

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**FACULDADE DE LETRAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS**

Cecília Valle Souza Toledo

**RELAÇÕES MÚLTIPLAS ENTRE ORALIDADE E ESCRITA:**  
**vogais médias e róticos**

Belo Horizonte - MG  
2023

Cecília Valle Souza Toledo

**RELAÇÕES MÚLTIPLAS ENTRE ORALIDADE E ESCRITA:  
vogais médias e róticos**

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Estudos Linguísticos.

Área de concentração: Linguística Teórica e Descritiva

Linha de pesquisa: Estudos Formais da Língua (1E)

Orientadora: Prof. Dra. Thaïs Cristófaro Alves da Silva

Belo Horizonte - MG  
2023

T649r

Toledo, Cecília Valle Souza.

Relações múltiplas entre oralidade e escrita [manuscrito] :  
vogais médias e róticos / Cecília Valle Souza Toledo. – 2023.  
1 recurso online (198 f. : il., tabs., color., p&b., color.) : pdf.

Orientadora: Thaís Cristófaros Alves da Silva.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudos Formais da Língua.

Tese (doutorado) – Universidade Federal de Minas Gerais,  
Faculdade de Letras.

Bibliografia: f. 151-161.

Apêndices: f. 162-198.

Exigências do sistema: Adobe Acrobat Reader.

1. Língua portuguesa – Fonologia – Teses. 2. Língua portuguesa – Ortografia e silabação – Teses. 3. Língua portuguesa – Variação – Belo Horizonte (MG) – Teses. 4. Língua portuguesa – Vogais – Teses. 5. Língua portuguesa – Português falado – Belo Horizonte (MG) – Teses I. Silva, Thaís Cristófaros. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 469.15



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
FACULDADE DE LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

### FOLHA DE APROVAÇÃO

**RELAÇÕES MÚLTIPLAS ENTRE ORALIDADE E ESCRITA: vogais médias e róticos**

**CECÍLIA VALLE SOUZA TOLEDO**

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Fonologia.

Aprovada em 06 de março de 2023, pela banca constituída pelos membros:

Prof(a). Thais Cristofaro Alves da Silva - Orientadora

UFMG

Prof(a). Daniela Mara Lima Oliveira Guimarães

UFMG

Prof(a). Maria Mendes Cantoni

UFMG

Prof(a). José Sueli de Magalhães

UFU

Prof(a). Christina Abreu Gomes

UFRJ

Belo Horizonte, 06 de março de 2023.



Documento assinado eletronicamente por **Daniela Mara Lima Oliveira Guimaraes, Professora do Magistério Superior**, em 09/03/2023, às 08:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Thais Cristofaro Alves da Silva, Professora do Magistério Superior**, em 11/03/2023, às 09:38, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **José Sueli de Magalhães, Usuário Externo**, em 13/03/2023, às 10:47, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Christina Abreu Gomes, Usuária Externa**, em 13/03/2023, às 13:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Maria Mendes Cantoni, Professora do Magistério Superior**, em 14/03/2023, às 20:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2091049** e o código CRC **F4EE80A3**.

---

## AGRADECIMENTOS

Seria impossível elaborar esta tese sozinha. A conclusão deste trabalho se deu graças ao apoio fundamental de pessoas especiais. Agradeço a todos que me ajudaram a seguir em frente nesses anos de pesquisa, em especial:

*À minha querida orientadora Professora Thaís Cristófaró Silva* por exercer com excelência sua função; por me ensinar fonologia, ética, cidadania e rigor científico; por me mostrar, na prática, que a carreira acadêmica deve ser traçada com desejo, com seriedade, com comprometimento e com muito trabalho; por ser acolhimento, incentivo e inspiração ao longo de toda minha formação.

*À Ana Luiza Couto.* “Ô Ana”, você deixou esse percurso muito mais leve e imensamente mais fácil. Você foi amparo e conforto. Muito obrigada por todas as trocas; por todo apoio emocional; pela revisão atenta do trabalho; por todos os cafés e conversas; por todo carinho e amizade.

*Ao Matheus Freitas Gomes* pela parceria de sempre; por me motivar a seguir em frente desde o mestrado; por ser meu fiel companheiro nos congressos e disciplinas; por ser um grande amigo.

*Aos membros da escola parceira desta pesquisa.* Agradeço pela confiança da equipe pedagógica, pela autorização dos pais e responsáveis das crianças e adolescentes e pela animada participação dos alunos.

*À banca.* Obrigada *Christina Gomes* pela leitura atenta do texto e por todos os comentários enriquecedores. Obrigada *José Magalhães* por acompanhar de perto toda a elaboração do trabalho e por me motivar a continuar. Obrigada *Daniela Oliveira Guimarães* por ser minha mentora nos tópicos sobre linguística aplicada, por me acolher no seu grupo de estudos e por me encorajar a entrar na escola. Obrigada *Maria Cantoni* pelas brilhantes colocações na arguição e por todos os ensinamentos ao longo dos seis anos de pós-graduação. Obrigada *Raquel Fontes Martins* e *Larissa Ciriaco* por aceitarem participar da banca como suplentes.

*Ao corpo docente do Poslin.* Obrigada a todos os professores por compartilharem comigo grandiosos saberes. Em especial, agradeço à *Mahayana Godoy* por me ajudar a aprender estatística e à *Janayna Carvalho* por me orientar no estágio e me aconselhar em diversos assuntos.

*Ao Wagner Teixeira*, meu amigo e professor de longa data, por ser meu fiel conselheiro sobre diversos tópicos relacionados à academia; por todo o apoio emocional e profissional.

*À minha mãe* por ser a raiz que me dá forças para poder crescer. Seu amor me move, mãe.

*Ao meu pai* por ser colo e cuidado; por se orgulhar dos meus voos, mas se colocar como um lugar de pouso caso eu queira voltar.

*Ao Caique* por todo o amor, generosidade, paciência, incentivo e companheirismo ao longo do percurso de elaboração deste trabalho.

*Aos meus colegas e amigos*. Obrigada Yasmin, Flora, Thaís Bechir, Larissa, Cássio, Magnum, Nívia, Wellington, Amanda, Kely, Marcus, Gabi por compartilharem experiências comigo, por me ouvirem e por me incentivarem.

*À CAPES* pelo apoio financeiro. O presente trabalho foi fomentado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasi (CAPES) – Código de Financiamento 001.

- *Vocês sabem a diferença entre a fala e a escrita?*  
— *Eu sei, professora. A fala vai embora pelo ar; a escrita fica para sempre no papel.*

*(Miguel, aluno do 6º ano do EF)*



## RESUMO

O objetivo desta tese é investigar as relações múltiplas que se estabelecem entre a fala e a escrita e entre a escrita e a fala na variedade de Belo Horizonte - MG. Quatro categorias linguísticas são consideradas: (1) vogais médias pretônicas, (2) vogais médias postônicas finais, (3) róticos finais em verbos, (4) róticos finais em nomes. Tradicionalmente, as categorias linguísticas analisadas são representadas por uma das seguintes letras: <e, o, r>. Ademais, todas as categorias estão envolvidas em fenômenos fonológicos variáveis que motivam erros ortográficos no decorrer do percurso de escolarização (ALVARENGA *et al.*, 1989; NÓBREGA, 2013). As vogais médias pretônicas e postônicas finais podem ser pronunciadas como vogais altas devido aos fenômenos de alçamento e redução vocálica (VIEGAS, 1995; DIAS; SEARA, 2013). Esses fenômenos fonológicos tendem a motivar erros ortográficos na escrita. Ex: <\*minino>, <\*patu>. Os róticos finais de verbos e nomes podem ser cancelados por consequência do fenômeno fonológico de cancelamento do R-final (OLIVEIRA, 1997; HUBACK, 2003). Esse fenômeno tende a motivar erros ortográficos na escrita. Ex: <\*cantá>, <\*celulá>. Considerando-se as aparentes semelhanças entre as categorias (1) e (2) e (3) e (4), buscamos responder a seguinte questão: *categorias linguísticas, aparentemente semelhantes, se realizam de forma análoga ou diferente na fala e na escrita de alunos do Ensino Fundamental?* Quatro hipóteses foram testadas: (1a) as trocas de <e, o> por <i, u> na escrita persistem, em índices significativamente diferentes, até estágios distintos da escolarização, a depender do contexto acentual: pretônico ou postônico final; (1b) a omissão da letra <r> na escrita persiste, em índices significativamente diferentes, até estágios distintos da escolarização, a depender da classe gramatical: verbo ou nome; (2a) o aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônico ou postônico final; (2b) o aumento do rótico final na fala ao longo dos anos escolares depende da classe gramatical: verbo ou nome. Como se pode ver, esperamos encontrar diferenças na fala e na escrita infantojuvenil, a depender das categorias linguísticas. As hipóteses foram testadas em um conjunto de dados de fala e de escrita de cinquenta crianças do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental da cidade de Belo Horizonte – MG. Os resultados alcançados confirmaram as hipóteses de que os erros ortográficos relacionados às vogais pretônicas e postônicas finais, bem como aos róticos finais de verbos e de nomes tendem a perdurar até diferentes estágios da escolarização. Além disso, os resultados trouxeram indícios de que o aprendizado da escrita interfere na produção oral dos indivíduos. Contudo, as interferências da escrita na fala dependem da categoria linguística. Com o apoio dos Modelos de Exemplares (JOHNSON, 1997; BYBEE, 2001; PIERREHUMBERT, 2001; FOULKES; DOCHERTY, 2006; PORT, 2007) e da Teoria dos Sistemas Dinâmicos e Complexos (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009), esses resultados foram interpretados como indícios de que as crianças do Ensino Fundamental gerenciam e reorganizam esquemas múltiplos ao longo da escolarização. A expressão ‘esquema múltiplo’ será utilizada para se referir aos padrões probabilísticos que emergem de múltiplas redes de conexões fonéticas, fonológicas, lexicais, morfológicas, sociais, etc. e podem ser usados para produzir novos padrões gramaticais (LANGACKER, 2000; TOMASELLO, 2000; BYBEE, 2001; CRISTÓFARO-SILVA; GOMES, 2004). Argumentamos que, ao longo da escolarização, os aprendizes acessam os esquemas múltiplos, o que motiva os diferentes estágios de aprendizado da ortografia. Além disso, argumentamos que o conhecimento ortográfico adquirido motiva a reorganização do sistema linguístico. Tal reorganização pode ser mais ou menos evidente, a depender da categoria linguística. Em síntese, pretendemos defender a tese de que *as letras são símbolos discretos que se associam a esquemas múltiplos, os quais emergem da interação complexa entre palavras, seqüências fonéticas, morfológicas e ortográficas.*

**Palavras-chave:** oralidade; escrita; esquemas múltiplos, erros ortográficos, variação.

## ABSTRACT

This thesis aims to investigate the multiple links established between speech and spelling and between spelling and speech for the Brazilian Portuguese spoken in Belo Horizonte. Four linguistic categories in Brazilian Portuguese are assessed: (1) pretonic mid vowels, (2) final posttonic mid vowels, (3) final rhotics in verbs, (4) final rhotics in nouns. The analyzed linguistic categories are traditionally spelt with one of the following letters: <e, o, r>. Additionally, all categories are prone to variable phonological phenomena which may motivate spelling errors during the school years (ALVARENGA et al., 1989; NÓBREGA, 2013). The final pretonic and posttonic mid vowels may be pronounced as high vowels due to vowel raising and reduction (VIEGAS, 1995; DIAS; SEARA, 2013). These phonological phenomena tend to motivate spelling errors in which <e, o> are replaced by <i, u>. Ex: <\*minino> ‘boy’, <\*patu> ‘duck’. Final rhotics in verbs and nouns may be lost (OLIVEIRA, 1997; HUBACK, 2003). This phenomenon tends to trigger misspelled forms that lack a final <r>. Ex: <\*cantá> ‘to sing’, <\*celulá> ‘cellphone’. Considering the apparent similarities between categories (1) and (2) and (3) and (4), we seek to answer the following research question: do apparently similar linguistic categories manifest analogous or different behavior in speech and in spelling patterns of elementary and middle school students? Four hypotheses were tested: (1a) the replacement of <e, o> for <i, u> persists in spelling, at significantly different rates, across different phases of schooling, depending on the syllabic position: pretonic or final posttonic; (1b) the loss of letter <r> in spelling persists, at significantly different rates, across different phases of schooling, depending on word class: verb or noun; (2a) the increase in production of mid vowels in speech across the school years depends on the syllabic position: pretonic or final posttonic; (2b) the increase in the production of the final rhotic in speech across the school years depends on word class: verb or noun. We predicted differences in the children's speech and spelling depending on the linguistic categories. The hypotheses were tested on a set of spoken and written data from fifty children from the 1st, 3rd, 5th, 7th and 9th years of elementary and middle school classes in the city of Belo Horizonte, Minas Gerais, Brazil. Results confirmed the hypotheses that spelling errors related to posttonic and final pretonic vowels as well as to final rhotics in verbs and nouns tend to persist at different rates across different stages of schooling. The results evidence that learning to spell affects speech production. However, the effects of spelling in speech depend on the linguistic category. Results are discussed in light of Exemplar Theory (JOHNSON, 1997; BYBEE, 2001; PIERREHUMBERT, 2001; FOULKES; DOCHERTY, 2006; PORT, 2007) and Complex Dynamic Systems Theory (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN -FREEMAN, 2009), which highlight that elementary and middle school aged children manage and reorganize multiple schemas during schooling. The term 'multiple schemas' is adopted to refer to the probabilistic patterns that emerge from multiple networks of phonetic, phonological, lexical, morphological, social connections etc. These relations can generate new grammatical patterns (LANGACKER, 2000; TOMASELLO, 2000; BYBEE, 2001; CRISTÓFARO-SILVA; GOMES, 2004). We argue that students access multiple schemas during schooling, which motivates the different spelling learning stages. We also argue that the learned orthographic knowledge motivates the reorganization of the language system. Such reorganization can be more or less evident, depending on the linguistic category. In summary, we intend to defend the thesis that *letters are discrete symbols associated with multiple schemas, which emerge from the complex interaction between words and phonetic, morphological and orthographic patterns.*

**Keywords:** speech; spelling; multiple schemas, spelling errors, variation.

## LISTA DE FIGURAS, QUADROS, TABELAS E GRÁFICOS

Figura 1: Relações entre sílabas pretônicas .....	26
Figura 2: Categoria dos erros relacionados às vogais.....	42
Figura 3: Conexões lexicais entre palavras terminadas em -aram.....	62
Figura 4: Representação do conjunto de exemplares.....	64
Figura 5: Associações lexicais e esquemas múltiplos .....	66
Figura 6: Exemplo de estímulo experimental.....	76
Figura 7: Exemplos de estímulos experimentais com perguntas-guia.....	77
Figura 8: Percursos evolutivos das vogais pretônicas e postônicas finais .....	94
Figura 9: Relações entre oralidade e escrita no contexto postônico final .....	113
Figura 10: Relações entre oralidade e escrita no contexto pretônico.....	114
Figura 11: Percursos evolutivos dos róticos finais .....	125
Figura 12: Relações múltiplas entre oralidade e escrita em verbos e nomes .....	138
Figura 13: Relações múltiplas entre oralidade e escrita .....	147
Figura 14: Atividade didática sobre a letra <r> .....	149
Quadro 1: Categorias linguísticas analisadas .....	15
Quadro 2: Exemplos de palavras com vogais pretônicas.....	25
Quadro 3: Estrutura morfológica dos nomes terminados em <r> final .....	30
Quadro 4: Definição de fonema, grafema, letra e som .....	33
Quadro 5: Relações entre letras e sons na ortografia do português .....	37
Quadro 6: Exemplos de erros ortográficos em contextos pretônico e postônico final.....	41
Quadro 7: Exemplos de erros ortográficos em final de verbos e de nomes.....	44
Quadro 8: Categoria dos erros ortográficos relacionados aos róticos.....	45
Quadro 9: Exemplos de hipercorreções.....	49
Quadro 10: Relações entre a fala e a escrita.....	54
Quadro 11: Critérios utilizados na seleção das palavras-teste.....	73
Quadro 12: Palavras-teste.....	74
Quadro 13: Palavras-teste para análise da hipercorreção .....	74
Quadro 14: Variantes orais identificadas no conjunto de dados.....	80
Quadro 15: Categorias dos dados de fala .....	81
Quadro 16: Planilha de organização dos dados .....	81
Quadro 17: Variáveis dependentes e independentes.....	83
Quadro 18: Respostas dos professores – vogais médias.....	117
Quadro 19: Observações dos professores sobre a omissão da letra <r> .....	140
Quadro 20: Motivação dos erros ortográficos pela visão do professor.....	141
Tabela 1: Quantidade total de dados de escrita .....	79
Tabela 2: Quantidade total de dados de fala.....	80
Tabela 3: Erros ortográficos por vogal anterior e posterior .....	102

Tabela 4: Hipercorreções em contextos pretônico e postônico final .....	103
Tabela 5: Quantidade de erros vocálicos observados pelos professores .....	118
Tabela 6: Hipercorreções em final de verbos e nomes .....	130
Tabela 7: Quantidade de erros ortográficos observados pelos professores.....	140
Gráfico 1: Variação das vogais anteriores pretônicas e postônicas finais .....	87
Gráfico 2: Variação das vogais anteriores por item lexical .....	88
Gráfico 3: Variação das vogais anteriores pretônicas por indivíduo .....	89
Gráfico 4: Variação das vogais anteriores postônicas finais por indivíduo .....	89
Gráfico 5: Variação das vogais posteriores pretônicas e postônicas finais.....	91
Gráfico 6: Variação das vogais posteriores por item lexical .....	92
Gráfico 7: Variação das vogais posteriores pretônicas por indivíduo .....	93
Gráfico 8: Variação das vogais posteriores postônicas finais por indivíduo .....	93
Gráfico 9: Trocas de <e> por <i> em contextos pretônico e postônico final .....	96
Gráfico 10: Trocas de <e> por <i> por item lexical.....	97
Gráfico 11: Trocas de <e> por <i> pretônicas por indivíduo .....	98
Gráfico 12: Trocas de <e> por <i> postônicas finais por indivíduo .....	98
Gráfico 13: Trocas de <o> por <u> em contextos pretônico e postônico final .....	99
Gráfico 14: Trocas de <o> por <u> por item lexical.....	100
Gráfico 15: Trocas de <o> por <u> pretônicas por indivíduo .....	101
Gráfico 16: Trocas de <o> por <u> postônicos finais por indivíduo.....	101
Gráfico 17: Hipercorreções vocálicas por ano escolar .....	104
Gráfico 18: Trocas de <e, o> por <i, u> por ano escolar e contexto acentual .....	106
Gráfico 19: Probabilidades de erros em contextos pretônico e postônico final.....	107
Gráfico 20: Índices de vogais médias na fala por ano escolar .....	108
Gráfico 21: Probabilidade de vogais médias pretônicas na fala por ano escolar .....	109
Gráfico 22: Variação segmental do rótico final em verbos e nomes .....	121
Gráfico 23: Variação do R-final por item verbal.....	122
Gráfico 24: Variação do R-final por item nominal.....	123
Gráfico 25: Variação do R-final em verbos por indivíduo .....	124
Gráfico 26: Variação do R-final em nomes por indivíduo .....	124
Gráfico 27: Omissão do <r> final em verbos e em nomes .....	126
Gráfico 28: Omissão do <r> final em itens verbais.....	127
Gráfico 29: Omissão do <r> final em itens nominais.....	127
Gráfico 30: Omissão do <r> final em verbos por indivíduo.....	128
Gráfico 31: Omissão do <r> final em nomes por indivíduo .....	129
Gráfico 32: Hipercorreções do <r> em final de verbos e de nomes .....	131
Gráfico 33: Omissão da letra <r> por ano escolar e por contexto acentual.....	132
Gráfico 34: Probabilidade de omissão da letra <r> final.....	133
Gráfico 35: Ocorrência do R-final por ano escolar e classe gramatical.....	134
Gráfico 36: Probabilidade de produção oral do rótico final .....	135

# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>15</b>
1.1 <i>Motivação e justificativas</i> .....	17
1.2 <i>Escopo teórico</i> .....	18
1.3 <i>Pergunta de pesquisa</i> .....	19
1.4 <i>Hipóteses e Tese</i> .....	19
1.6 <i>Método</i> .....	22
1.7 <i>Organização do texto</i> .....	22
<b>2. REVISÃO BIBLIOGRÁFICA .....</b>	<b>24</b>
2.1 <i>A Fala</i> .....	24
2.1.1 <i>Vogais médias na oralidade</i> .....	25
2.1.2 <i>Róticos finais na oralidade</i> .....	28
2.2 <i>A Escrita</i> .....	31
2.2.1 <i>A relação entre letras e sons: a essência da escrita alfabética</i> .....	32
2.2.2 <i>Ortografia</i> .....	35
2.2.3 <i>O erro ortográfico</i> .....	38
2.2.3.1 <i>A troca de &lt;e, o&gt; por &lt;i, u&gt;</i> .....	41
2.2.3.2 <i>A omissão do &lt;r&gt; em final de verbos e de nomes</i> .....	44
2.2.3.3 <i>As hipercorreções ortográficas</i> .....	46
2.4 <i>O aprendizado da escrita</i> .....	50
2.5 <i>As influências da escrita na fala</i> .....	54
2.6 <i>Resumo do Capítulo 2</i> .....	57
<b>3. QUADRO TEÓRICO .....</b>	<b>58</b>
3.1 <i>Representações fonológicas discretas: uma visão geral</i> .....	58
3.2 <i>Modelos de Exemplares</i> .....	62
3.2.1 <i>Esquemas</i> .....	65
3.3 <i>Teoria dos Sistemas Dinâmicos e Complexos</i> .....	68
3.3 <i>Resumo do Capítulo 3</i> .....	72
<b>4. METODOLOGIA .....</b>	<b>73</b>
4.1 <i>Procedimentos pré-coleta</i> .....	73
4.1.1 <i>Seleção de palavras</i> .....	73
4.1.2 <i>O campo de coleta e os participantes</i> .....	75
4.1.3 <i>Experimentos</i> .....	76
4.2 <i>Coleta dos dados</i> .....	78
4.3 <i>Procedimentos pós - coleta</i> .....	79

4.3.1 Organização dos dados.....	79
4.3.1.1 Organização da amostra dados de escrita.....	79
4.3.1.2 Organização da amostra dados de fala.....	80
4.3.2 Análise dos dados.....	81
4.3.2.1 Análise estatística.....	82
4.4 <i>Resumo do Capítulo 4</i> .....	85
<b>5. ANÁLISE DOS DADOS: VOGAIS MÉDIAS.....</b>	<b>86</b>
5.1 <i>PARTE I: Análise descritiva dos dados</i> .....	86
5.1.1 Vogais anteriores na fala.....	87
5.1.2 Vogais posteriores na fala.....	90
5.1.3 Diferenças entre vogais anteriores e posteriores na fala.....	94
5.1.4 As trocas de <e> por <i> na escrita.....	95
5.1.5 As trocas de <o> por <u> na escrita.....	99
5.1.6 Diferenças entre as letras <e> e <o>.....	102
5.1.7 Hipercorreções.....	102
5.2. <i>PARTE II: Hipóteses</i> .....	105
5.2.1 Hipótese (1a).....	105
5.2.2 Hipótese (2a).....	108
5.3. <i>PARTE III: Discussão teórica - vogais médias</i> .....	110
5.4. <i>PARTE IV: A visão dos professores sobre as trocas de &lt;e, o, i, u&gt;</i> .....	117
5.5 <i>Resumo do capítulo</i> .....	119
<b>6. ANÁLISE DOS DADOS: RÓTICOS.....</b>	<b>120</b>
6.1 <i>PARTE I: Análise descritiva dos dados</i> .....	120
6.1.1 Róticos finais na fala.....	120
6.1.3 A omissão do <r> final na escrita.....	126
6.1.5 Hipercorreções.....	130
6.2 <i>PARTE II: Hipóteses</i> .....	131
6.2.1 Hipótese (1b).....	132
6.2.2 Hipótese (2b).....	134
6.3 <i>PARTE III: Discussão teórica - Róticos</i> .....	136
6.4 <i>PARTE IV: A visão dos professores sobre a omissão do &lt;r&gt; final</i> .....	140
6.5 <i>Resumo do capítulo</i> .....	142
<b>7. CONSIDERAÇÕES FINAIS.....</b>	<b>144</b>
<b>8. REFERÊNCIAS.....</b>	<b>151</b>
<b>APÊNDICES.....</b>	<b>162</b>

## 1. INTRODUÇÃO

Esta pesquisa tem como objetivo investigar as relações múltiplas que se estabelecem entre a fala e a escrita e entre a escrita e a fala na variedade de Belo Horizonte – MG. Quatro categorias linguísticas são consideradas: (1) vogais médias pretônicas, (2) vogais médias postônicas finais, (3) róticos finais em verbos, (4) róticos finais em nomes. O Quadro 1 resume as categorias analisadas.

