-2 visualizan los resultados

de la cuantificación de la presencia de las siete variables en los Grupos 1 y 2 en valores relativos (porcentajes).

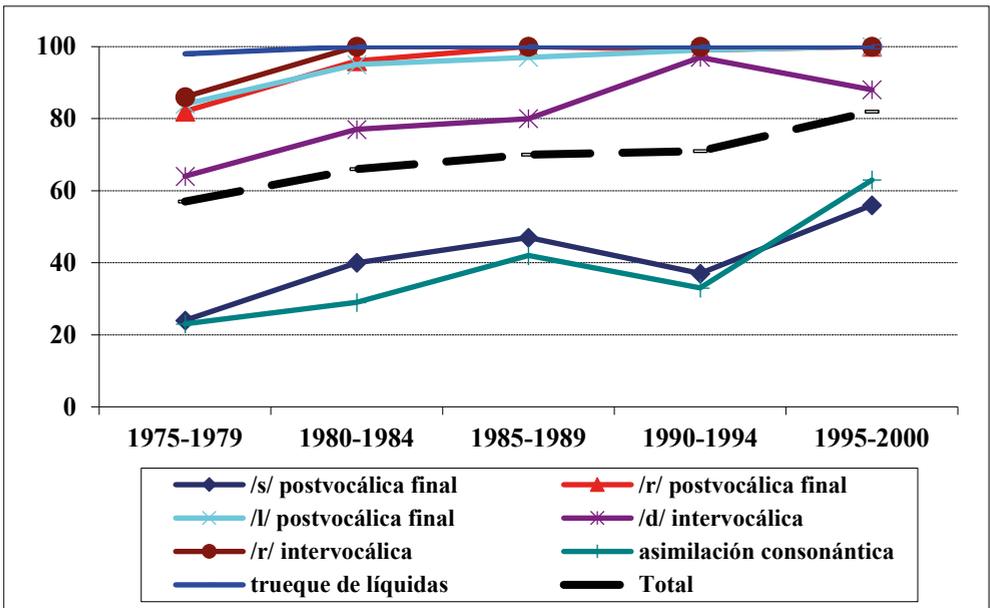
A tenor de las medias de los Grupos 1 y 2, ambos muestran un patrón constante de aproximación al modelo de habla estándar castellano, lo cual se manifiesta porcentualmente en nuestros datos con una disminución progresiva de la presencia de las variantes locales murcianas y un aumento en proporción diametralmente opuesta de las variantes estándares. Esto significa que el habla de la comunidad de hablantes murciana es más parecido al español del norte peninsular (castellano estándar) en el año 2000, con un 75%, que en 1975, con un 56,6% a tenor de estas siete variables. El aumento de la escolarización y la alfabetización —predominantemente normativista— durante esos 26 años, con el correspondiente mayor acceso a los estudios de formación superior, así como la estrecha relación existente entre la pronunciación estándar y la ortografía de su forma escrita, tuvieron una considerable influencia en el desarrollo de este modelo convergente con el estándar.

Atendiendo a las variables, si bien hay una tendencia general a la aproximación al estándar, ésta se manifiesta en distintas proporciones no sólo en el comportamiento sociolingüístico de los informantes sino también, dentro del sistema, en el de las propias variables. Desde una perspectiva diacrónica, las formas estándares son adoptadas por las clases más acomodadas y prestigiosas (clase política) antes que el resto de grupos sociales: variables cuya variante estándar aparece ya totalmente establecida y estabilizada en el Grupo 1 a principios de los años ochenta, por ejemplo, no se comportan como tal en el Grupo 2 hasta mediados o finales de la misma década, como es el caso de la variable (para) y el trueque de líquidas. La forma estándar en las variables (l) y (r) no se estabiliza hasta mediados de la década de los noventa, la presencia de la /d/ intervocálica se encuentra en fase avanzada de estandarización, y ya, a gran distancia, en el umbral del uso preponderantemente vernáculo, todavía quedan la /s/ postvocálica y la asimilación consonántica, que tienden a resistirse.

Desde una perspectiva sincrónica, afecta de forma desigual a los diferentes rasgos lingüísticos que componen y definen al dialecto murciano, iniciándose con aquellos rasgos menos característicos (/r/ postvocálica, /l/ postvocálica, /r/ intervocálica, y trueque de líquidas) y avanzando posteriormente por los más sobresalientes (/s/ postvocálica y la asimilación consonántica). Pero están

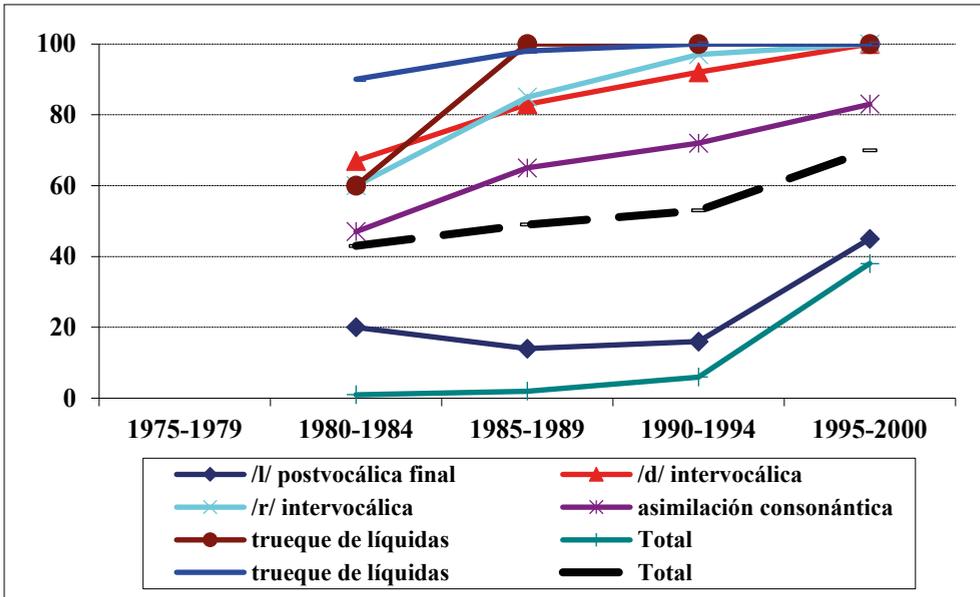
siendo estos últimos, los más prominentes, como la /s/ postvocálica, los que se están resistiendo con más energía a sucumbir a la uniformización por ser rasgos eminentemente sureños tan enraizados en la comunidad de habla murciana que forman parte de sus señas de identidad local frente a los hechos diferenciales de otras geografías peninsulares. La forma estándar de la /d/ intervocálica, por su parte, suele adoptar una posición intermedia-alta en su media de uso. Esto indica que: i) hay variables con un comportamiento similar porque a su vez son evaluadas de forma parecida por los hablantes y pueden encontrarse rigiéndose por algún tipo de covariación lingüística, probablemente estilística; y ii) aunque los diferentes grupos de clases sociales tienen distintos niveles de uso de las variantes, su evaluación de las dos variantes es exactamente la misma —los hablantes de todas las clases tienden a modificar su pronunciación exactamente en la misma dirección, elevando el porcentaje de uso de la forma estándar, perteneciente al estatus social prestigioso, conforme aumenta el nivel de formalidad del contexto estilístico, y viceversa.

Figura 1 – Evolución de las variables lingüísticas en el seguimiento de políticos murcianos.



Fuente: adaptada de Hernández-Campoy y Jiménez-Cano 2003: 3104.

Figura 2 – Evolución de las variables lingüísticas en el seguimiento de no-políticos murcianos.



Fuente: adaptada de Hernández-Campoy y Jiménez-Cano 2003: 315.

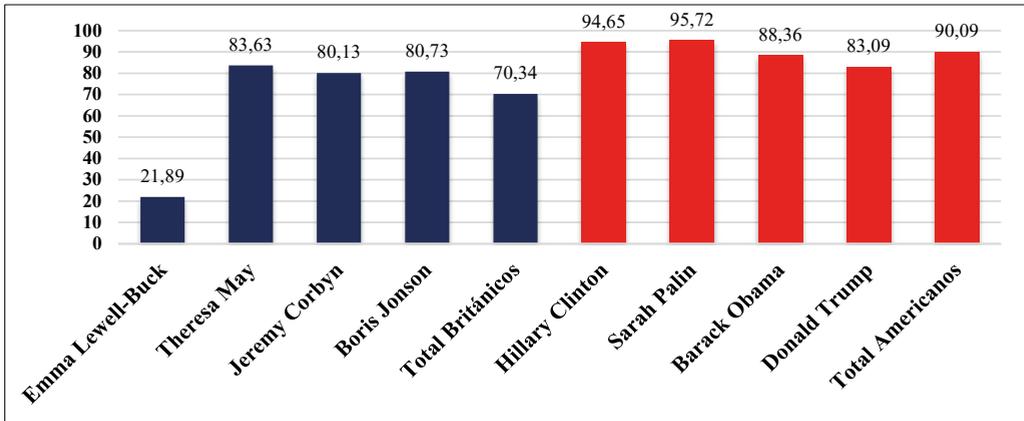
En su reciente estudio, Zapata-Barrero (2020) se sirve de muestras obtenidas en los medios de comunicación digitales (YouTube, CBS, CNN, C-SPAN, CNBC, web del Parlamento británico etc.) para observar el comportamiento sociolingüístico de políticos británicos (Emma Lewell-Buck, Theresa May, Jeremy Corbyn y Boris Johnson) y norteamericanos (Hillary Clinton, Sarah Palin, Barack Obama y Donald Trump) con el fin de identificar posibles diferencias en el uso de estrategias estilísticas en el habla relacionadas con la gestión de identidad cuando estos informantes participan en distintos contextos políticos. Al considerar las desventajas que implicarían la implementación de otras técnicas en la obtención de datos fonológicos para su estudio, Zapata-Barrero (2020) indica que los medios de comunicación online destacan por diversas razones: i) nos permiten eliminar la participación del investigador y su consiguiente efecto en el proceso de obtención de datos (Labov 1972: 209); ii) nos permiten acceder a una gran variedad de contextos políticos; iii) nos facilitan la obtención, así como el manejo de datos; iv) nos brindan la posibilidad de analizar distintas estrategias de estilo que un mismo político pueda utilizar través de los distintos contextos en

los que opera; y v) nos permiten comparar las estrategias de estilo utilizadas por distintos políticos en contextos parecidos.

Así, Zapata-Barrero (2020) analiza los datos lingüísticos obtenidos con métodos cualitativos y cuantitativos, prestando atención a la frecuencia de uso que cada informante hace de las variantes “mainstream” y “no mainstream” de las variables fonológicas objeto de estudio, que a su vez, constituyen rasgos prominentes del Inglés británico (FACE vowel, GOAT vowel, MOUTH vowel, /ʊ/-/ʌ/ Split, Glottalisation of /p, t, k/ y H-Dropping) y norteamericano (PRICE vowel, PIN-PEN merger, Progressive consonant assimilation, R-Dropping, T-Voicing y Yod-Dropping). Del mismo modo, aborda el efecto potencial que algunos factores extralingüísticos pueden tener en el estilo de habla de los políticos, tales como: sistemas sociales dentro de los cuales operan los informantes, región geográfica de procedencia, antecedentes educativos, estatus socioeconómico, género, ocupación y las características socio-contextuales que rodean los contextos políticos analizados. La Figura 3 visualiza la cuantificación de la presencia de las variantes “mainstream” y “no mainstream” de las variables fonológicas mencionadas anteriormente en el habla de los políticos británicos y norteamericanos en valores relativos (porcentajes).

Atendiendo a los principios establecidos por Coupland (1985, 2001, 2007) en su modelo de Diseño de Hablante, y teniendo en cuenta que la mutabilidad indexical de los procesos fonológicos se materializa en prácticas estilísticas (Eckert 2012, 2018), que los significados sociales surgen a través de diferentes contextos de interacción y, por lo tanto, que las ideologías del lenguaje se originan en la experiencia social (Eckert 2008), el estudio de Zapata-Barrero (2020) sugiere que los políticos británicos y norteamericanos se involucran en prácticas estilísticas o procesos de autoconstrucción motivados o influenciados por aspectos identitarios e ideológicos. Es decir, los hablantes pueden reelaborar creativamente los significados de las variables fonológicas a través de sus interacciones sociales. Finalmente, concluye que el estudio de la identidad y los fundamentos ideológicos del cambio de estilo es crucial para una explicación adecuada de cómo los políticos diseñan estratégicamente su estilo de habla a la hora de posicionarse socialmente en contextos comunicativos.

Figura 3 – Porcentajes de uso totales obtenidos por políticos británicos y norteamericanos.



Fuente: adaptada de Zapata-Barrero (2020).

### 3 FUENTES DE CORPUS HISTÓRICOS

La estratificación ontogenética de las variables lingüísticas refleja, a nivel macroscópico, cambios en el habla de la comunidad conforme evoluciona a lo largo del tiempo, o bien, a nivel microscópico, cambios en el habla del individuo a lo largo de su ciclo vital. El tratamiento del tiempo, por tanto, es fundamental en los estudios de variación y cambio lingüísticos, y para poder investigarlos los estudios necesariamente deben ser comparativos: transversales (tiempo aparente) o longitudinales (tiempo real) (véase Sankoff 2006). Sin embargo, las limitaciones de las aproximaciones transversales y las dificultades para realizar adecuadamente las longitudinales al observar el cambio lingüístico han constituido siempre dos preocupaciones esenciales en el tratamiento del tiempo en la sociolingüística variacionista (véanse Labov 1966, 1994; Trudgill 1988; Eckert 1997; Tillery y Bailey 2003; Sankoff 2006; Gertenberg y Voeste 2015; Wagner y Buchstaller 2018; o Beaman y Buchstaller 2021).

Sin embargo, los avances en la tecnología computacional junto con el desarrollo de recursos digitalizados y grandes corpus textuales electrónicos para uso como fuente de datos están transformando la investigación lingüística. La digitalización de las Humanidades particularmente y el desarrollo de corpus escritos históricos están permitiendo sumergirnos en periodos remotos de una lengua y reconstruir con mayor precisión su funcionamiento interno y el comportamiento sociolingüístico de sus hablantes en la interacción social comunicativa de la época con mayor precisión (véanse Romaine 1982; Nevalainen y Raumolin-Brunberg

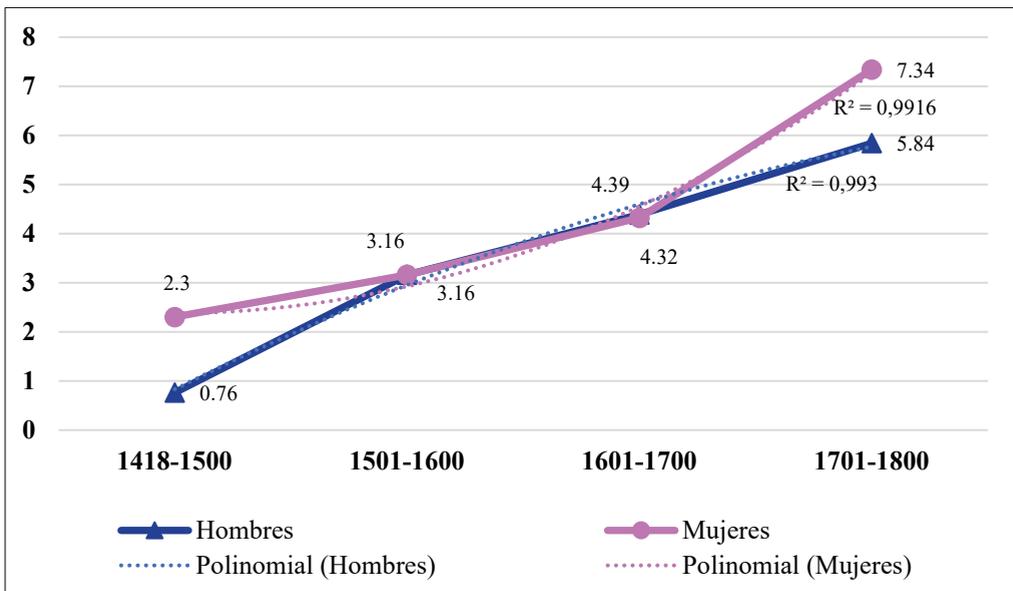
2003; Hernández-Campoy y Conde-Silvestre 2012; or Säily et al. 2017; entre otros). En este sentido, las colecciones de correspondencia privada, implicando a autores de diferente procedencia social y geográfica, nos ofrecen una fuente de datos muy útil para desarrollar análisis sociolingüísticos cuantitativos y cualitativos transversal y longitudinalmente sobre periodos remotos de la historia de una lengua (véanse Nevalainen y Tanskanen, 2007; Dossena y Del Lungo Camiciotti, 2012; van der Wal y Rutten, 2013; o Auer, Schreier y Watts, 2015; entre otros).

En el estudio de García-Vidal (2020) sobre el desarrollo de la formación del comparativo con formas sintéticas tradicionales en inglés y analíticas procedentes del latín, resulta crucial el uso de colecciones de materiales históricos digitalizados en forma de corpus que cuenten con metadatos para facilitar la reconstrucción de los factores extralingüísticos que podrían haber estado relacionados con la variación en el uso de estas formas de comparativo. Para tal fin, las colecciones de correspondencia privada histórica con autores de diferente caracterización socio-demográfica son fundamentales para la reconstrucción del contexto histórico-social que podrían haber afectado a los fenómenos de cambio y variación lingüística en el seno de las comunidades de habla, como hasta ahora se había estudiado desde perspectivas macro. Además, las cartas privadas pueden arrojar luz sobre las motivaciones y mecanismos para el desarrollo y aplicación de la variación intra-hablante y sus opciones estilísticas en los individuos para transmitir significado social en sociedades del pasado remoto, ya desde perspectivas micro.

Atendiendo a factores sociodemográficos como el género, la edad o estatus social y haciendo uso de los corpus de textos históricos *Parsed Corpus of Early English Correspondence (PCEEC)* y del *Corpus of Early English Correspondence Extension (CEECE)*, García-Vidal (2020) concluye que, la forma analítica del adjetivo comparativo en inglés era usada en mayor medida por informantes femeninos desde el siglo XV al XVIII, normalmente en cartas de mujeres pertenecientes a estatus sociales de rango superior, como la realeza, la nobleza o el clérigo (ver Figura 4). Asimismo, el análisis de las formas de comparativo correlacionadas con el estatus social del emisor y receptor señala que la forma analítica era más frecuente en cartas de emisores de rangos sociales más altos y preferible con adjetivos procedentes del latín o francés. Además, el uso de la forma inflexiva, mayoritariamente en relación con adjetivos de procedencia nativa o germana, era más frecuente cuando estos emisores se dirigían a sus familiares o miembros de clases más bajas. De la misma manera, se han encontrado patrones de variación estilística de emisores correspondientes a clases sociales más bajas

(como gente no perteneciente a la burguesía), ya que el uso de la forma analítica es mayor que la sintética cuando se dirigían a miembros de clases sociales más altas. Igualmente, el uso de las formas dobles de comparativo es más frecuente en cartas de emisores pertenecientes a grupos sociales superiores, apareciendo también en alguna carta de miembros no pertenecientes a la burguesía cuando se dirigen a miembros de clases sociales más altas. Además, el análisis de las formas de comparativo correlacionadas con el factor edad refleja un cambio generacional en el uso de la forma analítica a través de los siglos XV al XVIII, siendo más relevante durante el siglo XVIII, cuya forma era liderada por mujeres a principios de siglo para posteriormente ser frecuente en cartas de informantes masculinos. Todo ello apunta hacia el hecho de que la forma analítica podría haber sido introducida en la lengua inglesa a través de grupos sociales más altos, lo cual conlleva a pensar que esta forma gozaba de un carácter mas prestigioso.

Figura 4 – Distribución del adjetivo comparativo analítico en cartas escritas por hombres y mujeres (n.f.).



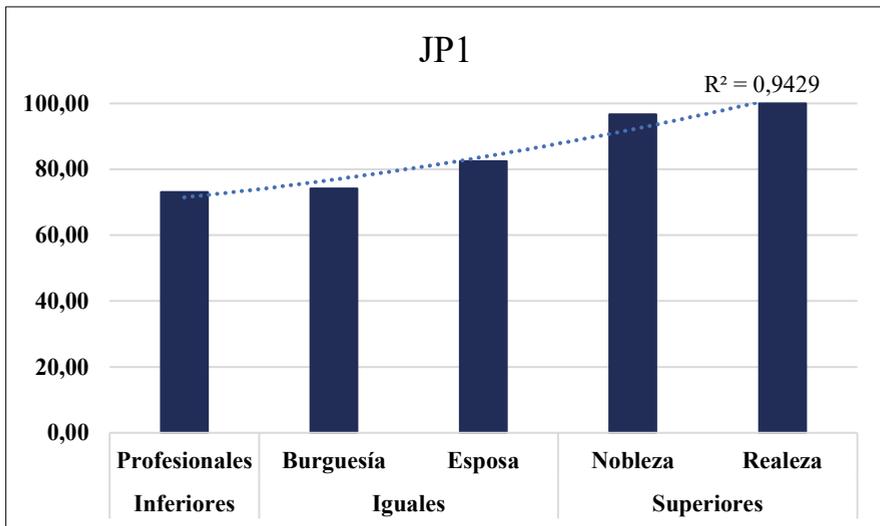
Fuente: adaptada de García-Vidal (2020: 179).

El análisis de un cambio de prestigio en curso en la transición del Inglés Medio al Inglés Moderno Temprano, como es la expansión del diacrítico latino <th> en detrimento del rúnico autóctono <b> en las Paston Letters, permite la observación del uso de esta variable (th) a lo largo del ciclo vital de miembros de la familia de los Pastons, que, en la línea de lo encontrado por Stenroos (2004), Conde-Silvestre

y Hernández-Campoy (2003) diagnosticaron como un cambio laboviano desde arriba. Dicho seguimiento permite seguir su evolución desde las primeras cartas conservadas de 1425 en el patriarca de la familia (William Paston I) hasta las últimas (William IV) a través de tres generaciones. La difusión de esta práctica ortográfica innovadora continental en la Inglaterra medieval tardía se observa como cambio en curso conducente a la configuración de la entonces incipiente norma ortográfica del Inglés Estándar.

Igualmente, haciendo uso del mismo corpus de correspondencia un ejemplo que pone en valor el principio de uniformidad laboviano, es el funcionamiento de la teoría del Diseño de Audiencia de Bell (1984) en los patrones de conducta sociolingüística de los Pastons en sus redes sociales epistolares del medievo tardío, como las analizadas por Hernández-Campoy y García-Vidal (2018). El análisis variacionista de dicho cambio permite la observación del uso de esta variable (th) en conexión con las características sociohistóricas y de estatus de los autores y destinatarios. En el caso concreto de John Paston I, como muestra la Figura 5, acomoda su uso de la variable (TH) según el estatus de sus destinatarios, con mayor presencia de la variante de prestigio continental <th> en los grupos más elevados y menor en los más bajos, en inversamente funcional en la variante autóctona en recesión <þ> a modo de acomodación ascendente, simétrica y descendiente.

Figura 5 – Correlación del uso de la variante de prestigio <th> de la variable (TH) y el estatus de los destinatarios de las cartas de John I.



Fuente: adaptada de Hernández-Campoy y García-Vidal (2018: 398).

## 4 CONCLUSIÓN

Las limitaciones de las aproximaciones transversales y las dificultades para realizar adecuadamente las longitudinales al observar el cambio lingüístico han constituido siempre dos preocupaciones esenciales en el tratamiento del tiempo en la sociolingüística variacionista. Sin embargo, los avances en la tecnología computacional junto con el desarrollo de recursos digitalizados y grandes corpus textuales están transformando la investigación lingüística. Así, se ha demostrado que el uso de corpus orales electrónicos, como los procedentes de muestras de habla radiofónica, son materiales excelentes para la detección y análisis cuantitativo y cualitativo de la variación estilística y el cambio lingüístico en curso. Igualmente, la digitalización de las humanidades y el desarrollo de corpus escritos históricos están proporcionando nuevas oportunidades de investigación en sociolingüística histórica, al permitir sumergirnos en periodos remotos de una lengua y reconstruir con mayor precisión su funcionamiento interno y el comportamiento sociolingüístico de sus hablantes en la interacción social comunicativa de la época. Ambos recursos digitales, por tanto, nos proporcionan una perspectiva privilegiada en la investigación para explorar lenguas actuales o del pasado: metodológicamente, al implicar a informantes de diferentes perfiles socio-demográficos, biológicos, geográficos, y circunstancias personales, son extremadamente útiles para analizar su comportamiento sociolingüístico inter-hablante e intra-hablante a lo largo de periodos prolongados de tiempo y, consiguientemente, para rastrear la naturaleza, evolución, dirección y difusión de una innovación lingüística transversalmente en tiempo aparente –generando variabilidad–, y longitudinalmente en tiempo real a través de grupos generacionales homogéneos de hablantes –fomentando cambio. A su vez, procedimentalmente, actúan como aproximaciones virtuales donde el problema del ‘tiempo real’ en su ejecución propiamente longitudinal se neutraliza convenientemente en términos prácticos. En tiempos de pandemia por contingencia COVID-19, como la situación actual, también resultan un recurso muy socorrido de fuente de datos lingüísticos cuando las restricciones sociales impiden realizar trabajos de campo por razones sanitarias.

## REFERENCIAS

AUER, Anita; SCHREIER, Daniel; WATTS, Richard J. (Eds.). *Letter Writing and Language Change*. Cambridge: Cambridge University Press, 2015.

- BEAMAN, Karen; BUCHSTALLER, Isabelle. (Eds.). *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies*. Londres: Routledge, 2021.
- BELL, Allan. Radio: The Style of News Language. *Journal of Communication* 32, p. 150-164, 1982.
- BELL, Allan. Language style as audience design. *Language in Society* 13, p. 145-204, 1984.
- CHAMBERS, Jack; TRUDGILL, Peter. *Dialectology*. Cambridge: Cambridge University Press, 1980.
- CHAMBERS, Jack; TRUDGILL, Peter; SCHILLING-ESTES, Natalie. (Eds.). *The Handbook of Language Variation and Change*. Oxford. Blackwell, 2002.
- CONDE-SILVESTRE, Juan Camilo; HERNÁNDEZ-CAMPOY, Juan Manuel. Tracing the generational progress of language change in fifteenth century English: the diffusion of <th> in the Paston Letters. *Neuphilologische Mitteilungen* 114(3), p. 279-299, 2013.
- COUPLAND, Nikolas. “Hark, Hark, the Lark”: Social Motivations for Phonological Style-shifting. *Language and Communication* 5(3), p. 153-171, 1985.
- COUPLAND, Nikolas. Dialect stylization in radio talk. *Language in Society* 30(3), p. 345-375, 2001.
- COUPLAND, Nikolas. *Style: Language Variation, and Identity*. Cambridge: Cambridge University Press, 2007.
- CUTILLAS-ESPINOSA, Juan Antonio; HERNÁNDEZ-CAMPOY, Juan M. Script design in the media: Radio talk norms behind a professional voice. *Language & Communication* 27(2), p. 127-152, 2007.
- DOSSENA, Marina; DEL LUNGO CAMICIOTTI, Gabriella. (Eds.). *Letter Writing in Late Modern Europe*. Amsterdam: John Benjamins, 2012.
- ECKERT, Penelope. Age as a sociolinguistic variable. En: COULMAS, Florian. (Ed.). *The Handbook of Sociolinguistics*. Oxford: Blackwell, 1997, p. 151-167.
- ECKERT, Penelope. Variation and the indexical field. *Journal of Sociolinguistics* 12(4), p. 453-476, 2008.
- ECKERT, Penelope. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology* 41, p. 87-100, 2012.

ECKERT, Penelope. *Meaning and Linguistic Variation: The Third Wave in Sociolinguistics*. Cambridge: Cambridge University Press, 2018.

FRANCIS, Winthrop Nelson. *Dialectology: An Introduction* New York: Longman, 1983.

GARCÍA-VIDAL, Tamara. *A Historical Sociolinguistic Approach to the Development of Adjective Comparison in English: Synthetic and Analytic Patterns from 1418 to 1800*. Murcia: Universidad de Murcia (Tesis Doctoral), 2020.

GERTENBERG; Annette; VOESTE, Anja. (Eds.). *Language Development: The Lifespan Perspective*. Amsterdam: Benjamins, 2015.

HERNÁNDEZ-CAMPOY, Juan M. *Research Methods in Sociolinguistics*. *AILA Review* 27, p. 5-29, 2014.

HERNÁNDEZ-CAMPOY, Juan M. *Sociolinguistic Styles*. Malden: Wiley-Blackwell, 2016.

HERNÁNDEZ-CAMPOY, Juan M. *Corpus-based Individual Lifespan Change in Late Middle English*. En: BEAMAN, Karen; BUCHSTALLER, Isabelle. (Eds.). *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies*. Londres: Routledge, 2021, p. 164-182.

HERNÁNDEZ-CAMPOY, Juan M.; CONDE-SILVESTRE, Juan Camilo. (Eds.). *The Handbook of Historical Sociolinguistics*. Malden: Wiley-Blackwell, 2012.

HERNÁNDEZ-CAMPOY, Juan M.; CUTILLAS-ESPINOSA, Juan Antonio. *Speaker Design Practices in Political Discourse: A Case Study*. *Language and Communication* 30, p. 297-309, 2010.

HERNÁNDEZ-CAMPOY, Juan M.; CUTILLAS-ESPINOSA, Juan Antonio. (Eds.). *Style-Shifting in Public: New Perspectives on Stylistic Variation*. Amsterdam: John Benjamins, 2012.

HERNÁNDEZ-CAMPOY, Juan M.; GARCÍA-VIDAL, Tamara. *Style-shifting and accommodative competence in late Middle English written correspondence: Putting Audience Design to the test of time*. *Folia Linguistica Historica* 39, p. 383-420, 2018.

HERNÁNDEZ-CAMPOY, Juan M.; JIMÉNEZ-CANO, José M. *Broadcasting Standardisation: An Analysis of the Linguistic Normalisation Process in Murcia*. *Journal of Sociolinguistics* 7(3), p. 321-347, 2003.

LABOV, William. (1966). *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics, 1966.

LABOV, William. *Sociolinguistic Patterns*. Filadelfia: University of Pennsylvania, 1972.

LABOV, William. *Principles of Linguistic Change. Volume I: Internal Factors*. Oxford: Blackwell, 1994.

LABOV, William, ASH, Sharon; BOBERG, Charles. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlín: Mouton de Gruyter, 2006.

NEVALAINEN, Terttu; RAUMOLIN-BRUNBERG, Helena. *Historical Sociolinguistics. Language Change in Tudor and Stuart England*. Londres: Routledge, 2003.

NEVALAINEN, Terttu; TANSKANEN, Sanna-Kaisa. (Eds.). *Letter Writing*. Amsterdam; Philadelphia: John Benjamins, 2007.

ROMAINE, Suzanne. *Socio-Historical Linguistics: Its Status and Methodology*. Cambridge: Cambridge University Press, 1982.

RUSSI, Cinzia. (Eed.). *Current Trends in Historical Sociolinguistics*. Berlín: Open De Gruyter, 2016.

SÄILY, Tanja; NURMI, Arja; PALANDER-COLLIN, Minna; AUER, Anita. (Eds.). *Exploring Future Paths for Historical Sociolinguistics*. Amsterdam: John Benjamins, 2017.

SANKOFF, Gillian. Cross-sectional and longitudinal studies. En: AMMON, Ulrich; DITTMAR, Norbert; MATTHEIER, Klaus J.; TRUDGILL, Peter. (Eds.). *Sociolinguistics: An International Handbook of the Science of Language and Society (Vol. 2)*. Berlín: Walter de Gruyter, 2006, p. 1003-1013.

SCLAFANI, Jennifer. *Talking Donald Trump: A Sociolinguistic Study of Style, Metadiscourse, and Political Identity*. Londres: Routledge, 2018.

SOUKUP, Barbara. Austrian listeners' perceptions of standard-dialect style-shifting: An empirical approach. *Journal of Sociolinguistics* 15(3), p. 347-365, 2011.

STENROOS, Merja. Regional dialects and spelling conventions in late Middle English. Searches for (th) in LALME data. En: DOSSENA, Marina; LASS, Roger. (Eds.). *Methods and Data in English Historical Dialectology*. Berna: Peter Lang, 2004, 257-285.

TILLERY, Jan; BAILEY, Guy. Approaches to real time in dialectology and sociolinguistics. *World Englishes* 22, p. 351-365, 2003.

TRUDGILL, Peter. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press, 1974.

TRUDGILL, Peter. Norwich revisited: Recent linguistic changes in an English urban dialect. *English World-Wide* 9, p. 33-49, 1988.

VAN DE VELDE, Hans; GERRITSEN, Merinel; VAN HOUT, Roeland. The Devoicing of Fricatives in Standard Dutch: A Real-Time Study Based on Radio Recordings. *Language Variation and Change* 8, p. 149-175, 1996.

VAN DE VELDE, Hans; VAN HOUT, Roeland; GERRITSEN, Merinel. Watching Dutch change: A real time study of variation and change in standard Dutch pronunciation. *Journal of Sociolinguistics* 1(3), p. 361-391, 1997.

VAN DER WAL, Marijke; RUTTEN, Gijsbert. (Eds.). *Touching the Past: studies in the Historical Sociolinguistics of Ego-documents*. Amsterdam: John Benjamins, 2013.

WAGNER, Suzanne Evans; BUCHSTALLER, Isabelle. (Eds.). *Panel Studies of Variation and Change*. Nueva York: Routledge, 2018.

YUS, Francisco. *Smartphone Communication: Interactions in the App Ecosystem*. Londres: Routledge, 2021.

ZAPATA-BARRERO, Belén. *Creating Political identities and Reflecting Social Values: Strategic Style-Shifting in Political Discourse in the USA and the UK*. Murcia: Universidad de Murcia (Tesis Doctoral), 2020.

# PORTVIX

## ESTRATÉGIAS DE BANCOS DE DADOS PARA A CONTINUIDADE DA DOCUMENTAÇÃO EM PESQUISAS SOCIOLINGUÍSTICAS

*Leila Maria Tesch*  
*Lilian Coutinho Yacovenco*  
Universidade Federal do Espírito Santo

### 1 INTRODUÇÃO

O PortVix (Português falado na cidade de Vitória), em duas décadas de atuação, tem contribuído significativamente para a descrição linguística da comunidade de fala capixaba e também para a descrição do português brasileiro. Inicialmente coordenado por Lilian Coutinho Yacovenco, hoje conta com outras duas coordenadoras: Maria Marta Pereira Scherre, que inicia sua participação no grupo em 2008, e Leila Maria Tesch, que participou da fundação do grupo como pesquisadora de Iniciação Científica e desde 2013 se torna uma de suas coordenadoras.

O primeiro banco de dados, denominado PortVix, com 46 entrevistas de representantes da comunidade de fala de Vitória, capital do Espírito Santo, elaborado nos primeiros anos da década de 2000, buscou atingir as seguintes metas: 1) organizar uma amostra da fala da cidade de Vitória, uma das mais antigas do Brasil, que não possuía, até então, um banco de dados sociolinguísticos; 2) analisar fenômenos variáveis da variedade capixaba. Essa variedade não é reconhecida

pelos brasileiros e o próprio capixaba afirma que sua fala não apresenta marcas características, considerando-a, portanto, não marcada, diferentemente do que se considera das variedades vizinhas – a baiana, a mineira e a carioca.

Ao longo desses 20 anos, outras amostras foram elaboradas e analisadas. O objetivo deste capítulo é apresentar as amostras elaboradas pelo grupo de pesquisa PortVix, detalhando suas características, seus objetivos e os trabalhos até o momento realizados. Para tanto, inicialmente comentamos as amostras de fala do PortVix de Santa Leopoldina e de Caravelas/BA, seguidas das baseadas na escrita, do jornal *A Gazeta*, de cartas pessoais e cartões postais e de jornais antigos, além de revistas em quadrinhos. Outras amostras a serem introduzidas serão as de telejornais capixabas e de outras abordagens da Sociolinguística, voltadas para a variação estilística e estudos de percepção linguística.

## **2 PORTVIX: AMOSTRAS**

Durante os quase 20 anos de desenvolvimento de pesquisas do PortVix, muitos estudos foram realizados com base no banco de dados que recebe o mesmo nome – e o pioneiro no estado do Espírito Santo (cf. seção 2.1). Novos integrantes do grupo, instigados por diversos questionamentos sociolinguísticos, se propuseram a organizar outros bancos de dados sob outras perspectivas. Neste capítulo, apresentamos esses diversos bancos de dados analisados ao longo de quase duas décadas de existência do grupo do PortVix.

### **2.1 O banco de dados do PortVix**

No início dos anos 2000, sob a coordenação da professora Lilian Coutinho Yacovenco, com o enfoque da Sociolinguística Variacionista, nasce o projeto PortVix, cujo objetivo maior é a descrição e análise da fala de pessoas nascidas e residentes na capital do Espírito Santo – Vitória. Nesse momento, em parceria com as professoras Maria da Conceição Paiva e Christina Abreu Gomes, do Programa de Estudos sobre o Uso da Língua (PEUL/RJ), que ministraram curso sobre formação de bancos de dados sociolinguísticos, um grupo de alunos junto à professora Lilian deram início aos trabalhos para a constituição desse corpus, dedicando-se à organização das características do banco de dados e aos procedimentos de coleta.

Durante esse período, o grupo elaborou e aprimorou roteiros a serem explorados nas entrevistas, métodos de abordagens aos informantes, verificação das variáveis extralinguísticas mais relevantes para a caracterização da fala capixaba, observação de possíveis problemas técnicos e metodológicos que poderiam ocorrer durante as gravações e elaboração de critérios para as transcrições das entrevistas.

Entre os anos 2001 e 2003, o grupo gravou 46 entrevistas com falantes que nasceram ou se mudaram para Vitória até os quatro anos de idade, de ambos os sexos, de três níveis de escolaridade – Ensino Fundamental, Médio e Universitário – e de quatro faixas etárias – 07 a 14 anos, 15 a 25 anos, 26 a 49 anos e 50 anos ou mais, conforme apresentado no Quadro 1:

Quadro 1 – Estratificação dos falantes da amostra PortVix.

Faixa etária	07-14		15-25		26-49		50-...		
	H	M	H	M	H	M	H	M	
Ensino Fundamental	4	4	2	2	2	2	2	2	20
Ensino Médio	-	-	3	3	2	2	2	2	14
Ensino Universitário	-	-	2	2	2	2	2	2	12
Número total de entrevistados									46

Fonte: Yacovenco (2002, p. 108) e Yacovenco et al. (2012, p. 777).

Os 46 falantes foram gravados em dois contatos, sempre por dois entrevistadores. No primeiro contato, buscava-se verificar se o participante atendia aos critérios estabelecidos e quais temas favoreciam o uso do vernáculo<sup>1</sup> pelo falante. No segundo contato, realizava-se a entrevista propriamente dita, cuja duração média é 60 minutos. Nesse momento, realizou-se uma entrevista semidirigida, tipicamente laboviana, e os entrevistadores adotaram procedimentos diversos para que o vernáculo emergisse e, assim, o falante prestasse o mínimo de atenção ao modo de falar e ficasse mais atento sobre o que dizer.

Algumas estratégias foram adotadas para minimizar o problema do paradoxo do observador<sup>2</sup> e favorecer o uso do vernáculo durante a entrevista, como: i)

<sup>1</sup> “Estilo em que se presta o mínimo de atenção ao monitoramento da fala” (LABOV, 2008, p. 244).

<sup>2</sup> Segundo Labov (2008 [1972], p. 244), ao abordar a questão do paradoxo do observador, discute que “o objetivo da pesquisa linguística na comunidade deve ser descobrir como as pessoas falam quando não estão sendo sistematicamente observadas, no entanto, só podemos obter tais dados por meio de observações sistemáticas. O problema evidentemente, não é insolúvel: ou achamos maneiras de suplementar as entrevistas formais com outros dados, ou mudamos a estrutura da situação de entrevista de um jeito ou de outro. (...) Uma maneira de superar o paradoxo é romper os constrangimentos da

não revelar, de imediato, a real motivação das entrevistas, mencionando que o objetivo era traçar o perfil social do capixaba; ii) conduzir a entrevista de forma descontraída, com base em um roteiro anteriormente elaborado, com perguntas claras e precisas; iii) instigar os participantes a falarem de fatos de suas vidas; iv) uso de gravadores pequenos (nesse período foram utilizados gravadores de fitas cassetes, por isso as entrevistas são de 60 minutos, duração máxima da fita) e microfone de lapela no entrevistado; v) realização da gravação no dia, local e hora de conveniência do participante (na maior parte das vezes na residência do participante); e vi) todos os entrevistadores eram alunos de graduação da Ufes e se apresentavam como tal, buscando uma maior empatia com o entrevistado.

Após as gravações, essas entrevistas foram transcritas pelo grupo e estão à disposição de pesquisadores que queiram realizar estudos com base nesse banco de dados. Atualmente, estamos em fase de organização desse material para divulgação em um site.

Conforme mencionado anteriormente, a composição do banco de dados de fala de Vitória tinha por objetivo o conhecimento da variedade capixaba. A intenção do grupo era buscar a identidade linguística do capixaba por meio de fenômenos que, por um lado, o individualizassem e, por outro, o aproximassem do português usado por todos os brasileiros. Para isso, diversas pesquisas foram realizadas e, até a presente data, foram defendidas 19 dissertações e 1 tese que objetivavam descrever e caracterizar a variedade capixaba, com base nas gravações do PortVix. Destacamos que, além desses trabalhos, foram realizadas outras pesquisas de iniciação científica e trabalhos de conclusão de curso que não serão aqui apresentadas, pois, em sua maioria, tiveram continuidade após a graduação, estando, conseqüentemente, aqui listadas.

Com o objetivo de traçar um panorama dos trabalhos realizados pelo grupo de pesquisadores ligados ao PortVix, listamos a seguir os títulos dos trabalhos desenvolvidos e o(a) pesquisador(a) responsável:

- 1) “A variação no âmbito do irrealis entre as formas de futuro do pretérito e pretérito imperfeito do indicativo na fala capixaba” (TESCH, 2007) – Leila Maria Tesch;

---

situação de entrevistas com vários procedimentos que desviem a atenção do falante e permitam que o vernáculo emerja”.

- 2) “A gramaticalização do verbo IR e a variação de formas para expressar o futuro do presente: uma fotografia capixaba” (BRAGANÇA, 2008) – Marcela Langa Lacerda Bragança;
- 3) “Nós/a gente em Vitória: uma análise sociolinguística da fala capixaba” (MENDONÇA, 2010) – Alexandre Kronemberger de Mendonça;
- 4) “Fala, Vitória! – a variação do imperativo na fala de Vitória e sua posição no cenário nacional” (EVANGELISTA, 2010) – Elaine Meireles Evangelista;
- 5) “Ponte da Passagem: você e cê transitando na fala de Vitória (ES)” (CALMON, 2010) – Elba Nusa Calmon;
- 6) “A expressão do tempo futuro no uso capixaba: variação e gramaticalização” (TESCH, 2011) – Leila Maria Tesch;
- 7) “Variação sintática das orações adverbiais finais: similaridades e diferenças entre fala e escrita” (DEOCLÉCIO, 2011) – Carlos Eduardo Deoclésio;
- 8) “A intercambialidade modo-temporal: análise sociolinguística da alternância tempo e modo verbal na fala dos moradores de Vitória/ES” (BARBOSA, 2011) – Astrid Franco Barbosa;
- 9) “A variação morfossintática do artigo definido na capital capixaba” (CAMPOS JÚNIOR, 2011) – Heitor da Silva Campos Júnior;
- 10) “Expansão das perífrases de gerúndio no português brasileiro” (BASÍLIO, 2011) – Jucilene Oliveira Sousa Basílio;
- 11) “A negação no português falado em Vitória/ES” (NASCIMENTO, 2014) – Cristiana Aparecida Reimann do Nascimento;
- 12) “As vogais médias pretônicas na fala de Vitória” (LEITE, 2014) – Melina de Figueiredo Leite;

- 13) “Não o vejo mais em Vitória: a substituição do clítico acusativo de terceira pessoa na fala capixaba” (LAUAR, 2015) – Aline Berbert Tomaz Fonseca Lauar;
- 14) “A Concordância Verbal na Fala de Vitória” (BENFICA, 2016) – Samine de Almeida Benfica;
- 15) “A regência variável do verbo de movimento ir na cidade de Vitória-ES” (CITELI, 2017) – Bárbara Gomes Citéli;
- 16) “A expressão do sujeito pronominal no português falado em Vitória/ES” (GENUINO, 2017) – Wladimir Ricardi Alves Genuino;
- 17) “Análise da concordância nominal na fala de Vitória/ES: o linguístico, o social e o estilístico” (SCARDUA, 2018) – Juliana Rangel Scardua;
- 18) “Tá Mudando? - uma análise Sociofuncionalista sobre a mudança do verbo estar na fala de Vitória/ES” (PINHEIRO, 2019) – Frederico Pitanga Pinheiro;
- 19) “A expressão da obrigação nas construções ter que + infinitivo, dever + infinitivo e precisar + infinitivo: uma análise variacionista” (PINTO, 2020) – Tarsila Machado Pinto;
- 20) “As Construções Relativas na Fala de Vitória/ES: uma Perspectiva Sociolinguística” (SANTOS, 2020) – André Poltronieri Santos.

As análises permitiram que se observasse que, em sua maioria, os fenômenos morfossintáticos analisados na variedade capixaba têm comportamento ao visto em outras pesquisas sobre o português brasileiro. Entretanto, alguns dos fenômenos permitiram que se notassem características da fala capixaba que a individualizam, entre os quais o da variação de segunda pessoa. Calmon (2010) verificou que não há ocorrências na amostra PortVix da variante *tu*, e que *você* é a mais frequente na fala capixaba. Notou, ainda, que *cé*, diferentemente do que ocorre em Belo Horizonte, é menos frequente, e que *ocê* possui raras ocorrências, nenhuma na função de sujeito.

As pesquisas referentes à variação de artigo definido antes de antropônimos e pronomes possessivos (CAMPOS JÚNIOR, 2011), da negação sentencial

(NASCIMENTO, 2014) e das vogais médias pretônicas (LEITE, 2014) mostraram que o falante capixaba possui características que o inserem no mapa sociolinguístico brasileiro numa variedade que se localiza entre a baiana e a carioca, tanto linguística como geograficamente. Campos Júnior mostra que os artigos definidos não são tão frequentes na fala capixaba, havendo, inclusive, ausência de artigos após preposições, como em “casa de Heitor”. Em relação ao uso de artigos definidos antes de antropônimos, a variedade capixaba apresenta índices intermediários entre os do Rio de Janeiro e da Bahia (CAMPOS JÚNIOR, 2011, p. 75). Entretanto, quando diante de pronomes possessivos, há menor frequência de uso de artigos definidos que a encontrada em Recife, capital brasileira investigada cujos usos são os menores entre as cidades estudadas (CAMPOS JÚNIOR, 2011, p. 76).

Nos fenômenos da negação sentencial e das vogais médias pretônicas também se observa o mesmo comportamento. Nascimento (2014) constata maior frequência de uso de dupla negação ou de negação pós-verbal na variedade capixaba que as encontradas no sul do Brasil e em Mariana/MG, porém menor que as obtidas em capitais nordestinas (NASCIMENTO, 2014, p. 70). Leite (2014) também verifica maior frequência de uso de vogais pretônicas médias-baixas na fala capixaba que no Rio de Janeiro, porém índices bem inferiores aos vistos em Salvador (LEITE, 2014, p. 59).

Os estudos baseados no banco de dados sociolinguísticos de Vitória permitiram que se situasse a variedade capixaba no cenário linguístico do Brasil. Pode-se dizer que, assim como pensam seus falantes, parece ser uma variedade menos marcada, sem grandes traços linguísticos que a diferenciam das demais. Entretanto, alguns fenômenos linguísticos podem caracterizar o capixaba, como são os casos dos já especificados.