**Quadro 1:** Categorias linguísticas analisadas

	<b>Categorias linguísticas</b>	<b>Variação sonora</b>	<b>Erros ortográficos</b>
VOGAIS MÉDIAS	(1) Vogais pretônicas	vogais médias ~vogais altas	troca de <e, o> por <i, u>
	(2) Vogais postônicas finais	vogais médias ~vogais altas	troca de <e, o> por <i, u>
RÓTICOS	(3) Verbos com R-final	R-final ~ cancelamento	omissão do <r> final
	(4) Nomes com R-final	R-final ~ cancelamento	omissão do <r> final

**Fonte:** elaboração própria.

As categorias (1) e (2) englobam vogais médias. No alfabeto do português brasileiro (PB), as vogais médias são representadas pelas letras <e> e <o>. No sistema oral, essas vogais podem alternar com vogais altas em algumas palavras. Exemplo: [me'ni. nũ] ~ [mi'ni. nũ] 'menino' (VIEGAS, 2001); ['fa. ve] ~ ['fa. vi] 'chave' (VIEIRA, 2002). A variação que ocorre na fala faz com que haja mais de um corresponde sonoro – [e, i, o, u] – para as letras <e, o> em contexto pretônico e postônico final.

Pesquisas dedicadas ao estudo do percurso de aprendizado da ortografia do PB já demonstraram que os aprendizes podem cometer trocas entre as letras <e, o> e <i, u> tanto no contexto pretônico (MIRANDA, 2006), como por exemplo em <\*pirigo>, quanto no postônico final (MIRANDA, 2008) como por exemplo em <\*penti>. Tradicionalmente, assume-se que os erros ortográficos relacionados às vogais pretônicas e postônicas finais são motivados pela oralidade: fala-se m[i]nino e b[u]neca e, por isso, escreve-se <\*minino> e <\*buneca> (ALVARENGA *et al.*, 1989).

De modo geral, as vogais médias pretônicas e as vogais postônicas finais são descritas, por vezes, como padrões semelhantes na fala e na escrita: ambas podem ser pronunciadas com vogais altas na fala e ambas podem ser grafadas, erroneamente, com <i, u> em vez de <e, o>. Considerando-se as semelhanças entre as categorias (1) e (2), é relevante avaliar, comparativamente, as relações entre oralidade e escrita nas duas posições acentuais. Se as amostras de dados de fala e de escrita das vogais pretônicas e postônicas finais se mostrarem semelhantes, será possível dizer que essas vogais operam de forma similar no português brasileiro. Por outro lado, se os dados indicarem diferenças significativas entre as vogais pretônicas e postônicas finais,

poderemos pensar que os indivíduos gerenciam categorias linguísticas diferentes para grafarem as letras <e, o> em diferentes contextos acentuais.

As categorias (3) e (4), descritas no Quadro 1, englobam róticos finais. O termo *rótico* é utilizado para se referir aos *sons de R* (LADEFOGED; MADDIESON, 1996). Na escrita, os róticos são representados pela letra <r> e, na fala, esses sons podem ser cancelados quando ocorrem no final de verbos e de nomes no PB. Exemplo: [va'he] ‘varrer’, [selu'la] (*celular*) (OLIVEIRA, 1997; HUBACK, 2003). Trabalhos precedentes demonstraram que aprendizes da ortografia do PB tendem a omitir a letra <r> que ocorre no final de verbos e de nomes (COSTA, 2009; TORRES; OLIVEIRA, 2015; FREIRE; HORA, 2019). Tradicionalmente, assume-se que a omissão do <r> na escrita infantil é motivada por um único fenômeno fonológico variável, geralmente denominado de ‘cancelamento do rótico pós-vocálico’ (NÓBREGA, 2013).

Como se pode ver, os verbos e os nomes terminados em R-final são descritos como padrões semelhantes na fala e na escrita: ambos estão envolvidos em variação sonora – cancelamento do rótico – e ambos podem ser grafados, erroneamente, sem a letra <r> final. Considerando-se as semelhanças entre as categorias (3) e (4), faz-se relevante compará-las no que diz respeito à oralidade e à escrita. Se as amostras de dados de fala e de escrita dos verbos e dos nomes com R-final se mostrarem semelhantes, será possível dizer que o rótico final opera de forma similar em ambas as classes gramaticais. Por outro lado, se os dados se mostrarem significativamente diferentes, poderemos dizer que os indivíduos gerenciam categorias distintas para falar e escrever o rótico final em diferentes classes gramaticais.

Em resumo, este trabalho pretende avaliar categorias aparentemente semelhantes, em busca de verificar se há diferenças entre elas no que diz respeito à produção oral e à escrita. Especificamente, busca-se cumprir os seguintes objetivos:

- Comparar os índices de erros ortográficos em contextos pretônico e postônico final na escrita de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.
- Comparar os índices de erros ortográficos em verbos e em nomes terminados com R-final na escrita de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.
- Avaliar a variação oral das vogais pretônicas e postônicas finais na fala de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.
- Avaliar a variação oral dos róticos finais de verbos e nomes na fala de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.
- Discutir a natureza abstrata das vogais médias e dos róticos finais à luz de Modelos múltiplos e dinâmicos.
- Discutir as relações entre oralidade e escrita à luz de Modelos múltiplos e dinâmicos.



## 1.1 Motivação e justificativas

Um consenso já estabelecido na linguística é o fato de que os sistemas oral e escrito se relacionam em algum nível, embora se diferenciem em muitos aspectos. Sabe-se que, no decorrer do percurso escolar, os aprendizes podem tentar transcrever a forma falada no papel (CAGLIARI, 1989). Além disso, há evidências recentes de que a forma escrita das palavras pode interferir na escolha de variantes orais (CHEVROT *et al.*, 2000; CRISTÓFARO-SILVA E GRECO, 2010; ADAMOLI, 2013). Assim sendo, pode-se perceber, pelo que já está documentado na literatura, que a fala e a escrita estabelecem uma relação de mão dupla – fala ↔ escrita. Sabendo-se disso, surgiu o primeiro interesse desta tese: avaliar as influências da fala na escrita, bem como da escrita na fala ao longo do percurso escolar.

As influências da fala na escrita são expressas, sobretudo, nos erros ortográficos cometidos ao longo da escolarização. Isto é, os aprendizes grafam palavras da forma como falam e, então, acabam errando a ortografia da língua. Os erros ortográficos motivados pela fala são, comumente, agrupados em uma mesma categoria. Alvarenga *et al.* (1989), por exemplo, agruparam os erros como <\*minino>, <\*buneca>, <\*penti> e <\*patu> em uma categoria denominada de ‘alçamento vocálico’. Nóbrega (2013), por sua vez, agrupou erros como <\*cantá> e <\*professo> na categoria chamada de ‘omissão do r em final de palavras’.

A interpretação implícita no método de categorização é a de que os erros ortográficos são decorrentes de uma mesma motivação: a oralidade. Entretanto, embora sejam todos reflexos da fala, há indícios de que as vogais médias e os róticos se manifestam de forma específica na língua portuguesa, a depender do contexto acentual e da classe gramatical. No contexto pretônico, por exemplo, há variação sonora lexicalmente motivada (OLIVEIRA, 1992). Já no contexto postônico final, as vogais altas são sistematicamente produzidas e, em alguns casos, há redução e cancelamento vocálico (VIEIRA, 1994; MAZZAFERRO; MATZENAUER, 2019). Ademais, é sabido que as vogais pretônicas e postônicas finais têm diferentes correlatos morfológicos na língua portuguesa (CÂMARA JUNIOR, 2015 [1970]). Nos verbos, por sua vez, o R-final é sistematicamente cancelado na fala, enquanto que, nos nomes, o cancelamento do rótico é variável (HUBACK, 2003). Além disso, é sabido que o R-final ocupa diferentes funções gramaticais na língua portuguesa, a depender da classe gramatical (CÂMARA JUNIOR, 2015 [1970]).

Considerando-se as diferenças linguísticas entre as vogais pretônicas e postônicas finais, bem como entre os róticos finais de verbos e de nomes, este trabalho pretende verificar se há também diferenças quanto ao índice e a persistência de erros ortográficos em cada categoria linguística avaliada. Com o apoio de modelos teóricos múltiplos e dinâmicos, pretendemos

argumentar que uma mesma letra – <e, o, r> – em diferentes contextos envolve o gerenciamento de diferentes categorias linguísticas. Assim sendo, este trabalho se justifica por:

- *Analisar erros ortográficos motivados pela fala a partir de perspectivas teóricas alternativas às tradicionais, que permitem refletir sobre os múltiplos percursos de aprendizado de uma mesma letra em diferentes contextos.*

Além de analisar os erros ortográficos motivados pela fala, este trabalho também se propõe a avaliar as influências da escrita na fala em quatro categorias linguísticas distintas. Alguns trabalhos já demonstraram que a forma escrita é um fator que pode influenciar<sup>1</sup> na produção oral de certos fenômenos fonológicos variáveis (CHEVROT *et al.*, 2000; SCHIWINDT *et al.*, 2007; CRISTÓFARO-SILVA E GRECO, 2010; ADAMOLI, 2013). Os trabalhos sobre esse tema, contudo, se dedicaram a avaliar fenômenos fonológicos específicos. Nenhum dos trabalhos teve a intenção de avaliar se a escrita pode afetar a produção da fala em diferentes contextos linguísticos. É, neste ponto, que este estudo pretende avançar: analisaremos a produção oral e escrita de diferentes categorias linguísticas, em busca de verificar as influências do aprendizado da escrita em diferentes contextos: vogais médias e altas pretônicas, vogais médias e altas postônicas finais; róticos finais de verbo; róticos finais de nomes. Assim sendo, esta tese também se justifica por:

- *Avaliar as influências da escrita na fala em diferentes categorias linguísticas.*

## 1.2 Escopo teórico

Esta pesquisa se fundamenta em duas abordagens teóricas: os Modelos de Exemplos (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e a Teoria dos Sistemas Dinâmicos e Complexos (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Os Modelos de Exemplos partem da premissa de que a variação linguística é parte inerente das categorias abstratas. Além disso, esses modelos propõem que os padrões gerais da linguagem (esquemas) emergem de associações complexas entre múltiplas informações linguísticas e não linguísticas. Nessa concepção, as letras têm representação discreta, mas os sons são parte de categorias múltiplas e detalhadas. Assim, postulamos, que uma mesma letra pode estar relacionada a diferentes categorias múltiplas. A Teoria dos Sistemas Dinâmicos e Complexos (TSDC), por sua

---

<sup>1</sup> Na literatura, utiliza-se, comumente, a expressão ‘retroalimentação da escrita sobre fala’ para se referir às influências da forma gráfica na oralidade (CRISTÓFARO-SILVA; GRECO, 2010). Pensamos, porém, que essa terminologia poderia causar interpretações equivocadas. A expressão “retroalimentação da escrita” pode passar a ideia de que a escrita está sendo retroalimentada, quando, na verdade, queríamos dizer o contrário. Em busca de diminuir os riscos de problemas de interpretação do texto, optamos por utilizar a expressão “influências/ interferências da escrita na fala”.

vez, propõe que linguagem é um sistema aberto, dinâmico, complexo e não linear. Apoiando-se na TSDC, argumentamos que o aprendizado do sistema ortográfico ocorre de forma não linear e motiva a instabilidade e, conseqüentemente, a auto-organização e adaptação do sistema linguístico. Pretendemos argumentar, em suma, que o aprendizado da escrita ortográfica envolve o gerenciamento e a auto-organização de categorias linguísticas múltiplas, dinâmicas e probabilísticas.

### 1.3 Pergunta de pesquisa

A pergunta de pesquisa a ser respondida nesta tese é a seguinte: *categorias linguísticas, aparentemente semelhantes, se manifestam de forma análoga ou diferente na fala e na escrita de alunos do Ensino Fundamental?* Considerando-se que as categorias linguísticas são dinâmicas, múltiplas e probabilísticas, como prediz o Modelo de Exemplares e a TSDC, é coerente pensar que haja diferenças entre elas no que diz respeito à realização gráfica e oral. Busca-se argumentar que os alunos do EF acessam esquemas linguísticos múltiplos para pronunciarem e escreverem as vogais pretônicas, as vogais postônicas finais, os róticos finais de verbos e os róticos finais de nomes.

### 1.4 Hipóteses e Tese

Quatro hipóteses foram formuladas neste estudo. Duas delas serão testadas em amostras de dados de escrita; as outras duas, em amostras de dados de fala. Vejamos, primeiro, as duas hipóteses relacionadas à escrita:

*(1a) as trocas de <e, o> por <i, u> na escrita persistem, em índices significativamente diferentes, até estágios distintos da escolarização, a depender do contexto acentual: pretônico ou postônico final;*

*(1b) a omissão da letra <r> na escrita persiste, em índices significativamente diferentes, até estágios distintos da escolarização, a depender da classe gramatical: verbo ou nome.*

Como se pode ver, as duas hipóteses visam testar se existem diferenças nos índices e na persistência dos erros ortográficos relacionados a cada categoria linguística. De modo geral, esperamos que os erros ortográficos persistam em diferentes índices na escrita infantojuvenil a depender da categoria linguística avaliada. As hipóteses (1a) e (1b) foram fundamentadas nas seguintes evidências já documentadas na literatura:

- erros ortográficos relacionados às vogais pretônicas são mais recorrentes do que os erros ortográficos relacionados às vogais postônicas finais (MIRANDA, 2010; CRISTÓFARO-SILVA E GUIMARÃES, 2013).

- os erros ortográficos relacionados aos róticos finais são mais recorrentes em verbos do que em nomes (COSTA, 2009).
- aspectos fonético-fonológicos (MIRANDA, 2010; 2020), morfológicos (MOTA, 2012), lexicais (SANTOS, 2000) e contextuais (TREIMAN, 2017) interferem no aprendizado da ortografia.

Como mostram os dois primeiros tópicos, há evidências de que os aprendizes de escrita cometem diferentes índices de erros ortográficos em contextos pretônico e postônico final, e em verbos e nomes terminados em <r>. Queremos testar se, além de diferirem quanto aos índices de ocorrência, erros ortográficos diferem-se no que diz respeito à persistência ao longo da escolarização. O tópico três indica que aspectos fonológicos, morfológicos, contextuais e lexicais podem interferir no aprendizado da ortografia. Considerando-se que as categorias avaliadas neste trabalho se diferem quanto à fonologia, à morfologia e ao léxico, esperamos que os alunos aprendam a representá-las na escrita em diferentes estágios da escolarização.

Se as hipóteses (1a) e (1b) se confirmarem, teremos evidências de que os aprendizes traçam diferentes percursos de aprendizagem para grafarem uma mesma letra em contextos distintos. Com o apoio dos Modelos de Exemplares (BYBEE, 2001), esses resultados poderão ser interpretados como evidências de que os aprendizes de escrita gerenciam esquemas linguísticos múltiplos ao longo do processo de escolarização. Além disso, os diferentes estágios de aprendizado da ortografia poderão trazer evidências da não linearidade do processo de ortografização, característica comum de sistemas dinâmicos, segundo a TDSC (BOT, LOWIE; VESPOOR, 2007).

Além de analisar os dados de escrita, analisaremos dados de fala, em busca de testar se o avanço da escolarização interfere na produção de variantes orais. Vejamos as hipóteses testadas:

*(2a) o aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônico ou postônico final.*

*(2b) o aumento do rótico final na fala ao longo dos anos escolares depende da classe gramatical: verbo ou nome.*

Como já foi mencionado na seção 1.1, as vogais pretônicas e postônicas finais podem ser pronunciadas com vogais médias ou altas em algumas variedades do PB. As vogais médias – [e, o] – são variantes orais que se aproximam da forma ortográfica em palavras como ‘*menino*’ e ‘*pato*’, por exemplo. Ou seja, os alunos podem escrever as letras <e, o> e falar as vogais médias [e, o] em contextos pretônico e postônico final. Sabendo-se disso, a hipótese (2a) visa testar se o contexto acentual interfere no aumento de vogais médias (variantes próximas à ortografia) ao longo dos anos escolares.

O rótico final, por sua vez, pode ser ora produzido ora cancelado em verbos e em nomes. A produção do rótico é a variante que mais se aproxima da forma gráfica. Ou seja, escreve-se a letra <r> e pronuncia-se o R-final. Sabendo-se disso, a hipótese (2b) testa se a classe gramatical das palavras interfere no aumento de róticos finais (variante próxima à ortografia) ao longo dos anos escolares. As hipóteses (2a) e (2b) foram fundamentadas nas seguintes evidências disponíveis na literatura:

- A escrita interfere na produção oral de fenômenos fonológicos variáveis (CHEVROT *et al.*, 2000; CRISTÓFARO-SILVA E GRECO, 2010; ADAMOLI, 2013).
- Em fenômenos fonológicos regulares, como o cancelamento do R-final em verbos (CALDAS; CALLOU, 2014) e a vocalização da lateral pós-vocálica (HORA, 2006), o aumento da escolaridade não motivou alterações na fala dos indivíduos.
- Vogais altas variam com vogais médias em contexto pretônico; em contexto postônico final, as vogais altas são produzidas com regularidade (VIEGAS, 1995; BISOL, 2003).
- O fenômeno de cancelamento do rótico final ocorre de forma regular nos verbos (OLIVEIRA, 1997) e de forma variável nos nomes (HUBACK, 2003).

Como demonstrado nos tópicos, a literatura já documentou que a escrita pode influenciar na forma de falar das pessoas. No entanto, há indícios de que essas influências dependem da categoria linguística. Análises sociolinguísticas já demonstraram que o avanço da escolaridade nem sempre motiva alterações na fala (CALDAS E CALLOU, 2014; HORA, 2006). Considerando-se que o avanço da escolaridade está relacionado ao aumento do contato formal com a escrita, se pode inferir que nem sempre o aumento do acesso a materiais escritos influencia na oralidade. Ademais, há evidências de que as vogais pretônicas e os nomes terminados em R-final variam na fala, mas as vogais altas postônicas finais e o cancelamento dos róticos finais de verbos no infinitivo ocorrem de forma categórica. Sabendo-se disso, esperamos que as influências da escrita na fala dependam da categoria linguística avaliada.

Se as hipóteses (2a) e (2b) forem confirmadas, haverá indícios de que as categorias linguísticas se organizam de forma diferente na língua portuguesa. Por exemplo: se, ao longo dos anos escolares, os alunos passarem a falar as vogais [e, o] em contexto pretônico, mas não em contexto postônico final, haverá evidências de que os alunos gerenciam categorias múltiplas ao pronunciarem as vogais em diferentes posições acentuais. Do mesmo modo, se houver aumento na produção de [h] finais em nomes, mas não em verbos, haverá indícios de que os verbos e nomes terminados em <r> constituem categorias linguísticas distintas na língua portuguesa. Essas evidências confirmarão a premissa de multiplicidade das representações mentais sustentada pela

Teoria de Exemplos. Além disso, se as hipóteses forem confirmadas, haverá indícios de que o aprendizado da forma ortográfica pode causar efeitos mais ou menos evidentes na oralidade, o que confirma a premissa de não linearidade e de dinamicidade do sistema linguístico, sustentada pela TSDC.

De forma geral, as hipóteses (1a), (2a), (1b) e (2b) serão testadas em busca de evidências que corroborem a seguinte tese: *as letras são símbolos discretos que se associam a esquemas múltiplos, que emergem da interação complexa entre palavras, sequências fonéticas, morfológicas e ortográficas.*

## 1.6 Método

Os dados desta pesquisa foram coletados por meio de metodologia experimental. O experimento desenvolvido envolveu perguntas-guias associadas a figuras e foi dividido em três etapas subsequentes. A primeira etapa visou familiarizar os participantes aos estímulos da pesquisa. A segunda etapa teve a finalidade de coletar dados de fala; a terceira etapa envolveu a coleta de dados de escrita. Participaram da pesquisa 50 alunos do EF de uma escola pública de Belo Horizonte – MG. Os anos escolares selecionados foram 1º, 3º, 5º, 7º e 9º anos do EF. Ao todo, foram analisados 2.907 dados de fala e 3.057 dados de escrita. A análise dos dados foi feita por meio da linguagem R (R CORE, 2022). Utilizamos métodos estatísticos descritivos e inferenciais, bem como modelos lineares generalizados múltiplos em busca de descrever as amostras de dados para tentar encontrar evidências sobre as diferentes relações entre a fala e a escrita ao longo da escolarização.

## 1.7 Organização do texto

O texto está subdividido em sete capítulos que discutem os seguintes aspectos:

**Capítulo 2:** revisa a literatura sobre oralidade e escrita e busca demonstrar como o presente estudo pretende avançar em relação às pesquisas precedentes. O conjunto de obras revisadas abarca, principalmente, as interações entre os sistemas de fala e de escrita. Primeiramente, ressalta-se a natureza variável do sistema oral e aborda-se, especificamente, a variação sonora das vogais médias e dos róticos. Em seguida, serão descritas as relações entre letras e sons; enfoque será dado ao sistema alfabético-ortográfico do português brasileiro. Além disso, é discutida a importância do erro ortográfico no processo de desenvolvimento do conhecimento gráfico, bem como os efeitos do aprendizado da escrita no desenvolvimento da linguagem. Por fim, serão apresentadas evidências sobre as influências da fala na escrita, bem como da escrita na fala no percurso escolar.

**Capítulo 3:** apresenta o quadro teórico que fundamenta a análise realizada nesta pesquisa. Na primeira seção, apresentam-se alguns pressupostos dos modelos teóricos tradicionais. Na

segunda, os pressupostos tradicionais são contrapostos às premissas dos Modelos de Exemplos (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007). Pretendemos demonstrar que, diferentemente das abordagens tradicionais, os Modelos de Exemplos assumem a multiplicidade das representações mentais. Na terceira seção, as premissas da Teoria dos Sistemas Dinâmicos e Complexos são apresentadas (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Resumidamente, essa teoria acomoda a ideia de que os sistemas oral e escrito interagem entre si, uma vez que se postula que a linguagem é um sistema aberto, dinâmico e complexo.

**Capítulo 4:** aborda os procedimentos metodológicos da pesquisa. Primeiro, a seleção das palavras, o perfil dos participantes escolhidos e o desenho experimental serão descritos. Em segundo lugar, aborda-se como foi feita a realização da coleta dos dados. Por fim, os princípios metodológicos utilizados na organização e na análise dos dados serão apresentados.

**Capítulo 5:** analisa os dados de fala e de escrita relacionados às vogais médias pretônicas e postônicas finais. Primeiro, os dados são descritos. E segundo lugar, as hipóteses (1a) e (2a) são testadas. Em seguida, os resultados são teoricamente discutidos. Por fim, a visão dos professores sobre os erros ortográficos em contextos pretônico e postônico final é apresentada.