## 2.2 O banco de dados de Santa Leopoldina

O banco de dados “Português falado na zona rural de Santa Leopoldina”, desenvolvido pelas alunas-pesquisadoras Camila Candeias Foeger e Lays Joel de Oliveira Lopes, sob a coordenação das professoras Lilian Coutinho Yacovenco e Maria Marta Pereira Scherre, foi constituído no período de 2011 a 2013. Atualmente, esse banco de dados integra o banco de dados do projeto PortVix, sendo base para análise de diversos fenômenos linguísticos.

A amostra da zona rural de Santa Leopoldina é composta por 44 entrevistas, de 50 a 60 minutos de gravação, realizadas com moradores da zona rural

do município, estratificados de acordo com sexo (masculino e feminino), faixa etária (07 a 14 anos, 15 a 25 anos, 26 a 49 anos e 50 anos ou mais) e grau de escolaridade (Ensino Fundamental I, Ensino Fundamental II e Ensino Médio), conforme ilustrado no Quadro 2.

Quadro 2 – Estratificação dos falantes da amostra Santa Leopoldina/ES.

Faixa etária	07-14		15-25		26-49		50-...		
Sexo	H	M	H	M	H	M	H	M	
Ensino Fundamental I	3	3	2	1	2	2	2	2	17
Ensino Fundamental II	3	3	2	2	2	2	2	2	18
Ensino Médio	-	-	2	2	2	2	1	0	09
Número total de entrevistados									44

Fonte: Lopes (2020, p. 113).

Segundo Lopes (2014, p. 60),

Para seleção dos informantes, estabelecemos os seguintes critérios: [...] o falante deveria ser natural de Santa Leopoldina e residente da zona rural do município. Além disso, não poderia ter se afastado da região por mais de um terço de sua vida, ter pais e cônjuges leopoldinenses e não falar outra língua, além do português (LOPES, 2014, p. 60).

Lopes (2020, p. 113-114) esclarece que

inicialmente, a composição da amostra visava a distribuição equilibrada de informantes. Almejava-se um quantitativo de 03 informantes por célula na primeira faixa etária – uma vez que, no Brasil, a pretensão é que os alunos concluam o ensino fundamental aos 14 anos – e 02 falantes a partir da segunda faixa etária – em decorrência da inserção dos alunos de ensino médio. Todavia, no período de coleta da amostra, alguns perfis foram mais difíceis de serem encontrados. Por exemplo, na faixa etária de 15-25 anos, dispomos de apenas 01 falante do sexo masculino com ensino fundamental 01. Sendo assim, buscamos equilibrar a amostra inserindo 03 informantes homens da mesma faixa etária com ensino fundamental 02. Outra dificuldade foi encontrar moradores da zona rural com ensino médio em idade superior a 49 anos. Isso porque, na idade escolar desses falantes, as escolas na zona rural do município eram escassas. A informante dessa faixa etária que cursou o ensino médio concluiu os estudos na vida adulta com muita dificuldade, como relata em sua entrevista, sendo a única da família com essa formação (LOPES, 2020, p. 113-114).

Vale destacar que a escolha desse município se deu com base nos dados populacionais divulgados pelo IBGE, no Censo de 2010, em que Santa Leopoldina aparece como o município com maior quantitativo percentual de habitantes moradores da zona rural.

As entrevistas seguiram as mesmas estratégias adotadas pelo PortVix para amenizar o paradoxo do observador e fazer emergir o vernáculo, seguindo as postulações de Labov (2008 [1972], p. 243-244). Atualmente, todas as entrevistas estão transcritas e podem ser utilizadas para análise de fenômenos.

Até o presente momento, foram defendidas duas dissertações e uma tese. Além disso, também foi desenvolvido uma pesquisa de iniciação científica, listadas a seguir:

- 1) “A concordância nominal de número no português falado na zona rural de Santa Leopoldina/ES” (LOPES, 2014) – Lays de Oliveira Joel Lopes;
- 2) “A primeira pessoa do plural no português falado em Santa Leopoldina” (FOEGER, 2014) – Camila Candeias Foeger;
- 3) “Pronomes de segunda pessoa: uma análise variacionista do português falado na região rural de Santa Leopoldina-ES” (DETONI, 2017) – Marliny Carla Detoni Caetano;
- 4) “‘As histórias são boas. As pessoas, maravilhosa’: análise da variação da concordância verbal e nominal na zona rural de Santa Leopoldina/ES” (LOPES, 2020) – Lays de Oliveira Joel Lopes.

Com o banco de dados de Santa Leopoldina, constatou-se que há muitas semelhanças com a comunidade de fala de Vitória, como o observado nas pesquisas sobre concordância nominal e verbal e variação de segunda pessoa. A variação de primeira pessoa do plural é semelhante à vista na cidade de Vitória, contudo não há uma curva de mudança como na capital, mas, sim, de gradação etária, como afirmado por Foeger (FOEGER, 2014, p. 109).

### 2.3 Os bancos de dados de outras comunidades de fala – Caravelas/BA

O banco de dados “Variação Linguística em Caravelas”, desenvolvido por Jares Gomes Lima, sob a coordenação da professora Maria Marta Pereira Scherre, foi constituído entre 2015 e 2019, e compõe-se de uma amostra de falantes nascidos em Caravelas, na Bahia.

Caravelas é uma das cidades baianas da costa do descobrimento do Brasil, sendo assim uma cidade histórica. A base econômica da cidade é a agricultura e a pesca de peixes e mariscos. Conforme o último censo do IBGE, em 2010, possuía 21.414 habitantes.

A amostra, com 36 entrevistados, está estratificada de acordo com a faixa etária dos falantes (14 a 25 anos, 26 a 49 anos, e 50 anos ou mais), de ambos os sexos (homens e mulheres) e de três níveis de escolaridade (Ensino fundamental, Ensino Médio e Universitário), com dois falantes para cada célula, assim distribuídos:

Quadro 3 – Estratificação dos falantes da comunidade de Caravelas/BA.

Faixa etária	15-25		26-49		50-...		
	H	M	H	M	H	M	
Sexo							
Ensino Fundamental	2	2	2	2	2	2	12
Ensino Médio	2	2	2	2	2	2	12
Ensino Universitário	2	2	2	2	2	2	12
Número total de entrevistados							36

Fonte: Baseado em Lima (2017, p. 41).

Jares Lima (2017, p. 38), responsável pela elaboração e realização dessas entrevistas, nascido e vivido na comunidade até sua ida para Vitória para realizar o mestrado no PPGEL (Programa de Pós-Graduação em Linguística) da Ufes, afirma que “a amostra não foi totalmente aleatória, pois para preencher as células sociais com os informantes necessários foram utilizados alguns que já eram conhecidos e que logicamente satisfaziam as exigências para estarem na amostra”.

Até o presente momento, em 2021, foi defendida uma dissertação de mestrado e o mesmo aluno está no doutorado e, em breve, teremos uma tese sobre essa mesma comunidade.

- 1) “O jogo na comunidade de Caravelas/BA: variação de fricativa coronal pós-vocálica” (LIMA, 2017) – Jares Gomes Lima.

Lima aponta para a importância do estudo da comunidade de fala de Caravelas e constata a presença de um traço característico: a realização da fricativa alveo palatal pós-vocálica antes de oclusivas alveolares surdas, como em [‘paʃta].

## 2.4 O banco de dados do jornal A Gazeta

O PortVix, além de amostras de fala, também possui bancos de dados de textos na modalidade escrita. Um material bastante interessante que temos utilizado para análises de alguns fenômenos linguísticos são as edições do jornal A Gazeta, digitalizados e organizados sob a coordenação das professoras Lilian Coutinho Yacovenco e Leila Maria Tesch.

O jornal A Gazeta foi fundado em 1928, sendo sua primeira edição datada de 11 de setembro de 1928. Segundo Lindenberg Neto (2008, p. 7), surgiu com o propósito de ser um jornal contra o “atraso” e a “inércia da mentalidade capixaba”.

Não é mais possível folhear as coleções dos primeiros anos do jornal. Entretanto, é possível realizar o levantamento desse material na Biblioteca Pública Estadual do Espírito Santo, já que todo o material está registrado em microfilme. Embora tenhamos acesso a esse riquíssimo material, vale ressaltar que o registro desses primeiros exemplares em microfilme é de baixa qualidade. Em diversos momentos, não é possível a leitura de trechos de reportagens, às vezes, até páginas inteiras.

Entretanto, tivemos acesso a um material distribuído internamente entre os funcionários da Rede Gazeta – um livro em comemoração aos 80 anos do jornal A Gazeta (LINDENBERG NETO, 2008). O livro está organizado por décadas, de 1920 a 2000, com a imagem das principais páginas do jornal desse período e, assim, pudemos ampliar o número de dados do jornal. Estamos atualmente em fase de digitalização desse material.

Segundo Tesch (2011, p. 69), a importância de se analisarem as edições desse jornal também se deve ao fato de

A Gazeta [ser] considerado o jornal de formação de opinião no Estado do Espírito Santo. É o único jornal capixaba que possui editorial, além de colunistas que são referência para os capixabas em diversas áreas, da economia e política à esportiva e cultural. Outro ponto a se destacar é sua semelhança a jornais como O Globo, do Rio de Janeiro: é dividido em “Cadernos”, como o 2º Caderno, que aborda questões culturais. A Rede Gazeta, responsável pelo jornal, é uma afiliada da Rede Globo, ou seja, apresenta o chamado “padrão Globo de qualidade” (TESCH, 2011, p. 69).

Alguns trabalhos já foram realizados baseados nesse material e em outras edições do jornal A Gazeta, listados a seguir:

- 1) “A expressão do tempo futuro no uso capixaba: variação e gramaticalização” (TESCH, 2011) – Leila Maria Tesch;
- 2) “A variação dos verbos ter e haver existenciais na modalidade escrita capixaba” (MACHADO; TESCH, a sair) – Amanda Henriques Machado e Leila Maria Tesch;
- 3) “A expressão do sujeito pronominal no domínio jornalístico capixaba” (SOUSA, 2018) – Vinícius Afonso Catazano de Sousa;

- 4) “A posição dos clíticos pronominais em textos jornalísticos capixabas: uma investigação do continuum de monitoramento inter e intragenérico” (BENINCÁ, 2019) – Ludimilla Rupf Benincá.

A amostra baseada no jornal A Gazeta permitiu que se observassem, em solo capixaba, fenômenos linguísticos amplamente abordados na literatura sociolinguística. Também permitiu que se constatasse a variação e mudança linguística no decorrer do tempo, conforme visto por Tesch (2011).

## 2.5 O banco de dados de cartas, cartões postais e jornais antigos

O banco de dados de cartas, cartões postais e jornais antigos, desenvolvido pelas alunas-pesquisadoras Caroliny Batista Massariol, Juliana Rangel Scardua e Karina Correa Conceição, sob a coordenação das professoras Lilian Coutinho Yacovenco e Maria Marta Pereira Scherre, foi constituído no período de 2015 a 2018. A amostra de cartas e cartões postais é composta por textos escritos por dois missivistas capixabas, Oswald Cruz Guimarães e Vicente Caetano, que escreveram para destinatárias mulheres.

Para ter acesso a esse material e montar essa amostra, as alunas Caroliny e Juliana recorreram ao Arquivo Público do Estado do Espírito Santo e encontraram cartas e cartões postais de Vicente Caetano em um álbum e todas as missivas estavam em perfeito estado e todas eram manuscritas. As demais missivas foram encontradas por meio da rede social Facebook, como relata Massariol (2018, p. 61-62),

Para montar a amostra, tivemos de buscar cartas pessoais e cartões postais em acervos públicos e privados. O acervo privado foi possível por meio do Facebook, uma rede social. No Facebook, entramos em um grupo destinado a fotos antigas do Espírito Santo e procuramos, através de uma publicação, alguém que tivesse cartas de cunho pessoal escritas por capixabas no início do século XX. Em seguida, uma pessoa comentou nossa publicação, dizendo que tinha algumas cartas. A partir disso, entramos em contato com Mariza Guimarães, neta de Oswald Cruz Guimarães, e falamos sobre o interesse que tínhamos ao pesquisar essas cartas. Mariza Guimarães nos disponibilizou as cartas de seus avós, entretanto não nos deixou usá-las como um todo em publicações. Sob sua autorização, podemos publicar, no máximo, três linhas de conteúdo. Dessa forma, podemos analisar todas as correspondências, mas só tornamos públicas pequenas partes, de até três linhas (MASSARIOL, 2018, p. 61-62).

Vale destacar que esse material é todo manuscrito, tendo havido casos em que não foi possível se entender todas as palavras por conta da caligrafia dos

missivistas. Esse banco de dados foi fotografado, digitalizado e está constituído da seguinte forma:

Quadro 4 – Distribuição das cartas e postais de Oswald Cruz Guimarães e Vicente Caetano por décadas.

Década – Correspondência - missivista	Quantidade de cartas
Década de 1910 –cartas - Oswald	28
Década de 1920 –cartas- Oswald	5
Década de 1950-postais- Vicente	56

Fonte: Massariol (2018, p. 63).

Até o momento, temos as seguintes pesquisas baseadas nesse banco de dados:

- 1) “Expressão do sujeito pronominal em cartas e postais capixabas do séc. XX” (MASSARIOL, 2018) – Caroliny Batista Massariol;
- 2) “A alternância tu e você: cartas capixabas” (SCHERRE; YACOVENCO; SCARDUA, 2018) – Maria Marta Pereira Scherre, Lilian Coutinho Yacovenco e Juliana Rangel Scardua;
- 3) “Objeto direto anafórico: reflexões e perspectivas para o ensino da mudança linguística na escola regular” (CONCEIÇÃO, 2016) – Karina Corrêa Conceição;
- 4) “A condição de distintividade na variação do sujeito pronominal de primeira pessoa do singular em cartas escritas por um capixaba” (MASSARIOL; YACOVENCO, 2020).

Essa amostra possibilitou que se observasse a existência do pronome tu, ausente na fala capixaba, em cartas pessoais escritas por capixabas (SCHERRE; YACOVENCO; SCARDUA, 2018). Também se pôde observar o aumento do uso de sujeitos pronominais expressos nas cartas pessoais (MASSARIOL, 2018) e a diminuição do uso de clíticos acusativos de terceira pessoa em jornais capixabas (CONCEIÇÃO, 2016).

## 2.6 O banco de dados de revistas em quadrinhos

Com o interesse em se observar o registro de alguns fenômenos linguísticos em outro gênero discursivo, foram realizadas pesquisas com base nas revistas em quadrinhos da Turma da Mônica, como as a seguir listadas.

- 1) “A variação da primeira pessoa do plural nas revistas em quadrinhos de Chico Bento” (SILVA, 2020) – Jessyca Christyna Soares da Silva;
- 2) “Já vejo ele nos quadrinhos: uma análise em tempo real da variação do objeto direto anafórico de terceira pessoa em revistas da Turma da Mônica” (ZANELATO, 2020) – Carolina Amorim Zanellato;
- 3) “Variação do imperativo em revistas em quadrinhos da Turma da Mônica: um estudo em tempo real” (CONCEIÇÃO, 2018) – Ednaildes Bispo da Conceição.

As pesquisas proporcionaram uma reflexão sobre a variação linguística no gênero história em quadrinhos. Todas apresentam análises em tempo real, uma vez que foram digitalizadas revistas desde a década de 1970, o que permitiu que se observasse a implementação de novas variantes linguísticas.

## 2.7 O banco de dados de telejornais capixabas

Com o intuito de se observarem variáveis relacionadas ao estilo, foram compostas amostras baseadas em telejornais capixabas. O banco de dados coletado também se mostrou interessante para a análise de outros fenômenos linguísticos. É importante destacar que os telejornais, apresentados no horário do almoço por duas emissoras distintas, são bastante populares, havendo, inclusive, grande interação com os telespectadores, que participam ativamente do telejornal. Esses fatos são relevantes na construção do cenário discursivo que envolve cada um dos jornais, destinados a públicos distintos.

No momento, em 2021, são realizadas duas pesquisas sobre a variação de primeira pessoa do plural:

- 1) “ESTV 1ª edição, é a gente com você: análise sociolinguística da variação da primeira pessoa do plural em jornal televisionado” (DIAS, 2020) – Kaio Rangel da Silva Dias;

- 2) “Variação da primeira pessoa do plural nós e a gente no telejornal Balanço Geral/ES” (SANTOS, 2020) – Tamilly Costa Santos.

Esse banco de dados foi organizado em tempos de pandemia, quando havia impossibilidade de se realizarem entrevistas sociolinguísticas tipicamente laboratoriais. As pesquisas apontam para a importância de variáveis relativas ao locutor, ao interlocutor, ao tema, às sequências discursivas e à situação comunicativa.

## 2.8 Pesquisas com outras abordagens

Mais recentemente, o grupo também tem trabalhado com outras abordagens da Sociolinguística – mais especificamente, com estudos de percepção e de variação estilística. Citamos os trabalhos por ora desenvolvidos:

- 1) “Sotaque capixaba: um estudo de percepção” (TESCH, 2021) – Leila Maria Tesch;
- 2) “Análise da variação estilística na concordância nominal de número de uma falante pouco escolarizada” (SOUZA, 2017) – Elaine Cristina Borges de Souza;
- 3) “A expressão do sujeito pronominal na fala de uma universitária capixaba: uma análise baseada no estilo” (MASSARIOL; YACOVENCO, 2017) – Caroliny Batista Massariol e Lilian Coutinho Yacovenco.

O projeto “O sotaque capixaba: um estudo de percepção”, foi desenvolvido pela professora Leila Maria Tesch, em período de Pós-Doutorado, realizado em 2020-2021. A pesquisa teve por objetivo realizar estudos de percepção para verificar se o capixaba teria uma identidade linguística bem definida, verificando se os capixabas conseguem se reconhecer capixabas pela fala e também se pessoas de outros estados conseguem reconhecer capixabas pela fala. Tencionamos, também, dessa forma, contribuir com os estudos de percepção no Brasil, ainda bastante incipientes.

Estudos de percepção linguística buscam observar como certos significados sociais podem se associar a usos linguísticos, tendo em vista que buscam verificar como diferentes formas linguísticas são ouvidas e processadas pelos membros de uma comunidade (CAMPBELL-KIBLER, 2006). Essa área da Sociolinguística

ainda tem sido pouco explorada, mas pode contribuir significativamente para um maior entendimento dos usos linguísticos.

Nesta pesquisa, foi aplicado um questionário online e os resultados estão em fase de análise. O questionário é composto por cinco seções: (i) apresentação da pesquisa e acesso ao Termo de Consentimento Livre e Esclarecido;<sup>3</sup> (ii) perguntas sobre informações pessoais, com a ressalva de que a identificação não será divulgada e a pessoa não precisa dizer o nome pessoal; (iii) questões ligadas ao reconhecimento de sotaques brasileiros; (iv) perguntas ligadas à avaliação de sotaques brasileiros; e (v) questões específicas sobre o sotaque capixaba.

Outra abordagem também explorada no grupo do PortVix são os estudos de variação estilística, como a pesquisa “Análise da variação estilística na concordância nominal de número de uma falante pouco escolarizada”, de Elaine Cristina Borges de Souza, sob coordenação da professora Maria Marta Pereira Scherre. Nesse estudo, a pesquisadora analisará a concordância nominal de número na fala de uma mulher, de 28 anos, que nasceu e sempre morou em Goiânia, na mesma casa. Essa mulher possui mestrado, atua como professora da rede municipal de ensino, é atriz de teatro e ativista. A análise se dará em gravações das seguintes situações: i) aula; (ii) conversas com amigos e família; (iii) ensaios de teatro; e (iv) apresentações de teatro.

Também com o intuito de observar a variação estilística na fala de uma jovem universitária, Massariol e Yacovenco (2017) analisaram a expressão do sujeito pronominal na fala dessa jovem em quatro situações comunicativas: assembleia geral estudantil da Universidade Federal do Espírito Santo (Ufes), reunião de uma força política estudantil no campus de Goiabeiras, reunião de uma força política estudantil no campus de Alegre, e uma conversa informal entre amigos. Foi constatado que, quanto maior intimidade e interação entre a jovem e os interlocutores, maior presença de sujeitos explícitos.

Essas novas perspectivas de análise possibilitam um olhar voltado para questões pouco abordadas no projeto PortVix e abrem horizontes para novas linhas de pesquisa, voltadas não somente para macrocategorias que atuam sobre a variação e mudança linguística, como sexo, escolaridade e faixa etária do falante, mas, também, para questões relativas à percepção linguística e para o próprio falante.

---

<sup>3</sup> Para continuar participando da pesquisa, é obrigatório clicar no campo “Sim, aceito participar”.

## CONSIDERAÇÕES FINAIS

Em seus 20 anos de existência, o projeto PortVix proporcionou o conhecimento de características da comunidade de fala capixaba, características essas que, ora a individualizavam, ora a aproximavam das demais variedades do português brasileiro. O presente artigo mostrou a evolução das pesquisas desenvolvidas pelo grupo PortVix, inicialmente preocupadas com a identificação da comunidade de fala por meio de banco de dados voltados para a produção linguística, levando em conta macrovariáveis sociais, como sexo, escolaridade e faixa etária do falante. No momento atual, as pesquisas também se debruçam sobre a variação estilística e os estudos de percepção. Podemos afirmar que o Projeto PortVix contribuiu para a inserção da variedade capixaba no cenário sociolinguístico brasileiro e para a discussão dessa variedade entre os próprios capixabas.

## REFERÊNCIAS

- CAMPOS JÚNIOR, Heitor da Silva. A variação morfossintática do artigo definido na capital capixaba. 2011. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal do Espírito Santo, Vitória, 2011.
- DIAS, Kaio Rangel da Silva. ESTV 1ª edição, é a gente com você: análise sociolinguística da variação pronominal de primeira pessoa do plural em jornal televisionado. Texto de qualificação de mestrado – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória, 2020.
- FOEGER, Camila Candeias. A primeira pessoa do plural no português falado em Santa Leopoldina. 2014. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal do Espírito Santo, Vitória, 2014.
- LEITE, Melina de Figueiredo. As vogais médias pretônicas na fala de Vitória. 2014. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal do Espírito Santo, Vitória, 2014.
- LIMA, Jares Gomes. O jogo na comunidade de Caravelas-BA: variação da fricativa coronal pós-vocálica. 2017. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória, 2017.

LINDENBERG NETO, Carlos Fernando, CORREA, José Carlos, LEITE, Antônio Carlos, MEDEIROS, Sandra. *A Gazeta – 80 anos de história*. Vitória: AS. A Gazeta, 2008.

LOPES, Lays de Oliveira Joel. *A concordância nominal de número no português falado na zona rural de Santa Leopoldina*. 2014. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória, 2014.

LOPES, Lays de Oliveira Joel. “As história são boas. As pessoas, maravilhosas”: análise da variação da concordância verbal e nominal na zona rural de Santa Leopoldina/ES. 2020. Tese (Doutorado em Estudos Linguísticos) – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória, 2020.

MASSARIOL, Carolyn Batista. *A expressão do sujeito pronominal em cartas e postais capixabas do século XX*. 2018. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal do Espírito Santo, Vitória, 2018.

NASCIMENTO, Cristiana Aparecida Reimann do. *A negação no português falado em Vitória*. 2014. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal do Espírito Santo, Vitória, 2014.

SCHERRE, Maria Marta Pereira; Yacovenco, Lilian Coutinho; SCARDUA, Juliana Rangel. *A alternância tu e você: cartas capixabas*. *Confluência*, v. 1, p. 9-25, 2018.

SILVA, Jessyca Christyna Soares da. *A variação de primeira pessoa do plural nas revistas em quadrinhos de Chico Bento*. 2020. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória, 2020.

TESCH, L. M. *A expressão do tempo futuro no uso capixaba: variação e gramaticalização*. 2011. 192f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

TESCH, L. M. *A variação no âmbito do irrealis entre as formas de futuro do pretérito e pretérito imperfeito do indicativo na fala capixaba*. 2007. 153f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.

YACOVENCO, Lilian Coutinho; SCHERRE, Maria Marta Pereira; TESCH, Leila Maria; BRAGANÇA, Marcela Langa L.; EVANGELISTA, Elaine Meireles; MENDONÇA, Alexandre Kronemberger de; CALMON, Elba Nusa; CAMPOS JÚNIOR, Heitor da Silva; BARBOSA, Astrid Franco; BASÍLIO, Jucilene Oliveira Sousa; DEOCLÉCIO, Carlos Eduardo; SILVA, Janaína Biancardi da; BERBERT, Aline Fonseca; BENFICA, Samine de Almeida. Projeto PortVix: a fala de Vitória/ES em cena. Alfa: Revista de Linguística (UNESP. Online), v. 56, p. 771-806, 2012.