**Capítulo 6:** analisa os dados de fala e de escrita relacionados aos róticos finais de verbos e de nomes. Primeiro, os dados são descritos. E segundo lugar, as hipóteses (1b) e (2b) são testadas. Em seguida, os resultados são teoricamente discutidos. Por fim, os erros ortográficos envolvendo o <r> final são discutidos pela visão dos professores.

**Capítulo 7:** retoma os objetivos da pesquisa, resume os resultados encontrados, apresenta as limitações encontradas e destaca a relevância deste estudo.

## 2. REVISÃO BIBLIOGRÁFICA

Neste capítulo, parte da literatura sobre oralidade e escrita é revisada, com o objetivo de elucidar o que já se sabe sobre esse campo de pesquisa e de demonstrar como o presente estudo avança em relação à bibliografia precedente. Na primeira seção, ressalta-se a natureza variável do sistema oral e descreve-se, especificamente, as vogais médias e os róticos na oralidade. Na segunda seção, discutem-se alguns aspectos da escrita, sobretudo, da escrita alfabética, enfatizando a relação entre as letras e os sons. Na terceira seção, aborda-se a importância da ortografia e apresentam-se algumas descrições da norma ortográfica do português brasileiro. Na quarta seção, a noção de erro ortográfico é apresentada. Além disso, revisam-se trabalhos que analisaram erros ortográficos cometidos nas categorias linguísticas de interesse desta pesquisa: vogais médias pretônicas, vogais médias postônicas finais, róticos finais em verbos, róticos finais em nomes. Na quinta seção, discute-se o processo de aprendizado do sistema alfabético-ortográfico, sobretudo, no que diz respeito ao acesso, à conscientização e ao desenvolvimento do conhecimento linguístico. Por fim, na sexta seção, são revisados trabalhos que demonstraram que a escrita também pode influenciar na fala.

### 2.1 A Fala

A fala pode ser definida como

o meio principal pelo qual a língua é expressa em todas as culturas humanas, exceto para pessoas surdas. A fala é uma modalidade da linguagem. A comunicação da fala é comum a quase todos os humanos em qualquer cultura, em qualquer parte da terra. O produto final da fala é um sinal acústico. Esse sinal representa a mensagem comunicativa do falante. Sob circunstâncias comuns, o sinal acaba rapidamente à medida que as vibrações sonoras são amortecidas pelo mundo físico. (KENT; READ, 2015, p. 11)

O sinal da fala é resultado de um contínuo de sons com correlato articulatório e físico. Tradicionalmente, os sons da fala são expressos por meio de símbolos fonéticos transcritos entre colchetes, por exemplo: [f], [k], [ʒ], [ɣ]. A transição fonética é um método muito utilizado na linguística para documentar a pronúncia de um enunciado (CRISTÓFARO-SILVA, 2017). Vale ressaltar que a ideia de transcrever a fala em símbolos fonéticos transmite a falsa sensação de que as línguas são compostas por um pequeno inventário de segmentos discretos e invariáveis. No mundo empírico, contudo, há uma imensa variabilidade de sons produzidos por falantes diferentes, em diferentes sequências, palavras e expressões (PORT, 2007).

Os sons de interesse desta pesquisa – as vogais médias e os róticos – são transcritos pelos seguintes símbolos fonéticos na variedade de Belo Horizonte – MG: [e, o] – vogais médias; [h] – rótico final. Embora sejam simbolicamente semelhantes, as vogais médias e os róticos ocorrem em diferentes contextos, cada qual com suas particularidades. Nas próximas seções, descrevemos



as vogais médias pretônicas e postônicas finais, bem como os róticos finais de verbos e de nomes, em busca de destacar alguns aspectos que diferenciam essas categorias na oralidade. Descrevem-se, a seguir, as vogais e os róticos, respectivamente.

### 2.1.1 Vogais médias na oralidade

A descrição oral de segmentos vocálicos é feita de acordo com suas posições acentuais. As vogais podem ser tônicas, pretônicas e postônicas. As vogais tônicas ocorrem em sílabas portadoras do acento primário<sup>2</sup>: *pecado*. As vogais pretônicas ocorrem em sílabas que antecedem a tônica – *pecado* – e as postônicas em sílabas que sucedem a tônica – *pecado* (CRISTÓFARO-SILVA, 2012). Uma diferença explícita entre as vogais pretônicas e postônicas finais é o contexto/ambiente em que ocorrem. As vogais pretônicas podem ocorrer em início e/ou meio da palavra, em sílabas próximas ou distantes da tônica, em mais de uma sílaba da mesma palavra, próximas de vogais e/ou consoantes (CRISTÓFARO-SILVA; GUIMARÃES, 2013). Por exemplo: ‘leão’, ‘inseguro’, ‘buraco’, ‘paralelepípedo’. Por outro lado, as vogais postônicas finais ocorrem, necessariamente, em posição final de palavras. Exemplo: ‘pato’, ‘dente’, ‘gata’. Ou seja, no que diz respeito ao contexto fonético, as vogais pretônicas são mais heterogêneas do que as vogais postônicas finais (CRISTÓFARO-SILVA; GUIMARÃES, 2013).

O caráter heterogêneo das vogais pretônicas pode ser evidenciado também na distribuição fonética. Na variedade de Belo Horizonte, algumas palavras são pronunciadas com vogais altas pretônicas – [i, u]; outras ocorrem regularmente com vogais médias pretônicas – [e, o]. Há também palavras pronunciadas ora com vogais altas, ora com médias (OLIVEIRA, 1992). O Quadro 2 expõe alguns exemplos.

**Quadro 2:** Exemplos de palavras com vogais pretônicas

Palavras com [i, u]	Palavras com [e, o]	Palavras com [e, i, u, o]
m[i]séria	c[e]gonha	m[e]nino ~ m[i]nino
p[i]rata	s[e]mana	t[e]soura ~ t[i]soura
b[u]raco	t[o]mada	f[o]rmiga ~ f[u]rmiga
t[u]barão	p[o]der	c[o]ruja~ c[u]ruja

Fonte: elaboração própria.

A primeira coluna do Quadro 2 lista palavras pronunciadas, regularmente, com vogais altas pretônicas. A segunda coluna apresenta as palavras pronunciadas, sistematicamente, com vogais médias na variedade de Belo Horizonte – MG (OLIVEIRA, 1992; VIEGAS, 1995). Na terceira coluna, estão as palavras em que a vogal pretônica ocorre variavelmente. Tradicionalmente, a

<sup>2</sup> Principal proeminência de uma vogal em relação às demais vogais do enunciado (CRISTÓFARO-SILVA, 2017).

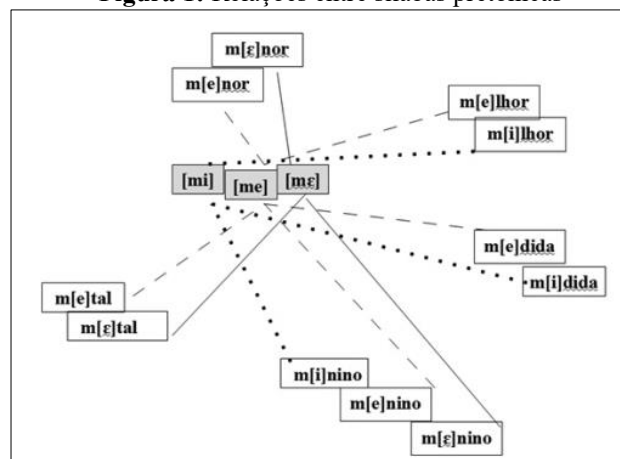
variação no contexto pretônico é explicada como resultado do fenômeno de ‘alçamento vocálico’, que ocorre quando vogais médias passam a ser pronunciadas como vogais altas (VIEGAS, 1995). Muitos trabalhos se dedicaram em descrever tal fenômeno no PB, o que possibilitou uma rica e minuciosa descrição dessas vogais (BISOL, 1989; SILVA, 1989; BATTISTI, 1993; SCHIWINDT, 1995; VIEGAS, 2001; CHAVES, 2013; MAGALHÃES; CAMPOS, 2022).

Bisol (1981) propôs que o alçamento das vogais pretônicas é motivado por um outro fenômeno, denominado de harmonia vocálica, que ocorre quando a pronúncia de uma vogal é alterada por influência de outra vogal da mesma palavra (CRYSTAL, 1988). Bisol (1981) percebeu que, no PB, as vogais médias pretônicas se realizam como altas, quando a vogal tônica é alta. Por exemplo: m[e]nino > m[i]nino. A autora propôs que o processo de harmonia vocálica no contexto pretônico do PB se realiza de maneira variável; isto é, mesmo em ambientes propícios, o alçamento da vogal pretônica ora ocorre, ora não.

Questionando a proposta de Bisol (1981), Oliveira (1992) demonstrou que a harmonia vocálica se aplica em apenas parte do léxico do PB e nem sempre há variação. No português de Belo Horizonte – MG, por exemplo, a palavra [mi. 'fĩdɐ] ‘metida’ é regularmente pronunciada com [i] pretônico, enquanto [me. 'dʒi. tɐ] ‘medita’ se realiza com [e] na primeira sílaba (OLIVEIRA, 1992, p. 33). O autor levantou evidências de que a variação das vogais pretônicas tem motivação lexical. Isso significa que cada item lexical se comporta de um jeito diante do fenômeno de alçamento vocálico: uns são produzidos sistematicamente com vogais altas pretônicas; outros apresentam flutuações entre vogais médias e altas, e alguns se realizam com vogais médias. Para Oliveira (1992), apenas as palavras envolvidas em flutuação podem ser consideradas casos legítimos de variação sonora.

Cristófaros-Silva e Guimarães (2013) ilustraram as complexas relações estabelecidas entre as vogais pretônicas com o seguinte diagrama:

**Figura 1:** Relações entre sílabas pretônicas



**Fonte:** Cristófaros-Silva e Guimarães (2013, p. 319)

Segundo as autoras, o diagrama exposto na Figura 1 ilustra relações entre as sílabas pretônicas [me, mi, me] que ocorrem em início de palavras da língua portuguesa (CRISTÓFARO-SILVA; GUIMARÃES, 2013, p. 319). Essas sílabas representam generalizações fonológicas que emergiram de relações lexicais. É possível notar, na Figura 1, que as palavras se associam a mais de uma sílaba. Isso porque tais palavras são pronunciadas de forma variável na oralidade. O diagrama proposto por Cristófaros-Silva e Guimarães (2013) mostra que a categoria das vogais pretônicas é heterogênea. Em consonância com as autoras, o presente trabalho sugere que a variação lexical das vogais pretônicas deve ser incorporada aos modelos de gramática. Concomitantemente, busca-se defender que a heterogeneidade das vogais pretônicas é parte do conhecimento linguístico acessado pelos aprendizes de escrita ao longo da escolarização.

As vogais postônicas finais estão envolvidas em um outro tipo de fenômeno fonológico denominado, geralmente, de redução vocálica (ASSIS, 2017). A redução vocálica pode ser definida como um fenômeno de enfraquecimento das vogais, que, em estágio avançado, pode resultar no cancelamento vocálico. Embora haja evidências de que as vogais pretônicas também podem reduzir (SOUZA, 2012), o fenômeno parece ser mais recorrente em contexto postônico final (DIAS; SEARA, 2013). A redução vocálica é variável no contexto postônico final, o que significa que o fenômeno ora ocorre, ora não. Quando as vogais postônicas finais são pronunciadas, elas ocorrem sistematicamente como [ɪ, ə, ʊ]. Por exemplo: ['mɔɪɪ] 'mole'; ['balə] 'bala', ['patɔ] 'pato', na grande maioria das variedades do PB (BISOL, 2003). Cristófaros-Silva e Vieira (2015) argumentam que o cancelamento das vogais postônicas finais é um padrão inovador na língua portuguesa. Tomando as autoras como referência, pode-se dizer que as vogais médias postônicas finais estão em estágios mais avançados do que as vogais pretônicas no percurso evolutivo da língua portuguesa.

Até o momento, vimos que as vogais pretônicas e postônicas finais se diferenciam quanto à heterogeneidade silábica e quanto à variação fonética. Há ainda, outro aspecto que as distingue: a morfologia. As vogais pretônicas podem pertencer a prefixos – morfemas derivacionais inseridos antes do radical (AZEREDO, 2018, p. 155). Por exemplo: '(des)cansa', '(des)necessário'; '(re)organizar', '(re)capitular'; (pre)determinado, (pre)disposto, etc. Além disso, essas vogais podem também constituir radicais das palavras. O radical é a parte da palavra que carrega o seu significado externo (BECHARA, 2000). Por exemplo: a vogal pretônica [e] é parte do radical *menin-* que, juntamente aos sufixos, forma diferentes palavras da língua portuguesa, tais como: 'menino', 'meninice', 'meninada', etc.

As vogais postônicas finais, por outro lado, apresentam certa regularidade morfológica por poderem corresponder à flexão de gênero. Segundo Azeredo (2018, p. 170), "o gênero é uma propriedade gramatical inerente aos substantivos e que serve para distribuí-los em dois grandes

grupos: nomes masculinos e nomes femininos”. Alguns nomes ocorrem no gênero masculino e no gênero feminino, por exemplo, ‘pato - pata’. Outros não variam em gênero, sendo unicamente masculinos – como ‘(o) planeta’ – ou unicamente femininos: ‘(a) terra’ (CÂMARA JUNIOR, 2015 [1970]). O correspondente fonético do gênero masculino é tipicamente a vogal [u] postônica final e o correspondente fonético do gênero feminino é tipicamente a vogal [ə] (CRISTÓFARO-SILVA, 2013). Cristófaros-Silva e Guimarães (2013) argumentam, com o apoio dos Modelos de Exemplares, que os esquemas morfológicos relacionados ao gênero são produtivos do português, uma vez que todas as palavras da língua estão, obrigatoriamente, associadas ou ao masculino ou ao feminino. Segundo as autoras, a conexão entre vários substantivos e adjetivos utilizados na língua possibilita a emergência de generalizações probabilísticas, como [ə] marcando o feminino e [u] marcando masculino. Neste trabalho argumentamos que as informações morfológicas são acessadas pelos aprendizes de escrita ao longo da escolarização.

A revisão feita nesta seção descreveu as diferenças fonético-fonológicas, lexicais e morfológicas das vogais médias pretônicas e postônicas finais do PB. A intenção da seção foi demonstrar que um mesmo símbolo fonético – [e, o] – pode ocorrer em diferentes contextos, pode estar envolvido em diferentes fenômenos de variação sonora e pode ocupar diferentes funções morfológicas na língua. Ao longo desta tese, argumentaremos que as particularidades das vogais médias pretônicas e postônicas finais devem ser levadas em consideração no estudo das relações entre oralidade e escrita.

### 2.1.2 Róticos finais na oralidade

Além das vogais médias, esta tese analisa o rótico que ocorre em final de verbos e de nomes como, por exemplo, ‘cantar’ e ‘amor’. O R-final pode ser realizado com inúmeras variantes fonéticas, como vibrantes, tepe, fricativas, aproximantes e aproximantes aspiradas, etc. (RENNICKE, 2015). A análise de todas as variantes, contudo, ultrapassa o escopo deste trabalho. Nesta tese, as variantes fonéticas relacionadas aos róticos serão agrupadas em duas categorias: (1) R-final – [f, h, ɹ, ʁ, r, ʁ, x]; (2) cancelamento – [∅].

Na fala, o padrão silábico CVR está relacionado a um fenômeno geralmente denominado ‘cancelamento do R-final’. Quando o R-final é cancelado, uma vogal tônica passa a ocupar a posição final das palavras: [kã'ta] ‘cantar’ e [a'mo] ‘amor’. O cancelamento do R-final ocorre tanto em nomes quanto em verbos, mas há diferenças nas trajetórias de implementação do fenômeno em cada uma das classes gramaticais (OLIVEIRA, 1997).

Muitos trabalhos trouxeram evidências de que o cancelamento do rótico em final de verbos é antigo no português do Brasil (CALLOU; MORAES; LEITE, 1996; OLIVEIRA, 1997; MONARETTO, 2000; RENNICKE, 2015). Inicialmente estigmatizado, o fenômeno era utilizado

para caracterizar o falar de pessoas de classes mais baixas da sociedade, como escravos, ciganos e mulçumanos. Gradativamente, o fenômeno se expandiu e, hoje, afeta, categoricamente, os verbos de infinitivo na fala da maioria dos brasileiros (OLIVEIRA, 1997; RENNICKE, 2016).

Oliveira (1997) encontrou 95,4% de róticos cancelados no final de verbos e assumiu que o fenômeno pode ser considerado quase categórico nessa classe gramatical. Partindo do pressuposto de que todas as mudanças sonoras ocorrem por difusão lexical<sup>3</sup>, Oliveira (1997) argumentou que a regularidade do cancelamento do R-final reflete o estágio avançado da mudança sonora em verbos.

Rennicke (2016), em um estudo detalhado sobre a variação dos róticos do português mineiro, encontrou o cancelamento do R-final em 98,3% dos verbos analisados, o que confirma a regularidade do fenômeno em tal classe gramatical. A autora propôs, apoiando-se nos Modelos de Exemplares, que:

a representação fonológica dos róticos no PB tenha origem lexical. A forma fonética de uma palavra é armazenada junto com o seu significado, o que cria conexões fonéticas, morfológicas e semânticas com outras palavras da língua que tenham semelhanças fonotáticas ou de significado. A partir dessas conexões, o falante forma generalizações probabilísticas sobre variantes posicionais. (RENNICKE, 2016, p. 87)

Em consonância com Rennicke (2016), este trabalho parte do pressuposto de que as categorias dos róticos emergem das relações entre um grande estoque de palavras atestadas na experiência linguística (BYBEE, 1998). Considerando-se que, no final de verbos, o rótico é regularmente cancelado (RENNICKE, 2016, p. 87), é coerente pressupor que formas lexicais, como [kã'ta] '*cantá*', [fa'ze] '*fazê*' e [so'hi] '*sorri*', sejam parte da representação.

No caso dos nomes, o cancelamento do rótico ocorre em menores índices do que em verbos (OLIVEIRA, 1997; HUBACK, 2006; RENNICKE, 2016). Para Oliveira (1997), o menor índice do cancelamento do rótico final em nomes decorre do tempo de existência do fenômeno na língua portuguesa. O fenômeno de cancelamento é mais recente em nomes do que em verbos e parece se implementar lexicalmente. Nos dados de Oliveira (1997), alguns itens, como '*mulher*' e '*qualquer*', favoreceram o cancelamento do R-final, enquanto outros, como '*interior*' e '*vestibular*', inibiram-no.

Huback (2006), em busca de evidências adicionais que corroborassem a abordagem difusionista da mudança sonora, demonstrou que o cancelamento do R-final em nomes se dá paulatinamente, de palavra a palavra, e que a frequência lexical pode explicar o porquê de algumas

---

<sup>3</sup> A difusão lexical é uma perspectiva teórica sobre mudança linguística. Entre os pressupostos desta teoria, encontra-se o seguinte: a mudança é foneticamente abrupta e lexicalmente gradual; ou seja, o contexto fonético muda de uma só vez, mas as palavras não mudando lentamente e, nem sempre, todas são afetadas (BORTONI; GOMES; MALVAR, 1992; OLIVEIRA, 1992; VIEGAS, 2001).

palavras serem mais suscetíveis à mudança do que outras. Itens frequentes, como ‘*a partir de*’ e ‘*qualquer*’ favoreceram o cancelamento do rótico, nos dados de Huback (2006). Para a autora, esses resultados indicam que a representação mental armazena palavras inteiras utilizadas frequentemente no uso linguístico e que a organização do léxico se dá por conexões semânticas e morfológicas. A proposta de Huback (2006) condiz com as ideias defendidas na presente pesquisa. Argumentamos que as relações entre os itens lexicais possibilitam a emergência de padrões fonéticos, morfológicos e ortográficos.

As evidências levantadas até o momento mostram que os róticos de verbos e de nomes têm diferentes realizações na fala: nos verbos, o R-final é regularmente cancelado, enquanto nos nomes o fenômeno é variável (OLIVEIRA, 1997; HUBACK 2006). Considerando-se que as representações fonológicas emergem do uso, pode-se dizer que as categorias abstratas dos róticos são múltiplas e, conseqüentemente, diferentes a depender da classe gramatical.

Além das diferenças fonéticas, os verbos e os nomes se diferenciam quanto à morfologia. Nos nomes, o R-final pode fazer parte do radical da palavra, além de compor sufixos. Consideremos os exemplos no Quadro 3:

**Quadro 3:** Estrutura morfológica dos nomes terminados em <r> final

Estrutura morfológica dos nomes terminados em R-final	Exemplos
sufixo -dor e -tor agentivo	orientador, agricultor
sufixo -dor e -tor não agentivos	computador, diretor
outros sufixos	circular, popular
parte do radical da palavra	mulher, colher, lugar

Fonte: adaptado de Huback (2003).

Huback (2003) demonstrou que o cancelamento do R-final é favorecido em palavras terminadas pelo sufixo *-dor* e *-tor* agentivo e não agentivo. Apoiando-se no Modelo de Exemplares, Huback (2003, p. 111) argumenta que “pode ser que os sufixos *-dor* e *-tor*, independentemente de serem agentivos, estejam formando uma rede que, como um todo, é mais propensa ao cancelamento do (r)”.

Nos verbos, o R-final corresponde ao morfema de infinitivo verbal. Todo verbo tem, obrigatoriamente, sua forma infinitiva. Assim, pode-se dizer que o infinitivo verbal é altamente produtivo no PB, uma vez que todos os verbos da língua devem, obrigatoriamente, se conectar a ele. Nesta tese, argumentamos que as informações morfológicas particulares de cada categoria linguística podem ser acessadas pelos aprendizes ao longo da escolarização.

Em síntese, esta seção teve o objetivo de descrever as diferenças entre verbos e nomes, visando evidenciar que sons aparentemente semelhantes – como [h] – podem estar envolvidos em diferentes percursos de variação e mudança sonora, além de ocupar diferentes funções

morfológicas na língua. Ao longo desta tese, argumentaremos que as particularidades do róticos finais de verbos e de nomes devem ser consideradas no estudo das relações entre oralidade e escrita.

## 2.2 A Escrita

Em culturas letradas, a escrita está presente em inúmeras práticas sociais. Desde a primeira infância, antes mesmo da alfabetização formal, crianças observam textos escritos e, cotidianamente, se deparam com pessoas lendo e escrevendo a sua volta (TREIMAN; KESSLER, 2014). A grande familiaridade com a escrita pode despertar a falsa sensação de que escrever é uma habilidade natural da espécie humana. Até mesmo na teoria linguística, há quem assume que a aprendizagem da escrita ocorre com a mesma naturalidade da aquisição da fala (GOODMAN; GOODMAN, 1979; SMITH, 1989). Goodman e Goodman (1979, p. 457), por exemplo, assumem que “as crianças aprendem a ler e a escrever do mesmo modo como aprendem a falar e a ouvir e pela mesma razão [...] a razão é a necessidade. A aprendizagem da língua, seja oral, seja escrita, é motivada pela necessidade de comunicação, de compreender e de ser compreendido<sup>4</sup>”.