ZANELATO, Carolina Amorim. Já vejo ele nos quadrinhos: uma análise em tempo real da variação do objeto direto anafórico de terceira pessoa em revistas da Turma da Mônica. 2020. Texto de qualificação de mestrado – Programa de Pós-Graduação em Linguística, Universidade Federal do Espírito Santo, Vitória, 2020.



# EFEITOS DAS MÁSCARAS FACIAIS NA INTERAÇÃO E A COMPENSAÇÃO NA FALA

*Raquel Meister Ko. Freitag*  
*Julian Tejada*  
Universidade Federal de Sergipe

## 1 INTRODUÇÃO<sup>1</sup>

A pandemia de Covid-19 afetou significativamente o comportamento humano (ALMEIDA et al., 2020; IBARRA et al., 2020; SEPÚLVEDA-LOYOLA, 2020), e mesmo quando cessar o isolamento imposto, efeitos ainda serão sentidos, inclusive na língua. A ausência de interação face a face tem resultado em novas formas de comunicação e pode levar a novos modos de mudança linguística (DONG, 2021). Circunstâncias novas demandam um léxico novo; além disso, a limitação da interação face a face em função do isolamento e o uso de máscaras faciais podem levar a mudanças na língua.

Máscaras instauram significados indexicais; usar máscara pode configurar novas identidades associadas a um conjunto de valores indexados pela máscara, em oposição aos que não usam máscaras (POLLOCK, 1995), o que é construído socialmente. Pollock compara a indexicalidade promovida pela máscara com a indexicalidade de traços linguísticos a grupos, tipos de pessoas ou atitudes.

---

<sup>1</sup> Este texto resume a proposta do projeto “Efeitos das máscaras faciais na interação: linguagem e emoções”, desenvolvido pelos autores na Universidade Federal de Sergipe, junto ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC/UFS).

O uso de máscaras faciais é um dos recursos de baixo custo mais efetivos para mitigar o contágio (EIKENBERRY et al., 2020). No entanto, o seu uso ainda enfrenta diferentes tipos de objeção, desde questões identitárias, associadas ao negacionismo da pandemia, até questões fisiológicas e linguísticas, como a dificuldade para respirar e falar. Até que ponto máscaras podem produzir mudanças na língua? Com essa questão, discutimos premissas de mudança linguística da perspectiva da sociolinguística, evidenciando o papel das barreiras físicas na língua, especialmente no campo dos gestos. Nessa trajetória, ampliamos a concepção de gramática da língua para incluir não só gestos linguísticos, mas também gestos faciais que expressam emoções.

## 2 MÁSCARAS FACIAIS E A MUDANÇA NA LÍNGUA

O que em português tem sido comumente denominado de “máscara facial” envolve uma gama de dispositivos, como máscaras faciais de produção caseira ou industrial, feitas com camadas de tecido, ou de natureza médica, como máscaras cirúrgicas, até respiradores semifaciais (recomendados nas situações mais críticas), que contém uma peça facial filtrante (PFF), que filtra o agente antígeno, em diferentes níveis de filtragem (sendo o nível 2 o recomendado para o vírus SARS-CoV-2).

Independentemente do tipo, a máscara facial se comporta como uma barreira física à língua. O efeito de barreiras físicas sobre a língua é aludido no Curso de Linguística Geral como desencadeador de mudanças. Mais recentemente, ao estudar o efeito do isolamento oceânico no léxico no Japão, incluindo efeitos geográficos, como rios e montanhas, Lee e Hasegawa (2014) demonstraram, do ponto de vista evolutivo, que a diversidade linguística pode ter sido ainda mais ampliada e mantida por barreiras políticas que permitiram o contato linguístico exclusivamente entre indivíduos geneticamente próximos que vivem dentro das fronteiras do seu grupo. Gorenflo e Romaine (2021) afirmam que a diversidade cultural correlaciona com a biodiversidade. Eles examinaram 48 áreas protegidas dentro de sítios de patrimônio mundial (WHS nas suas siglas em inglês), declarados assim pela UNESCO, e descobriram que nos sítios onde se encontra maior diversidade linguística também se encontra maior biodiversidade, e que na medida que as fronteiras entre esses grupos se diluem, perdendo assim a diversidade linguística, se perde também biodiversidade. Evidências desse tipo, que consideram barreiras entre grupos, são bem conhecidas na literatura, inclusive embasando a tese do português brasileiro como conservador. No entanto, ainda

há poucos estudos que considerem o efeito de barreiras no nível individual, tais como as máscaras faciais.

Nem todas as pessoas que não usam máscara facial são negacionistas da gravidade da pandemia; muitas alegam o desconforto para a respiração e para a fala. Em um estudo composto por um conjunto de 500 postagens no Twitter durante o início da pandemia nos Estados Unidos, 30,6% das postagens reportam o desconforto físico ou efeitos negativos decorrentes da percepção ou experiência como resultado do uso de máscara, como erupção cutânea, acne, falta de ar ou desmaio; ou, ainda, da crença de que o uso de máscara causaria danos ao sistema imunológico (HE et al., 2021). Em situações de obrigatoriedade de usos, frequentemente pessoas retiram as máscaras para falar, ou mesmo para respirar. Muitos alegam não poder falar normalmente porque têm a percepção de que os sons saem abafados e isso atrapalharia a comunicação. A barreira física da máscara demanda ajustes na produção linguística para a compreensão do conteúdo, o que pode catalisar processos de mudança em curso.

O uso de máscara facial cobre a boca e o nariz, o que abafa o som e torna difícil compreender a fala e a voz mais aguda (MHEIDLY et al., 2020). A cobertura oro-facial pode interferir nos movimentos necessários para a produção de consoantes labiais /p, b, f, v, m/ e de vogais com o traço de arredondamento, interferindo na sua produção. O resultado disso pode ser visto pela necessidade de repetição de determinadas palavras, que demanda estratégias de articulação oral mais controlada. O pedido por repetição por parte do interlocutor, presente na coleta de dados da realização do /r/ nas lojas de departamento de Nova Iorque por Labov (1966[2006]) pode ser entendido como uma sinalização de que o segmento não foi bem compreendido, causando dúvida, ou que o segmento realizado não foi o esperado para aquela situação; em ambas as situações, a repetição se configura como uma situação de maior atenção à fala, com cuidado consciente na articulação do traço. O constante monitoramento da produção, em função da barreira física da máscara, demandaria um estado de consciência sociolinguística ativo: a maior sensibilidade à comunicação e compreensão também implicaria em maior sensibilidade e atenção aos traços linguísticos variáveis, que são objeto de avaliação social. O estado de monitoramento tem potencial de catalisar processos de mudança linguística no nível da consciência.

A dificuldade na compreensão pode estar associada ao fato de que a barreira física das máscaras faciais atua na filtragem de frequências acústicas altas, resultando em um som “abafado”. Ao dobrar a distância da fonte sonora, a intensidade é reduzida pela metade. Isso significa que a diretriz de distanciamento social

recomendada de dois metros entre as pessoas resulta em um som que é aproximadamente metade da intensidade (COREY; JONES; SINGER, 2020). Como efeito prático, as pessoas precisam falar mais alto, e este é um comportamento consciente, associado ao monitoramento. Além disso, o distanciamento social diminui a intensidade dos sons que chegam ao ouvinte, reforçando ainda mais a necessidade de aumento da intensidade sonora da fala.

Estudos prévios acerca da limitação das máscaras na articulação de sons decorrente de interferências nos movimentos da boca apontam para o fato de que os movimentos labiais necessários para as consoantes com traço labial /p, b, m, f, v/, o glide /w/ e as vogais arredondadas ficam restritas, assim como a possibilidade de abaixar a mandíbula ao produzir vogais abertas também ficam mais limitadas (MERCER; LOWELL, 2020).

Os diferentes tipos de máscara interferem na inteligibilidade da fala (COREY; JONES; SINGER, 2020), com atenuação em frequências acima de 1000 Hz. Em um estudo sobre os efeitos do uso de máscaras faciais na comunicação em sala de aula, Bottalico (2020) avaliou os efeitos de três tipos diferentes de máscaras faciais (de tecido, cirúrgica e N95) sobre a inteligibilidade da fala apresentada a estudantes universitários em situação de sala de aula, constatando que o uso de máscaras de tecido produziu uma redução significativamente maior na inteligibilidade da fala em comparação com as outras máscaras. A manipulação de estilos também é afetada pelas máscaras: Cohn, Pycha e Zellou (2021) investigaram o efeito do estilo da fala – casual, claro (com ênfase e articulação de todos os segmentos), e positivo-emocional (tom de voz alegre) – em condições de visualização total da face (sem máscara) e com face encoberta (com máscara) e constataram que, com a presença de máscara, a discriminação da fala é mais precisa quando o estilo de fala é claro. Na condição positivo-emocional, a fala com máscara facial foi menos inteligível do que a fala sem máscara facial, e na condição casual, nenhuma diferença foi observada, sugerindo que os estilos emocional-positivo e casual não são estilos produzidos com a intenção explícita de serem inteligíveis e colaborativos para os ouvintes. Esses efeitos têm potencialidade para catalisar processos de variação fonológica em curso.

No entanto, máscaras são usadas há muito tempo. Em países do oriente médio, é um costume religioso, com niqabs e burkas, e até mesmo um costume sanitário na Ásia, não só por conta de síndromes respiratórias, mas também pelos efeitos de poluição. Não há relatos da interferência da máscara na comunicação a ponto de inviabilizá-la.

O que, então, faz com que as pessoas percebam as máscaras como barreiras para a comunicação? As barreiras acústicas identificadas em estudos que comparam os diferentes tipos de máscara podem ser superadas com o reforço sonoro, mas parece que esta medida ainda é insuficiente. É necessário considerar então que recursos corporificados apresentam propriedades que são características da gramática da língua (FREITAG; CRUZ; NASCIMENTO, 2021), e que a barreira física da máscara impede a visualização de pistas articulatórias que completam aquilo que os ouvidos não captam, como as pistas de articulação que nos ajudam a distinguir consoantes oclusivas, como /p, d, g/. MacLeod e Summerfield (1987) apontam que ver a face de um falante pode trazer à percepção da fala um aumento na relação sinal/ruído de 8-10 dB. Embora pareça pequeno, esse aumento intensifica as taxas de reconhecimento de palavras. Os gestos faciais relacionados à articulação da fala somam-se aos gestos acústicos para a articulação; e a dissociação desses pode prejudicar o resultado: considerada uma medida inclusiva, máscaras transparentes, que permitem a leitura labial, são as que têm o pior desempenho acústico (MERCER; LOWELL, 2019), o que também prejudica a comunicação.

Como vimos, a adoção de máscaras faciais tem efeitos documentados em dois domínios cognitivos: o da linguagem (COREY; JONES; SINGER, 2020; BOTTALICO, 2020) e o das emoções (CALBI et al., 2020; CARBON, 2020; FREUD et al., 2020; MHEIDLY et al., 2020), que, na interação são inter-relacionados.

### 3 COMPENSAÇÃO DA FALA

Com o campo oro-facial parcialmente recoberto com máscaras, parte da informação dos gestos faciais é perdida, levando as pessoas a mobilizarem outros recursos para sinalizarem as suas emoções no processo interacional, em um processo de compensação da fala.

Assumindo a proposta de tipologização de Ekman para a categorização de emoção, consideramos que as emoções básicas são demonstradas por meio de um padrão fisiológico de contração dos músculos da face. O seu reconhecimento e análise pode se dar por meio do *Facial Action Coding System* (FACS) (EKMAN, 1997), uma taxonomia das expressões faciais humanas a qual classifica os movimentos individuais dos músculos envolvidos nas mesmas. Essa taxonomia se baseia na identificação da contração ou o relaxamento dos músculos faciais, e classifica pelo menos 46 unidades de ação ou AU (na sua sigla em inglês) que servem para descrever qualquer tipo de expressão facial. Por exemplo, a expressão facial de alegria envolve o movimento do levantamento das sobrancelhas

externas, provocado pelos músculos frontais pares laterais (unidade de ação 2) e puxamento do canto do lábio, provocado pelo músculo zigomático maior (unidade de ação 12), movimentos que se opõem, por exemplo, à expressão neutra, na qual os músculos da face estão em repouso.

Estudos experimentais evidenciam que há efeito das máscaras faciais no reconhecimento das emoções e da sua valência. Em um estudo com adultos (n = 41), Carbon (2020) identificou menor precisão no reconhecimento de emoções em uma tarefa experimental na condição de presença de máscara, além de padrões de erros na identificação de erros sistemáticos como a identificação de faces de nojo, como raiva, e o julgamento de faces que expressavam emoções como sendo neutras. Resultados similares foram encontrados no estudo de Freud et al. (2020), também com adultos (n=496), Marini et al. (2021) (n=122) e Noyes et al. (2021) (n=159), que também reportaram que as máscaras faciais geram problemas não só na identificação das emoções, mas também no reconhecimento, na atribuição de confiança e no reconhecimento de rostos familiares. Calbi et al. (2020) também encontraram que a utilização de máscaras, sejam de pano ou aquelas utilizadas pelas equipes médicas, provocam atribuições de valências mais negativas, mas não afetam a identificação das emoções, trabalhando com uma amostra também de adultos (n=96). Por outro lado, em um estudo com crianças em idade escolar (n = 81), Ruba e Pollak (2020) evidenciam que o fato de partes do rosto estarem cobertas, seja por óculos escuro para a área dos olhos, seja por máscara facial para a região oro-facial, não interferiu na capacidade dos participantes em inferir as emoções. Os autores concluem que, em combinação com outras pistas contextuais, é pouco provável que as máscaras prejudiquem as interações sociais das crianças em seu cotidiano. No entanto, fatores socioculturais podem interferir nesse processo, o que demanda a realização de estudos em outras realidades.

Carbon (2020) sugere ações compensatórias para manter a interação social efetiva em situações em que a informação visual relevante para a identificação de emoções é reduzida, como no caso da adoção de máscaras faciais. Por exemplo, as pessoas podem compensar a capacidade limitada de leitura das pistas de identificação facial por meio do uso de diferentes fontes de informação, como a modulação do tom e do ritmo da fala (gestos vocais), movimentos da cabeça, ou a incorporação de gestos corporais que expressem emoções. Mheidly et al. (2020) listam oito medidas para melhorar a interação enquanto se usa máscara facial, dentre as quais orientações diretamente relacionadas aos gestos faciais e vocais, como “utilizar e reconhecer a face superior através das sobrancelhas,

olhos e bochechas superiores durante a comunicação interpessoal” e “falar mais alto e mais devagar em ambientes mais silenciosos”, assim como recomendam a realização de pesquisas transversais que explorem o efeito das máscaras faciais na interação.

## 4 PARADIGMAS PARA O ESTUDO DA COMPENSAÇÃO DA FALA

Considerando que recursos corporificados (gestos faciais e gestos vocais) apresentam propriedades que são características da gramática da língua, tais como a regularidade sistemática, a variação e a apreciação social (FREITAG; CRUZ; NASCIMENTO, 2021), mudanças linguísticas decorrentes da adoção de máscaras faciais, e sua relação com as respostas emocionais, são potenciais campos latentes para o estudo do processo de compensação da fala.

A constituição de amostras sociolinguísticas por meio de entrevistas em que os interactantes usam máscara facial é uma abordagem para a descrição do efeito da compensação na fala na produção linguística; protocolos de documentação sociolinguística que integrem gestos verbais e gestos faciais têm sido implementados para a descrição de traços variáveis que são sensíveis ao componente afetivo, como os diminutivos (PINHEIRO, 2021) e a modalização de certeza/incerteza/dúvida (CARDOSO, 2021). No entanto, esse tipo de abordagem pode ficar restrito neste primeiro momento por imposições decorrentes do distanciamento social sanitário. Abordagens experimentais, particularmente por meio de tarefas que articulem gestos vocais e gestos faciais, e que possam ser realizadas a distância, podem auxiliar no desvelamento de valores de base para estudos de compensação na fala, a serem aplicados posteriormente em abordagens de produção baseada em amostras linguísticas.

A tarefa experimental do estudo de Carbon (2020) para compensação foi baseada no paradigma de “bolhas” (GOSELIN; SCHYNS, 2001), que possibilita identificar as informações visuais necessárias para expressar e ler emoções a partir de expressões faciais. Nessa tarefa, as faces são apresentadas em bolhas, para padronização. Em outros estudos, Marini et al. (2021) e Calbi et al. (2020) utilizaram imagens de um conhecido banco de expressões faciais, The Karolinska Directed Emotional Faces (KDEF) (LUNDQVIST; FLYKT; ÖHMAN, 1998), nas quais colocaram digitalmente as máscaras faciais. E Noyes et al. (2021) trabalharam também com reconhecimento de rostos familiares, usando fotos de pessoas tanto com máscara facial quanto com óculos escuros.

O paradigma de “leitura da mente”, estabelecido para a avaliação neuropsicológica das funções sociocognitivas, que mensura a capacidade de identificar estados emocionais a partir do campo dos olhos, pode estabelecer um valor de base para a identificação de emoções, e que pode ser articulado com a pista vocal. *Reading the mind in eyes* (RTME) refere-se a uma tarefa de identificação das emoções de conjunto de 36 frames da região dos olhos de uma pessoa (BARON-COHEN; WHEELWRIGHT; JOLLIFFE, 1997, 2001). E *reading the mind in the voice* (RMV) refere-se a uma tarefa de discriminação auditiva de 25 estados emocionais a partir da gravação em áudio (RUTHERFORD; BARON-COHEN; WHEELWRIGHT, 2002; GOLAN et al., 2007).

A tarefa experimental para identificar o efeito da máscara em processos de variação linguística consiste em um teste de decisão lexical visando a discriminação linguística que considere os efeitos da máscara na pista auditiva, com fenômenos linguísticos variáveis já documentados na comunidade dialetal, e os efeitos das expressões faciais de FACS do campo dos olhos neste processo. Uma segunda versão da tarefa consiste em inserir o parâmetro campo visual no teste de discriminação, incorporando o paradigma *Reading the mind*. Assim, além do efeito da máscara na mudança linguística, os resultados da tarefa têm potencial para contribuir para a avaliação dos estados emocionais e o seu efeito no processo de compensação de informações decorrentes do encobrimento do rosto pela máscara.

## 5 FUTURO DA INTERAÇÃO COM MÁSCARA

As implicações em longo prazo do uso da máscara facial na interação são já sentidas no ambiente de saúde (MARLER; DITTON, 2021) e também da escola (BOTTALICO, 2020). Nesse sentido, investigações que considerem a interface entre linguagem e emoções podem contribuir para o desenvolvimento de estratégias de comunicação particularizadas para as diferentes realidades socioculturais.

Do ponto de vista teórico, uma abordagem que incorpora as expressões visual-corporais humanas no domínio da expressão linguística permite ampliar o poder explanatório da análise para o estudo da gramática em especial, e dos processos cognitivos em geral. O pareamento dos gestos faciais que expressam emoções com os gestos vocais que expressam linguagem em situações de interação é afetado pela adoção de máscaras porque a imposição de uma barreira física no corpo, especificamente na área oro-facial, impacta em demandas de adaptabilidade, tanto no reconhecimento de emoções, como na produção linguística.

Por um lado, o uso de máscaras faciais limita a capacidade de ver e compreender as expressões faciais das pessoas durante as interações, o que diminui o impacto do conteúdo linguístico. Por outro lado, a barreira física da máscara demanda ajustes na produção linguística para a compreensão do conteúdo, o que pode catalisar processos de mudança em curso. Mesmo com o abrandamento da pandemia, é certo que o uso de máscara continuará fazendo parte do comportamento das comunidades ocidentais, assim como aconteceu na Ásia. Assim, processos de compensação da fala, articulando gestos faciais e verbais e associados às emoções, são campo latente na pesquisa sociolinguística.

## REFERÊNCIAS

- ALMEIDA, M. *et al.* The Impact of the COVID-19 Pandemic on Women’s Mental Health. *Archives of Women’s Mental Health*, v. 23, n. 6, p. 741-748. 2020.
- BARON-COHEN, S.; WHEELWRIGHT, S.; JOLLIFFE, T. Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual cognition*, v. 4, n. 3, p. 311-331, 1997.
- BARON-COHEN, S. *et al.* The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry*, v. 42, n. 2, p. 241-251, 2001.
- BOTTALICO, P. *et al.* Speech intelligibility in auralized classrooms when the talker is wearing a face mask. *The Journal of the Acoustical Society of America*, v. 148, n. 4, p. 2631-2631, 2020.
- CALBI, M. *et al.* I See How You Feel: Facial Expressions’ Recognition and Distancing in the Time of COVID-19. *Research Square*, p. 1-20, 13. 2020. <https://doi.org/10.21203/rs.3.rs-107145/v1>.
- CARBON, C. C. Wearing face masks strongly confuses counterparts in reading emotions. *Frontiers in Psychology*, v. 11, p. 2526, 2020.
- CARDOSO, P. B. Efeitos paralinguísticos na inferência de sentidos indicados por (eu) acho que em entrevistas sociolinguísticas. *Dissertação (Mestrado em Letras)*. Universidade Federal de Sergipe. Pós-Graduação em Letras, 2021.
- COHN, M.; PYCHA, A.; ZELLOU, G. Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. *Cognition*, v. 210, p. 104570, 2021.

COREY, R. M.; JONES, U.; SINGER, A. C. Acoustic effects of medical, cloth, and transparent face masks on speech signals. *The Journal of the Acoustical Society of America*, v. 148, n. 4, p. 2371-2375, 2020.

DONG, J. Language and globalization revisited: Life from the periphery in COVID-19. *International Journal of the Sociology of Language*, v. 2021, n. 267-268, p. 105-110, 2021.

EIKENBERRY, S. E. et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modelling*, v. 5, p. 293-308, 2020.

EKMAN, R. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. New York, Oxford University Press, 1997.

FREITAG, R. M. K.; CRUZ, R.; NASCIMENTO, T. A gramática no corpo: dos recursos corporificados na construção e negociação dos sentidos. *Cadernos de Linguística*, v. 2, n. 2, p. 1-20, 2021.

FREUD, E. *et al.* The COVID-19 pandemic masks the way people perceive faces. *Scientific reports*, v. 10, n. 1, p. 1-8, 2020.

GOLAN, O. *et al.* The 'Reading the Mind in the Voice' test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of autism and developmental disorders*, v. 37, n. 6, p. 1096-1106, 2007.

GORENFLO, L. J.; ROMAINE, S. Linguistic Diversity and Conservation Opportunities at UNESCO World Heritage Sites in Africa. *Conservation Biology*. 2021. <https://doi.org/10.1111/cobi.13693>.

GOSSELIN, F.; SCHYNS, P. G. Bubbles: a technique to reveal the use of information in recognition tasks. *Vision research*, v. 41, n. 17, p. 2261-2271, 2001.

HE, L. et al. Why do people oppose mask wearing? A comprehensive analysis of US tweets during the COVID-19 pandemic. 2021. UC Irvine. <http://dx.doi.org/10.1093/jamia/ocab047> Retrieved from <https://escholarship.org/uc/item/01d909qs>.

IBARRA, F. P. *et al.* Impact of the COVID-19 Pandemic on the Sexual Behavior of the Population. *The Vision of the East and the West. International Braz J Urol: Official Journal of the Brazilian Society of Urology*, v. 46, n. suppl.1, p. 104-112, jul. 2020.