Há de se ressaltar, todavia, que todos os seres humanos – sem afasias – têm a capacidade de falar, mas nem todos fazem o uso da escrita. As populações humanas eram inicialmente ágrafas e, com o passar do tempo, desenvolveram a tecnologia da escrita. Quando a escrita surgiu, apenas alguns poucos indivíduos eram letrados como, por exemplo, escribas, copistas, classes altas. Hoje em dia, ainda existem línguas ágrafas, como, por exemplo, a língua Yaahte (COSTA, 2016). Falantes dessa língua não utilizam os meios gráficos de comunicação por não possuírem sistemas de escrita. Além disso, há indivíduos que não sabem escrever, mesmo em sociedades letradas. No Brasil atual, por exemplo, cerca de 6,6% da população é analfabeta, de acordo com o IBGE (BRASIL, 2019). Vê-se, portanto, que é difícil sustentar a premissa de que escrever seja tão natural quanto falar. Se assim o fosse, todos os seres humanos, falantes de qualquer língua, deveriam ser capazes de desenvolver a língua escrita com a mesma espontaneidade com que desenvolvem a língua oral.

De acordo com Treiman e Kessler (2014), a escrita é uma ferramenta inventada pelo homem, e não uma habilidade natural da espécie. Antes da sua invenção, os pensamentos humanos eram transmitidos apenas oralmente e, por isso, eram momentâneos. A necessidade de preservar ideias motivou a criação de um sistema gráfico de representação linguística. Pode-se dizer que “as

---

<sup>4</sup> No original: “We believe that children learn to read and write in the same way and for the same reason that They learn to speak and listen [...] The reason is need. Language learning whether oral or written is motivated by the need to communicate, to understand and be understood” (GOODMAN, GOODMAN, 1979, p. 457).

raízes desse sistema se encontram na necessidade fundamental dos seres humanos de armazenar informação para comunicar, a si mesmos ou a outros, distantes no tempo e no espaço” (FISCHER, 2009, p. 13).

Embora tenham o comum objetivo de registrar graficamente as línguas humanas, os sistemas de escrita seguem diferentes princípios, a depender da língua que representam. Basicamente, os sistemas de escrita são agrupados em dois grandes grupos: os ideográficos e os fonográficos. Os sistemas ideográficos de escrita são formados por desenhos chamados de ideogramas. Esses desenhos registram, primordialmente, o significado (CARVALHO, 2014). Exemplos de escritas ideográficas são a egípcia (hieroglífica), a mesopotâmica (suméria) e a chinesa (CAGLIARI, 1989, p. 108). Por outro lado, sistemas fonográficos registram o significante<sup>5</sup>; isto é, eles dependem essencialmente das unidades sonoras das línguas. Os sistemas fonográficos podem representar as sílabas (sistemas fonográficos silábicos) ou os sons individuais (sistemas fonográficos alfabéticos) (CAGLIARI, 1989). Neste trabalho, ênfase é dada ao sistema alfabético – sistema do português brasileiro.

### 2.2.1 A relação entre letras e sons: a essência da escrita alfabética

Na cultura ocidental, os sistemas de escrita são, em sua maioria, alfabéticos. Um alfabeto é formado por um conjunto finito de símbolos gráficos que representam os símbolos fonológicos (MAN, 2002; VELOSO, 2005; SOARES, 2018). Embora existam diferentes tipos de alfabetos – o grego, o latino, o semítico, o indiano, o fenício, etc. – todos partem do mesmo princípio, denominado de *princípio alfabético*. Este princípio diz respeito ao fato de que cada símbolo gráfico deve representar, ao menos, um som da fala. De acordo com Man (2002, p. 12), o alfabeto é uma das mais importantes ideias da humanidade devido a sua singularidade, a sua simplicidade e a sua adaptabilidade. Com algumas adequações, mas sem muito esforço, um alfabeto pode se adaptar a qualquer língua natural.

O alfabeto do português brasileiro provém do alfabeto latino. Este, por sua vez, surgiu da adaptação do sistema de escrita etrusco (FISCHER, 2009). Durante o século I a.C, o alfabeto latino tinha, praticamente, uma letra para cada som (FISCHER, 2009, p. 132). No português brasileiro, uma letra do alfabeto pode representar mais de um som, assim como um som pode ser representado por várias letras. Por exemplo: a letra <s> representa o som de [z] em *casa* e de [s] em *cassar*. Já o som [i] pode ser registrado pelas letras <e>, em *menino*, ou <i>, em *miséria*. As relações entre letras e sons são complexas devido ao fato de que, nem sempre, a forma gráfica é espelho da forma

---

<sup>5</sup> O significante e o significado são as duas partes do signo linguístico para a teoria saussuriana. O significante é a imagem acústica; ou a impressão psíquica do som. O significado é o conceito (SAUSSURE, 2012, p. 106).



falada (CAGLIARI, 1989). Como a língua falada é muito mais variável e dinâmica do que a língua escrita, é comum haver incompatibilidades entre os dois sistemas (TREIMAN; KESSLER, 2014). Tais incompatibilidades são discutidas em mais detalhes na seção 2.2.2, que aborda a norma ortográfica do PB. Na presente seção, pretende-se, principalmente, ressaltar que os alfabetos se caracterizam pela relação entre os sons e as letras.

Recorrentemente, o termo *som* é utilizado como sinônimo de *fonema* e o termo *letra* como sinônimo de *grafema*. O uso intercambiável da terminologia, no entanto, pode causar interpretações equivocadas, posto que, teoricamente, cada um dos termos se refere a uma unidade específica, como demonstrado no Quadro 4:

**Quadro 4:** Definição de fonema, grafema, letra e som

<b>Termo</b>	<b>Definição</b>	<b>Exemplos</b>
Som/Fone	elemento sonoro concreto, que possui correlato acústico e articulatorio	[p], [dʒ], [k], [u]
Fonema	unidade gramatical da sonoridade, não pronunciável, abstrata e categórica	/p/, /d/, /k/, /u/
Letra	unidades representacionais gráficas, que podem possuir diferentes caligrafias e tamanhos	p, P, p, A, a, A
Grafema	unidade funcional do sistema de escrita. Várias letras podem ser representadas pelo mesmo grafema. <sup>6</sup>	P, A

**Fonte:** elaboração própria.

Os sons/fones e os fonemas são parte do sistema oral, que é inerentemente variável. Já as letras e os grafemas são parte do sistema gráfico, que é conservador e só muda com autorização de órgãos reguladores. Embora uma mesma letra possa ter diferentes tamanhos e formas, a sua interpretação sonora se mantém. Por outro lado, os sons/fones são produzidos variavelmente e assumem diferentes funções, a depender de fatores como o item lexical, o contexto fonético-fonológico, a morfologia, o contexto de uso, etc. (CRISTÓFARO-SILVA, 2011).

Os sons/fones são objeto de estudo da fonética, enquanto os fonemas são parte da fonologia. A fonética se dedica à descrição dos mecanismos articulatorios e acústicos envolvidos na produção e percepção dos sons da fala. A fonologia, por sua vez, analisa a organização dos sons em um sistema linguístico e a função que esses sons desempenham em diferentes línguas (CAGLIARI, 2002). Para teorias fonológicas tradicionais, o fonema é a unidade mínima do sistema sonoro. A motivação de se postular o conceito de fonema se originou, segundo Crystal (1988, p. 112),

<sup>6</sup> Existe ainda a diferença entre letra e grafema no que se refere à unidade: um grafema pode ser composto por mais de uma letra: <CH>, <SS>

do interesse em estabelecer padrões de organização dentro de uma gama indefinidamente grande de sons ouvidos nas línguas. As especificações fonéticas dos sons (ou fones) contêm muito mais detalhes do que o necessário para identificar a maneira como as línguas fazem contrastes de significação. A noção de fonema permitiu que os linguistas agrupassem os fones semelhantes foneticamente como variantes, ou membros, da mesma unidade subjacente. Segundo os linguistas, os fones eram realizações de um mesmo fonema e as variantes eram os alofones dos fonemas.

Pode-se perceber, pelas palavras de Crystal (1988), que, tradicionalmente, o sistema fonológico é caracterizado como sendo de natureza discreta, na medida em que apenas entidades contrastivas o constituem. Toda a gama de variantes identificadas no uso linguístico está fora do inventário sonoro das línguas naturais, na vertente tradicional.

Abordagens teóricas alternativas – como os Modelos de Exemplares (ver Capítulo 3) – têm questionado a natureza discreta do sistema fonológico, ao postularem a relevância dos alofones (variantes de um mesmo fonema) para a organização da representação mental (BYBEE, 2001; PIERREHUMBERT, 2003; CRISTÓFARO-SILVA; GOMES, 2004). Ao invés de se agrupar as variantes em uma mesma unidade subjacente, propõe-se que a representação mental é composta de categorias múltiplas que armazenam toda a gama de sons encontrada no uso linguístico (CRISTÓFARO-SILVA, 2011). Nessa perspectiva, os sons da fala (vogais e consoantes) são compreendidos como rotinas motoras com correlato acústico, articulatorio e cognitivo. Assim sendo, os variados movimentos motores (motricidade) feitos na produção de um som estão registrados na representação cognitiva (CRISTÓFARO-SILVA, 2011).

Se, por um lado, o sistema oral é composto por inúmeras variantes sonoras, por outro, o sistema de escrita alfabético é formado por um conjunto restrito de unidades: as letras e os grafemas. As letras podem variar quanto à caligrafia e à tipografia (CARVALHO, 2014). Já os grafemas são unidades abstratas e invariáveis. Dialogando com a noção de fonema, o grafema é definido como a unidade mínima contrastiva do sistema de escrita (CRYSTAL, 1998, p. 128). Ou seja, o grafema cumpre a função de atribuir um mesmo valor às diferentes caligrafias e traçados das letras (CARVALHO, 2014). Por exemplo: as letras <e, E, e, E, e, E> são representadas pelo mesmo grafema: <E>.

Embora haja variação quanto à forma das letras, estas possuem uma única interpretação sonora (CRISTÓFARO-SILVA, 2011). Por exemplo: as letras <e, E, e, E, e, E> têm como correlato sonoro [e], seja oral (mesa), seja nasal (pente). Um indivíduo alfabetizado associa, rapidamente, a letra <e> ao som [e]. Há, no entanto, casos em que a letra <e> se associa a outro som. Por exemplo: na palavra <menino>, a letra <e> pode ser pronunciada como [i]. Casos como esse são consequência da variação do sistema oral. A letra <e> permanece estática na escrita, embora, na fala, possa ocorrer, variavelmente, os sons [i] ~ [e]. Vê-se, portanto, que os sons e as letras têm características distintas e são parte de sistemas diferentes (CRISTÓFARO-SILVA, 2011).

As noções de som, fonema, letra e grafema se encontram na essência dos alfabetos e, por isso, são pilares das abordagens de desenvolvimento da escrita, entre as quais se destacam três: a fonológica, a psicogenética e a aprendizagem estatística (POLLO, TREIMAN, KESSLER, 2015). Embora tenham diferenças substanciais – cuja discussão detalhada ultrapassaria o escopo desta pesquisa – essas teorias concordam que uma etapa fundamental da aprendizagem da escrita é a descoberta do princípio alfabético; isto é, a descoberta de que “as letras que constituem nossa língua escrita representam sons individuais que constituem nossa língua falada” (BYRNE, 1998, p. 1). Pode-se ver, portanto, que a construção do saber alfabético envolve, necessariamente, a compreensão das relações entre as letras e os sons. Faz-se relevante, então, avaliar essas relações para que se compreenda o processo de aprendizagem da escrita alfabética como um todo.

Tradicionalmente, assume-se que as relações entre oralidade e escrita se estabelecem pela relação entre fonemas e grafemas. Na presente tese, por outro lado, argumentamos que as relações entre oralidade e escrita se estabelecem pelas associações entre letras – símbolos discretos – e categorias linguísticas múltiplas. Nessa concepção, uma mesma letra, como <e, o, r>, pode se relacionar a diversas categorias da língua portuguesa, que armazenam variadas informações fonéticas, fonológicas, morfológicas e lexicais.

As teorias de desenvolvimento da escrita reconhecem que, além de compreenderem as relações letra-som, os aprendizes precisam transitar do saber alfabético para o saber ortográfico ao longo da escolarização (FERREIRO, 1990; MASSINI-CAGLIARI, CAGLIARI, 2004; SEYMOUR, 2005, SOARES, 2018). Ou seja, para além do princípio alfabético, os aprendizes de escrita precisam aprender as convenções ortográficas de sua língua. A seção a seguir se dedica a este aspecto: apresentam-se algumas especificidades das ortografias de diferentes línguas, ressalta-se a sua importância para a manutenção do sistema de escrita e descreve-se a norma ortográfica do português brasileiro.

### 2.2.2 Ortografia

Para que o sistema de escrita seja decifrável ao longo do tempo, é necessário haver normas que mediem seu uso. As normas, também chamadas de ortografia, cumprem a função de unificar a forma escrita das palavras, para que estas possam ser lidas por falantes de variedades linguísticas diferentes (GOMES DE MORAIS, 1998). Se cada falante pudesse escrever conforme fala, uma mesma palavra teria diferentes representações gráficas, haja vista que a fala é variável. Por exemplo: a palavra *fósforo* poderia ser grafada como \*fósfru, \*fósfuru, \*fósfu, uma vez que todas essas variantes são passíveis de ocorrer na fala. Porém, se escrever fosse livre de normas, haveria o enorme risco de a escrita se tornar ilegível (MASSINI-CAGLIARI; CAGLIARI, 1999). A norma ortográfica, portanto, tem o objetivo de unificar a escrita para que esta seja legível por diversos

indivíduos ao longo de muitas gerações. Como bem define Gomes de Moraes (1998, p. 10),

a ortografia funciona como um recurso capaz de “cristalizar” na escrita as diferentes maneiras de falar dos usuários de uma mesma língua. Escrevendo de forma unificada, podemos nos comunicar mais facilmente. E cada um continua tendo a liberdade de pronunciar o mesmo texto à sua maneira quando, por exemplo, o lê em voz alta.

Foi visto, na seção anterior, que sistemas de escrita de diferentes línguas podem seguir o mesmo princípio: o princípio alfabético. Convém destacar, no entanto, que embora os alfabetos sejam semelhantes na essência, eles se diferenciam no que diz respeito à ortografia. Isso porque os sistemas alfabéticos são fundamentados em um mesmo princípio abstrato, enquanto os sistemas ortográficos são regulados por intervenções conscientes, fruto de um acordo social (COULMAS, 1989; GOMES DE MORAIS; TEBEROSKY, 1994; SOARES, 2018).

Entre as características que diferenciam ortografias destacam-se a estrutura silábica e as correspondências entre letras e sons (SEYMOUR, ARO, ERSKINE, 2003; RUSSAK, ZARETSKY, 2022). No que diz respeito à estrutura silábica, as línguas românicas – português, francês, italiano, espanhol – têm poucos encontros consonantais; predominam-se, nessas línguas, as sílabas formadas por uma consoante e uma vogal (CV). Já as línguas germânicas, como o inglês, o dinamarquês e o alemão, têm estruturas silábicas mais complexas, com um alto número de encontros consonantais e com a presença de sílabas formadas por várias consoantes (SOARES, 2018). Seymour, Aro e Erskine (2003) encontraram indícios de que a complexidade da estrutura silábica dificulta a decodificação dos padrões ortográficos.

Outra diferença fundamental entre as ortografias diz respeito às correspondências entre as letras e os sons. Os sistemas de escrita alfabéticos são classificados pelo critério de maior ou menor complexidade nas correspondências entre letras e sons. Esse critério é chamado de *profundidade ortográfica* (KATZ; FROST, 1992). São profundas as ortografias que apresentam grandes inconsistências entre letras e sons; ou seja, em ortografias profundas, um mesmo grafema pode representar mais de um fonema, assim como diferentes grafemas podem representar um mesmo fonema. Por outro lado, em ortografias transparentes, cada grafema corresponde a um único fonema (SOARES, 2018). Em um extremo do *continuum*, há sistemas alfabéticos consideravelmente transparentes, como o finlandês e o turco; em outro, há sistemas muito profundos, como o inglês (CARAVOLAS, 2004).

Estudos sobre profundidade ortográfica tendem a comparar ortografias de diferentes línguas alfabéticas, visando verificar como se dá o desenvolvimento do conhecimento ortográfico em línguas mais e menos profundas (CARAVOLAS *et al.*, 2012; CARAVOLAS; SAMARA, 2015; MOLL *et al.*, 2014). Seymour *et al.* (2003) fizeram um estudo comparativo com treze línguas e concluíram que há diferenças no processo de aprendizagem da ortografia em função da

profundidade ortográfica: aprendizes de ortografias profundas podem levar o dobro de tempo para desenvolver as mesmas habilidades de aprendizes de ortografias transparentes (SEYMOUR, ARO, ERSKINE, 2003). Sprenger-Charolles (2004) comparou o processo de aprendizagem da ortografia do inglês, do francês, do alemão e do espanhol e concluiu que a ortografia mais difícil de se aprender é a da língua inglesa – que possui o sistema ortográfico extremamente profundo. Para a autora, o aprendizado da ortografia “depende parcialmente das características específicas de cada língua, e não apenas de princípios gerais comuns a todas as línguas” (SPRENGER-CHAROLLES, 2004, p. 43). Sabendo-se disso, é necessário considerar a profundidade de cada ortografia para compreender o processo de aprendizado. Vejamos, portanto, as características da ortografia do português brasileiro: língua de interesse desta pesquisa.

De acordo com Soares (2018), o português brasileiro tem um sistema ortográfico relativamente transparente, haja vista que há relações diretas com o sistema oral, mas também apresenta irregularidades. Os estudos sobre a ortografia do PB, geralmente, categorizam as correspondências letra-som como regulares, regulares contextuais e irregulares (LEMLE, 1987; GOMES DE MORAIS, 1998; NÓBREGA, 2013; FARACO, 2015; SOARES, 2018). O Quadro 5 resume essas relações:

**Quadro 5:** Relações entre letras e sons na ortografia do português

Relação letra-som	Descrição	Som/ Contexto	Letra
Regulares diretas	um som é representado por uma letra	[b]	<b>bola</b>
Regulares contextuais	um som em diferentes contextos é representado por letras diferentes	[h] em final de sílaba [h] intervocálico	<b>mar</b> <b>carro</b>
Regulares contextuais	dois ou mais sons em contextos diferentes são representados pela mesma letra	[e] tônico [i] átono final	<b>seca</b> <b>pente</b>
Irregulares/arbitrárias	um som é representado por duas ou mais letras em um mesmo contexto	[i] pretônico	<b>perigo</b> <b>pirata</b>

**Fonte:** Lemle, 1987; Gomes de Moraes, 1998; Soares, 2018.

As correspondências regulares diretas expressam a relação “um para um” entre letras e sons. Há pouquíssimos casos de regularidades diretas na ortografia do PB (LEMLE, 1987) e os aprendizes raramente apresentam dificuldades para compreendê-las (SOARES, 2018).

As regularidades contextuais exigem do aprendiz noções dos sons, das letras e dos contextos em que os segmentos ocorrem. Por exemplo: a vogal <o> ocorre, sistematicamente, em sílabas átonas finais que, na fala, terminam com [ʊ]. De acordo com Gomes de Moraes (1998), os padrões regulares contextuais podem ser aprendidos rapidamente, se a escola ajudar os aprendizes a compreenderem as regras ortográficas aplicadas em cada contexto.

As correspondências letra-som irregulares, por outro lado, são as mais difíceis de se

aprender e, em muitos casos, deixam dúvidas por toda a vida (LEMLE,1987). Isso acontece porque, de acordo com Lemle (1987, p. 32), as irregularidades dependem da memorização da ortografia de cada palavra individualmente. Gomes de Moraes (1998) sugere que as crianças sejam expostas à escrita correta de palavras irregulares para que possam memorizar sua imagem visual.

Há de se apresentar ainda as regularidades ortográficas relacionadas à morfologia da língua portuguesa. Como destacado por Lemle (1987), as regularidades morfológicas podem ajudar os aprendizes na escolha dos padrões corretos da escrita do PB. Por exemplo: palavras terminadas pelo sufixo -agem são todas grafadas com <g> (LEMLE, 1987, p. 37). Soares (2018, p. 303) sintetiza a importância da morfologia para o aprendizado da ortografia do português: “em ortografias próximas da transparência, como o português brasileiro, a consciência morfológica e a sua tradução em regras ortográficas tornam-se um recurso de facilitação para o domínio da grafia de correspondências irregulares grafema-fonema.”

As relações letra-som das categorias linguísticas avaliadas no presente estudo são descritas da seguinte maneira nos manuais de ortografia do PB: as vogais postônicas finais e os róticos em final de palavras são padrões regulares contextuais; as vogais pretônicas envolvem correspondências irregulares entre letras e sons (LEMLE, 1987). Se as irregularidades são mais difíceis de se aprender do que as regularidades contextuais, como apontou Lemle (1987), então a ortografia das vogais médias pretônicas deve ser mais difícil do que a ortografia do <r> final e das vogais postônicas finais. De fato, há evidências de que as crianças cometem erros ortográficos em contexto pretônico (LEMOS, 2001; MIRANDA, 2010). No entanto, as crianças também erram bastante a escrita do <r> final de verbos no infinitivo (COSTA, 2009; TORRES; OLIVEIRA, 2015). Esse fato contraria a ideia de que as irregularidades são as mais difíceis de se aprender; há casos em que os aprendizes demonstram dificuldades tanto em padrões irregulares quanto em padrões regulares contextuais. Cada padrão, portanto, parece envolver trajetórias específicas de aprendizado. Assim, é esperado que o aprendizado de cada padrão ortográfico – seja ele regular, contextual ou irregular – envolva percursos de aprendizado distintos.

Veremos, na próxima seção, que o erro ortográfico é parte fundamental dos percursos de aprendizado da ortografia. Além disso, veremos evidências de que os aprendizes de escrita cometem diferentes índices de erros ortográficos durante o aprendizado de uma mesma letra em diferentes contextos, o que indica, mais uma vez, que há diferentes percursos de aprendizado da ortografia.

### 2.2.3 O erro ortográfico

No dicionário eletrônico Houaiss (2009), o termo ‘erro’ é definido como: “1. Ato ou efeito de errar; 2. Juízo ou julgamento em desacordo com a realidade observada; engano; 3. Qualidade

daquilo que é inexato, incorreto; 4. Desvio do caminho considerado correto, bom apropriado; desregramento”. Além das definições, o dicionário apresenta os seguintes sinônimos para a palavra ‘erro’: “confusão, desacerto, desvio, engano, equívoco, falha, falta, imprecisão, impropriedade, incorreção, inexatidão, lapso”. Como se pode ver, a maioria dos termos relacionados ao item ‘erro’ apresenta conotação negativa. Errar, no sentido dicionarizado, significa cometer uma ação inapropriada, desviante do caminho bom. Tomando esse significado como referência, o erro ortográfico é, por vezes, interpretado como algo a ser evitado.

A interpretação preconceituosa do erro ortográfico é pautada, sobretudo, nos princípios estabelecidos pela gramática normativa. Esse tipo de gramática é responsável por prescrever o que é correto e o que errado na língua. Entende-se por correto aquilo que segue a norma culta padrão, falada pelas classes dominantes e firmada na escrita. Errado é tudo o que desvia dessa norma (TRAVAGLIA, 2009). A tradição escolar, orientada pela vertente normativa da gramática, costuma priorizar o ensino da norma, descartando a imensa variabilidade da língua (CAGLIARI, 1989). Em relação ao ensino da ortografia, a escola tradicional se preocupa mais em identificar e corrigir os erros ortográficos do que em tomar a ortografia como objeto de reflexão. Aprender a ortografia é, de fato, uma necessidade primordial para que o indivíduo se torne um escritor e leitor bem sucedido (TREIMAN; KESSLER, 2014). Contudo, esse aprendizado precisa envolver atividades reflexivas e analíticas. Estigmatizar os erros ortográficos, tratando-os como algo a ser rapidamente combatido, desestimula o aluno, esvazia as possibilidades de reflexão sobre a língua e dificulta o processo de aprendizagem da escrita.