- LABOV, W. The social stratification of English in New York city. Cambridge: Cambridge University Press, 1972.
- LEE, S.; HASEGAWA, T. Oceanic barriers promote language diversification in the Japanese Islands. *Journal of Evolutionary Biology*, v. 27, n. 9, p. 1905-1912, 2014.
- LUNDQVIST, D.; FLYKT, A.; ÖHMAN, A. The Karolinska Directed Emotional Faces – KDEF. [s.l.] [CD ROM] from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.
- MACLEOD, A.; SUMMERFIELD, Q. Quantifying the contribution of vision to speech perception in noise. *British journal of audiology*, v. 21, n. 2, p. 131-141, 1987.
- MARINI, M. *et al.* The Impact of Facemasks on Emotion Recognition, Trust Attribution and Re-Identification. *Scientific Reports*, v. 11, n. 1, p. 5577, 10 mar. 2021.
- MARLER, H.; DITTON, A. “I’m smiling back at you”: Exploring the impact of mask wearing on communication in healthcare. *International journal of language & communication disorders*, v. 56, n. 1, p. 205-214, 2021.
- MERCER, E.; LOWELL, S. Y. The low mandible maneuver: Preliminary study of its effects on aerodynamic and acoustic measures. *Journal of Voice*, v. 34, n. 4, p. 645. e1-645. e9, 2020.
- MHEIDLY, N. *et al.* Effect of Face Masks on Interpersonal Communication During the COVID-19 Pandemic. *Frontiers in Public Health*, v. 8, p. 898, 2020.
- NOYES, E. *et al.* The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *Royal Society Open Science*, v. 8, n. 3, p. 201169, 2021.
- PINHEIRO, B. F. M. Pistas linguísticas e paralinguísticas para os sentidos dos diminutivos. Dissertação (Mestrado em Letras). Universidade Federal de Sergipe; Pós-Graduação em Letras, 2021.
- POLLOCK, D. Masks and the Semiotics of Identity. *Journal of the Royal Anthropological Institute*, p. 581-597, 1995.
- RUBA, A. L.; POLLAK, S. D. Children’s emotion inferences from masked faces: Implications for social interactions during COVID-19. *Plos one*, v. 15, n. 12, p. e0243708, 2020.
- RUTHERFORD, M. D.; BARON-COHEN, S.; WHEELWRIGHT, S. Reading the mind in the voice: A study with normal adults and adults with Asperger

syndrome and high functioning autism. *Journal of autism and developmental disorders*, v. 32, n. 3, p. 189-194, 2002.

SEPÚLVEDA-LOYOLA, W. *et al.* Impact of Social Isolation Due to COVID-19 on Health in Older People: Mental and Physical Effects and Recommendations. *The Journal of Nutrition, Health & Aging*, v. 24, n. 9, p. 938-947, 2020.

# UMA EXPERIÊNCIA DE APLICAÇÃO REMOTA DO DISCOURSE COMPLETION TASK A ENTOAÇÃO DO PORTUGUÊS EM CONTATO EM TIMOR LESTE

*Marco Barone*

Universidade Federal de Pernambuco

*Davi Borges de Albuquerque*

Universidade Federal de Goiás

### 1 A FONOLOGIA DA ENTOAÇÃO

Nas línguas entoacionais a variação tonal é utilizada para modificar não os significados dos itens lexicais, mas sim a modalidade ou o valor pragmático da sentença, que também são chamados de ‘significados pós-lexicais’. A pesquisa sobre a entoação é escassa e relativamente recente, e concentrada nos aspectos fonéticos, devido sobretudo à relativa falta de ferramentas conceituais para a discretização e a transcrição da entoação e de suas unidades relevantes para a diferenciação de significados, diferente do que acontece com as unidades segmentais.

A inexistência de um sistema ortográfico para os fenômenos fonológicos da entoação também atrasou os estudos de variação e comparativos, e tal atraso contribuiu para que algumas observações ‘soltas’ de fenômenos em línguas individuais, por mais que recorrentes, fossem generalizados pelo ouvinte desprovido a toda língua e variedade. Com isso, passa-se a impressão errada de que os modos

de associação entre entoação e valor pragmático sejam universais e translinguísticos, construindo-se mitos como: ‘a entoação das afirmações é descendente e a das perguntas contém algum movimento ascendente’ ou ‘a ênfase se caracteriza por um tom alto na sílaba tônica da palavra enfatizada’.

O fato é que tais tendências de uso icônico, fonossimbólico e paralinguístico da entoação, que Gussenhoven (2015) chama de ‘códigos biológicos’, apesar de serem correlacionados com a maioria dos sistemas de entoação e de serem universalmente reconhecíveis pelos humanos, convivem com um uso arbitrário, convencional e língua-específico, derivado da estrutura linguística, que pode chegar a inverter a previsão ditada por tais códigos. Os estudos sobre a variação dos contornos melódicos associados, em diferentes línguas e variedades, a um mesmo tipo pragmático de sentença, têm apontado cada vez mais para a natureza convencional e língua-específica da entoação: é o caso dos *high-rise-terminals* em inglês, das declarativas e perguntas polares do Chickasaw (GORDON, 2003) e, em âmbito românico, dos contornos médio-descendentes dos focos contrastivos e das perguntas polares no contexto italiano (BARONE, 2020). As documentações recentes distinguem 35 tipos pragmáticos, chegando a multiplicar as chances de comportamentos entoacionais menos previsíveis.

O signo entoacional é, assim, a associação convencional, em dado repertório delimitado, entre significantes entoacionais (contornos melódicos e suas unidades menores) e significados pós-lexicais (modalidade e outras nuances pragmáticas, epistêmicas ou informativas) que definem o valor ou tipo pragmático da sentença. Enquanto a fonologia clássica, ou segmental, se constitui como correspondência entre um plano fonético e um plano semântico, a fonologia da entoação poderá assim ser chamada de ‘fono-pragmática’. Assim como para a fonologia clássica, o alcance do sistema linguístico considerado a cada vez é crucial para entender a natureza das alofonias: em um repertório com mais variedades presentes no inventário da percepção (por exemplo, na variação diatópica), as alofonias se dão na competência passiva, porém, se existir variação livre em produção (ex: variação de registro), também na competência ativa.

Devido a tais considerações, é razoável que os estudos de documentação da fonologia da entoação foquem em uma variedade local por vez, e que não faz muito sentido falar, por exemplo, da entoação ‘do português’ como um tudo, ou ter expectativa que as variedades de uma mesma língua, pertencendo a realidades geográficas distintas, apresentem macrocaracterísticas suprasegmentais

comuns que possam agrupar suas entoações em uma família única. A própria definição de ‘entoação brasileira’, utilizada nos primeiros estudos dentro e fora do Brasil, embora apresente algumas tendências comuns, a depender do tipo de frase, não é uma expressão adequada ao nível de análise variacionista atual. Gili-Fivela et al. (2015) mostram como a realidade de um território relativamente pequeno como a Itália, com uma língua segmental comum (porém heranças linguísticas muito diferenciadas) apresenta a beleza de 7 contornos entoacionais possíveis para as perguntas polares neutras em apenas 13 variedades locais analisadas para este tipo frasal.

Por ser transmitida de pais para filhos e incorporada, por hábito e frequência, na interação e exposição ao convívio local, a entoação deve espelhar uma continuidade de matriz genética. Mesmo assim, para certos tipos de frase, é possível achar usos da entoação muito diferentes em realidades geográficas relativamente próximas, assim como, línguas muito distantes que, por coincidência, acabam usando o mesmo contorno. A suportar tamanha variação microareal e falta de coerência com a variação segmental e seu sistema de agrupamentos genético-tipológicos, concorre a ausência de uma norma padrão da entoação na maioria das realidades nacionais e escolares. Devido à sua inconsciência como fenômeno sistêmico, a entoação não é sensível às mesmas forças de reação que delimitam as línguas segmentais, tais como a cristalização da escrita.

Para certo tipo de frase, pelo contrário, a resistência à mudança no tempo pode ser tamanha que o mesmo contorno sobrevive longamente nas línguas derivadas em situações de diáspora, dando informações valiosas acerca da história linguística, o que torna a entoação uma potencial aliada da linguística histórico-comparativa e um novo campo de estudos de reconstrução no futuro próximo.

Esse desatrelamento conceitual entre entoação e nível segmental na definição e no aprendizado linguístico também faz com que, em situações de contato ou aprendizado de L2, a entoação não seja aprendida juntamente à língua. Nesse caso, o falante de uma língua entoacional, aprendiz de outra língua entoacional, enxerga erroneamente seus próprios modos nativos de expressar a entoação, como subjetivos, livres e extrassistêmicos, aplicando-os à língua-alvo em um mecanismo chamado de transferência prosódica, e dificilmente presta atenção à entoação nativa da língua-alvo. Diferente é o caso dos falantes de línguas não entoacionais que, ao aprenderem uma língua entoacional precisam adquirir, pela

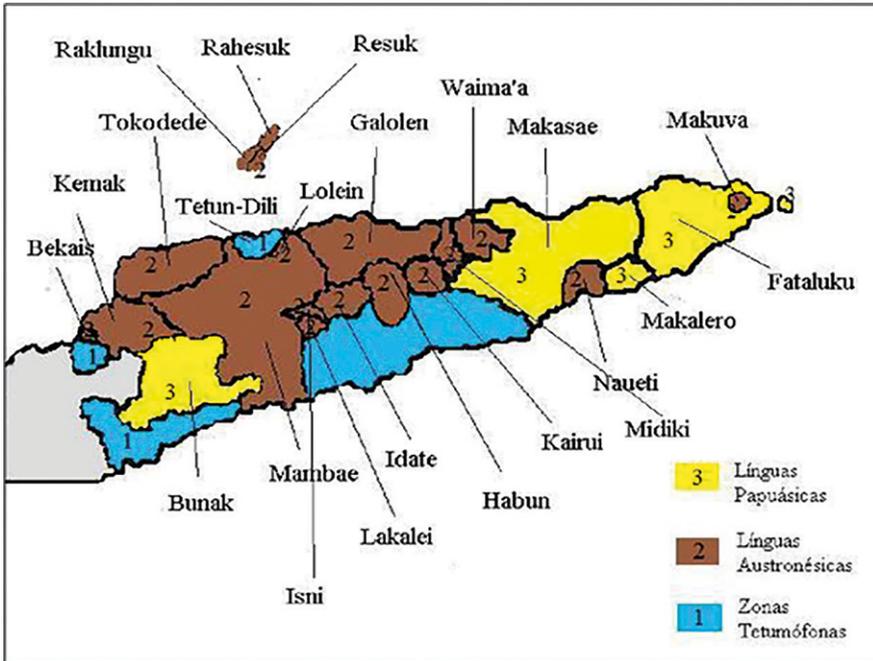
atenção e tentativa de reprodução, os mecanismos-alvo de produção de significados pós-lexicais, por não terem seus próprios para transferir e comunicar.

Se o falante-sujeito considera a entoação como um meio de expressão subjetivo, volátil e transponível de língua em língua, o sociolinguista, pelo qual o falante, com sua subjetividade, é um objeto de observação, diante da constatação que pessoas com herança linguística comum compartilham a entoação. Não se pode ignorar a natureza intrassistêmica da entoação e realizar o pulo epistemológico que permite vê-la não como um modo de expressão do material linguístico sujeito ao arbítrio do falante, mas sim como objeto de observação e descrição, parte do próprio material linguístico convencional, cujo conhecimento compartilhado garante a compreensão. Para tal descrição, como toda fonologia categoriza o que é distintivo, a fonologia da entoação precisará de uma ‘ortografia’ compatível que etiquete de forma distinta os contornos que produzem diferença de significado e suas unidades menores. Um tal sistema de etiquetagem, o ToBI (SILVERMAN, et al. 1992; FROTA, 2000), será apresentado na Seção 3, assim como a teoria métrica autosegmental, em cujos pressupostos ele foi desenvolvido. A seguir, fornecemos uma introdução histórica do nosso caso de estudo.

## 2 O CONTATO LINGUÍSTICO EM TIMOR-LESTE

A República Democrática de Timor-Leste ocupa a porção oriental da ilha de Timor, localizada no sudeste asiático, e sua única divisa terrestre é com a parte ocidental, que pertence à Indonésia. Colônia portuguesa durante 4 séculos, após a Revolução dos Cravos o território passou duas décadas sob a dominação indonésia, até ter sua independência reconhecida somente em 2002. No território são faladas diversas línguas originárias, de origem papuásia e austronésia, cujo número é uma questão debatida e que Hull (2001) estabeleceu em 16. Juntamente a essas, o Tétum-Dili, ou Tétum-Praça, é a língua franca nacional, austronésia, com superstratos português e malaio, L2 da maioria da população e L1 apenas da região da capital e de outras regiões meridionais da ilha (Figura 1). A constituição de 2002 designa como línguas oficiais o português e o Tétum-Praça. Contudo, o português não é falado fluentemente por todos os falantes, sendo língua L3 da maioria dos falantes e geralmente L2 dos falantes nativos de Tétum-Praça.

Figura 1 – As línguas de Timor-Leste.



Fonte: Albuquerque (2011, p. 66).

Estudos recentes (Albuquerque, 2011; Afonso & Goglia, 2015; Batoréo, 2016) sugerem que o português de Timor Leste (PTL, doravante) seja uma variedade não dominante emergente em ‘estágio inicial’, ou ‘em construção’ que poderá se estabelecer como variedade nacional. Porém, para que o português possa ir além do estágio de língua estrangeira estudada pelos falantes locais, e se passe a falar em ‘português timorense’ ocorrem características de homogeneidade e coesão, certos traços que os falantes timorenses, e apenas estes, compartilham. É com esse intuito que o presente estudo se insere no debate acerca do estado de variedade do PTL, na esperança de localizar na entoação um conjunto de traços que seja comum e exclusivo de seus falantes, ou seja, de caracterizá-la como variedade entoacional. A ideia é que a língua compartilhada por todos os falantes, o Tétum, já tendo se amalgamado como língua mista de diferentes substratos originários, possa transferir seus contornos entoacionais para o português dos timorenses, tornando assim a entoação do português timorense um sistema fonológico bem definido, compartilhado e coerente.

O contato com o português em Timor-Leste teve como primeiros protagonistas os colonos portugueses. Assim sendo, a variedade europeia do português é a única que realizou um contato significativo com os locais. Reforçou-se

outrossim, depois do período da colonização, por meio da instrução, e os professores nativos que fisicamente se encontraram em contato com os aprendizes timorenses foram preponderantemente de origem portuguesa. A variedade europeia se apresenta, portanto, como uma forte candidata a ter contribuído para influenciar a entoação do PTL mediante um mecanismo de escuta e reprodução pela imitação. Todavia, devido ao estado de aprendizado ainda imperfeito junto à maioria dos falantes, à lenta internalização e à capacitação de docentes locais, podemos esperar que, se o Tétum tiver um sistema fonológico entoacional coeso, este tenha sido transferido ao português dos Tétum-falantes, pelo menos seus falantes L1.

## **2.1 Fundamentação teórica e metodologia: O modelo métrico autossegmental, o sistema de anotação ToBI e a metodologia do Discourse Completion Task**

O modelo métrico autossegmental (Pierrehumbert, 1980) se baseia na noção intuitiva que a camada dos eventos tonais seja sincronizada com a camada segmental, de modo que certos eventos tonais estariam alinhados temporalmente com outros ‘eventos’ da camada segmental (sílabas tônicas) e de tal alinhamento dependeriam os significados pós-lexicais.

Em primeiro lugar, é estabelecida a estrutura hierárquica do fraseamento prosódico, que determina qual é a maior unidade dotada de um único significado pós-lexical (o sintagma entoacional) e suas unidades menores, o sintagma intermédio, o grupo prosódico, o sintagma fonológico, a palavra prosódica, e, no nível sublexical, o pé, a sílaba e a mora. Ao sintagma entoacional é associado um contorno entoacional, que é a curva que a frequência fundamental desenha ao longo da emissão vocal de tal sentença, com suas subidas e descidas correspondendo às impressões acústicas de subidas e descidas da altura melódica (nota musical) associada a cada instante da pronúncia.

Todo contorno entoacional seria assim construído como uma interpolação substancialmente linear (a menos de fenômenos biológicos como a declinação) entre alvos tonais altos (H), baixos (L) ou, mais raramente, médios (!H) alinhados em certos pontos específicos da linha temporal e pelos quais a curva da frequência fundamental é obrigada a passar.

Em geral, os alvos tonais estão posicionados no final do sintagma entoacional, intermédio e em correspondência das sílabas tônicas, podendo haver um

só ou mais de um alvo, desenhando um movimento complexo em proximidade de tal sílaba tônica. Tais movimentos, geralmente monotonais ou bitonais, são chamados de acentos de frequência. Eles podem ocorrer apenas em posições já acentuadas no nível da palavra com acento primário, ou seja, sílabas tônicas, também chamadas de *tone bearing units* (TBU), mas não é necessário que todas as sílabas tônicas possuam de fato um acento frasal. Em correspondência com a última sílaba tônica do sintagma intermédio (que muitas vezes coincide com o sintagma entoacional) deve ser realizado um acento de frequência, que se chama acento nuclear, e contribui a determinar a pragmática da frase.

O sistema de anotação ToBI (*'tone and break indices'*) estabelece etiquetas de 0 até 4 (*break indices*), como delimitadores das unidades da hierarquia do fraseamento prosódico na segmentação da enunciação, desde a palavra fonológica até o sintagma entoacional. Nas fronteiras esquerda e direita do sintagma entoacional podem ser situados tons ou movimentos tonais, chamados respectivamente tom inicial e final de fronteira que, juntamente aos acentos de frequência, contribuem com a pragmática. Nas variedades do português até agora estudadas, o acento nuclear e o tom de fronteira final são relevantes fonologicamente. Os acentos de frequência são etiquetados por combinações de símbolos, como H\*, H+L\*, L\*+H, H\*+!H, H+H\* etc., onde o asterisco indica o tom alinhado com a sílaba tônica nuclear, podendo ou não haver outro alvo tonal, logo antes ou depois desta, que cria um movimento em entrada ou em saída da sílaba nuclear. Já os tons de fronteira finais serão H%, L%, LH% etc., onde o último tom é realizado em final de frase, podendo ou não ser imediatamente precedido por outro alvo.

O nosso estudo tem como objetivo a documentação e a etiquetagem ToBI de uma porção do sistema fonológico entoacional do Tetum-Praça e do português de um grupo de falantes de Timor-Leste, assim como a comparação destes. Mais especificamente, analisaremos as realizações de 6 tipos pragmáticos de sentença-alvo nas duas línguas (declarativas neutras, perguntas polares neutras e com surpresa, perguntas disjuntivas, perguntas parciais, declarativas óbvias), para entender se, para cada um destes tipos frasais, 1) o Tétum possui padrões entoacionais coerentes entre os participantes e se 2) os contornos do Tétum são transferidos no português dos participantes.

Para tal finalidade, utilizaremos a metodologia do *Discourse Completion Task* (DCT, doravante) (BLUM-KULKA et al., 1989), uma coleta de dados correspondendo aos tipos de frase desejados, obtidos mediante o uso de um

questionário e de uma específica técnica de entrevista. Como a gravação foi realizada durante a pandemia de 2020, ela ocorreu de forma remota, principalmente mediante WhatsApp e outras ferramentas digitais e mídias sociais. Todas as questões referentes ao DCT, suas vantagens e limitações, desde os cuidados na preparação dos questionários até a contribuição dos participantes, no caso do Tétum, assim como os problemas da adaptação remota com os quais nos confrontamos, estão detalhadamente relatados em Barone e Albuquerque (2021).

Analisaremos os resultados de 3 participantes, uma falante nativa de Tétum, e L2 de português e dois falantes L2 de Tétum e L3 de português, sendo falantes nativos de Tokodede e Waimaha, respectivamente. Todos os participantes foram entrevistados em português e em Tétum. Para o português, foi desenvolvido um questionário prévio contendo as realizações dos 6 tipos de frases indicados. Em seguida, foi pedido que os participantes vertessem as frases para o Tétum, com alguns cuidados que emergiram de um estudo preliminar com outros falantes (BARONE; ALBUQUERQUE, 2021), como por exemplo, distanciá-las entre si para evitar um efeito de ‘eco’. No futuro, será interessante, com a ajuda dos participantes, a elaboração de outro questionário específico para o Tétum e possivelmente diferente do português, para confirmar os resultados.

No total, 60 sentenças, 23 em português e 37 em Tétum resultaram aceitáveis para análise e, com o auxílio do PRAAT (BOERSMA; WEENINCK, 2021), foi realizada sua etiquetagem ToBI e foi assim preparada uma tabela comparativa (veja Figura 8) para mostrar quais contornos utilizados em Tétum também ocorrem em português. Devido ao fato que existem, para cada tipo de frase, questões não triviais de comparabilidade segmental-sintática entre as línguas que justifique a comparabilidade estrutural dos contornos (ou seja, a correspondência entre as unidades menores que o compõem), discutiremos tais questões juntamente com os resultados, à medida que apresentamos os dados.

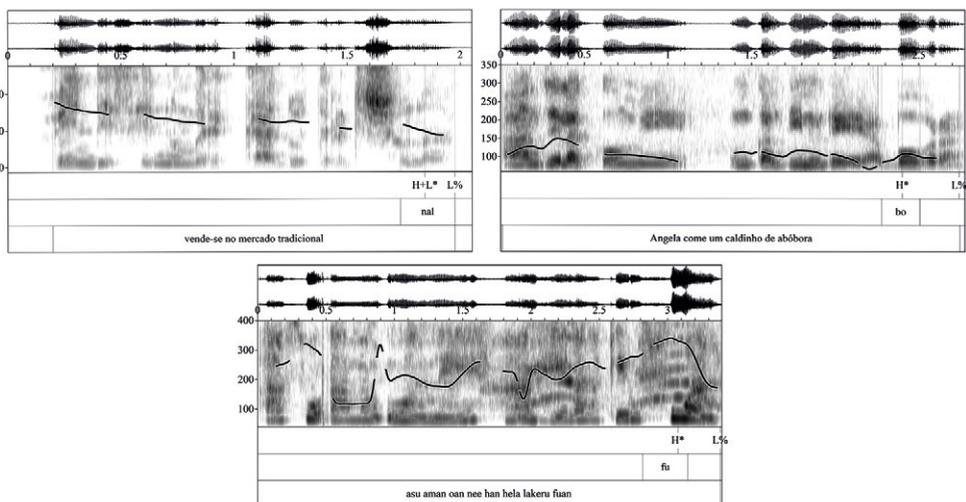
### **3 RESULTADOS E DISCUSSÃO: DECLARATIVA NEUTRA**

A declarativa neutra, ou declarativa de foco amplo, é uma frase declarativa onde a intenção informativa do falante é de não ressaltar nenhum elemento menor mais do que os demais e o foco é colocado no sintagma que representa a sentença inteira. Tal frase se distingue da declarativa de foco estreito, onde um constituinte menor é focalizado informativamente e costuma receber, no nível

prosódico, um movimento acentual saliente, que pode variar de língua em língua. O efeito prosódico da falta de foco informativo estreito, que também pode variar de língua para língua, não sempre consiste na mera ausência da proeminência que a língua atribuiria em caso de foco estreito. É possível que exista uma posição padrão cujo constituinte deva receber a mesma proeminência (foco prosódico), em caso de ausência de foco informativo. Nunca teremos a garantia de que o falante, por si só, não coloque um foco desnecessário sobre um constituinte menor, portanto o ideal seria usar, para este tipo de frase, uma amostra maior, descartando os contornos minoritários.

Os resultados da nossa documentação no PTL mostram dois contornos possíveis sendo utilizados para realizar as declarativas neutras, H+L\* L% e H\*L%, enquanto no TP encontramos apenas H\*L% (veja Figura 2). O contorno descendente H+L\* L%, é o mais comum entre as línguas do mundo para esse tipo frasal, utilizado também em todas as variedades de PE e PB observadas até agora. Já o segundo é um contorno que costuma salientar a sílaba tônica, geralmente encontrado nas exclamações ou para foco estreito. Podemos observar a seguir, exemplos de suas ocorrências.

Figura 2 – Os contornos das declarativas neutras em PTL (acima) e em Tétum (abaixo).



Fonte: Os autores.

A presença do contorno H\*L% no PTL pode significar 4 coisas: ou 1) os falantes nunca estão realizando uma declarativa neutra e a documentação é

falha, e, embora as chances deste tipo de erro se repetir em todos os falantes seja mínima, o único método para descartar isso é usar um número maior de ocorrências e compará-las como o foco estreito; ou 2) esse é o contorno do TP, e é transferido para o PTL, mas paralelamente os falantes de PTL aprendem, por imitação, e juntamente à língua, uma entoação que vem de outra variedade do português; ou 3) o contorno H+L\* L% também existe em TP, mas não o encontramos; ou 4) os falantes de PTL, ainda não fluentes, na tentativa de se certificar da correta marcação do acento de palavra, utilizam o *pitch* alto simultaneamente à intensidade e duração na sílaba tônica nuclear, realizando uma hipermarcação.