O erro é um elemento natural de todo processo de desenvolvimento (FERREIRO; TEBEROSKY, 1985; HAN; ADOLPH, 2020). Bebês, por exemplo, aprendem muito com seus erros frequentes – como tropeços e quedas ao caminhar. Han e Adolph (2021) mostraram que as quedas (ou erros motores) não atrapalham os bebês a adquirirem a prática necessária para andar. Pelo contrário, as quedas frequentes e de baixo impacto incentivam os bebês a continuarem treinando até alcançarem sucesso na ação de andar. Assim como cair ajuda o bebê a caminhar, errar ajuda o aprendiz a compreender a ortografia de sua língua materna. Os erros ortográficos encontrados nos textos infantis são, portanto,

compreensíveis. Eles revelam que o aprendiz precisa de ajuda para incorporar todas as facetas da escrita. Para o professor que olhe a produção da criança sem muito cuidado, pode parecer que ela apenas errou ou acertou. Para o aprendiz, o “erro” pode revelar diferentes níveis de conhecimento[...]. Incorporar a norma ortográfica é consequentemente um longo processo para quem se apropriou da escrita alfabética. Não podemos nos assustar e, em nome da correção ortográfica, censurar ou diminuir a produção textual no dia dia-a-dia. Enfatizo que o ensino sistemático de ortografia não pode se transformar em “freio” às oportunidades de a criança apropriar-se da linguagem escrita pela leitura e composição de textos reais (GOMES DE MORAIS, 1998, p. 22).

O estudo dos erros ortográficos se justifica tanto do ponto de vista pedagógico, quanto do ponto de vista teórico (ALVARENGA *et al.* 1989). No âmbito pedagógico, o erro ortográfico pode ser uma ferramenta para auxiliar o professor a avaliar criticamente as dificuldades do aluno. Em vez de apenas sinalizar os erros e os acertos, o professor pode descobrir o que motivou o aluno a errar, compreendendo, assim, as hipóteses cognitivas por trás de cada erro. Sabendo o porquê de cada tentativa ortográfica, o professor pode construir métodos efetivos de intervenção que possibilitem o aluno a avançar (NOGARO; GRANELLA, 2004). No âmbito teórico, o estudo do erro ortográfico pode trazer ricas informações sobre o conhecimento linguístico utilizado pelos aprendizes em suas escritas. Como bem coloca Miranda (2014, p. 47), o erro ortográfico é um

dado capaz de revelar as hipóteses das crianças sobre o sistema que elas estão a adquirir e, sobretudo, como um elemento revelador do conhecimento linguístico construído desde os primeiros anos de desenvolvimento da linguagem até o momento em que elas ingressam no processo de escolarização.

As hipóteses construídas pelos aprendizes estão, normalmente, de acordo com as possibilidades e restrições de sua língua (CAGLIARI, 1989; BOHN; SOUZA, 2017). Quando uma criança escreve algo como <\*minino>, por exemplo, ela está infringindo a norma ortográfica, mas dentro das possibilidades do sistema linguístico do PB. A sílaba <mi> ocorre em palavras do português, como, por exemplo, ‘*miséria*’, ‘*minuto*’, ‘*ministro*’, ‘*minhoca*’, *etc.* Assim sendo, é totalmente possível, embora seja não ortográfico, grafar a palavra ‘*menino*’ como <\*minino>. Além disso, tal palavra é pronunciada, em algumas variedades do português, como [mi'niɲo]. Assim, é explicável que o aprendiz escreva <\*minino> em busca de aproximar a forma gráfica da forma oral. Os erros ortográficos não são caóticos; pelo contrário, eles mostram a brilhante capacidade dos aprendizes de criar hipóteses dentro das possibilidades da sua língua materna. Cabe aos linguistas e aos professores interpretarem essas hipóteses.

Muitas das hipóteses formuladas na escrita inicial refletem a tentativa dos aprendizes de escreverem conforme falam. Isso acontece porque, como explica Cagliari (1989, p. 128), as crianças, que ainda não dominam o sistema ortográfico da língua, pensam mais na forma como falam do que na maneira como devem escrever. Assim, as escritas espontâneas infantis expressam ativamente o que as crianças estão pensando sobre a língua falada. O sistema oral é inerentemente variável e, por isso, a escrita inicial pode expressar a variação encontrada na oralidade. Uma criança que está acostumada a falar e a ouvir formas como [bisi'çretə] e ['bruzə], por exemplo, possivelmente escreverá <\*bicicreta> e <\*brusa> nos seus primeiros textos. Em vez de corrigir punitivamente os erros ortográficos decorrentes da variedade de fala da criança, o professor pode partir da análise dos erros para introduzir a discussão sobre a natureza variável do sistema



linguístico oral. Afinal, esses erros ortográficos têm muito a contribuir com a caracterização dos fenômenos variáveis presentes na língua oral.

Pesquisas interessadas em descrever e explicar os erros ortográficos motivados pela fala tendem a agrupá-los em diferentes categorias (ALVARENGA, *et al.*, 1989; CALGIARI, 1989; NÓBREGA, 2013). A partir da categorização, os estudiosos buscam sistematizar os erros ortográficos em busca de entender quais fenômenos variáveis podem interferir na escrita para que, com isso, seja possível auxiliar o professor a compreender quais são os possíveis desafios ao longo da trajetória de ortografização. Cada trabalho, entretanto, descreve um conjunto de dados específico e utiliza um método próprio para a categorização, o que causa grande divergência na literatura. Alvarenga *et al.* (1989), por exemplo, criaram dezenove categorias distintas para explicar os erros motivados pela fala. Nóbrega (2013), por sua vez, classificou as marcas orais em quinze categorias diferentes. Já Cagliari (1989, p. 138) agrupou um quarto de toda a sua amostra em uma categoria denominada de “transcrição fonética”.

A interpretação implícita no método de categorização é a de que os erros ortográficos agrupados em uma mesma categoria são decorrentes de uma mesma motivação: a oralidade. Entretanto, embora sejam todos reflexos da fala, os erros ortográficos podem revelar aspectos distintos do conhecimento linguístico infantil. Nesta tese, argumentamos que erros ortográficos aparentemente semelhantes podem ser cometidos por diferentes fatores. Assim, embora possam ser agrupados em uma mesma categoria por motivos de sistematização, deve-se ter clareza das múltiplas motivações por trás de cada erro ortográfico. Três tipos de erros são interesse desta pesquisa: (i) a troca de <e, o> por <i, u> em contextos pretônico e postônico final; (ii) a omissão da letra <r> em final de verbos e de nomes; <iii> as hipercorreções. Cada um dos tipos será abordado individualmente em uma seção a seguir.

### 2.2.3.1 A troca de <e, o> por <i, u>

Linguistas dedicados à análise do processo de aprendizado da escrita já documentaram que, no início do percurso escolar, crianças tendem a trocar as letras <e, o> por <i, u> em posição pretônica e postônica final (ALVARENGA *et al.*, 1989; LEMOS, 2001; MIRANDA, 2010; CRISTÓFARO-SILVA; GUIMARÃES, 2013). Vejamos alguns exemplos:

**Quadro 6:** Exemplos de erros ortográficos em contextos pretônico e postônico final

	trocas de <e> por <i>	trocas de <o> por <u>
<b>posição pretônica</b>	<*minino> (menino)	<*buneca> (boneca)
<b>posição postônica</b>	<*penti> (pente)	<*patu> (pato)

Fonte: elaboração própria.

De acordo com Miranda (2010), erros ortográficos como os exemplificados no Quadro 6

decorrem da pronúncia de vogais altas nas duas posições acentuais. Ou seja, as crianças escolhem as letras <i> e <u>, porque falam [i] e [u] pretônicas e postônicas finais e, no início da escolarização, tendem a estabelecer uma relação direta entre a forma falada e a forma escrita (CAGLIARI, 1989). Alvarenga e colegas (1989) agruparam os erros ortográficos relacionados às vogais pretônicas e às vogais postônicas finais em uma mesma categoria, como pode ser visto na Figura 2.

**Figura 2:** Categoria dos erros relacionados às vogais

<b>X - ALÇAMENTO DE VOGAIS</b>	
1. e →	i (pré-tônico)
2. e →	i (pós-tônico)
3. o →	u (pré-tônico)
4. o →	u (pré-tônico)

Fonte: Alvarenga, *et.al.*, (1989, p. 12).

Para os autores, erros como <\*minino>, <\*penti>, <\*buneca> e <\*patu> são explicados pelo fenômeno fonológico de alçamento vocálico, que ocorre quando vogais médias passam a ser pronunciadas como vogais altas, por exemplo: m[e]nino > m[i]nino (VIEGAS, 1995). Essa categorização dos erros ortográficos permite a interpretação de que os aprendizes criam hipóteses semelhantes ao grafarem as vogais em diferentes contextos acentuais. Entretanto, há evidências na literatura que confrontam a ideia de que os erros ortográficos em contextos pretônico e postônico final decorrem na mesma motivação (LEMONS, 2001; MIRANDA, 2010; CRISTÓFARO-SILVA; GUIMARÃES, 2013). Revisaremos, de agora em diante, algumas dessas evidências e destacamos como esta tese pretende avançar em relação à literatura já publicada.

Em primeiro lugar, revisam-se os resultados de Lemos (2001). O autor analisou os reflexos do fenômeno de alçamento vocálico na escrita de crianças dos quatro primeiros anos do Ensino Fundamental (EFI), em busca de investigar se o mesmo erro ortográfico seria cometido em posições silábicas diferentes – pretônica, postônica medial e postônica final. Lemos (2001) demonstrou, por meio de uma descrição qualitativa, que as vogais pretônicas e postônicas mediais são escritas de formas variadas pelos aprendizes, como, por exemplo, <\*minino \*pérula, \*bússula, \*cósigas ~ \*cosgas, etc.>. Já em relação às vogais postônicas finais, o autor verificou regularidade na ortografia, na medida em que todas as palavras testadas foram escritas sem erros ortográficos. De forma geral, Lemos (2001) demonstrou que as crianças do EFI enfrentam desafios distintos ao escrever as vogais médias em diferentes posições acentuais. O presente trabalho dialoga com o de Lemos (2001) ao analisar a escrita das vogais médias pretônicas e postônicas finais. Por outro lado, esta tese se diferencia pelos seguintes motivos: (i) a análise dos dados de escrita será quantitativa; (ii) analisaremos dados de escrita de alunos dos anos iniciais e finais do EF; (iii) os dados serão interpretados à luz de modelos múltiplos e dinâmicos. Em síntese, este trabalho pretende avaliar,

quantitativamente, o percurso de aprendizado das vogais médias pretônicas e postônicas finais, em busca de evidências sobre a multiplicidade das representações mentais.

Outro trabalho que merece destaque é o artigo de Miranda (2010). A autora abordou a grafia de vogais pretônicas e postônicas finais, a partir da análise de resultados documentados em trabalhos anteriores (MIRANDA, 2006, 2008). Os resultados mostraram que a maior incidência de erros ortográficos se deu no contexto pretônico (MIRANDA, 2010). Para a autora, as crianças erram a ortografia das vogais pretônicas em altos índices por dois motivos: (i) há grande variação na fala no contexto pretônico; (ii) a ortografia das vogais pretônicas é arbitrária, no sentido de que não é possível prever, por via de regras, qual letra deve ocupar tal contexto. Em posição postônica final, Miranda (2008) demonstrou que a maior parte dos erros ortográficos se concentrou na grafia da letra <e>, em comparação à grafia de <o>. Para a autora, esses resultados podem estar relacionados ao fato do <o> átono final ser portador de generalizações morfológicas, como a marcação de gênero, o que não acontece com a vogal <e> átona final. De forma geral, o trabalho de Miranda (2010) traz indícios de que as vogais pretônicas são mais suscetíveis a erros ortográficos do que as vogais postônicas finais. Além disso, a autora sugere que, para além da fonologia, a morfologia pode auxiliar o aprendizado da ortografia.

Embora Miranda (2010) tenha avançado na investigação das vogais na escrita infantil, a autora não pretendeu analisá-las comparativamente. Miranda (2010) analisou os erros ortográficos em *corpora* diferentes. As grafias relacionadas às vogais pretônicas foram extraídas de cerca de mil textos espontâneos de alunos dos quatro primeiros anos escolares. Já as vogais postônicas finais foram analisadas em uma amostra de 400 textos espontâneos de crianças do 1º e 2º anos do EFI. No presente trabalho, a coleta dos dados será controlada, para que haja, em um único *corpus*, uma quantidade equivalente de dados referentes à escrita de vogais pretônicas e de postônicas finais. O controle metodológico permitirá comparar os índices de erros ortográficos em ambas as posições acentuais e em diferentes anos escolares. Além disso, esta tese analisará dados de escrita de alunos dos anos iniciais e finais do EF. Esperamos, com isso, avaliar o percurso de aprendizado da ortografia ao longo de todo o Ensino Fundamental, em busca de evidências sobre o conhecimento implícito das crianças a respeito das vogais pretônicas e postônicas finais.

Por fim, Cristófaros-Silva e Guimarães (2013) citam alguns resultados de Greco (2009) sobre o percurso de aprendizado das vogais pretônicas e postônicas finais a fim de discutirem as relações entre a fala e a escrita. Greco (2009) analisou as influências da fala na escrita no contexto pretônico e concluiu que os índices de erros ortográficos nesse contexto diminuíram, mas não foram completamente solucionados ao longo do tempo. Os resultados da autora demonstraram que dados como <\*minino> ‘*menino*’ e <\*curuja> ‘*coruja*’ perduram na escrita de alunos do 4º e do 5º anos. No que diz respeito às vogais postônicas finais, Cristófaros-Silva e Guimarães (2013)

citam, brevemente, alguns dados encontrados – mas não reportados – por Greco (2009). As autoras sugeriram que os erros ortográficos relacionados às vogais postônicas finais se resolvem no 2º ano do EFI. A partir desses resultados, Cristóforo-Silva e Guimarães (2013, p. 321) formulam a seguinte questão: “*por que, em posição átona final, a troca das letras “e, o” por “i, u” é resolvida nas séries iniciais, mas, em posição pretônica, tal troca persiste ao longo da vida escolar do indivíduo?*”. Como resposta, as autoras argumentam que há grande regularidade no contexto postônico final: marca morfológica de gênero, tonicidade, fim de palavra. Para as autoras, a regularidade permite “um alto grau de generalização por efeitos probabilísticos”, o que pode ter influenciado na aprendizagem da escrita (CRISTÓFARO-SILVA; GUIMARÃES, 2013, p. 321). Esta tese dialoga com o trabalho de Cristóforo-Silva e Guimarães (2013) por discutir o aprendizado das vogais pretônicas e postônicas finais à luz de modelos múltiplos e dinâmicos. Porém, este trabalho avança por analisar, experimentalmente, a escrita de alunos dos anos iniciais e finais do EF, além de aprofundar a discussão a respeito das várias relações entre oralidade e escrita.

Como se pode ver, a literatura revisada nesta seção mostra que os aprendizes de escrita tendem a construir diferentes estratégias para grafarem as vogais em diferentes contextos acentuais. No entanto, nenhum dos trabalhos se concentrou em discutir a relação entre os diferentes percursos de aprendizado da ortografia e o conhecimento linguístico internalizado dos aprendizes. É neste âmbito que esta pesquisa visa contribuir: pretende-se analisar a escrita das vogais médias pretônicas e postônicas finais, em busca de evidências sobre a organização do conhecimento linguístico dos alunos. Se as trajetórias de aprendizado da escrita das vogais pretônicas e postônicas finais se confirmarem diferentes nos dados desta pesquisa, haverá evidências para sustentar a ideia de que os aprendizes de ortografia precisam gerenciar esquemas linguísticos múltiplos para grafarem uma mesma letra – <e, o> – em diferentes posições acentuais.

### 2.2.3.2 A omissão do <r> em final de verbos e de nomes

Além de os erros ortográficos relacionados às vogais pretônicas e postônicas finais, os aprendizes de escrita também tendem a omitir a letra <r> que ocorre em final de verbos e de nomes (COSTA, 2009; TORRES; OLIVEIRA, 2015; FREIRE; HORA, 2019). Vejamos alguns exemplos do Quadro 7:

**Quadro 7:** Exemplos de erros ortográficos em final de verbos e de nomes

Omissão do <r> final em verbos	Omissão do <r> final em nomes
<*canta> (cantar)	<*elevado (elevador)
<*corre> (correr)	<*celula> (celular)
<*sorri> (sorrir)	<*flo> (flor)

**Fonte:** elaboração própria.

Pesquisas dedicadas a descrever os erros ortográficos tendem a dizer que o fenômeno fonológico denominado de ‘cancelamento do R-final’ é o responsável por motivar erros como os exemplificados no Quadro 7 (ALVARENGA *et al.*, 1989; NÓBREGA, 2013). Propõe-se que os aprendizes de escrita omitem a letra <r> no final das palavras, porque eles cancelam o rótico na oralidade. Fala-se [kã'ta] e, por isso, escreve-se <\*canta>. Uma vez que a omissão da letra <r> final ocorre tanto em verbos quanto em nomes, os manuais de ortografia tendem a agrupar palavras das duas classes gramaticais em uma mesma categoria de erros ortográficos (LEMLE, 1987; ALVARENGA *et al.*, 1989; CAGLIARI, 1989). Vejamos, por exemplo, a categoria proposta por Nóbrega (2013).

**Quadro 8:** Categoria dos erros ortográficos relacionados aos róticos

12. Omissão do R em final de palavras	encontrar → ‘encontra’ pescador → ‘pescado’
---------------------------------------	--

**Fonte:** Nóbrega, 2013, p. 32.

Como se pode ver, Nóbrega (2013) propõe que erros ortográficos como <\*encontra> e <\*pescado> são parte da mesma categoria. Logo, a autora assume que esses erros têm a mesma motivação: o cancelamento do R-final na oralidade. Se o mesmo fenômeno fonológico motivasse o mesmo tipo de erro em verbos e em nomes, os índices de erros ortográficos deveriam ser semelhantes em ambas as classes gramaticais. No entanto, há evidências de que os aprendizes de escrita cometem uma quantidade maior de erros ortográficos relacionados aos verbos do que aos nomes (COSTA, 2009; TORRES; OLIVEIRA, 2015; FREIRE; HORA, 2019), o que confronta a ideia de que esses erros são decorrentes de uma mesma motivação. Vejamos alguns desses indícios ao longo desta seção.

O primeiro trabalho revisado nesta seção é o de Costa (2009). A autora avaliou a escrita de nomes e de verbos por alunos do 2º ao 4º anos, de uma escola de Catu – BA. Os alunos do 2º ano omitiram o R-final em altos índices, tanto nos nomes (75%), quanto nos verbos (90%). No 3º e no 4º anos, houve redução do apagamento do <r> na escrita nas duas classes gramaticais, mas os alunos erraram mais a escrita de verbos (40%), do que de nomes (20%). Para a autora, a omissão da letra <r> é preferivelmente feita em verbos, porque nessa classe gramatical há altos índices de cancelamento do rótico na fala. Embora tenha encontrado diferenças na escrita dos verbos e dos nomes, Costa (2009) apenas descreveu os resultados, sem a intenção de discutir o percurso de aprendizado da ortografia, ou de refletir sobre o conhecimento implícito das crianças. A presente pesquisa pretende avançar ao comparar os índices de erros ortográficos de crianças de anos escolares diferentes, visando discutir sobre o conhecimento linguístico infantil implícito envolvido na escrita de palavras de diferentes classes gramaticais.

Outro trabalho que motivou a presente pesquisa é o de Torres e Oliveira (2015). Os autores analisaram a omissão do <r> em final de palavras escritas por alunos do 6º e 9º anos do Ensino Fundamental II (EFII), em uma escola de Feira de Santana – BA. A hipótese de Torres e Oliveira era a de que o <r> seria cancelado em maiores índices em verbos do que em não-verbos, tendo em vista que os verbos favorecem o cancelamento do R-final na fala. Os dados confirmaram a hipótese dos autores. Houve maiores índices de erros ortográficos em verbos de infinitivo (32%), do que em substantivos (14%). Para Torres e Oliveira (2015), os resultados contribuem com a descrição das dificuldades encontradas ao longo do aprendizado da escrita e, com isso, podem colaborar para a melhorias no ensino de língua portuguesa. Além disso, o trabalho de Torres e Oliveira (2015) contribui com a descrição das particularidades da escrita dos verbos e dos nomes terminados em R-final. O presente trabalho pretende avançar em relação ao de Torres e Oliveira (2015) ao comparar os índices de omissão da letra <r> final em diferentes classes gramaticais e em cinco anos escolares do EF, em busca de avaliar como se dá o percurso de aprendizado destas classes gramaticais. Ademais, pretende-se, nesta tese, discutir de que forma as diferenças entre verbos e nomes expressam a organização das representações linguísticas dos aprendizes de escrita.

Por fim, mais um trabalho que ressalta as diferenças entre os verbos e os nomes na escrita é o de Freire e Hora (2019). Os autores analisaram textos de alunos do 7º ao 9º ano do EF, de uma escola pública do interior do Rio Grande de Norte, e encontraram 12,4% de omissão da letra <r> no *corpus*. Além disso, os resultados dos autores indicaram que a omissão do <r> é maior em verbos (13,6%) do que em nomes (2,8%). Para Freire e Hora (2019), esses resultados são explicados pelo fato de que os verbos estão mais sujeitos ao cancelamento do R-final em verbos do que os nomes. O trabalho de Freire e Hora (2019) mostra que os alunos dos anos finais do EF criam diferentes hipóteses para grafarem itens verbais e nominais. No entanto, assim como os demais trabalhos, Freire e Hora (2019) cumpriram o objetivo de descrever os erros ortográficos cometidos no EF, mas sem a intenção de discuti-los à luz da teoria fonológica. Esta pesquisa avança em relação ao de Freire e Hora (2019) por avaliar os erros ortográficos infantojuvenis em busca de evidências sobre o conhecimento cognitivo das crianças sobre a língua. Se o percurso de aprendizado da escrita do <r> final em verbos e em nomes se confirmar diferente, teremos indícios de que os aprendizes acessam categorias linguísticas múltiplas ao longo do período escolar.

### 2.2.3.3 *As hipercorreções ortográficas*

Para que se compreenda o conceito de hipercorreção, é necessário abordar, primeiro, a valoração social por trás das estruturas linguísticas. Como já foi mencionado anteriormente neste capítulo, as línguas faladas são inerentemente variáveis. Isso significa que os indivíduos que compartilham uma mesma língua a falam de diferentes maneiras, a depender de fatores como a

idade, a região onde moram, a classe social, a intenção por trás da comunicação, etc. (MOLLICA; BRAGA, 2004). Linguistas, cada dia mais, reconhecem todas as manifestações das línguas como legítimas e necessárias. A sociedade, porém, ainda expõe as variantes linguísticas a julgamentos sociais, de modo que algumas são prestigiadas e outras estigmatizadas.

As variantes de prestígio são, na maioria das vezes, aquelas faladas pela parcela culta da sociedade, que possui alto nível de escolaridade e conhece a norma padrão da língua. Já as variantes estigmatizadas são aquelas faladas por pessoas sem instrução formal, residentes de zonas rurais ou de periferias das grandes cidades, etc. (BAGNO, 2007, p. 76). As variantes estigmatizadas são avaliadas, preconceituosamente, como sinal de ignorância; já as variantes de prestígio são socialmente relacionadas à sabedoria, à forma ‘correta’ de se falar. A avaliação positiva ou negativa das formas linguísticas tem, portanto, relação com a posição socioeconômica e sociocultural que os falantes ocupam.

A noção de hipercorreção tem íntima relação com a valorização social de determinadas formas linguísticas, tendo em vista que indivíduos tendem a hiper corrigir a fala na tentativa de alcançar a variedade de prestígio da língua. Por exemplo: fala-se ‘mer[e]tíssimo’ e ‘s[e]nosite’ por se achar que essas formas de falar são prestigiadas na língua. São várias as definições de hipercorreção encontradas na literatura (JOTA, 1976; DUBOIS *et al.*, 1995; LABOV, 2008; BAGNO, 2011; BORTONE; ALVES, 2014). Para Jota (1976, p. 166), por exemplo, a hipercorreção decorre da pretensa correção naquilo que se julga errado. Já para Dubois *et al.* (1995, p. 323), o fenômeno de hipercorreção “consiste na busca excessiva de correção, que acaba por levar a pronúncias e a usos incorretos por temor de incidir em erros populares”. Labov (2008) destaca que a hipercorreção é consequência da aplicação equivocada de uma regra por influência da valoração social. Bagno (2011, p. 949), por sua vez, relaciona hipercorreção à “insegurança linguística”. De acordo com o autor, a hipercorreção ocorre “quando um(a) falante ou uma comunidade de falantes, ao tentar se aproximar de um padrão ideal imaginário de língua ‘boa’, acaba ‘acertando demais’ e se desviando tanto da gramática intuitiva da língua quanto da gramática normativa” (BAGNO, 2011, p. 949). Como se pode ver, nas definições apresentadas, a noção de hipercorreção está atrelada à tentativa de alcance de um padrão socialmente prestigiado.

Além de ser descrita como um fenômeno oral, a hipercorreção também é documentada como um fenômeno presente na escrita (BAGNO, 2011; BORTONE; ALVES, 2014). A escola tradicional sempre orientou os alunos a escreverem conforme a norma padrão da língua portuguesa. Ensinos mais rígidos tendem, até mesmo, a punir os alunos que não seguem a norma. Para Bortone e Alves (2014), a busca excessiva pela norma padrão das palavras pode fazer com que os alunos produzam hipercorreções. Nesse sentido, a hipercorreção na escrita é entendida

como uma correção exagerada, que decorre da suposição enganosa de que se está escrevendo de maneira correta (BAGNO, 2011).

Vale lembrar que a fala e a escrita são sistemas distintos: o primeiro é inerentemente variável; o segundo é conservador. Como a fala varia, é natural que os indivíduos julguem as variantes conforme estas são utilizadas na sociedade. Como bem coloca Bagno (2007, p. 77), “onde tem variação sempre tem também avaliação”. Por outro lado, o sistema de escrita é estável e deve seguir uma norma ortográfica. Uma palavra pode ser falada de diferentes formas, mas deve ser grafada de uma única maneira: a maneira correta. Em outros termos, os indivíduos podem escolher uma dentre um *continuum* de variantes linguísticas na oralidade. Na escrita ortográfica, porém, há apenas duas maneiras de se escrever: a maneira correta e a maneira errada. Sabendo-se disso, podemos questionar a ideia de hipercorreção ortográfica: *como identificar uma correção excessiva em um sistema que permite apenas erro e acerto?* A expressão ‘correção excessiva’ só pode fazer sentido em um *continuum* que parte do menos correto até alcançar o excessivamente correto. Na ortografia, porém, não existem formas mais corretas do que outras. Há apenas duas possibilidades: ou se escreve corretamente, ou se escreve errado. Assim é, no mínimo, desafiador identificar correções excessivas, ou hipercorreções, em um sistema categórico como o ortográfico.

A escola cobra diariamente que seus alunos escrevam de acordo com a norma padrão. Assim sendo, pode-se dizer que aprendizes de escrita estão, quase sempre, na busca por alcançar a grafia correta das palavras. Então, os erros ortográficos – sejam ou não hipercorreções – não seriam, em sua maioria, tentativas de se escrever corretamente? Utilizaremos o termo hipercorreção no decorrer deste texto para que o leitor saiba claramente sobre a qual erro ortográfico estamos nos referindo. No entanto, tentaremos defender que os erros ortográficos tradicionalmente chamados de hipercorreções decorrem da tentativa de associar, na escrita, palavras que têm alguma semelhança fonética. Por exemplo: quando o aprendiz escreve <\*boraco> em vez de <buraco> ele está aplicando o mesmo padrão ortográfico de palavras como <boneca>, <bonita> e <botina>, por exemplo. Curiosamente, todas essas palavras podem ser faladas, em algumas variedades, com a vogal [u] pretônica. Assim, é totalmente explicável que o aprendiz tente aproximar, na escrita, palavras que são faladas de forma semelhante. Em outros termos, se o aprendiz fala [bu'raku], [bu'nekə], [bu'nitə] e [bu'tinə], é possível que ele crie a hipótese de que todas essas palavras são grafadas com a mesma letra em contexto pretônico. Sendo assim, não há correções ortográficas exageradas (ou hipercorreções); pelo contrário, há uma rica reflexão sobre a relação entre os padrões fonéticos e ortográficos da língua.

Ênfase será dada aos casos de hipercorreção ortográfica que acometem as vogais pretônicas e postônicas finais, bem como os róticos finais de verbos e de nomes. Vejamos alguns exemplos no Quadro 9.



**Quadro 9:** Exemplos de hipercorreções

Hipercorreção em vogais átonas	Hipercorreções em verbos e nomes
<*fotebol> (futebol)	<*aprender> (aprendi)
<*boraco> (buraco)	<*estar> (está)
<*taxe> (táxi)	<*sofar> (sofá)
<*safare> (safári)	<robor> (robô)

**Fonte:** elaboração própria.

Do lado esquerdo do Quadro 9, encontram-se alguns exemplos de hipercorreções ortográficas documentadas em contextos pretônico e postônico final. Essas hipercorreções ocorrem porque os aprendizes escrevem as letras <e, o> em palavras que deveriam ser escritas com <i, u> pretônico e postônico final. Bohn e Souza (2017) mostraram que as hipercorreções são mais frequentes em contexto pretônico do que em contexto postônico final. Para as autoras, esse resultado pode ser interpretado como evidência de que as crianças são sensíveis à maior instabilidade das vogais pretônicas em relação às postônicas. Belga (2019) verificou que as hipercorreções em contexto pretônico estão presentes até mesmo na escrita acadêmica. Considerando-se que hipercorreções podem contribuir com a investigação sobre o percurso de aprendizado das vogais pretônicas e postônicas finais, faz-se necessário analisá-las nesta pesquisa.

Do lado direito do Quadro 9, encontram-se exemplos de hipercorreções em final de verbos e de nomes. Nesse contexto, as hipercorreções decorrem da tentativa de inserção da letra <r> em final de algumas palavras terminadas por vogal. Cesar (2017) analisou casos de hipercorreção envolvendo a letra <r> final na escrita de alunos do 9º ano. A autora demonstrou que os alunos inseriram o <r> de forma indevida no final de palavras de diferentes classes gramaticais. Por exemplo: <\*vocêr> – ‘você’, <\*estar> – está, <\*meninor> – ‘menino’. Castelo Branco (2020) analisou erros ortográficos relacionados ao rótico final na escrita de alunos do 8º ano e encontrou vários casos de hipercorreção. Por exemplo: <\*aprender> – aprendi, <\*resolver> – resolvi, <\*seguir> – segui, <\*estar> – está. Curiosamente, Castelo Branco (2020) encontrou hipercorreções apenas em verbos, sobretudo, em verbos de 3º conjugação. Considerando-se que hipercorreções podem contribuir com a investigação sobre o percurso de aprendizado do <r> final em nomes e em verbos, faz-se necessário analisá-las nesta pesquisa.

Segundo Tenani e Reis (2011), as hipercorreções decorrem da tentativa dos aprendizes em generalizar os padrões ortográficos. Por exemplo: ao entender que a letra <e> representa a vogal [ɛ] postônica final, o aprendiz pode grafar <\*taxe> em vez de <táxi> em busca de generalizar o padrão ortográfico. Vale ressaltar, porém, que a explicação de Tenani e Reis (2011) se aplica a vários tipos de erros ortográficos e não apenas às hipercorreções. Quando um aluno escreve <\*minino> e <\*avestrus>, por exemplo, ele também pode estar tentando encontrar generalizações ortográficas. Na escrita padrão do português, a letra <i> pode ocupar o contexto pretônico em palavras como <miséria> e <ministro>, por exemplo. Então, o aprendiz pode escrever <\*minino>

em busca de generalizar a letra <i> para outras palavras do PB. Além disso, o aluno pode escrever <\*avestrus> em busca de acomodar tal palavra no mesmo padrão ortográfico de <bambus> e <jesus>, por exemplo. De maneira geral, queremos defender que as hipercorreções ortográficas são erros ortográficos como outros quaisquer. Quase todos os erros ortográficos refletem a tentativa dos aprendizes em acomodar a palavra na ortografia da língua. Assim, não há, na escrita ortográfica, correções excessivas; pelo contrário, há a constante tentativa de adequar a forma gráfica das palavras à norma ortográfica da língua. Nesse processo de aprendizado da ortografia, os aprendizes acessam as representações linguísticas, se conscientizam sobre os elementos da língua e, então, desenvolvem o conhecimento linguístico. Os erros ortográficos – sejam ou não hipercorreções – são, portanto, ricas evidências da reflexão dos aprendizes sobre as línguas oral e escrita.

## **2.4 O aprendizado da escrita**

O processo de aprendizado da escrita se dá de maneira distinta do processo de aquisição da fala (KATO, 1990). Para falar, a criança precisa não só aprender inúmeras palavras e expressões da sua língua, como também todos os padrões gramaticais compartilhados pelas pessoas ao seu redor (TOMASELLO, 2005). Este processo ocorre naturalmente, “bastando que a criança esteja imersa em ambiente em que ouve e fala a língua materna” (SOARES, 2018, p. 45). Já para escrever, o indivíduo precisa ser instruído a desenvolver consciência das unidades linguísticas para que consiga compreender o que cada símbolo gráfico representa. Além disso, o escritor precisa se atentar às peculiaridades estruturais que diferenciam os textos escrito e oral (SOARES, 2018, p. 124).

Em primeiro lugar, o aprendiz de escrita precisa compreender as propriedades básicas do sistema gráfico (TREIMAN; KESSLER, 2014). É preciso, por exemplo, entender que o papel ou a tela, o teclado ou a tinta são alguns dos objetos necessários no ato de escrever. Ademais, o aprendiz tem que se atentar para o fato de que a escrita se configura em linhas retas e horizontais, que se formam a partir de sequências de unidades, como letras ou outros tipos de símbolos (TREIMAN; KESSLER, 2014). Outro ponto a se observar é a finitude dos símbolos gráficos. Os aprendizes precisam notar que há uma quantidade finita de unidades que se repetem – com mais ou menos frequência – ao longo de um texto escrito (TREIMAN; KESSLER, 2014). Todas essas propriedades da escrita podem ser aprendidas apenas pela observação da aparência do sistema. Ou seja, o aprendiz pode observar a forma externa da escrita sem que se compreenda o que cada símbolo representa (TREIMAN; KESSLER, 2014). Nesse âmbito, o aprendiz ainda será incapaz de compreender a mensagem escrita no texto; ele só conseguirá identificar que determinado artefato se trata de um objeto escrito.

Para se tornar de fato um escritor eficiente, o indivíduo precisa saber quais são as funções internas dos símbolos gráficos. É preciso “tomar a língua como objeto de reflexão e análise” (SOARES, 2018, p. 125), se atentando para as relações entre as representações oral e gráfica. No caso da escrita alfabética, os aprendizes precisam entender que as letras representam sons. Nesse processo, cabe ao professor instruir os alunos de modo a levá-los a refletir sobre a língua e, conseqüentemente, a se conscientizar sobre os padrões linguísticos.

A reflexão sobre a língua materna é uma das habilidades cruciais para o desenvolvimento da escrita. Como afirma Osion (1999, p. 32), “aprender a escrever é aprender a ouvir e a pensar sobre a sua própria língua de uma nova maneira”. O aprendiz de escrita precisa ser capaz de acessar as representações sobre sua língua materna (MIRANDA, 2006) e, mais do que isso, precisa fazer com que o conhecimento tácito sobre a linguagem passe a ser explícito (GOMBERT, 2003; SEYMOUR, 2005). De forma geral, o sucesso no aprendizado da escrita depende do acesso à gramática e do desenvolvimento da consciência metalinguística, que se caracteriza como “a habilidade de refletir sobre e manipular os aspectos estruturais da língua” (NAGY; ANDERSON, 1999, p. 155)

No uso cotidiano da língua falada, não há a necessidade de monitorar conscientemente as propriedades linguísticas. Quando uma criança aprende uma nova palavra de sua língua materna, ela é capaz de utilizá-la em diferentes contextos sem precisar refletir sobre os aspectos fonológicos, sintáticos, semânticos, etc., por trás do item lexical. A aquisição da linguagem oral envolve, portanto, mecanismos inconscientes da produção e compreensão da língua (KARMILOFF-SMITH *et al.*, 1996). Já o aprendizado da escrita exige do aprendiz a manipulação intencional das estruturas linguísticas (GOMBERT, 1992). Para Soares (2018, p. 128), “a consciência metalinguística é, ao mesmo tempo, condição e resultado do aprendizado da leitura e da escrita”.

Pesquisadores dedicados ao estudo da consciência metalinguística, como Tunmer e Harriman (1984) e Gombert (1992), subdividem-na em diferentes tipos: consciência fonológica, consciência morfológica, consciência sintática, consciência lexical, etc. Essa subdivisão é pautada nos diferentes níveis de análise linguística propostos por modelos tradicionais da linguagem. Como as categorias linguísticas avaliadas nesta pesquisa se diferenciam, sobretudo, no âmbito da fonologia e da morfologia, vejamos, em mais detalhes, a que se refere às consciências fonológica e morfológica.

A consciência fonológica envolve a capacidade de segmentar/contar/selecionar palavras, sílabas e os sons (SOARES, 2018). Essa habilidade é fundamental para que se compreenda o princípio alfabético; ou seja, o aprendiz precisa desenvolver consciência da sonoridade da língua para entender as relações entre letras e sons (SOARES, 2018, p. 167). Segundo Stanovitch (2000, p. 63), a consciência fonológica é desenvolvida em um *continuum*, que envolve desde a reflexão

sobre unidades maiores, como palavras, até a reflexão sobre unidades menores, como os sons. Morais *et al.*, (1979) testaram a consciência fonológica de indivíduos alfabetizados e analfabetos. Somente os indivíduos alfabetizados conseguiram realizar o teste. Morais *et al.*, (1979) concluíram que a capacidade de manipular os sons da fala é resultado da alfabetização. Read *et.al* (1986) verificaram que apenas adultos chineses conhecedores do alfabeto foram capazes de excluir e adicionar consoantes em palavras faladas. Para os autores, esses resultados mostram que o aprendizado da escrita alfabética ajuda os indivíduos a desenvolver a habilidade de segmentação dos sons da fala. Diferentemente, Gomes de Morais (2004, 2015) assume que há habilidades relacionadas à fonologia que precedem a alfabetização. Em Gomes de Morais (2015), há indícios de que as crianças precisam desenvolver consciência fonológica para avançar nas etapas de desenvolvimento da escrita. Para o autor, as habilidades relacionadas à fonologia “tendem a ser dominadas em diferentes momentos, antes, durante ou após a alfabetização inicial, numa relação interativa com a instrução escolar” (GOMES DE MORAIS, 2004, p. 176).

Já a consciência morfológica se caracteriza como a habilidade de refletir sobre os raízes/radicais e os afixos (prefixos e sufixos) das palavras (SOARES, 2018). Algumas pesquisas indicam o desenvolvimento precoce de habilidades morfológicas (TREIMAN; CASSAR; ZUKOWSKI, 1994), mas há também evidências de que as crianças se conscientizam da morfologia da língua tardiamente (PAULA, 2007). Treiman, Cassar e Zukowski (1994), por exemplo, mostraram que crianças jovens, de 5 a 9 anos, acertaram mais a ortografia quando as letras pertenciam a raízes morfológicas do que quando não pertenciam. Paula (2007), por outro lado, mostrou que a consciência sobre a morfologia é mais aguçada em alunos de séries escolares mais avançadas. Carlise (1995) solicitou a crianças do ensino infantil que completassem lacunas com palavras flexionadas e derivadas. Os resultados indicaram que as tarefas sobre as flexões morfológicas foram mais fáceis, para os participantes, do que as tarefas sobre derivações. Deacon e Bryant (2005) demonstraram que crianças inglesas, de anos escolares iniciais, também têm mais facilidade na ortografia das flexões morfológicas do que das derivações. Para Casalis e Louis-Alexandre (2000), as crianças adquirem aspectos da morfologia flexional antes de começarem a alfabetização, enquanto que a morfologia derivacional se desenvolve até o final do Ensino Fundamental.

Como se pode ver, a análise separada de cada dimensão da consciência metalinguística possibilita a apreensão de características específicas do processo de aprendizado da escrita e permite a descrição detalhada das etapas de desenvolvimento do conhecimento gráfico. No entanto, é válido pontuar que os elementos linguísticos interagem entre si de forma complexa ao longo de todo o processo de desenvolvimento da linguagem. Assim, embora seja metodologicamente interessante que se segmente a língua em diferentes níveis, deve-se lembrar

que todos os elementos – sons, sílabas, raízes, afixos, palavras, frases – são simultaneamente acessados ao longo do desenvolvimento da língua: seja oral, seja escrita.

A reflexão sobre a interação complexa entre as partes do sistema linguístico ajuda a situar a proposta desta tese. Parte-se do pressuposto que, durante o processo de aprendizado da escrita, os indivíduos acessam categorias linguísticas múltiplas, que armazenam informações lexicais, morfológicas, fonético-fonológicas, semânticas, sociais, etc.. Assim, argumenta-se que o aprendizado de uma mesma letra em diferentes contextos pode precisar do acesso a diversificadas categorias. A letra <r> em final de verbos, por exemplo, pode ser aprendida com o auxílio da consciência da morfologia verbal do PB. Por outro lado, o <r> em final de nomes pode ser aprendido por outros meios não relacionados à flexão de infinitivo. Do mesmo modo, a letra <o> em contexto postônico final pode ser aprendida com acesso à morfologia de gênero do português, o que não ocorre com a letra <o> em contexto pretônico. Em síntese, busca-se defender que o aprendiz precisa gerenciar e tomar consciência sobre um detalhado e complexo conhecimento linguístico ao longo do aprendizado do sistema gráfico.

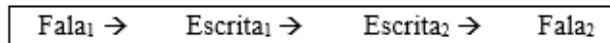
Para além do gerenciamento da linguagem, o aprendizado da escrita alfabética parece modificar as formas de uso da língua falada. Brown (1981) distingue dois tipos de fala: a que antecede o letramento<sup>7</sup> e a que o sucede; ou seja, a fala pré-letrada e a fala pós-letrada. Para o autor, a língua falada por indivíduos altamente alfabetizados em sociedades letradas é fortemente influenciada pela língua escrita. Por outro lado, a fala de indivíduos pré-alfabetizados é muito pouco afetada pelo sistema gráfico (BROWN, 1981, p. 168). Apresentamos, de forma resumida, duas das diferenças entre as falas pré e pós-letramento listadas por Brown (1981):

- A fala pós-letramento é caracterizada por uma sintaxe rica e complexa, em que há presença regular de subordinações, além do uso de pronomes relativos e conectivos. Já a fala pré letramento é muito mais simples no que diz respeito à ordem sintática (BROWN, 1981, p. 169).
- A fala altamente letrada contém um léxico mais rico e mais específico do que a fala pré-letrada (BROWN, 1981, p. 169).

Em consonância com Brown (1981), Kato (1990) assume que após o aprendizado da escrita, a fala é alterada. Para a autora, a escrita inicial representa a fala, mas, posteriormente, é a fala que tenta simular a escrita. Kato (1990) propõe, então, o seguinte esquema de representação das relações entre oralidade e escrita:

---

<sup>7</sup> Letramento: “desenvolvimento de habilidades de uso desse sistema em atividades de leitura e escrita, nas práticas sociais que envolvem a língua escrita” (SOARES, 2003, p.14).

**Quadro 10:** Relações entre a fala e a escrita

Fonte: Kato, 1990, p. 11

De acordo com Kato (1990), a Fala<sub>1</sub> é a que antecede o letramento. A Escrita<sub>1</sub>, por sua vez, é resultado da tentativa dos aprendizes de representarem visualmente a fala. A Escrita<sub>2</sub> é aquela que já se desvinculou da fala. Por fim, a Fala<sub>2</sub> é a fala pós-letramento. De maneira geral, Kato (1990) defende que a fala e a escrita são sistemas particulares que se relacionam ao longo do processo de alfabetização.

A análise da escrita inicial pode trazer ricas evidências sobre a reflexão dos aprendizes quanto à língua materna. Como bem define Miranda (2008, p. 152), “a escrita inicial das crianças é uma fonte extraordinária de dados, que, ao serem analisados e interpretados, podem contribuir para uma melhor compreensão a respeito dos conhecimentos que elas já possuem sobre a língua materna”. Além disso, a análise da fala de aprendizes de escrita pode nos trazer indícios sobre o papel do letramento na organização do conhecimento linguístico oral. É sobre este aspecto que se trata a próxima seção. Revisamos, a seguir, alguns trabalhos que mostraram que o aprendizado da escrita interfere na maneira de falar dos indivíduos.

## 2.5 As influências da escrita na fala

Até este ponto do capítulo, foi visto que a fala pode influenciar na escolha das formas ortográficas. Pontuamos, especificamente, que os aprendizes podem escrever algo como <\*minino>, <\*patu>, <\*canta> e <\*celula>, porque falam [mi'ninũ], ['patũ], ['kãtã] e [selu'la]. Há de se ressaltar, contudo, que não só a fala pode influenciar na escrita, como também a escrita pode interferir na forma como falamos as palavras. Embora sejam ainda pouco investigadas, as interferências da escrita na fala já foram documentadas, sobretudo, em trabalhos dedicados à análise do processo de aquisição/aprendizagem da linguagem oral e escrita (CHEVROT *et al.*, 2000; CARRERA-SABATÉ, 2006; SCHWINDT *et al.*, 2007; CRISTÓFARO-SILVA; GRECO, 2010; ADAMOLI, 2012). Revisamos, nesta seção, alguns desses trabalhos.

Em uma pesquisa sociolinguística sobre a aquisição de variáveis sonoras, Chevrot *et al.*, (2000) formularam a seguinte hipótese: *o contato com palavras escritas é um fator que influencia na aquisição de unidades fonológicas variáveis*. Para testar a hipótese, os autores analisaram a pronúncia do R pós-consonantal do francês – como em *sucre* ‘açúcar’ – na fala de crianças em fase de alfabetização. O R pós-consonantal é altamente cancelado na fala de adultos franceses. Assim, uma palavra como *sucre* é, recorrentemente, pronunciada como [syk] – suc(ʁe) por adultos. Os autores testaram se, após terem o contato com o <r> ortográfico, as crianças francesas aumentariam os índices de produção da variante [r], ou se o cancelamento prevaleceria na fala

infantil. Os resultados indicaram que o aprendizado da ortografia motivou o aumento da pronúncia do rótico na fala das crianças. Para os autores, esses resultados indicam que “a representação ortográfica de uma palavra permite que as crianças complementem a representação fonológica com um fonema que é frequentemente excluído no ambiente linguístico” (CHEVROT *et al.*, 2000, p. 321, tradução nossa<sup>8</sup>). Na visão de Chevrot *et al.*, (2000), a forma escrita atualizaria a representação fonológica dos indivíduos.