Nossa intuição, observando esses dados, se inclina em favor da hipótese 2), mas não é uma evidência suficiente, e precisamos documentar uma amostra maior para confirmar o achado. Também um só tipo de frase é pouco para afirmar a existência de mecanismos de transferência e sequer a existência fonológica de uma ‘imagem acústica’ de contorno entoacional que seja objeto de transferência. Vamos, portanto, corroborar nossa análise pelos outros tipos de sentença.

## 4 RESULTADOS E DISCUSSÃO: INTERROGATIVA POLAR NEUTRA

A interrogativa polar neutra, ou pergunta sim/não informativa, é a pergunta que prevê as opções ‘sim’ ou ‘não’ como resposta, sem pressuposição (afirmativa ou negativa) por parte do falante, com a intenção de obter uma informação sobre uma ocorrência, e sem qualquer suspeita acerca do seu grau de verdade. À ocorrência, apresentada textualmente, se acrescenta a informação contextual de que o falante está querendo conhecer seu grau de veracidade. Em português, tal informação contextual (o ato da interrogação), diferente de línguas como o inglês ou o chinês, com raras exceções de inversão no PE, é inteiramente extratextual e expressa unicamente mediante entoação. Portanto, o material segmental de uma pergunta informativa é idêntico, em português, ao de uma sentença declarativa, e a entoação possui o papel crucial de distinguir as duas.

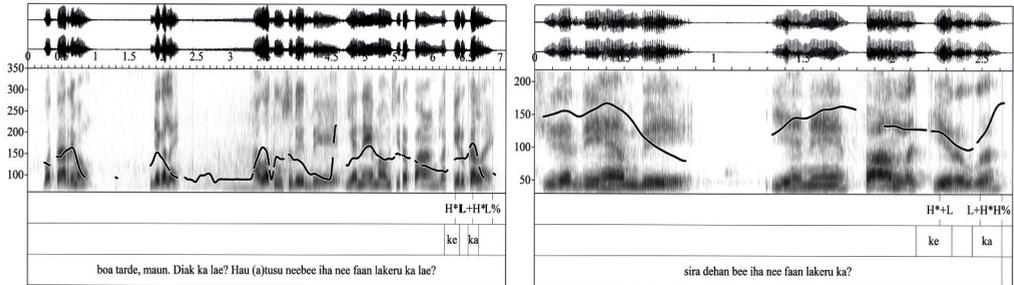
Já no Tétum, a interrogativa polar é realizada colocando opcionalmente, ao lado da afirmação declarativa, uma partícula interrogativa, ou marcador interrogativo polar, a saber, a expressão ‘ka lae?’ (literalmente ‘ou não?’) ou sua forma abreviada ‘ka?’ (lit. ‘ou...?’). As realizações de tais formas em

Tétum confirmam que a partícula ‘ka’ é portadora de acento primário na expressão ‘ka lae’ e não cliticiza com a palavra anterior quando abreviada.

Ex: Iha nee faan lakeru ka (lae)?

Aqui vendem abóbora, OU (NÃO)?

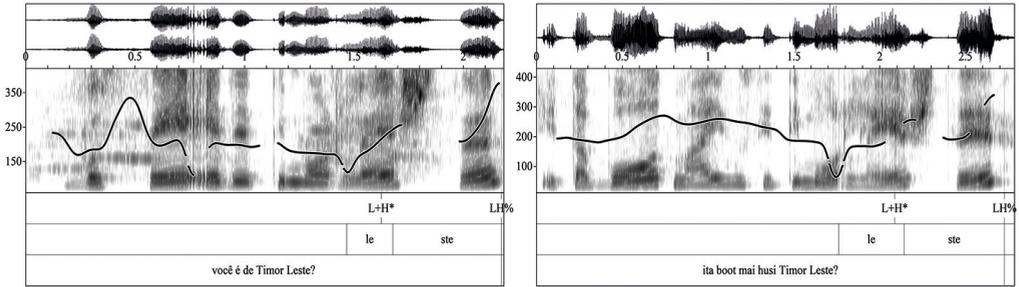
Figura 3 – Contornos das perguntas polares com ‘ka lae’(L+H\*L%) e com ‘ka’(L+H\*H%).



Fonte: Os autores.

Do ponto de vista da entoação, podemos esperar que a mesma divisão informativa seja refletida na entoação, ou seja, que o falante produza o primeiro trecho antes de chegar ao ‘ka’ exatamente como uma declarativa, desvelando somente em seguida sua intenção interrogativa. À primeira vista, isto ocorre, já que a primeira parte das interrogativas que analisamos possuem um acento H\* como o acento nuclear da frase declarativa (H\*L%). Ao reanalisar a sentença como interrogativa, contudo, tal acento se torna pré-nuclear, e o que contribui à nova pragmática (interrogativa) é o acento L+H\* na partícula ‘ka’. Porém, na primeira forma, a pergunta ‘ou não?’, que recebe um acento descendente, pode ser considerada como uma verdadeira pergunta disjuntiva, em quanto a forma abreviada, mais frequente, pode estar sofrendo reanálise como pergunta polar, tanto no nível segmental, quanto no suprasegmental. Devemos, portanto, analisar as perguntas polares do Tétum sem a expressão ‘ka’/‘ka lae’ e comparar estas com as perguntas polares em português. A pergunta ‘Você é de Timor Leste?’ é facilmente comparável porque, se em Tétum não for usada a forma ‘ka’, a sílaba nuclear é ‘Le’ como em português.

Figura 4 – Os contornos polares em Tétum e Português: L+H\*(L)H%.



Fonte: Os autores.

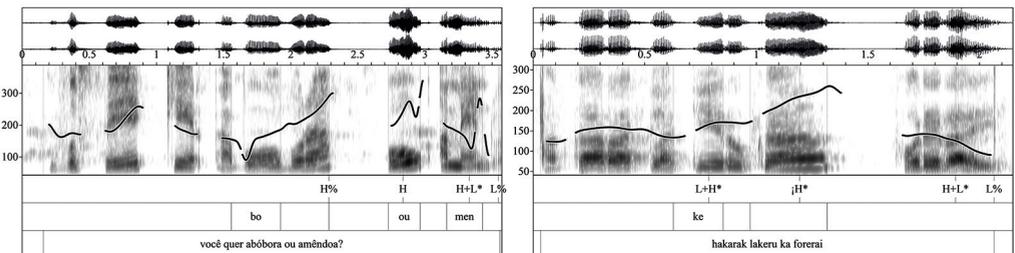
Em ambas as línguas, as perguntas polares se caracterizam por um acento ascendente. O tom de fronteira exibe uma variação, podendo ser ascendente (como mostrado em figura) ou alto, também encontrado e idêntico à forma em ‘ka’ na figura anterior. Podemos concluir, portanto, que no tocante a esse tipo de sentença, há evidência, seja da transferência prosódica, como da reanálise suprasegmental da forma declarativa+‘ka’ como pergunta polar.

## 5 RESULTADOS E DISCUSSÃO: PERGUNTA DISJUNTIVA

Como já mencionamos, perguntas disjuntivas são expressas por duas sentenças unidas pelo marcador ‘ka’ que, por esta razão, foi traduzido aproximativamente com ‘ou’ e que vimos que é reanalisado como marcador polar. Observe-mos a sentença a seguir nas suas línguas:

Você quer abóbora ou amêndoa? / Ita hakerak lakeru ka forerai?

Figura 5 – Os contornos das perguntas disjuntivas.



Fonte: Os autores.

A correspondência dos contornos é muito patente: nas duas línguas podemos notar um tom particularmente alto no conector disjuntivo ‘ou’/‘ka’ e uma

pausa saliente (área branca), logo após este. Notamos também que a primeira parte da frase em Tétum, com seu contorno e fronteira alta, corresponde exatamente a uma pergunta polar (forma com marcador ‘ka’). Tudo faz pensar em uma transferência do contorno entoacional disjuntivo como um todo. Todavia, tal transferência não é realizada com a consciência da função que cada parte exerce na frase, sintaticamente como prosodicamente: diante da semelhança dos contornos, há uma assimetria segmental entre o uso de ‘ou’ em português e o de ‘ka’ em Tétum: enquanto em português é possível quebrar uma pergunta disjuntiva em duas perguntas, sendo a primeira polar, e a segunda uma ‘retomada’ disjuntiva, como ‘Você quer abóbora? Ou prefere batata?’ (mas certamente não como ‘Você quer abóbora ou? Prefere batata?’), em Tétum, tal forma é ausente, porque é impossível começar uma sentença por ‘ka’. Por isso, se a primeira pergunta foi posta em Tétum como polar ‘Você quer abóbora?’, não podemos esperar uma ‘continuação’ disjuntiva.. De fato, o elemento ‘ka’ na Figura 5 é ancorado ao sintagma que o precede, com o qual forma uma pergunta polar, até pela entoação da frase truncada, sendo assim difícil acreditar que ele possa atuar como o núcleo sintático de um ConjP e formar um sintagma juntamente à frase seguinte, diferente do que ocorre em português.

Esse é um exemplo muito singular de duas estruturas, nas duas línguas, que parecem comparáveis pela ordem das unidades correspondentes, mas sua constituição sintática é incomparável, gerando transferência do contorno total sem que seja claro quais são as unidades menores e qual sua função.

## 6 RESULTADOS E DISCUSSÃO: PERGUNTAS PARCIAIS

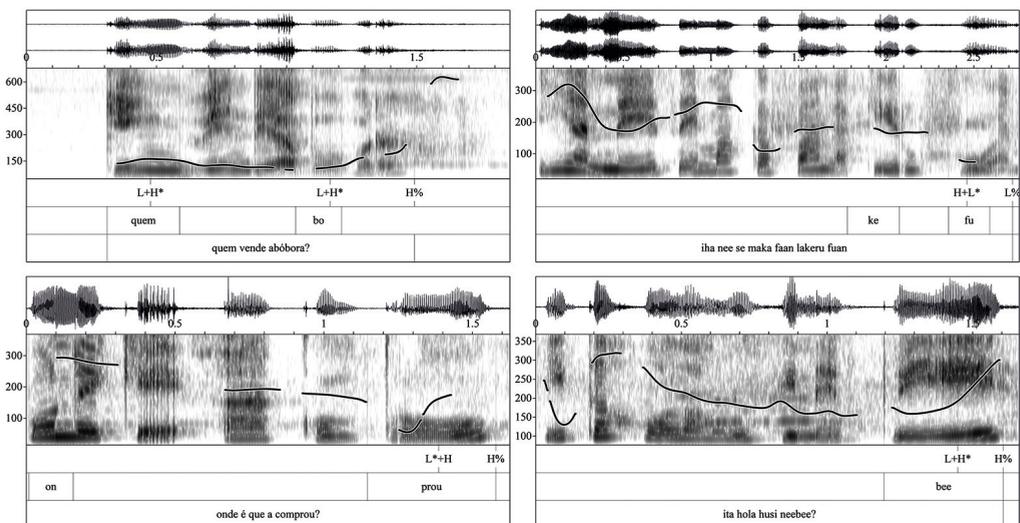
Nas línguas entoacionais como o português, as perguntas parciais (ou perguntas abertas, ou perguntas qu-) costumam levar um acento prénuclear na palavra qu- (assim é chamado o pronome interrogativo ‘quando’, ‘quem’, ‘o quê’, ‘como’, ‘onde’, ‘por quê’...). Na nossa situação, existe uma assimetria que torna trabalhosa a comparação das estruturas entoacionais. Com efeito, enquanto em português a estrutura mais frequente de pergunta parcial coloca o pronome interrogativo no começo da sentença, em Tétum não é possível realizar o movimento do pronome interrogativo qu- para a periferia esquerda, e este é realizado *in situ*, ou seja no lugar onde é gerado sintaticamente, sendo tal lugar o final da sentença, se se tratar de objeto (o quê (ACC)?, quem (ACC)?, a quem?), PP ou

adjunto (por quê?, quando?, onde?, como?...) e à esquerda somente se for sujeito (o quê(NOM)?, quem(NOM)?).

<u>Quem</u> vende abóbora?	<u>Se mak</u>	faan	lakeru?
	<u>Quem</u> que	vende	abóbora
<u>Onde</u> (se) vende abóbora?	Lakeru	faan	<u>neebee</u> ?
	Abóbora	vende	<u>onde</u> ?

Existem, portanto, duas maneiras de tornar as estruturas nas duas línguas comparáveis: 1) utilizar pronomes interrogativos de sujeito em Tétum; ou 2) forçar sua realização in situ em português. O segundo método, porém, é pouco compatível com a estrutura metodológica do DCT, onde o entrevistador não pode mencionar trechos da frase para não condicionar o falante, e uma instrução metatextual que force a realização in situ em português se revela difícil de construir sem mencionar explicitamente as partes do discurso envolvidas. Portanto, resolvemos incluir no questionário em Tétum perguntas parciais que utilizam o pronome interrogativo de sujeito ('quem?').

Figura 6 – Os contornos das perguntas parciais neutras.



Fonte: Os autores.

Os resultados mostram que um tom ascendente é sempre realizado no pronome *qu-*, porém, em posição nuclear, enquanto o português (onde a palavra nuclear não é a palavra *qu-*) sempre exibe outro contorno ascendente com fronteira

alta (LH H%), o Tétum somente produz contorno nuclear ascendente quando a palavra nuclear coincide com a palavra *qu-*, ou seja, pela maioria dos pronomes (adjuntos e objetos), mas não no único caso que justamente corresponde à ordem sintática portuguesa. Na frase ‘*se maka faan lakeru fuan?*’ (‘quem vende abóbora?’) o contorno nuclear é H+L\*. Isso pode ser explicado da forma seguinte: como a realização final da palavra *qu-* é de longe a mais frequente em Tétum, o contorno ascendente é associado à pragmática interrogativa parcial. Em seguida, tal contorno é aplicado ao português ‘na sua totalidade’ sem que os falantes façam a correspondência entre unidades segmentais das duas línguas durante o mecanismo de transferência.

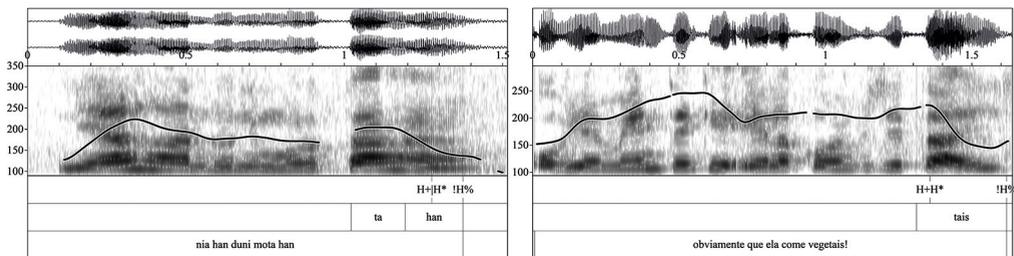
## **7 RESULTADOS E DISCUSSÃO: DECLARATIVAS ÓBVIAS E PERGUNTAS POLARES COM SURPRESA**

As perguntas polares com surpresa são perguntas que expressam a incredulidade do falante a respeito de uma possível resposta positiva. É difícil situar o valor epistêmico da expectativa do falante acerca da resposta, quando realmente este não acredita na possibilidade do que pergunta, ou aceita o improvável como verdade, mostrando apenas sua surpresa. As declarativas óbvias, ou de obviedade, são aquelas em que o falante expressa, acerca do texto afirmado, uma implicatura epistêmica de que não só o julga verdadeiro, mas que o julga obviamente verdadeiro, e sobretudo, a respeito do interlocutor, sua própria surpresa que este possivelmente não o considere igualmente óbvio. As declarativas de obviedade podem ser realizadas como meras exclamativas, outras vezes com um foco contraditório, ou com uma entoação de surpresa, pois elas possuem um viés de contraexpectativa acerca do não compartilhamento com o falante da mesma postura epistêmica.

Com respeito aos dois tipos de frase mencionados, é tarefa delicada e técnica induzir os falantes a produzir as sentenças que incluam todas essas nuances, distinguir entre elas e se certificar que estejam sendo colocadas. Certamente, não esperamos uma classificação exaustiva pela entrevista de apenas 3 falantes. Contudo, devemos anotar os resultados positivos, ou seja, os contornos salientes que foram encontrados, no sentido que não são previsíveis e são distintos de outras modalidades investigadas (declarativa ou pergunta polar neutra) e porventura correspondendo entre as duas línguas. Por isso, não entraremos nos pormenores da análise e apenas apresentaremos, na figura a seguir, um exemplo de par de

contornos das declarativas óbvias, nas duas línguas que sugere transferência, e diremos que as perguntas com surpresa também apresentaram contornos muito semelhantes. O contorno em figura, muito peculiar, mostra uma queda incompleta ou ‘suspensa’ de um tom alto para um tom médio, realizada entre a sílaba nuclear e a fronteira.

Figura 7 – As declarativas óbvias: H+(!)H\* !H%.



Fonte: Os autores.

## 8 CONCLUSÕES

Na tabela comparativa a seguir, os contornos realçados são encontrados em Tétum e, em forma idêntica ou muito parecida, também em português. Embora, em alguns casos, as correspondências segmentais entre as duas línguas não se deem de forma completamente transparente, e uma minoria de contornos sejam previsíveis pelos códigos biológicos ou compartilhadas pelas demais variedades de português, julgamos que os dados fornecem evidência suficiente para postular um mecanismo de transferência sistemático. Em particular, são consideráveis as semelhanças entre os contornos das perguntas polares e parciais, com surpresa e a declarativa óbvia, nas duas línguas, assim como o tom alto no marcador ‘ka’/‘ou’, seja nas perguntas polares, como nas disjuntivas. Resta estabelecer se, para os tipos de sentença onde falta correspondência segmental, a transferência envolve o contorno na sua totalidade fonética, ou há um reconhecimento e consciência das unidades menores de fraseamento correspondentes nas duas línguas.

Quadro 1 – Os contornos do português timorense e do Tétum por tipo de sentença.

	PTL	TP
Declarativa Neutra	H*L%, H+L* L%	H*L%
Pergunta parcial neutra	(LH)H%	(LH)H%
		H+L* L%
Pergunta polar neutra	(LH)H%	(LH)H%
	(LH)LH%	(LH)LH%
Pergunta com surpresa	H+L* H%	H+!H* H%, H+L* ;H%
	(L+H)* !H% (neutra?)	L+H* L-H% (neutra?)
	(LH) L% (exclamativa?)	
Declarativa óbvia	H+H*!H%	H+!H*!H%, (H+)H*!H%
	H*+L L% (neutra?)	H*L% (neutra?)
Pergunta disjuntiva (1º item)	L+H* H%	L+H* ;H%
Pergunta disjuntiva (2º item)	H+L* L%	H+L* L%
	L+H* L% (Hipermarcação da tônica?)	

Fonte: Os autores.

## REFERÊNCIAS

AFONSO, S.; GOGLIA, F. Linguistic innovations in the immigration context as initial stages of a partially restructured variety: Evidence from SE constructions in the Portuguese of the East Timorese diaspora in Portugal. *Studies in Hispanic and Lusophone Linguistics*, v. 8, n. 1, p. 1-33, 2015.

ALBUQUERQUE, D. B. O Português de Timor Leste: contribuição para o estudo de uma variedade emergente. *Papia*, v. 21, n. 1, p. 65-82, 2011.

BARONE, M.; ALBUQUERQUE, D. B. Adaptando o Discourse Completion Task: a documentação da fonologia prosódica em tempos de pandemia. *Enlaces - Revista de Estudos Linguísticos e Literários*, v. 2, 2021.

BARONE, M. On only pragmatically driven intonational change. *In: Proceedings of 11<sup>th</sup> international conference on experimental linguistics. Ex-ling 2020. Aténas, 10-12/10/2020, 2020, p. 54-57, 2020.*

BATORÉO, H. J. The contact induced partial restructuring of the non-dominating variety of Portuguese in East Timor. *In: MUHR, R. (ed.). Pluricentric*

Languages and Non-Dominant Varieties Worldwide: the pluricentricity of Portuguese and Spanish: new concepts and descriptions. Vol. II. Viena/ Frankfurt: Peter Lang Verlag, 2016. p. 137-53.

BLUM-KULKA, S.; HOUSE, J.; KASPER, G. Investigating cross-cultural pragmatics: an introductory overview. *In*: BLUM-KULKA, S.; HOUSE, J.; KASPER, G. (orgs.). *Cross-Cultural Pragmatics: Requests and Apologies*. Norwood, NJ: Ablex, 1989, p. 1-34.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer [Programa de Computador]. Version 6.1.39, 2021. Disponível em: <http://www.praat.org>. Acesso em: 24 fev. 2021.

FROTA, S. *Prosody and Focus in European Portuguese: Phonological Phrasing and Intonation*. New York: Garland, 2000.

GILI FIVELA, B.; AVESANI, C.; BARONE, M.; BOCCI, G.; CROCCO, C.; D'IMPERIO, M.; GIORDANO, R.; MAROTTA, G.; SAVINO, M.; SORIANELLO, P. Varieties of Italian and their intonational phonology. *In*: FROTA, S.; PRIETO, P. (orgs.). *Intonational phonology of the regional varieties of Italian*. Oxford: Oxford University Press, 2015, p. 140-197.

GORDON, M. The phonology of pitch accents in Chickasaw. *Phonology* 20, p. 173-218, 2003.

GUSSENHOVEN, C. Foundations of intonational meaning: Anatomical and Physiological Factors. *Topics in Cognitive Science*, 8(2), p. 425-434, 2016.

HULL, G. A Morphological Overview of the Timoric Sprachbund. *Studies in Languages and Cultures of East Timor*, 4, p. 98-205, 2001.

PIERREHUMBERT, J. *The phonetics and phonology of English intonation*. Tese de Doutorado, Massachusetts Institute of Technology, 1980.

SILVERMAN, K.; BECKMAN, M.; PITRELLI, J.; OSTENDORF, M.; WIGHTMAN, C.; PRICE, P.; PIERREHUMBERT, J.; HIRSCHBERG, J. TOBI: a standard for labeling English prosody. *In*: J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge and G. E. Wiebe (eds.). *Proceedings of the 1992 International Conference on Spoken Language Processing*, vol. 2. Department of Linguistics, University of Alberta, 1992, p. 867-870.

# DESVIOS ORTOGRÁFICOS A SEMIVOCALIZAÇÃO DA LATERAL /L/ NO CONTEXTO DO ENSINO PRIVADO

*André Pedro da Silva*

Universidade Federal da Bahia

*Yasmin Maria Macedo Torres Galindo*

Universidade Federal de Pernambuco

## 1 INTRODUÇÃO

Os sujeitos, quando imersos no letramento, já nos primeiros anos, passam por um processo de quebra na familiaridade da língua. Estão entrando em contato com uma outra modalidade até então apenas testada, começam a imergir no campo da norma e os contatos tornam-se simbióticos e de embate à medida que avançam no conhecimento da escrita. A fala, que já está sob seu domínio, vaza para a escrita, e é por meio destes vazamentos que investigamos os processos fonéticos e fonológicos que emergem desse embate.

A criança passa por um grande processo de aprendizagem e, em um domínio ortográfico, percebe-se que o indivíduo, em fase de aquisição da ortografia, busca organizar suas próprias regras em um plano interiorizado a partir das impressões que capta do meio (MORAIS, 2003). A escrita inicial tem como ponto de partida, abundantemente, a fala; uma vez que essa representa uma instância mutável e de caráter individual, intrinsecamente marcada por vivências. Assim, a criança, nosso sujeito investigado, tende a espalhar as marcas da oralidade, que captou em seu convívio, na escrita, fazendo uma correlação entre estes dois modelos.

Partindo desse pressuposto, o do vazamento, cabe pontuar a figuração da língua falada na análise do desenvolvimento da criança enquanto sujeito. Porém, devemos ter e mente, enquanto pesquisadores da uma individualidade latente, que nem sempre foi assim, dessa forma. Faz-se necessário um primeiro panorama histórico: no século XX, com o desenvolvimento da linguística moderna, a fala ocupa o lugar de objeto de análise e a escrita é marginalizada enquanto fonte de coleta de dados para análise, por ser julgada como um constructo que simplesmente reproduzia a fala – o que hoje discordamos. Nos dias atuais, no entanto, sabemos, por meio de estudos sociolinguísticos desenvolvidos a partir da década de 1970, que a escrita e a fala possuem relações bastante complexas e que dizem muito mais do que uma mera reprodução da norma culta na língua. A escrita é um campo vasto a ser analisado.

As postulações de Luria (1986) já ressaltam que a palavra é a “célula da linguagem”, ou seja, é a base de compreensão para que os indivíduos estabeleçam interações entre si e consigo, tudo passa pela palavra para obter “existência”. Baseados nesse conceito, aludimos a Vygotsky (2009), que comprova que a palavra se desenvolve a partir da interação do sujeito com o mundo, ou seja, a partir das funções psíquicas superiores, tais como a memória, a percepção e o pensamento. De forma análoga, a palavra, no domínio da linguagem escrita, também não acontece de forma simples, não basta apenas a aquisição mecânica dos signos visuais, que aqui chamamos de letras, há uma necessidade da apropriação da subjetividade do domínio desse código, para que haja a correspondência eficiente entre grafema e fonema. Dentro dessa perspectiva, Martins (2011, p. 146) diz que “o domínio da linguagem escrita representa para a criança o domínio de um sistema simbólico altamente complexo e dependente, em alto grau, do desenvolvimento das funções psíquicas superiores do comportamento infantil”. Destarte, a escrita acompanha o desenvolvimento cognitivo da linguagem de cada indivíduo ao longo da vida. Assim, toda escolha na escrita conta a história de uma exclusão, e essa escolha diz muito da capacidade linguística de seu emissor.

Baseamo-nos ainda nos paradigmas instaurados por William Labov, na Teoria da Variação Linguística, ressaltamos o caráter heterogêneo da língua e a importância desta heterogeneidade linguística em uma sociedade que também se faz heterogênea. A língua, por sua vez, enquanto construção social, tende a se adaptar aos tempos e aos indivíduos. Aqui, interessa-nos a língua de caráter individual, a parole para Saussure, nos remotos tempos do Estruturalismo, a fala para nós; sendo assim, a língua caracteriza-se como mutante frente às diversas intervenções internas, como, por exemplo, no campo cognitivo

(psicológico) ou no campo de fatores exteriores às personas de um determinado grupo (fatores sociais).

A pesquisa aqui apresentada baseia-se, sobretudo, no método quantitativo-comparativo; ou seja, foram analisados dados estatísticos para que nossa hipótese norteadora fosse provada com totalidade. Buscamos como objetivo descrever estatisticamente a manifestação das variáveis presentes no fenômeno de semivocalização da lateral líquida /l/, com a finalidade de analisar, apreender e sistematizar essa variante nos diversos contextos, nos quais ela pode se manifestar, no nosso caso, o contexto do ensino fundamental particular; e discutir quais os fatores influem sobre seu aparecimento ou desaparecimento. Assim, foram analisados um total de 52 treinos ortográficos de frases e 52 de palavras. Para tanto, o colégio usado como campo de pesquisa, localizado no bairro do Ibura, Jaboatão dos Guararapes-PE,<sup>1</sup> foi escolhido previamente para a coleta de dados, oferecendo o total de 26 sujeitos femininos e 26 sujeitos masculinos, divididos igualmente para as séries do 2º ao 5º ano do Ensino Fundamental I.

Dividimos em dois níveis de coleta dos treinos ortográficos, frases e palavras, para poder observar o momento de atenção e suas variações possíveis dentro do contexto de sala de aula. Assim, no treino de palavras, a ideia era a de que o aluno se preocupasse apenas com a palavra em si, feito por intermédio de um ditado de vinte palavras, todas com a manifestação da lateral em posição de coda; e um momento de espontaneidade, presente na produção do treino de frases, momento em que o informante se preocupa com todas as palavras, uma vez que estas estão contextualizadas. Dessa forma foi possível estabelecer comparações entre estilos: menos monitorado (treino de frases) e mais monitorado (treino de palavras).

Antes de investigar o evento fonético em si, traçaremos um percurso por dentro da construção no indivíduo, Panorama Socioconstrutivista, e, posteriormente, mergulharemos no fenômeno em si com o tópico Panorama Fonético: a semivocalização da lateral /l/.

---

<sup>1</sup> O bairro do Ibura é uma grande periferia constituída de diversas unidades residenciais, sendo um bairro periférico, é formado em sua maioria por indivíduos pertencentes às classes C e D, e alguns lugares de maior precariedade que constituem a classe E. Esse não é caso da unidade residencial 6, onde fica localizado o colégio campo da pesquisa. Essa escola adota um método de ensino majoritariamente tradicionalista, os indivíduos em questão estão na faixa etária entre 07 e 11 anos, não apresentando destoamentos entre o ano escolar e a faixa etária correspondente.

## 1.1 Panorama Socioconstrutivista

É importante também que tenhamos uma visão para além da *Hard-Science Linguistic*, é necessário que investiguemos a construção desse sujeito em meio a um contexto determinado. A partir de uma perspectiva socioconstrutivista da linguagem, podemos nos apoiar no ponto de vista explanado por Vygotsky em seus estudos sobre a linguagem, para validar o caráter estritamente ligado a meios culturais da aquisição da língua de um determinado grupo. Mas ele não exclui os processos psicológicos naturais e inatos a todo e qualquer ser humano em desenvolvimento. Assim,

Vygotsky concebe o homem e seu desenvolvimento numa perspectiva sociocultural. Contudo, referindo-se aos estudos experimentais da formação dos conceitos, o autor afirmou que “[...] a tarefa cultural, por si só, não explica o mecanismo de desenvolvimento em si, [...]” (Vygotsky, 1998, p. 73). Por meio desta afirmativa explicitou a necessidade de compreender o desenvolvimento de um indivíduo de maneira global. Noutras palavras, salientou o quão importante é verificar todas as funções implicadas nesse processo, as quais continuamente são construídas, reconstruídas e incorporadas a uma já existente ou nova estrutura, porque durante a formação dos conceitos este indivíduo aprende “[...] a direcionar os próprios processos mentais com a ajuda de palavras e signos [...]” (MIRANDA, J.; SENRA, L., 2012, p. 6).

Observada notadamente nas pesquisas de campo que abordam a aquisição da linguagem e o letramento, inclusive numa pesquisa dessa temática, realizada pelo Projeto Relação entre Fala e Escrita – REFALES,<sup>2</sup> a teoria de Vygotsky faz-se presente junto à sociolinguística de Labov, já que um indivíduo “X”, ainda que dotado de suas totais capacidades psíquicas e cognitivas, tem na cultura em que é imerso o potencializador das descobertas da linguagem.

Esse ambiente propício ao treino e ao teste da língua com finalidade de aperfeiçoamento é encontrado nas séries iniciais do ensino básico e fundamental escolar. A criança imersa e participante de um ambiente escolar sadio passará pelos testes de linguagem e aperfeiçoará sua língua, tanto escrita como falada, aproximando-se cada vez mais da norma padrão vigente numa sociedade.

Língua também é cultura, sendo assim, aperfeiçoar-se tal qual a norma vigente, nesse caso, do Português Brasileiro (doravante PB), é adentrar na comunidade e tornar-se parte dela. No percurso que a criança faz até o aperfeiçoamento

---

<sup>2</sup> *O processo de Harmonização Vocálica Nominal na escrita de alunos do Ensino Fundamental de escola pública na cidade de Jaboatão dos Guararapes-PE*, pesquisa realizada sob o intermédio do programa de Bolsas de Incentivo Acadêmico (BIA), custeado pela FACEPE, realizada nos anos de 2014-2015.

de seu uso linguístico, ela flerta com diversas variantes linguísticas, conceituadas por Tarallo (1986, p. 8) como sendo “diversas maneiras de se dizer a mesma coisa em um mesmo contexto e com o mesmo valor de verdade”. Segundo Tasca (2002),

[...] a variação, embora aparentemente caótica e aleatória, constitui um objeto de estudo científico, uma vez que heterogeneidade das línguas é imanente às mesmas, podendo ser prevista e sistematizada. Uma das metas dos estudos da variação é descrever e explicar os usos alternantes, indicando seu caráter ou de mudança ou de progresso (TASCA, 2002, p. 17).

A variação é nosso objeto de observação e figura para Tasca (2002) como sinalizadora do progresso do desenvolvimento interno da língua, vemos claramente o intrínseco relacionamento da sociolinguística com a psicologia do desenvolvimento.

Nas séries iniciais, o indivíduo começa a testar a língua com a finalidade de acomodá-la em sua mente, ou seja, uma criança testa as diversas variantes que lhe são possíveis, cabendo ao professor e/ou instrutor de aprendizagem corrigi-la. A tendência de testagem nas séries iniciais é associar a fala à escrita. Logo, crianças menores que adquiriram a escrita recentemente tendem a harmonizá-la com a fala, sendo a fala a guia das escolhas variacionais dos indivíduos em processo de letramento. Novamente, fazendo um pouso na psicologia do desenvolvimento, Santos afirma que:

[...] o processo de aquisição da fala, nas crianças, é um processo de aprendizagem das regras de um jogo de linguagem. Nesse caso, o jogo é o próprio ato de pronunciar as palavras. Ele observa que existem formas corretas e erradas de pronunciar as palavras e que estas formas são estabelecidas pela cultura. Quando começa a falar, a criança geralmente pronuncia as palavras de um modo errado, do ponto de vista cultural. Por isso é preciso ensinar a criança a pronunciar as formas corretas das palavras. Portanto, desde o início do processo de aquisição da fala a criança encontra-se mergulhada em um jogo de linguagem – neste caso, o ato de falar – e na cultura (SANTOS, 2008 apud WITTGENSTEIN, 1978, p. 165).

Usando como base o pensamento de Wittgenstein acerca da aquisição dos mecanismos da fala, podemos concluir que a aquisição da escrita não se faz de forma muito diferente, o espelhamento fala-escrita é natural, como vai nos apontar Mollica (1998) ao dizer que:

As crianças são praticamente categóricas até dois anos mais ou menos. Os processos variáveis e cancelamento ou inserção de segmentos em estruturas silábicas quase não apresentam alternância nos primeiros estágios, porque os nativos em qualquer língua

realizam padrões silábicos simples do tipo V e CV no início da aquisição linguística, [...]. Em geral o processo de auto-correção se responsabiliza pela solução da maioria dos vestígios de oralidade que venham a permanecer na escrita, quanto maior o grau de experiência com a modalidade escrita é mais elevado o nível de escolarização. Mas cada regra variável tem peculiaridades quanto à sua gênese na língua falada, fato que deve ser levado em conta de forma particularizada em trabalho que considere a relação fala e escrita sob um enfoque aplicado (MOLLICA, 1998, p. 20).

Logo, a criança testa as diversas possibilidades de transcrever o que fala e ouve, e esse jogo da linguagem falada com a linguagem escrita resultará nas diversas formas de escrever uma mesma palavra, numa tentativa de decodificação desse som.

Sendo assim, as variantes, segundo Labov, existirão em dois contextos: na coocorrência, quando duas formas são usadas ao mesmo tempo, observada, sobretudo, em momentos do letramento mais iniciais, momento em que as crianças testam as diversas formas de escrever a mesma palavra; e na concorrência, quando as duas formas ocorrem tanto a variável quanto a padrão, ambas com o valor de verdade, mais comum quando o processo de letramento ou obteve êxito, no caso do uso da norma padrão, ou fracasso, no caso da insistência na variável.

Tendo esclarecidos os diversos processos pelos quais a criança passa ao longo da jornada de adquirir a língua falada e a escrita, o presente trabalho preocupar-se-á com o estudo das expressões variacionais encontradas no fenômeno de semivocalização da lateral /l/, transformando-a em /w/, natural na língua falada. A transformação escrita será expressa pela substituição da consoante líquida *l* pela vogal alta *u*.

## 1.2 Panorama Fonético: a semivocalização da lateral /l/

A semivocalização da lateral em coda silábica passa longe de ser uma manifestação variável recente, ela figura desde o latim na constituição da língua e supõe-se que a passagem do *l* para o *u* exista desde a passagem do latim clássico para o vulgar. A busca pela imitação do *sermo urbanus* pelo *sermo vulgaris* resulta pela primeira vez nessa semivocalização da lateral. Teyssier aponta para o ressurgimento desse movimento na língua quando o PB já é semiconstituído. Dessa maneira, a semivocalização passa a ser uma sinalização do distanciamento do PB do Português de Portugal (PP), que preserva a pronúncia velar da lateral.

A líquida /l/ é observada nas posições de *onset* (inicial e medial) e coda (medial e final), a qual, em geral, é produzida como o glide [w]. É a líquida /l/, na posição de coda, que vai nos interessar neste estudo, justamente por ser produzida como glide /w/ de sonoridade igualitária, a vogal alta /u/ tende a associar à vogal *u* os méritos da consoante líquida *l*, como, por exemplo: ca[l].da ~ ca[w].da; ca.na[l] ~ ca.na[w].

Essa consoante líquida obtém esse nome justamente pelo seu caráter mutável, sendo a única, segundo Dickey (1997), a possuir todos os pontos de articulação do aparelho fonador em diversas línguas do mundo. A líquida /l/, além de vocalizada, como em cana[w], pode ser velarizada, como em cana[ɣ], ou apagada, como em cana[Ø], como ressaltam De Pinho e Margotti (2010), ao analisarem as ocorrências da /l/ em coda silábica nas capitais brasileiras; dessa forma, há, em todo o Brasil, a manifestação da variante e do apagamento, porém a velarização do /l/ em coda final como em cana[ɣ] vai se restringir ao sudeste e, principalmente, ao sul. Crê-se, haja vista estudos anteriores, que as vogais anteriores à lateral é que ditarão o comportamento dessa lateral no contexto escrito. Acredita-se nisso porque, segundo Silva e Sedrins (s.d.), em um estudo sobre o comportamento variacionista do segmento lateral na escrita das séries iniciais de crianças de Serra Talhada-PE, apontaram que:

[...] entre os fatores linguísticos, o fator mais significativo que condiciona a vocalização da lateral é a variável tipo de vogal precedente, conforme evidenciado em Tasca (2002), onde tal variável mostrou-se bem relevante. A tendência revelada pelos dados da pesquisa evidencia que os alunos tendem a preservar o segmento e a vocalizá-lo diante das vogais baixas: baixa [a], da média baixa [E], e da vogal alta [i] confirmando a influência da qualidade dessas vogais na vocalização da lateral na coda. Já em relação à vogal média alta [o], a pesquisa mostrou que o segmento é favorável ao apagamento (SILVA, J.; SEDRINS, A., s.d., p. 1).

Nosso estudo, guiado à luz da literatura aqui explicitada, parte para a investigação da vocalização da lateral /l/, em posição de coda, na escrita de crianças participantes da rede do ensino particular, diferentemente da pesquisa citada anteriormente realizada por Silva e Sedrins (s.d.).

A seguir, apresentaremos os resultados quantitativos de nossa pesquisa. Ali, os dados obtidos serão expostos, comparados e analisados de modo a comprovar nossas hipóteses apresentadas no panorama introdutório.

## 2 RESULTADOS E ANÁLISES

Apresentar-se-ão aqui a catalogação dos resultados, bem como a análise desta pesquisa, a qual se dividiu em dois grupos: Fatores Fonológicos e Fatores Sociais, pois sabemos, anteriormente, que ambos se completam na aquisição da linguagem. Como bem pontua Tasca (2002), “a variação, embora aparentemente caótica e aleatória, constitui um objeto de estudo científico, uma vez que a heterogeneidade das línguas é imanente às mesmas, podendo ser prevista e sistematizada”. É justamente essa sistematização das congruências e pontos críticos que nos propomos aqui.

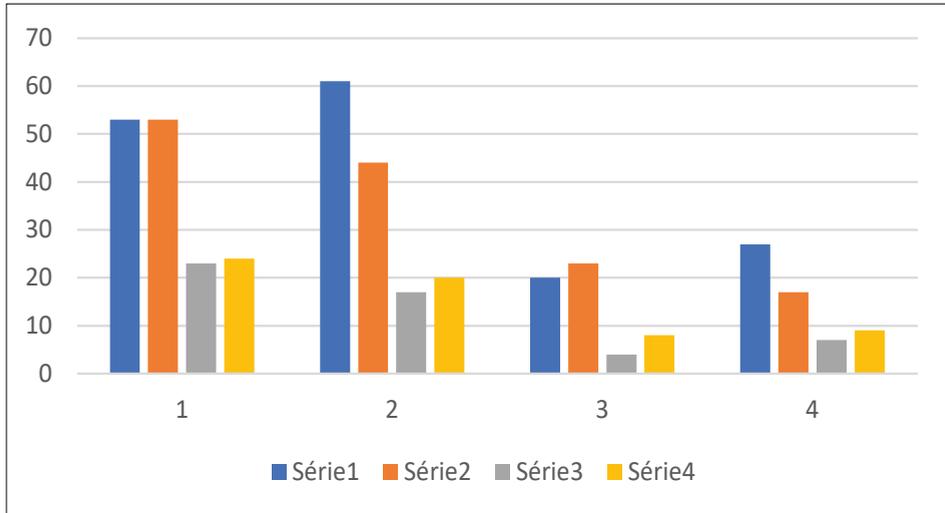
### 2.1 Fatores Fonológicos

Antes de direcionar nosso foco para a posição do /l/ em final de sílaba, posição de coda, é necessário observarmos a manifestação do surgimento da líquida lateral no falar infantil.

Em geral, as consoantes laterais surgem primeiro, e são, grosso modo, consoantes articuladas que, ao serem produzidas, emitem o ar pelas laterais da língua ao invés de pelo centro da língua que é o movimento de escape de ar mais comum. São líquidas laterais o /l/ e o /ʎ/. Segundo Wiethan, Melo e Mota (2011), ao tratarem das consoantes líquidas, o /l/ ainda vai apresentar sobre o /ʎ/ maior facilidade de produção, sendo a primeira das laterais líquidas a ser internalizada, entre os 2:8 e 3:0 anos de idade.

Prosseguindo debruçados aos estudos dos desvios fonológicos possíveis das consoantes líquidas, Wiethan, Melo e Mota (2011) perceberam, após o levantamento dos seus dados, que o fonema /l/ apresentará maior índice de produções corretas, sendo pouco ou nunca acometido pelo que chamaremos de estratégias de reparo. E, ao analisarmos treinos ortográficos que tratem da posição do /l/ em coda final e medial, o parâmetro mudará, dependendo de fatores linguísticos presentes na sílaba analisada, o /l/ pode ser vocalizado ou até mesmo, apagado. Desse modo, deparar-nos-emos com o que se chamará de estratégia de reparo vocalizada. Observemos a manifestação do fenômeno de semivocalização em escrita, quando levamos em conta o contexto, as palavras ou frases, a escolaridade e, sobretudo, a posição do /l/ na sílaba:

Gráfico 1 – Variante [w] na posição silábica (coda final ou medial) em contexto (palavra ou frase).



Fonte: Elaborado pelos pesquisadores, 2017.

Tasca (2002) pontua que a realização dessa semivocalização nos estudos, sobretudo, realizados no Sul, foi vista durante muito tempo como inexistente, apenas o *l* velar /ʎ/ e o dental /l/ eram entendidos como possíveis. Isso desponta para a visão da diferença dos falares das regiões, como sabido, as vogais no território nordestino são pronunciadas mais abertamente; o gráfico anterior afirma essa existência da variante [w] abundantemente e alinha-se com o pensamento de Câmara Jr:

[...] além do movimento da ponta da língua junto com os dentes, há um levantamento do dorso posterior da língua para junto do véu palatino, dando o que provavelmente os gramáticos latinos chamavam o *l* pinguis ou “gordo”. Daí decorre uma mutação, que em linguística diacrônica chama-se “vocalização” da consoante: cessa a elevação da ponta da língua junto aos dentes, a elevação posterior do dorso da língua não chega a interromper a corrente de ar e há um concomitante leve arredondamento dos lábios. O resultado é um /u/ assilábico, mal torna-se homônimo de mau, vil de viu e assim por diante (CÂMARA JR, 1973, p. 51).

Outro ponto a ser observado no gráfico é a predominância da semivocalização quando se trata de coda medial, pois a pronúncia da lateral fica acentuadamente vocalizada quando, após seu aparecimento, existe outra consoante, como bem exemplificamos nos exemplos retirados dos treinos de todas as séries analisadas:

- (1) Aluno 6, 2º ano, 6 anos: ta[w]co (sic)
- (2) Aluno 10, 3º ano, 7 anos: ca[w]deirão (sic)
- (3) Aluno 14, 4º ano, 8 anos: pa[w]co (sic)
- (4) Aluno 3, 5º ano, 11 anos: sa[w]chixa (sic)

Ao deparar-se com as palavras usadas nos treinos de coleta de dados como: soldado [sow'dadu] (coda medial) e jornal [joh'naw] (coda final), haverá a tentativa por parte da criança em teste de reparar vocalicamente a dúvida do uso da consoante /l/, trata-se, então, do /l/ entendido como glide /w/. Observemos a análise das sílabas que apresentam esse fenômeno<sup>3</sup>:

- em soldado: /sol/
- em jornal: /nal/

Em “soldado”, após o núcleo /o/ (vogal média baixa), há a expectativa do predomínio do apagamento do /l/ em posição de coda na transcrição ortográfica da palavra, resultando em “sodado” (Aluno 5, 2º ano, 6 anos) que pode também ser entendido como apagamento de coda, já que se forma, assim, um som homorgânico. Tal apagamento também foi observado por Silva e Sedrins (s.d., p. 6), quando afirmam que “[j]á em relação à vogal média alta [o], a pesquisa mostrou que o segmento é favorável ao apagamento”. Crê-se, com isso, que a proximidade sonora da vogal média alta /o/ com a vogal alta /u/ seja o maior fator causador desse apagamento, já que a estratégia de reparo vocalizada é a transformação ortográfica do /l/ em /u/, ambos elementos labiais.

Porém, no caso da nossa pesquisa, comparando o aparecimento da semi-vogal com o zero fonético, a primeira mostra-se muito mais presente, como em “soudado” (Aluno 10, 3º ano, 7 anos) e essa escolha acontece mais uma vez pela enunciação da vogal na região que é aberta. O apagamento do /l/ fica muito mais evidente quando tratamos da palavra “multidão”, exatamente pelo contexto precedente tratar-se de uma vogal alta, como afirmam Silva e Sedrins (s.d.):

- (5) Aluno 6, 2º ano, 6 anos: mu[ø]tidão
- (6) Aluno 10, 3º ano, 7 anos: mu[ø]tidão

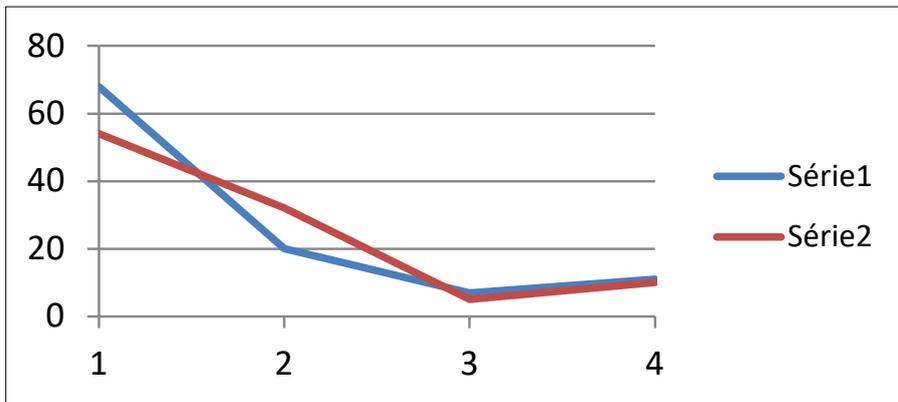
<sup>3</sup> Leia-se: A por Ataque; N por Núcleo e C por Coda.

(7) Aluno 14, 4º ano, 8 anos: mu[ø]tidão

(8) Aluno 3, 5º ano, 11 anos: mu[ø]tidão

A seguir, veremos o gráfico geral do apagamento, salientamos que o maior número de apagamento veio de sílabas formadas pela vogal alta *u*, como dito anteriormente nos exemplos. No gráfico, observaremos como se comporta o zero fonológico levando em consideração variantes, como o tipo de treino e o tipo de coda.