Adamoli (2012) também propôs que a forma escrita pode atualizar a gramática, após verificar a influência da escrita na pronúncia de ditongos variáveis. O autor analisou a variação dos ditongos [aɪ] e [eɪ] na fala infantil. Exemplo: ['faɪʃə]~ ['faʃə] 'faixa' e ['keɪʒʊ] ~ ['keʒʊ] 'queijo'. As crianças concluíram o 1º ano escolar falando e escrevendo formas monotongadas, como \*quejo e \*faja, em maiores índices. Ao final do 2º ano, as crianças aumentaram os índices de produção oral e escrita dos ditongos – como em *queijo* e *faixa*. O autor concluiu que a estrutura do ditongo (vogal + glide) é adquirida pelas crianças por meio do aprendizado da forma ortográfica.

Chevrot *et al.* (2000) e Adamoli (2012) explicam as influências da escrita na fala com o suporte de modelos formais da fonologia. Diferentemente, este trabalho pretende analisar dados de fala de aprendizes de escrita com o apoio dos Modelos de Exemplares (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e da Teoria dos Sistemas Dinâmicos e Complexos (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Ancorando-se na ideia de que o sistema linguístico é dinâmico e aberto às influências de outros sistemas (ELLIS; LARSEN-FREEMAN, 2009) – como o ortográfico, por exemplo – propõem-se que a escrita pode motivar a auto-organização das representações linguísticas. Esta proposta dialoga com a apresentada por Cristófar-Silva e Greco (2010). As autoras reportaram resultados da pesquisa de Greco (2009) sobre as influências da forma ortográfica na pronúncia de vogais pretônicas. Os resultados reportados mostraram que as crianças do 3º (62%) e 5º (52%) anos do EFI pronunciaram a vogal [e] – [meninʊ] – em maiores índices do que as crianças do 1º ano (30%); isto é, os alunos dos anos escolares mais avançados tenderam a pronunciar as vogais da forma mais próxima à representada pela escrita. A partir desses resultados, Cristófar-Silva e Greco (2010, p. 91) sugeriram que as representações fonológicas podem ser reorganizadas ao longo do tempo, devido às pressões sociais, como a escolarização. O presente trabalho também reflete sobre a natureza da representação mental a partir da análise das influências da escrita na fala, mas avança ao analisá-la em quatro categorias linguísticas distintas:

---

<sup>8</sup> [...] the orthographic representation of a word allows young speakers to complement the phonological representation with a phoneme that is often deleted in the linguistic environment (CHEVROT *et al.*, 2000, p.321)

vogais pretônicas, vogais postônicas finais, R-final em verbos, R-final em nomes. Especificamente, pretendemos verificar se, com o passar dos anos escolares, os alunos passam a produzir maiores índices de vogais médias e de róticos finais na fala. Retomemos as hipóteses (2a) e (2b):

*(2a) o aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônica ou postônica final.*

*(2b) o aumento do rótico final na fala ao longo dos anos escolares depende da classe gramatical: verbo ou nome.*

Essas hipóteses são fundamentadas nas evidências trazidas por Chevrot *et al.*, (2000), Adamoli (2012) e Cristófar-Silva e Greco (2010). Como já foi demonstrado, esses trabalhos mostraram que a escrita pode exercer influência sobre a fala. Então, esperamos que, nesta pesquisa, também haja efeitos da forma gráfica na produção de variantes orais. Contudo, este trabalho avança em relação à bibliografia precedente ao verificar se a categoria linguística interfere nas influências da escrita na fala. Como pode ser visto nas hipóteses (2a) e (2b), esperamos que o contexto acentual (pretônico ou postônico final) e a classe gramatical (verbos ou nomes) influenciem no aumento das variantes próximas à escrita. Essa hipótese se ancora na literatura sociolinguística que já reportou que a escolaridade afeta a pronúncia dos indivíduos, mas não em todos os contextos (HORA, 2006; CALDAS; CALLOU, 2014).

Hora (2006), por exemplo, analisou as influências do nível de escolaridade na realização da lateral pós-vocálica. A lateral pós-vocálica ocorre em palavras como ‘sal’ e pode ser realizada de forma velarizada [sal], vocalizada [saw] e, até mesmo, apagada [saø] (HORA, 2006; SCHWINDT *et al.*, 2007). Hora (2006) mostrou que falantes com níveis maiores de escolaridade pronunciam a lateral vocalizada – [w] – em maiores índices do que falantes analfabetos. A variante vocalizada é mais distante da representação ortográfica do que as variantes velarizada [ɣ] e alveolar [l] (SCHWINDT *et al.*, 2007). Ou seja, os resultados de Hora (2006) mostram que falantes mais escolarizados preferem pronunciar a variante não representada pela escrita. Isso significa que mesmo os falantes familiarizados com materiais escritos preferem pronunciar a variante [w], que é distante da forma ortográfica. No caso da lateral pós-vocálica, portanto, a forma escrita das palavras parece não influenciar a forma falada.

Caldas e Callou (2014) analisaram o efeito da escolaridade (falantes cultos x falantes não-cultos) no fenômeno de cancelamento do R-final em verbos em três capitais brasileiras: Salvador - BA, Recife - PB e Rio de Janeiro - RJ. Os resultados indicaram que o fenômeno ocorre de forma quase categórica em verbos, independentemente do nível de escolaridade dos falantes e da capital em que vivem. Considerando-se que a variante zero [ø] é mais distante da forma escrita do que as



demais variantes róticas, o trabalho de Caldas e Callou (2014) mostra que os falantes cultos preferem cancelar o R-final dos verbos, mesmo que isso afaste a forma oral da forma ortográfica das palavras.

As evidências levantadas pela literatura sociolinguística mostram que o efeito da escolaridade parece depender do fenômeno fonológico. Considerando-se que o aumento da escolaridade está diretamente relacionado ao maior contato formal com materiais escritos, podemos dizer que a escrita parece nem sempre interferir na forma de falar dos indivíduos. É esperado, portanto, que a escrita influencie a fala em diferentes índices, dependendo da categoria linguística: vogal pretônica, vogal postônica final, verbos com R-final, nomes com R-final.

## **2.6 Resumo do Capítulo 2**

Este capítulo revisou trabalhos sobre oralidade e escrita. A primeira seção descreveu as vogais médias e os róticos na oralidade. Buscou-se, sobretudo, destacar as particularidades fonéticas, lexicais e morfológicas de cada categoria avaliada nesta tese. Na segunda seção, ressaltamos que as relações entre letras e sons se encontram na essência da escrita alfabética. Além disso, mostramos que, embora se relacionem, os sistemas oral e escrito são diferentes, tendo em vista que o primeiro é variável e o segundo não. A terceira seção frisou a importância da ortografia para a manutenção dos sistemas de escrita e, além disso, descreveu alguns aspectos da norma ortográfica do PB. Pontuamos que, embora os padrões ortográficos possam ser agrupados em categorias, deve-se reconhecer as especificidades de aprendizado de cada um deles. A quarta seção abordou a noção de erro ortográfico como parte necessária de qualquer percurso de aprendizado. Ênfase foi dada em três tipos de erros ortográficos: (i) troca de <e, o> por <i, u>; (ii) omissão do <r> final; (iii) hipercorreções. Buscou-se demonstrar que, embora a literatura agrupe esses erros em categorias semelhantes, eles ocorrem em diferentes índices da escrita infantil. Além disso, problematizamos o conceito de hipercorreção ortográfica, argumentando que, em um sistema categórico como o ortográfico, não há como haver correções excessivas. A quinta seção discutiu o aprendizado do sistema alfabético-ortográfico, enfatizando três habilidades envolvidas nesse processo: o acesso às representações linguísticas; a consciência dessas representações e o desenvolvimento da linguagem. Por fim, a sexta seção revisou trabalhos que demonstraram que a escrita também pode influenciar na fala de forma específica a depender do tipo de fenômeno fonológico.

### 3. QUADRO TEÓRICO

A análise desenvolvida nesta tese se enquadra nos pressupostos teóricos dos Modelos de Exemplares (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e da Teoria dos Sistemas Dinâmicos e Complexos – TSDC (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Para que se compreenda a mudança de paradigma proposta por esses modelos, faz-se necessário discutir, inicialmente, algumas premissas das teorias tradicionais da linguística moderna. Assim sendo, este capítulo apresenta, na primeira seção, uma visão geral de duas teorias tradicionais: o Estruturalismo e o Gerativismo. Na segunda seção, as ideias de multiplicidade, emergência e dinamicidade da linguagem são discutidas com o apoio dos Modelos de Exemplares. Por fim, é discutida a natureza aberta, dinâmica, não linear e complexa do sistema linguístico, por meio dos pressupostos da TSDC. O objetivo do capítulo é destacar o potencial dos modelos teóricos múltiplos e dinâmicos para incorporar todos os subcampos da linguística – fonética, fonologia, léxico, morfologia e, até mesmo, a escrita – na teoria da linguagem.

#### 3.1 Representações fonológicas discretas: uma visão geral

Várias abordagens teóricas visam explicar a natureza da habilidade cognitiva que permite aos seres humanos converter uma sequência de ruídos em palavras compreensíveis. A discussão dessas muitas teorias, contudo, ultrapassaria o escopo deste trabalho. O intuito desta seção é apenas apresentar uma síntese das concepções teóricas tradicionais da linguística moderna para que, posteriormente, seja possível contrastá-las com as concepções inovadoras dos Modelos de Exemplares (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e da TDSC (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Assim sendo, esta seção apresenta, brevemente, algumas premissas de duas importantes correntes teóricas da linguística moderna: o Estruturalismo e o Gerativismo.

Na concepção estruturalista, a língua é um sistema fechado, regido por leis internas e compartilhado por indivíduos de uma mesma comunidade. A definição de língua, no estruturalismo, pressupõe a eliminação de tudo que seja designado como parte da linguística externa (SAUSSURE, 2012 [1916], p. 29). Para explicar as diferenças entre linguística interna e externa, Ferdinand Saussure (2012 [1916]), precursor do estruturalismo, compara o sistema linguístico ao jogo de xadrez:

Nesse jogo, é relativamente fácil distinguir o externo do interno; o fato de ele ter passado da Pérsia para a Europa é de ordem externa; interno, ao contrário, é tudo quanto concerne ao sistema e às regras. Se eu substituir as peças de madeira por peças de marfim, a troca será indiferente para o sistema; mas se eu reduzir ou aumentar o número de peças, essa mudança atingirá profundamente a “gramática” do jogo [...] é interno ao sistema tudo o que provoca mudança em qualquer grau. (SAUSSURE, 2012 [1916], p. 29)

A escrita, na perspectiva saussuriana, é um dos elementos externos à linguagem, que apenas cumpre a função de registrar a língua e nada contribui para o seu funcionamento (SAUSSURE, 2012 [1916], p. 33; BLOMFIELD, 1996 [1933], p. 21). Enquanto a língua oral é caracterizada como parte do conhecimento puramente linguístico, a língua escrita é indiferente ao sistema. Para Bloomfield (1996[1933], p. 33, tradução nossa), a “escrita não é língua, mas apenas uma maneira de registrar a língua por meio de marcas visíveis”<sup>9</sup>. É necessário destacar, porém, que trabalhos como os de Chevrot *et al.*, (2000), Adamoli (2012), Cristófar-Silva e Greco (2010) – já descritos no capítulo 2 – evidenciaram que a escrita pode provocar mudanças na fala; ou seja, há evidências de que a escrita contribui para o funcionamento do sistema linguístico. Assim sendo, embora tenha o mérito de sistematizar as línguas naturais, o estruturalismo deixa lacunas referentes às relações entre os sistemas oral e escrito, uma vez que secundariza o estudo da escrita ao priorizar a fala (VELOSO, 2003, p.133).

Dentre as noções fundamentais do estruturalismo, encontra-se o conceito de fonema. Como já foi mencionado no Capítulo 2, os fonemas podem ser definidos como unidades abstratas, independentes, autônomas e não pronunciáveis, que representam um conjunto de alofones (CRYSTAL, 1988). A noção de fonema é adotada não só em pesquisas de teoria fonológica, mas também em estudos sobre aprendizagem da escrita. Assume-se, tradicionalmente, que aprendizes de escrita precisam se conscientizar das unidades mínimas do sistema fonológico – os fonemas – para compreenderem o princípio alfabético (SOARES, 2018, p. 193). É necessário destacar, todavia, que há evidências de que, para além dos fonemas, os aprendizes de escrita utilizam de informações fonéticas (MIRANDA, 2010; 2020), morfológicas (MOTA, 2012), lexicais (SANTOS, 2000) e contextuais (TREIMAN, 2017) ao longo da escolarização. Assim sendo, pode-se dizer que os aprendizes de escrita se beneficiam de um sistema linguístico muito mais complexo e detalhado – que armazena mais do que fonemas – do que o postulado pela abordagem estruturalista da linguagem.

O caráter discreto da gramática pode ser notado também nos pressupostos da abordagem gerativista da linguagem. A Teoria Gerativa assume que os seres humanos são dotados de um dispositivo inato, chamado Gramática Universal (GU), que representa o conhecimento linguístico

---

<sup>9</sup> No original: “Writting is not language, but merely a way of recording language by means of visible marks.”

internalizado. Nas palavras de Noam Chomsky (1994, p. 43), precursor do Gerativismo, “a GU é uma caracterização de princípios inatos e biologicamente determinados, que constituem um componente da mente humana, a faculdade da linguagem.”

Segundo Boeckx (2010), o grande avanço da proposta chomskyana foi considerar o caráter mental da linguagem. A ênfase dada aos aspectos linguísticos mentais, contudo, motivou a rígida separação entre os domínios cognitivo e sociocultural da linguagem, priorizando-se o estudo do primeiro e dispensando-se o segundo. Segundo Veloso (2003, p. 138), o conhecimento linguístico, na vertente gerativa, é “radicalmente impermeável à influência de fatores de ordem social e cultural”. Então, a escrita não pode interferir na organização da GU, uma vez que os materiais gráficos são fruto de uma convenção social.

No gerativismo, denomina-se de língua-E (linguagem externa) a linguagem relacionada a aspectos sociais e culturais e de língua-I (linguagem interna) a linguagem que compõe o domínio cognitivo (LIGHGFOOT, 2010). A língua-E é composta pelas infinitas estruturas que são expressas na interação comunicativa, enquanto a língua-I armazena um conjunto finito de regras, capazes de gerar as infinitas estruturas observadas no uso (LIGHGFOOT, 2010). A Gramática Gerativa se interessa, primordialmente, pelo estudo da língua-I, uma vez que se dedica à investigação dos elementos finitos que compõem o módulo cognitivo da linguagem. Para Boeckx (2010), a língua-I representaria a linguagem comum à espécie humana:

[...] muitos de nós pensamos que existe apenas uma linguagem humana no planeta. Isso não significa negar a existência de muitas variantes linguísticas, mas esses não são os principais objetos de investigação do linguista cognitivo [...] Como qualquer outro cientista, o linguista terá que aprender a ir além da multiplicidade de idiomas para alcançar os princípios fundamentais que tornam possíveis a experiência linguística (BOECKX, 2010, p.5)

Tomando por referência as palavras de Boeckx (2010), pode-se dizer que, para a abordagem gerativa da linguagem, todos os seres humanos compartilham uma mesma gramática, dotada de princípios inatos e universais. Para explicar a natureza do conhecimento gerativo, Langacker (2008) faz uma analogia entre a gramática universal e o computador: o conjunto finito de regras da GU é análogo ao conjunto de comandos de programação computacional; as expressões usadas no uso linguístico são as imagens geradas na tela do computador (LANGACKER, 2008, p. 219). As regras gerativas, portanto, não representam exatamente às infinitas expressões encontradas no uso. Pelo contrário, as regras existem independentemente das formas que delas resultam (BYBEE, 2001). O papel das regras é, sobretudo, modelar o conhecimento coletivo que permite a todos os falantes produzir expressões bem formadas (LANGACKER, 2008). Vê-se, então, que o modelo gerativo tem potencial para formalizar generalizações universais e para explicar o porquê de todos os seres humanos serem capazes de se comunicar por meio de línguas

naturais. No entanto, a gramática universal se limita a modelar apenas os princípios entre as línguas naturais; as especificidades de cada língua e a variabilidade dialetal são abandonadas na vertente clássica do gerativismo.

Como se pode ver, um dos fundamentos da linguística tradicional é a máxima de que a linguagem é categórica (BOD, HAY, JANEDY, 2003). Mesmo que haja um consenso de que as línguas reais sejam altamente variáveis, gradientes e dinâmicas, muitos linguistas argumentam que a gramática é composta de categorias discretas. Nessas vertentes, o desempenho (o uso da língua) pode ser rico em variação, mas a competência (conhecimento abstrato) não (BOD, HAY, JANEDY, 2003). Os dados empíricos, no entanto, desafiam claramente a ideia de discretude do conhecimento linguístico. Muitos estudos têm evidenciado a natureza variável e contínua da linguagem (LABOV, 2008; MENDOZA-DENTON; HAY; JANNEDY, 2003; FOULKES; DOCHERTY, 2006). Em um trabalho referencial, Labov (2008) mostrou que novaiorquinos de classes sociais diferentes pronunciam o rótico pós-vocálico com diferentes intervalos de variação. Mendoza-Denton, Hay, Jannedy (2003) mostraram que os padrões de produção diferem entre os indivíduos. Foulkes e Docherty (2006) mostraram que as crianças de uma mesma comunidade de fala experimentam informações linguísticas diferentes. Isto é, cada criança está em contato com *inputs* linguísticos específicos e, por isso, é improvável que o desenvolvimento da linguagem se dê da mesma forma em duas ou mais crianças da mesma comunidade de fala (FOULKES; DOCHERTY, 2006). Como bem resume Langacker (2008, p. 217),

todo mundo fala um pouco diferente. Você não encontrará dois falantes, por exemplo, que controlem exatamente o mesmo vocabulário e atribuam exatamente o mesmo significado a cada item lexical. Se reificarmos as habilidades empregadas na fala, referindo-nos a elas individualmente como unidades linguísticas e coletivamente como um sistema linguístico, temos que reconhecer que o sistema linguístico de cada falante é diferente de todos os outros (tradução nossa<sup>10</sup>).

Se cada falante possui seu próprio jeito de falar, não há como identificar elementos linguísticos únicos e discretos; é preciso acomodar na teoria da linguagem a variabilidade dos inúmeros sistemas linguísticos individuais. Sabemos, contudo, que membros de uma mesma comunidade de fala exibem forte semelhança linguística, apesar das particularidades. Ao mesmo tempo em que há a nítida variação individual, há uma estrutura coletiva (LANGACKER, 2008). Os modelos de linguagem precisam, portanto, expressar a variabilidade linguística, sem desconsiderar a estrutura geral que possibilita aos falantes a interação comunicativa.

---

<sup>10</sup> Everybody talks a bit differently. You will not find any two speakers, for example, who control exactly the same vocabulary and ascribe exactly the same meaning to every lexical item. If we reify the skills deployed in talking, referring to them individually as linguistic units and collectively as a linguistic system, we have to recognize that every speaker's linguistic system is different from everyone else's (LANGACKER, 2008, p.217).

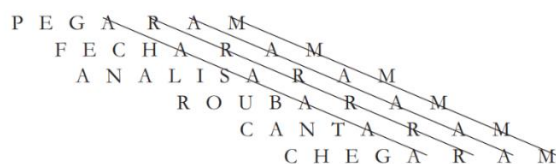
Dentre os modelos que conciliam experiência e estrutura abstrata encontram-se os Modelos de Exemplos (BYBEE, 2001; PIERREHUMBERT, 2001; JOHNSON, 2007; TOMASELLO, 2003). A seção a seguir se dedica a apresentar estes modelos em maiores detalhes: algumas premissas fundamentais para esta pesquisa são abordadas a fim de demonstrar o potencial dessas teorias para explicar a organização do conhecimento linguístico – seja oral, seja escrito.

### 3.2 Modelos de Exemplos

Pesquisas ancoradas nos Modelos de Exemplos compartilham o pressuposto de que as unidades linguísticas armazenadas na memória humana são unidades atestadas no uso real da língua (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; CRISTÓFARO-SILVA; GOMES, 2004; FOULKES; DOCHERTY, 2006; PORT, 2007). Diferentemente dos modelos linguísticos tradicionais – que assumem a invariância das representações mentais – os Modelos de Exemplos propõem que o conhecimento linguístico representa a totalidade de experiências que um indivíduo tem (FOULKES; DOCHERTY, 2006). Nesta abordagem, fonemas e regras fonológicas são substituídos por categorias fonéticas altamente detalhadas. Cada categoria contém um grande conjunto de exemplos dos variados itens experienciados no uso linguístico.

O *locus* das categorias são os itens lexicais. Isso significa que são as palavras e estruturas sintáticas cristalizadas – *chunks* – que compõem o conhecimento implícito do falante (CRISTÓFARO-SILVA; GOMES, 2004). Nessa perspectiva, as unidades internas das palavras são deriváveis de conjuntos de conexões feitas entre as palavras que têm partes relacionadas; ou seja, raízes e afixos não são unidades independentes, mas existem como relações de semelhanças entre palavras (BYBEE, 1998). Vejamos um exemplo:

**Figura 3:** Conexões lexicais entre palavras terminadas em -aram



**Fonte:** Huback (2003)

Os Modelos de Exemplos defendem que as palavras são organizadas por similaridade: memórias altamente semelhantes estão próximas umas das outras e memórias diferentes estão distantes (PIERREHUMBERT, 2001). Na Figura 3, por exemplo, os verbos terminados pela flexão modo temporal -aram estão associados entre si. As representações são, portanto, altamente redundantes, pois um mesmo morfema, como -aram, está registrado na memória muitas vezes,

haja vista o grande número de palavras que possui este padrão.

Vale ressaltar que há variações fonéticas na produção de itens lexicais com estruturas internas semelhantes. Oliveira (1992) mostrou que palavras com sílabas semelhantes, como *'metida'* e *'medita'* se comportam de forma distinta diante do fenômeno de alçamento vocálico no português de Belo Horizonte – MG. Bybee (2001), ao estudar o inglês americano, mostrou que palavras como *'every'*, *'memory'* e *'mammary'* variam de forma específica no que diz respeito à redução do *schwa*. Bybee (2002) mostrou que a frequência com que cada palavra ocorre na língua interfere no fenômeno de apagamento das consoantes [t] e [d] finais no inglês. Huback (2003) encontrou diferenças no que diz respeito ao cancelamento do R-final em nomes terminados pelo sufixo -dor. Fontes-Martins (2007) mostrou que uma mesma palavra pode ser produzida de forma distinta por diferentes indivíduos. Para os Modelos de Exemplares, todas essas variações encontradas na produção lexical são parte da representação mental. O grande potencial desses modelos é, então, fornecer

uma maneira de formalizar o conhecimento fonético detalhado que os falantes nativos têm sobre as categorias de sua língua. Como a Teoria de Exemplares armazena diretamente a distribuição dos valores dos parâmetros fonéticos associados a cada rótulo, ela nos fornece uma imagem do 'conhecimento fonético implícito do falante'. A aquisição desse conhecimento pode ser entendida simplesmente em termos da aquisição de um grande número de memórias de experiência. (PIERREHUMBERT, 2001, p. 6)

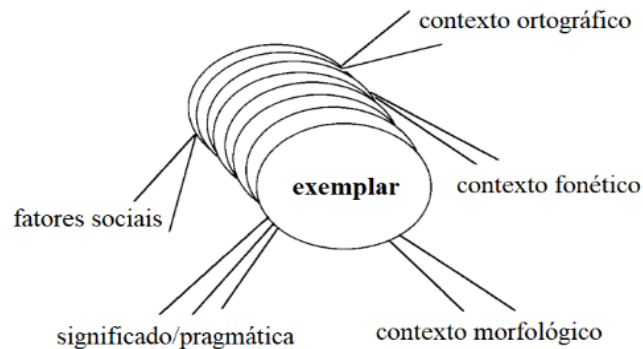
Como se pode ver, os Modelos de Exemplares propõem que o conhecimento fonético implícito representa, precisamente, as experiências vivenciadas pelo falante. Enquanto nos modelos tradicionais a fonética é dissociada da fonologia, nos Modelos de Exemplares a gradualidade fonética é fundamental para a organização do conhecimento fonológico (CRISTÓFARO-SILVA E GOMES, 2004). Considerando-se que a fonética e a fonologia são interrelacionadas, então aprendizes de ortografia precisam lidar com um conhecimento sonoro que ultrapassa a noção de fonema.