Gráfico 2 – Manifestação do zero fonológico – Apagamento da lateral quanto aos tipos de treinos ortográficos.

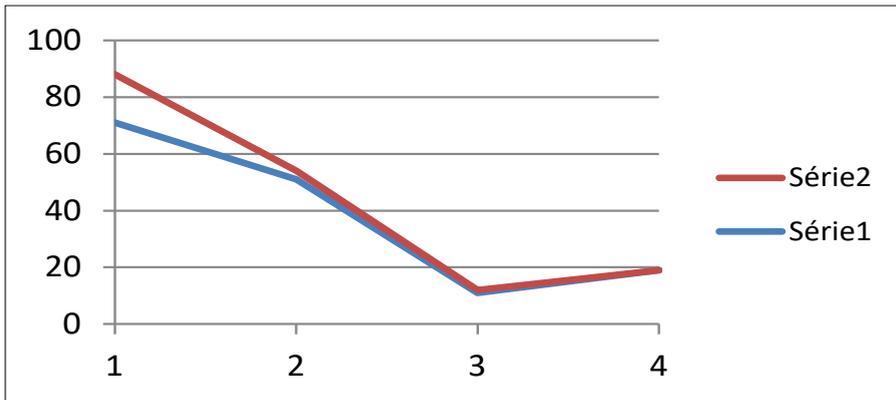


Fonte: Elaborado pelos pesquisadores, 2017.

O Gráfico 2 demonstra a diferença estatística da manifestação da variação dentro dos contextos palavra e frase, a saber. Dessa diferença, com a menor manifestação do apagamento quando no contexto espontâneo (frase), pode-se teorizar que as crianças usem mais de seus conhecimentos internalizados, quando estas se encontram à vontade, por exemplo, não há a repetição do vocábulo visando corrigi-lo. Dessa forma, o mecanismo da hipercorreção não é usado. Já em nível de palavras, há a repetição em busca do aprimoramento, e nessa tentativa de perfeição, pode-se perder ou ganhar fonemas, nesse caso, acarreta a perda.

Ainda embasados na questão do apagamento, passemos agora à observação da manifestação da variante zero levando em consideração sua posição na palavra, se em coda medial ou final.

Gráfico 3 – Manifestação do zero fonológico – Apagamento da lateral quanto ao tipo de coda.



Fonte: Elaborado pelos pesquisadores, 2017.

Já em palavras como “jornal”, o /l/ aparece em posição de coda final e é abraçado pela nossa hipótese da estratégia vocalizada. Ou seja, o /l/, uma vez que foneticamente transcrito, transforma-se em glide que ortograficamente transformar-se-á em vogal alta *u*, resultando na variável “jornau” (Aluno 9, 2º ano, 6 anos). A manifestação escrita da vogal alta *u* e seu potencialmente inexistente apagamento, diante de um núcleo silábico que contenha uma vogal baixa, dar-se-á, supõe-se, pela oposição sonora de *a – u* e pela inexistência de uma consoante após seu aparecimento, como no exemplo anteriormente mencionado.

Hora (2006), em pesquisa semelhante a esta aqui apresentada, chegou à conclusão de que o fator contexto fonológico precedente seria mais atuante na manifestação da variação do que os fatores sociais:

Embora o apagamento da lateral em posição final possa atingir qualquer uma das vogais que preenchem o contexto fonológico precedente, ele será estigmatizado entre as pessoas escolarizadas, exceto se essa vogal for [u]. Em interior de palavra, o apagamento só ocorre se o contexto fonológico precedente for uma vogal posterior, incluindo a vogal [o]. O mesmo não acontece com as demais vogais, que resultarão em itens mal-formados (HORA, 2006, p. 12).

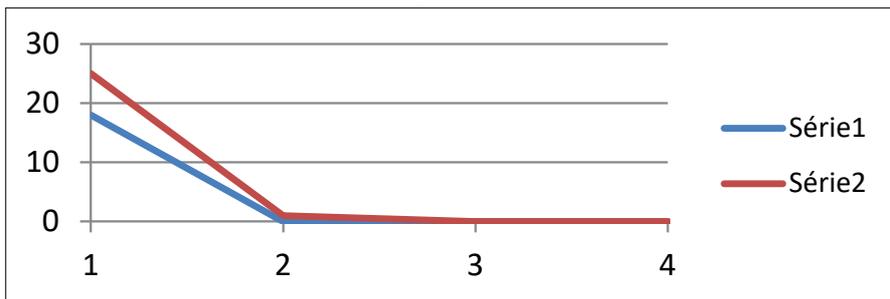
A tendência do /l/ pós-vocálico em regiões do nordeste (cf. HORA, 2006; SILVA; SEDRINS, s.d.) é ser vocalizado, uma vez que as crianças em fase de aprendizagem e letramento, sobretudo, em séries iniciais, tendem a adequar a escrita à sonoridade, fazendo uso da estratégia de reparo vocalizada. A criança ainda não tem cristalizada cognitivamente a noção linguística de que o /l/ pode representar não somente um fonema alveolar, mas também um fonema vocal, já

que /l/ em posições pós-vocálicas, de coda medial e final, em regiões do nordeste é pronunciado tal qual um *u*. Porém, essa regra não valerá para todo o território nacional, sobretudo na região sul do Brasil, em que se detecta outra manifestação do /l/, fato este que foi sabiamente observado por Câmara Jr.:

[...] o contraste entre /l/ e /w/ depois de vogal não deve ir ao ponto de se articular o /l/ depois de vogal exatamente como o /l/ antes de vogal. Salvo no extremo sul do país, esta pronúncia indiferenciada soa anômala, e dá a impressão de haver um ligeiro /i/ depois do /l/ de maneira que uma palavra como cal quase se confunde com cale ou mel com mele (CÂMARA JR., 1973, p. 31, grifos do autor).

Outro fenômeno encontrado na análise dos dados pesquisados foi o aparecimento da vogal /o/, majoritariamente em posição de coda, no final de palavras, sobretudo nas séries iniciais. Mas, essa tende a desaparecer rapidamente com o avanço do letramento e da internalização da Língua Portuguesa. Esse fenômeno mostrou-se bem específico do nosso estudo e, após a análise das suas manifestações, chegamos à conclusão de que acontece uma alofonia.<sup>4</sup> Observemos o Gráfico 4:

Gráfico 4 – Surgimento da variável /o/.



Fonte: Elaborado pelos pesquisadores, 2017.

Assim, a noção de que o que deve ser transcrito é a semivogal /w/ acaba levando as crianças a optarem pelo fonema /o/, que pode também possuir a manifestação /w/, como em ovo [‘ovw]. Logo, a criança transforma o /l/ em /w/ e opta por transcrevê-lo com o fonema /o/, representado pela vogal *o*, como em “varao” ~ varal e “paoco” ~ palco. Passemos agora aos Fatores Sociais.

<sup>4</sup> Fenômeno pelo qual diferentes variantes, ou alofones, apresentam equivalência funcional, podendo ocorrer em **distribuição complementar** ou variação livre. No português, um exemplo de alofonia seria a palatalização ou africacão das oclusivas dentais /t/ e /d/ antes da vogal [i] (SILVA, 2011, p. 53).

## 2.2 Fatores Sociais

Além dos fatores linguísticos referentes à posição do /l/ da sílaba fonética, são os fatores culturais notadamente influentes a esta variação, tal qual discutida anteriormente e apoiada em questões socioconstrutivistas (cf. MIRANDA; SENRA, 2012; SANTOS, 2008). Os fatores culturais manifestar-se-ão quando levados em conta fatores de contexto de aplicação do teste e de nível de alfabetização.

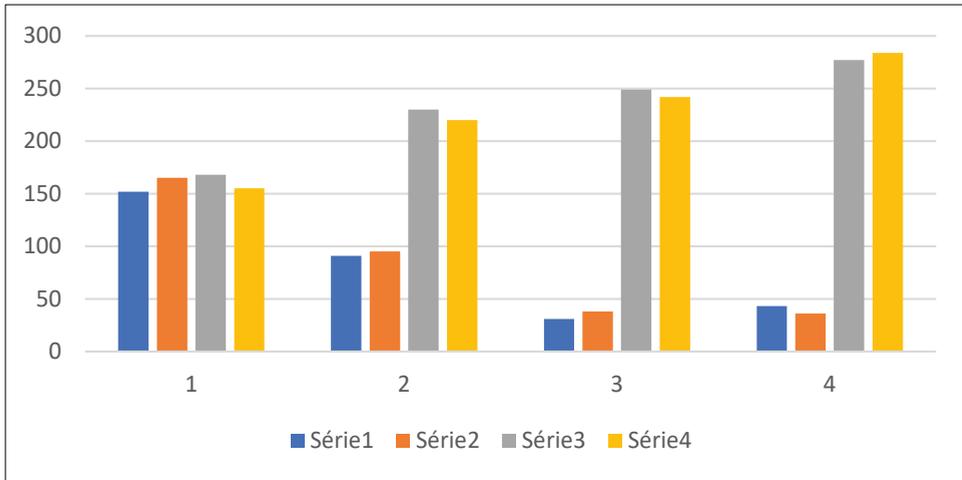
### 2.2.1 Variável Escolaridade

Sabe-se que a construção cognitiva da escrita visa esse gradual afastamento da fala, já discutido no presente artigo. Tais resultados obtidos resultam num entendimento que ainda que profundamente ligadas, a fala e a escrita muitas vezes não são correspondentes entre letra-som. A grafia corresponde a um sistema análogo, porém, não gêmeo da oralidade. A criança então, com a ajuda da escola, passa a compreender as duas formas de manifestação da língua, não como unas, mas como dependentes. Sobre esse amadurecimento gradual, Pedrosa e Nascimento (2014) são pontuais ao afirmarem que:

[...] o processo de escrita apresenta um grau considerável de dificuldade para a criança no início de sua aquisição. Isso porque o nosso sistema de escrita apresenta dois tipos de organização. O primeiro deles é baseado na proposta alfabética, em que há uma correlação entre a fala e a escrita e o segundo, que diz respeito a uma sistematização ortográfica, que busca anular a variação linguística e propõe uma normatização sem tomar por base a oralidade. Dessa forma, a criança precisa sistematizar dois tipos de conhecimentos: um em que pode fazer uso de conhecimentos prévios advindos da oralidade; outro do qual só terá conhecimento a partir da alfabetização (PEDROSA; NASCIMENTO, 2014, p. 3).

O fator escolaridade também está ligado à cultura. O tipo de escolaridade do informante fala não somente dele, mas também do grupo social a qual pertence. Esse trabalho propõe-se, sobretudo, comprovar que cada vez que uma criança ascende uma série do ensino básico, sua escrita afasta-se mais da fala, ou seja, nas séries iniciais, os dados de escrita estão mais próximos da fala e, na medida em que se dá o letramento, com o aumento da escolarização, tende a se afastar desta (cf. MORAIS, 2003, 2007; MARCUSCHI, 2008; FARACO, 2012). Como pode-se observar no gráfico a seguir:

Gráfico 5 – Variável Escolaridade.



Fonte: Elaborado pelos pesquisadores, 2017.

Idade e nível de escolarização, principalmente, no eixo social escolhido para a coleta, o eixo particular de ensino, estão mais interligados. As crianças do ensino fundamental I (2º a 5º ano) demonstram uma homogeneidade quanto à idade dos informantes. Todos seguem uma faixa etária parecida, sem grandes discrepâncias por grupo. E o desaparecimento do uso da variável segue o fluxo esperado sem maiores problemas de aquisição da língua padrão.

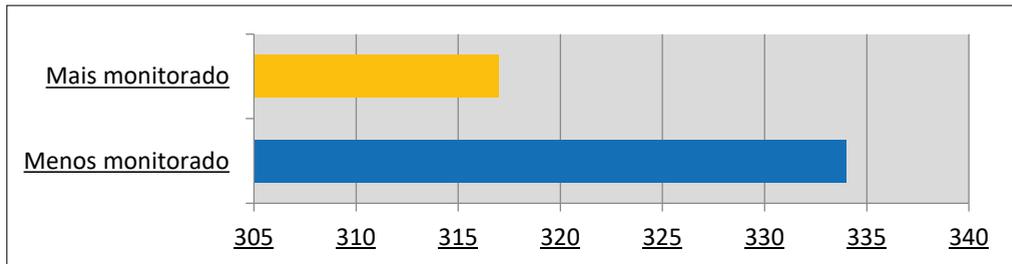
### 2.2.2 Contexto de Aplicação

A escolha pela aplicação dupla dos testes que geraram os resultados da pesquisa não foi feita de forma arbitrária. Sendo assim, a escolha por essa modalidade de coleta de dados teve como finalidade observar o comportamento das crianças frente às escolhas possíveis.

O pesquisador teve como ferramenta o uso de dois tipos de ditado: um que contava com uma atividade voltada para a completude de frases, contendo imagens, visando à criação de um ambiente mais espontâneo e livre para as escolhas, onde o pesquisador adotou uma abordagem mais lúdica; e o segundo tipo que denotava mais seriedade, logo, menos espontâneo, visto que as crianças careciam de um maior isolamento das interações com os outros, e no qual foi cobrado silêncio e concentração, além de não contar com o uso de figuras ou com a postura lúdica do pesquisador. Nesse caso, utilizamos o ditado de palavras, método já largamente usado pelos professores no processo de letramento das séries iniciais, inclusive, peculiarmente, método amplamente usado na escola pesquisada.

As palavras usadas em ambos os treinos foram as mesmas, todas com a possibilidade da manifestação da semivocalização do /l/, divididas em dois grupos, o primeiro com o /l/ em posição de coda medial: caldeirão, soldado, multidão, palco, filme, culto, salsicha, fralda, talco, algodão; e, em posição de coda final: jornal, cristal, temporal, varal, legal, hospital, cereal, farol, curral, anzol. Observemos a eficácia dos dois tipos de contexto no uso da variante:

Gráfico 6 – Contexto de Aplicação – Tipo de Ditado.



Fonte: Elaborado pelos pesquisadores, 2017.

Sabendo-se dos contextos abordados pelo pesquisador na execução do trabalho, nota-se que o ambiente “menos monitorado”, o ditado de frases, abriu mais espaço, embora com diferença de, mais ou menos, 5%, para o uso da variante. Sendo assim, as crianças, em sua maioria, cometeram mais variação (apagamento ou semivocalização) e, conseqüentemente, menos escolhas de acordo com a norma padrão. Vale frisar novamente que já havia o costume do uso de ditado como estratégia de fixação da aprendizagem linguística, logo, acredita-se que esse costume com o método tenha sido o responsável pela maior escolha pela norma padrão na modalidade de contexto mais monitorada.

Após a discussão e análise dos resultados com que trabalhamos, esse trabalho buscou provar todas as hipóteses aqui levantadas, a fim de validar os pressupostos teóricos debatidos à luz da literatura utilizada. Buscamos, sobretudo, a regularidade (*reliability*) e a intersubjetividade (*intersubjectivity*), características de trabalhos que lidam com o mesmo campo de coleta (cf. BAILEY; TILLERY, 2004 apud LEAL, 2015).

### 3 CONSIDERAÇÕES FINAIS

Ao término desta análise, concluímos, com êxito, o que diz respeito à nossa hipótese fundamentadora. Comprovamos, assim, que, em séries iniciais, os dados de escrita estão mais próximos da fala e na medida em que se dá o letramento, com o aumento da escolarização, tende a se afastar desta (cf. MORAIS,

2003, 2007; MASSINICAGLIARI; CAGLIARI, 2008; MARCUSCHI, 2008; FARACO, 2012; ZILLES; FARACO, 2015). Essa conclusão mostra-se geral no âmbito variacionista e foram também discutidos, teorizados e provados os pontos seguintes:

- Embora a líquida lateral seja um dos primeiros fonemas internalizados na infância no âmbito da fala, a questão da escrita do segmento esbarra no nível de maturidade linguística: os sons produzidos pelo /l/, tendem a variar nas séries iniciais, criando assim a forma [w], e, só com o aprimoramento da escolaridade, essa escrita tende a harmonizar-se com a norma culta.
- Como já constatado em pesquisas anteriores realizadas por nosso grupo de estudos (REFALES), o contexto mais monitorado obteve maior êxito no uso da norma padrão, porém, em contrapartida, em pontos críticos, como o do apagamento do /l/, obteve mais uso da variante zero, como em mutidão ~ multidão e anso ~ anzol. Quantitativamente, foram observados 68 casos no 2º ano no ditado de palavras (contexto mais monitorado) contra 54 casos de apagamento no mesmo ano no ditado de frases (contexto menos monitorado). Acusa-se o uso da hipercorreção resultante de um contexto onde se tem mais cobrança do acerto.
- Apontou-se aqui o aparecimento quase inédito da variante /o/ nas séries iniciais, que está intimamente próximo com a ligação do som da líquida lateral em posição de coda medial e final com a grafia do /u/, gerando assim um ambiente propício para a alofonia, como nas manifestações observadas em: varao ~ varal e em paoco ~ palco.
- Por fim, no contexto fonológico, provou-se que a manifestação da semivocalização mostra-se mais abundante quando se trata de coda medial, pois, de tal maneira, a consoante seguinte força essa semivocalização, como em: tauco ~ ta[w]co (talco), no 2º ano; caudeirão ~ ca[w]deirão (caldeirão), no 3º ano; pauco ~ pa[w]cu (palco), no 4º ano; augodam ~ a[w]godão (algodão), no 5º ano.

Chamemos a atenção também para a importância da incorporação desse material ao corpus da literatura que postula sobre as variações fonológicas, primeiramente, pela inovação no trato de determinadas variantes e a ligação delas ao aparecimento da semivocalização da líquida lateral /l/.

E, em segunda instância, a importância deste estudo, somado aos já existentes, pois o cânone bibliográfico variacionista tende a se repetir e a se limitar aos mesmos autores. É tamanha a importância de novas visões, de novos autores sobre as manifestações linguísticas; inclusive, quando se trata de estudos linguísticos realizados no nordeste brasileiro. Isso porque os falares daqui (Nordeste), em comparação aos do Sul, principalmente em relação ao nível fonológico, costumam ser demasiadamente diferenciados uns dos outros. Portanto, faz-se mais do que necessário esse debruçar-se sobre as manifestações variacionistas nordestinas.

## REFERÊNCIAS

BISOL, Leda. (org.). *Introdução a Estudos de Fonologia do Português Brasileiro*. 4. ed. Porto Alegre: EDIPUCRS, 2005.

CÂMARA JR., Joaquim Mattoso. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes, 1973.

DE PINHO, Antônio José e MARGOTTI, Felício. A variação da lateral pós-vocálica /l/ no português do Brasil. DOI:10.5007/1984-8420.2010v11n2p67. *Working Papers em Linguística*, Florianópolis, v. 11, n. 2, p. 67-88, mar. 2011. ISSN 1984-8420. Disponível em: <https://periodicos.ufsc.br/index.php/workingpapers/article/view/1984-8420.2010v11n2p67/17634>. Acesso em: 11 jun. 2017.

DICKEY, Laura Walsh. *The Phonology of Liquids*. Amherst: GLSA, 1997.

ESPIGA, Jorge. *O Português dos Campos Neutrais: um estudo sociolinguístico da lateral pós-vocálica nos dialetos fronteiriços de Chuí e Santa Vitória do Palmar*. Tese (Doutorado em Letras), Faculdade de Letras, PUCRS, Porto Alegre, 2001.

FARACO, Carlos Alberto. *Linguagem Escrita e Alfabetização*. São Paulo: Contexto, 2012.

HORA, Dermeval. *Variação da Lateral /l/: correlação entre restrições sociais estruturais*. *SCRIPTA*, Belo Horizonte, v. 9, n. 18, p. 29-44, 2006.

LABOV, William. *Principles of Linguistic Change: social factors*. Oxford: Blackwell, 2001.

LEAL, Eneida de Goes. *A Lateral em Coda no VARSUL: Generalização de resultados e possibilidades de comparação*. *ABRALIN*, v. 14, p. 271-294, 2015.

- LURIA, Alexandr Romanovich. Pensamento e linguagem: as últimas conferências de Luria. Porto Alegre: Artes Médicas, 1986.
- MARCUSCHI, Luiz Antônio. Da Fala para a Escrita: Atividades de retextualização. 9. ed. São Paulo: Cortez, 2008.
- MARTINS, Lígia Márcia. O desenvolvimento do psiquismo e a educação escolar: contribuições à luz da psicologia histórico-cultural e da pedagogia histórico-crítica. Campinas-SP: Autores Associados, 2013.
- MIRANDA, Josete; SENRA, Luciana. Aquisição e Desenvolvimento da Linguagem: contribuições de Piaget, Vygotsky e Maturana. *Psicologia.pt*, p. 1-16, 2012.
- MOLLICA, Maria Cecília. Influência da Fala na Alfabetização. Rio de Janeiro: Tempo Brasileiro, 1998.
- MORAIS, Artur Gomes. Ortografia: ensinar e aprender. São Paulo: Ática, 2003.
- MORAIS, Artur Gomes (org.). O aprendizado da Ortografia. Belo Horizonte: Autêntica, 2007.
- PEDROSA, Juliene Lopes; NASCIMENTO, Jefferson. Vocalização da consoante lateral /l/ em posição de coda na modalidade de escrita: um estudo sociolinguístico. UEPB, [s.d.].
- QUEDNAU, Laura. A Lateral Pós-Vocálica no Português Gaúcho: análise variacionista e representação não linear. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1993.
- SANTOS, Ivanaldo. Wittgenstein e a importância dos jogos de linguagem na educação infantil. *Educação & Linguagem*, v. 17, p. 160-167, 2008.
- SILVA, Thaís Cristóforo. Dicionário de Fonética e Fonologia. São Paulo: Contexto, 2011.
- SILVA, José Robson; SEDRINS, Adeilson Pinheiro. Comportamento Variacionista do Segmento Lateral na Escrita das Séries Iniciais de Crianças de Serra Talhada-PE. UAST/UFRPE, [s.d.].
- TARALLO, Fernando. A Pesquisa Sociolinguística. São Paulo: Ática, 1986.
- TASCA, Maria. A Lateral em Coda Silábica no Sul do Brasil. 1999. Tese (Doutorado em Linguística Aplicada) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1999.

TASCA, Maria. Interferências da Língua Falada na Escrita das Séries Iniciais: o papel de fatores lingüísticos e sociais. Porto Alegre: EDIPUCRS, 2002.

TEYSSIER, Paul. História da Língua Portuguesa. São Paulo: Martins Fontes, [s.d.].

VYGOTSKY, Lev. A construção do pensamento e da linguagem. São Paulo: Editora WMF Martins Fontes, 2009.

WIETHAN, Fernanda; MELO, Roberta; MOTA, Helena. Consoantes Líquidas: ocorrência de estratégias de reparo em diferentes faixas etárias e gravidades do desvio fonológico. Revista CEFAC, v. 13, n. 4, p. 607-616, jul./ago. 2011.

ZILLES, Ana Maria Stahl; FARACO, Carlos Alberto (orgs.). Pedagogia da Variação Linguística: língua, diversidade e ensino. São Paulo: Parábola, 2015.



Este livro só foi possível graças ao apoio do Edital CAPES/FAPITEC/PROMOB 10/2016 e traz questões sobre os desafios enfrentados na execução de pesquisas sociolinguísticas, principalmente os que se relacionam à constituição de banco de dados. Assim, o leitor encontrará uma breve retrospectiva sobre iniciativas que objetivaram debater questões sobre a constituição e disponibilização de corpora linguísticos; relatos sobre características de alguns bancos de dados; discussões sobre critérios que vêm sendo adotados na sua constituição; diferentes perspectivas de análises; propostas para a obtenção de dados, especialmente através de interação online como fonte de estudo da variação linguística; discussões de medidas para atenuar as dificuldades no que diz respeito à escassez de recursos para organizar corpora de grande e pequeno porte e disponibilizá-los à comunidade científica, às medidas de isolamento impostas pela pandemia do coronavírus e ao acesso a algumas comunidades.



[openaccess.blucher.com.br](http://openaccess.blucher.com.br)



**Blucher** Open Access