Tradicionalmente, os manuais de ortografia ressaltam que, para aprender a escrever, o aluno precisa se conscientizar dos fonemas da língua (SOARES, 2018). Essa perspectiva é apoiada na ideia de que a gramática fonológica é um inventário de unidades discretas. Contudo, vimos, no Capítulo 2, que a discretude da gramática é questionável, posto que a variação é inerente às línguas naturais. Apoiando-se na premissa de que a gramática fonológica é múltipla, argumentamos que os indivíduos gerenciam a variação fonética durante o aprendizado da ortografia. Prova disso são os variados erros ortográficos que refletem a sensibilidade dos aprendizes de escrita quanto a algumas variantes orais.

Além da variação fonética, as categorias linguísticas propostas pelos Modelos de Exemplares armazenam informações contextuais, semânticas, socioculturais, etc. (JOHNSON,

1997; FOULKES, DOCHERTY, 2006). Nos modelos tradicionais, as informações extralinguísticas que acompanham um enunciado são tidas como dispensáveis para a organização do conhecimento linguístico. Nos Modelos de Exemplos, por sua vez, as representações lexicais devem incluir, simultaneamente, uma vasta quantidade de informações linguísticas e não linguísticas (BYBEE, 2001).

**Figura 4:** Representação do conjunto de exemplares



**Fonte:** adaptado de Cristófar-Silva (2006, p. 172).

Os círculos ovais expostos na Figura 4 representam o conjunto de exemplares. As linhas que irradiam dos ovais ilustram as várias informações – linguísticas e não linguísticas – representadas na memória. Para exemplificar, pensemos na representação linguística da palavra ‘*mar*’. Para os modelos tradicionais, a representação de tal palavra é limitada a três fonemas. Para os Modelos de Exemplos, por outro lado, há registros detalhados de todos os exemplares da palavra ‘*mar*’ que o indivíduo experienciou no uso real da língua. Assim, a representação lexical inclui toda a variabilidade acústico-articulatória na produção dos sons da palavra e também todas as informações morfológicas, pragmáticas, semânticas relacionadas ao item lexical. Além dos fatores linguísticos, os exemplares registram todas as informações socioculturais: a voz dos falantes, as situações comunicativas em que a palavra é empregada, as variações diatópicas, diastráticas e diafásicas, etc.. A vantagem dos Modelos de Exemplos é, portanto, capturar o fato de que toda expressão carrega, simultaneamente, informações linguísticas e socioculturais (FOULKES; DOCHERTY, 2006).

Em sociedades grafocêntricas, a escrita pode ser o objeto sociocultural mais respeitado e mais utilizado no dia a dia. Livros são parte de nossa vida desde muito cedo. Antes mesmo de aprendermos a ler e a escrever, folheamos páginas cheias de letras (SOARES, 2010). Até indivíduos analfabetos estão em contato com a escrita, tendo em vista que esta é parte da paisagem urbana de sociedades letradas. Nas ruas, presenciamos, com naturalidade, as placas, os *outdoors*, os letreiros, os panfletos, etc.. Se a escrita está tão presente em nossas vidas, é necessário que pensemos em sua influência no nosso conhecimento linguístico.



Para Port (2006), nossas intuições sobre a oralidade são fortemente influenciadas pelas formas ortográficas que aprendemos ao longo do tempo. Para Morais (1996), a alfabetização estabelece uma nova relação entre língua e falante: não é mais possível ouvir uma palavra sem relacioná-la a sua forma ortográfica. Chevrot (2000) mostrou que a forma ortográfica influencia na escolha de variantes fonológicas. Paiva (2003) concluiu que o comportamento do indivíduo muda na medida em que há o aumento do contato com a escrita. Para Mattingly (1992), a consciência linguística do falante é moldada pelo sistema ortográfico. Todas essas conclusões documentadas na literatura nos motivam a pensar que a forma escrita é parte das representações lexicais. Por isso, adaptamos o diagrama proposto por Cristófar-Silva (2006) de modo a acomodar o contexto ortográfico como parte do conjunto de exemplares (Figura 4). É nítido, mais uma vez, o potencial dos Modelos de Exemplares para formalizar o conhecimento linguístico dos falantes em sua totalidade. A premissa de multiplicidade das representações mentais possibilita englobar todos os subcampos da linguística – inclusive a escrita – na teoria da linguagem.

É importante ressaltar que o fato de as categorias serem múltiplas não as isenta de generalizações. Afinal, são os padrões gerais que possibilitam aos indivíduos a interação por meio de uma mesma língua natural. A diferença dos Modelos de Exemplares, em relação aos modelos tradicionais, é que os padrões gerais são emergentes, e não inatos (BYBEE, 2001). Isto é, as estruturas gerais da linguagem são construídas à medida que as experiências são organizadas em esquemas (BYBEE, 2001, p. 85). A noção de esquema é fundamental para o presente trabalho, uma vez que a estamos tentando defender a tese de que as letras se associam a esquemas linguísticos múltiplos. Por isso, dedicamos uma seção específica, a seguir, para abordar tal conceito.

### 3.2.1 Esquemas

O termo *esquema* é encontrado com facilidade em publicações de psicólogos cognitivos que buscam caracterizar processos de memorização e aprendizado (BARTLETT, 1932; RUMELHART, 1980; IRAN-NEJAD, 1984). Para a psicologia cognitiva, o conhecimento prévio do indivíduo desempenha um papel muito importante no aprendizado de algo novo. Nesse viés, o esquema pode ser pensado como um organizador de todo o conhecimento de alguém sobre um tópico (RICHGELS, 1982). É um padrão cognitivo sistemático que organiza os conhecimentos prévios das pessoas e fornece uma estrutura para o entendimento futuro. São “blocos de construção da cognição” (RUMELHART, 1980), que classificam os conhecimentos e experiências prévias dos indivíduos em seções. Os esquemas organizam o conhecimento por semelhanças, o que facilita a recuperação de informações salvas e a incorporação de novas informações na cognição (IRAN-NEJAD, 1984). De modo simplificado, os esquemas são estruturas cognitivas que relacionam o

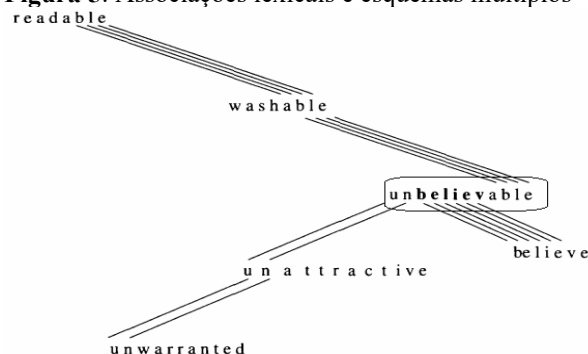
conhecimento novo ao conhecimento prévio do indivíduo, possibilitando à compreensão do mundo.

Na linguística, o conceito de esquema está presente em obras ancoradas nos Modelos de Uso (BYBEE, 1998; TOMASELLO, 2000; BYBEE, 2001; PIRREHUMBERT, 2001, LANGACKER, 2008). Uma das premissas desses modelos é a de que o conhecimento linguístico se desenvolve por meio de processos cognitivos gerais dentre os quais se encontra a esquematização. Para Langacker (2008, p. 17), a esquematização é “o processo de extrair semelhanças inerentes às múltiplas experiências para se chegar a uma concepção que represente um nível de abstração”.

No que diz respeito à sonoridade das línguas, Pierrehumbert (2003) sugere que esquemas sobre os padrões silábicos permitidos ou não nas línguas (fonotática) emergem das relações estabelecidas entre as palavras da língua. Para a autora, a fonologia de uma língua é um tipo de generalização que emerge das formas lexicais conectadas na representação. As formas lexicais, por sua vez, são generalizações da ampla gama de formas fonéticas encontradas na fala. Tomando a ideia de Pierrehumbert (2003) como referência, pode-se pensar que sílabas postônicas finais, como por exemplo [tʃi], emergem da rede de conexão entre palavras como lei[tʃi] ‘leite’, pot[tʃi] ‘pote’ e den[tʃi] ‘dente’. Além disso, é possível sugerir que a conexão entre essas mesmas palavras pode motivar a emergência do padrão ortográfico <te>.

Um mesmo elemento linguístico pode ocorrer em diferentes categorias e, conseqüentemente, pode contribuir com a emergência de diferentes tipos de esquemas (BYBEE, 1998, p. 422). Consideremos a Figura 5:

**Figura 5:** Associações lexicais e esquemas múltiplos



**Fonte:** Bybee (1998, p. 423).

A palavra ‘*unbelievable*’, na Figura 5, está associada às palavras terminadas pelo sufixo -able, à palavra primitiva *believe* e às palavras que se iniciam pelo prefixo -un. Como bem argumenta Bybee (2001, p. 7), existem muitos tipos de combinações entre os elementos linguísticos, como, por exemplo, a relação entre duas sequências fonéticas em uma mesma palavra;

a relação entre morfemas semelhantes em diferentes palavras; a relação entre dois fones semelhantes em diferentes palavras ou contextos, etc. Tomando por referência a proposta de Bybee (2001), é possível dizer que há, para além de relações fonéticas e morfológicas, relações fonortográficas; ou seja, a forma ortográfica das palavras pode contribuir com a emergência dos esquemas gerais da gramática. Especificamente, queremos defender que há diferentes esquemas linguísticos relacionados aos padrões ortográficos da língua. Por exemplo: a letra <e> em contexto pretônico se relaciona a esquemas linguísticos diferentes da letra <e> que ocorre em contexto postônico final. Isso porque o conhecimento desses padrões ortográficos emerge da relação entre diferentes itens lexicais.

Vale ressaltar, ainda, que, para os Modelos de Exemplares, os esquemas são probabilísticos (PIERREHUMBERT, 2003). Isso significa que o número de vezes que um indivíduo experiencia determinado prefixo, por exemplo, desempenha um papel importante na construção dos esquemas emergentes e, conseqüentemente, na replicação desse esquema em novos contextos. Além disso, o número de vezes que uma mesma palavra é produzida e percebida pelo falante pode afetar a probabilidade desta mesma palavra voltar a ocorrer no uso linguístico (BOD; HAY; JANNEDY, 2003). Palavras frequentes no uso linguístico possuem muitos exemplares e, conseqüentemente, possuem esquemas robustos. Por outro lado, palavras pouco frequentes, com poucos exemplares armazenados na gramática, podem desaparecer, assim como outras memórias. A frequência com que uma palavra (frequência de ocorrência) ou um padrão (frequência de tipo) ocorre na experiência afeta a sua produtividade na língua (BYBEE, 2001). O termo produtividade “se refere à tendência de generalização de determinado padrão como reflexo da experiência linguística” (CRISTOFARO-SILVA; GUIMARÃES, 2013, p. 318). Quanto mais itens são associados a um mesmo esquema, maior é sua produtividade; isto é, maior a possibilidade de um padrão geral se expandir para novos itens. Os esquemas são, portanto,

padrões recorrentes que podem ser usados para produzir novas combinações; [...] quanto maior o número de formas distintas a que um padrão se aplica, maior é a probabilidade de que ele se aplique a novos itens. Assim, alguns esquemas serão muito fortes e facilmente acessíveis para aplicação a muitos itens, e outros serão muito menos. (BYBEE, 1998, p. 422, tradução nossa<sup>11</sup>).

Uma vez que o conhecimento linguístico se organiza probabilisticamente, o conhecimento gráfico, armazenado na memória, também pode ser probabilístico. Isto é, na medida em que os indivíduos vivenciam experiências gráficas, eles vão construindo esquemas ortográficos. Assim,

---

<sup>11</sup> No original: “Recurrent patterns are the emergent generalizations or schemas that can be used to produce new combinations; [...] the greater the number of distinct stems a pattern applies to, the greater is the likelihood that it will apply to new items. Thus, some schemas will be very strong and easily accessible for applying to a great many items, and others will be much less so.” (BYBEE, 1998, p.422).

um padrão ortográfico que se repete muito – como as letras <e, o> postônicas finais – tem representação robusta e, por isso, são acessíveis para a aplicação em novas palavras. Por outro lado, padrões pouco frequentes – como as letras <i, u> postônicas finais – têm representação menos robusta e, conseqüentemente, são replicados com menos frequência ao longo do aprendizado da escrita.

Por fim, é relevante destacar que o conhecimento linguístico, nos Modelos de Exemplos, é plástico e dinâmico (CRISTÓFARO-SILVA; GOMES, 2004). Esse pressuposto tem potencial para explicar as diferenças no comportamento linguístico de pessoas com diferentes graus de escolaridade. Morais *et al.* (1979), por exemplo, mostraram que indivíduos alfabetizados julgam a linguagem de forma diferente de indivíduos analfabetos. Esse fato nos motiva a pensar que a gramática de indivíduos escolarizados armazena experiências distintas daqueles que não frequentaram a escola. Paiva (2003) analisou ditongos crescentes [ey] e [ay], que ocorrem em palavras como ‘madeira’ e ‘baixo’, na fala dos mesmos indivíduos adultos em diferentes momentos do tempo. A autora concluiu que

o comportamento linguístico do indivíduo sofre mudanças ao longo de sua vida para se ajustar aos diferentes momentos etários e às pressões sociais a que tem que fazer face em diferentes momentos de sua vida. Não está excluída, no entanto, a possibilidade de que essas mudanças reflitam também maior processo de escolarização e um maior contato com a forma escrita da língua em que a semivogal anterior é preservada (PAIVA, 2003, p. 46).

Sabendo-se que o comportamento de um mesmo indivíduo varia ao longo do tempo, pode-se pensar que os esquemas gerais, compartilhados pela comunidade de fala, emergem das experiências individuais. Nesse sentido, indivíduos alfabetizados podem agregar formas ortográficas aos esquemas linguísticos, ampliando o grau de generalização das categorias. Por outro lado, os indivíduos pré-alfabetizados ou analfabetos constroem suas generalizações por vias diferentes dos alfabetizados. Até mesmo o contato com pessoas escolarizadas pode alterar as representações mentais de pessoas com pouca vivência escolar. Considerando-se que cada indivíduo constrói a própria representação mental a partir da experiência, indivíduos com diferentes graus de escolarização compartilharão representações mentais diferentes, haja vista que vivem realidades diferentes. Em suma, os sistemas linguísticos alteram ao longo do tempo, uma vez que são abertos às múltiplas experiências individuais. Nesse sentido, o aprendizado da ortografia, ao longo da escolarização, pode influenciar na organização do sistema linguístico, o que explica as interferências da escrita na fala.

### 3.3 Teoria dos Sistemas Dinâmicos e Complexos

Ao definir a língua como um sistema, deve-se ter em mente que sistemas são constituídos de elementos que se relacionam formando um todo coerente. Existem sistemas fechados e abertos.

Um sistema fechado é determinado apenas por forças internas. Um sistema aberto se mantém pela interação das forças internas com o ambiente (OLIVEIRA, 2014). A TSDC tem como objeto de estudo sistemas abertos e, conseqüentemente, complexos. Essencialmente, a TSDC se enquadra no paradigma da complexidade, que é multidisciplinar e tem ancorado investigações de diversas disciplinas, como a biologia, a meteorologia, a medicina, a economia, a oceanografia, as engenharias, a administração, a educação e a literatura (LIMA JUNIOR, 2012). No âmbito da linguística, a TSDC assume que a linguagem é um sistema aberto, complexo, dinâmico e adaptativo, que evolui de forma não linear (BECKNER *et al.*, 2009). Enquanto o estruturalismo postula que a linguagem é um sistema fechado aos elementos externos, a TSDC assume que o funcionamento do sistema linguístico depende da interação mútua entre os vários elementos linguísticos e o ambiente (OLIVEIRA, 2014).

Para ilustrar a interdependência dos componentes de sistemas complexos, Oliveira (2014) recorre ao exemplo do sistema de trânsito. As ruas e avenidas que compõem o complexo tráfego de uma cidade são interdependentes; se uma avenida for retirada do sistema abruptamente, haverá efeitos caóticos e imprevisíveis (OLIVEIRA, 2014, p. 15). De maneira análoga, os elementos que compõem o sistema linguístico estão interligados entre si. Qualquer perturbação em um dos subsistemas da linguagem – indivíduo, comunidade de fala, fonologia, morfologia, sintaxe, ortografia, etc.. – pode causar efeitos em toda a sua estrutura.

Na TSDC, a linguagem é produto da interação entre os falantes de uma comunidade de fala e, portanto, a sua evolução depende fortemente do ambiente sociocultural (BECKNER *et al.*, 2009). Pesquisas como de Soares (2010), de Treiman (2017) e de Treiman e Zhang (2020) mostram que, em sociedades grafocêntricas, a escrita está presente no ambiente sociocultural de tal forma que mesmo os falantes analfabetos ou pré-alfabetizados estão em interação com práticas de leitura e escrita. Treiman (2017) mostrou que crianças norte-americanas distinguem desenhos de letras desde a tenra idade. Além disso, Treiman e Zhang (2020) evidenciaram que, desde os três anos de idade, as crianças conhecem os símbolos gráficos de sua cultura, à medida que julgam mais facilmente as letras do alfabeto latino do que símbolos de outras culturas, como o chinês. Considerando-se que, em culturas grafocêntricas, indivíduos interagem, constantemente, com materiais gráficos, é possível dizer, com o apoio da TSDC, que a escrita contribui para a evolução do sistema linguístico, uma vez que este é aberto às influências socioculturais.

Além de ser aberto, o sistema é dinâmico na perspectiva da TSDC. É consensual na linguística teórica que a língua está em constante evolução (FARACO, 1991). Tanto os idioletos<sup>12</sup>,

---

<sup>12</sup> Idioleto: a variedade linguística utilizada por um indivíduo concreto (BLOCH, 1948).

quanto a língua comunitária mudam de estado ao longo do tempo. O estado linguístico atual do português brasileiro, por exemplo, é resultado de trajetórias de mudanças passadas, assim como o estado futuro depende das interações existentes no sistema atual (FARACO, 1991). A TSDC tem potencial para explicar o caráter dinâmico da linguagem, uma vez que parte do pressuposto de que uma língua está em constante transição de fases. Há fases mais estáveis, em que o sistema se mantém por algum tempo, e há fases mais instáveis, em que o sistema muda rapidamente de um estado para outro (CAMERON; LARSEN-FREEMAN, 2007). Para explicarmos como ocorrem as mudanças no sistema, na perspectiva da TSDC, lançamos mão de alguns conceitos, que serão retomados no capítulo de análise dos dados:

- **Atrator:** é o estado estável e temporário que atrai e molda o comportamento do sistema por algum tempo (CAMERON; LARSEN-FREEMAN, 2007). Um único sistema pode conter mais de um estado atrator e a força de atração desses estados é variável (BYBEE; BECKNER, 2015). Segundo Bot, Lowie e Vespoor (2007), a trajetória do sistema rumo a estados atratores é análoga à trajetória de uma bola rolando em uma superfície com buracos e lombadas. Os buracos são como os atratores que tendem a atrair e estabilizar o rolar da bola (sistema). Quanto mais profundo é o buraco, mais energia é necessária para resgatar a bola e fazê-la seguir seu caminho; ou seja, quanto maior a força de atração, mais energia é necessária para impulsionar o sistema para outros estados (BOT; LOWIE; VESPOOR, 2007, p. 8). Nesta pesquisa, consideramos que formas ortográficas são atratores que moldam o sistema linguístico do aprendiz de escrita por um tempo. Nesta perspectiva, a ortografia aprendida tende a conduzir o sistema linguístico a um novo comportamento.
- **Auto-organização:** é uma propriedade que possibilita o reequilíbrio de um sistema dinâmico após algum estado de instabilidade (OLIVEIRA, 2016). Thelen e Smith (2003) descrevem o desenvolvimento locomotivo de bebês para explicarem o comportamento de sistemas complexos e dinâmicos. Inicialmente, bebês engatinham para se locomover de um lugar para outro. A ação de engatinhar é estável por vários meses (estado atrator). Quando os bebês começam a aprender a andar surge um novo padrão que desordena o sistema. A instabilidade do sistema é evidenciada pelo comportamento variável dos bebês: engatinham, levantam, andam, caem, sucessivamente. Em busca de resolver a instabilidade, o sistema precisa se **auto-organizar**. Após a auto-organização, o padrão de engatinhar é substituído por uma nova solução: o padrão de andar (THELEN; SMITH, 2003, p. 344). Nesta pesquisa, pretendemos argumentar que o aprendizado da escrita perturba o sistema linguístico e, conseqüentemente, há a auto-organização desse sistema rumo à adaptação do conhecimento ortográfico adquirido no conhecimento linguístico do indivíduo.

- **Adaptação:** reflete alterações no espaço fase que promovem mudanças no sistema como um todo (CRISTÓFARO-SILVA, 2016). A adaptação do sistema depende da interação entre subsistemas. No âmbito da linguagem, sistemas linguísticos de diferentes indivíduos se adaptam uns aos outros na interação comunicativa. Em um nível mais amplo, as interações entre os subsistemas causam adaptações no sistema linguístico de uma comunidade de fala (CAMERON; LARSEN-FREEMAN, 2009). Cameron e Larsen-Freeman (2009) caracterizam a aquisição de línguas como um constante processo de adaptação de padrões linguísticos que emergem dos atos comunicativos. Nesta pesquisa, propomos que o aprendizado da escrita envolve um processo de adaptação dos novos padrões ortográficos adquiridos no sistema linguístico.

De forma geral, a TSDC parte do pressuposto de que o desenvolvimento de sistemas complexos é dinâmico e envolve estados de estabilidade e instabilidade. Estados de estabilidade atraem o sistema (atratores). Em fases de instabilidade, os sistemas dinâmicos tendem a se auto-organizar, em busca de se estabilizarem novamente. É importante ressaltar que momentos de instabilidade são essenciais para que o sistema avance e se adapte a um novo estado.

O sistema aberto e dinâmico é também não linear na vertente da TSDC. Segundo Bot, Lowie e Vespoor (2007, p. 8, tradução nossa<sup>13</sup>),

o desenvolvimento de alguns sistemas dinâmicos parece ser altamente dependente de seu estado inicial; pequenas diferenças no início podem ter consequências drásticas a longo prazo. Isso é chamado de efeito borboleta, termo proposto pelo meteorologista Lorenz, que queria contabilizar o grande impacto de pequenos efeitos locais no clima global. Relacionado a isso está a noção de não linearidade; há uma relação não linear entre o tamanho de uma perturbação inicial em um sistema e os efeitos que pode ter a longo prazo. Pequenas mudanças podem levar a efeitos enormes, enquanto grandes perturbações podem ser absorvidas pelo sistema sem muita alteração.

No que diz respeito ao sistema linguístico, a não linearidade pode ser evidenciada em análises de aquisição de língua estrangeira e de língua escrita. A diversidade de características dos aprendizes – idade, motivação, nível de letramento, condição socioeconômica, tipo de instrução recebida, etc. – faz com que indivíduos de um mesmo ano escolar/nível utilizem rotas não lineares de aprendizado (LIMA JUNIOR, 2012). Além disso, o percurso de aprendizado oscila, de modo que, em alguns momentos, é possível pensar que turmas escolares retrocederam em habilidades

---

<sup>13</sup> the development of some dynamic systems appears to be highly dependent on their initial state; minor differences at the beginning may have dramatic consequences in the long run. This is called the butterfly effect, a term proposed by the meteorologist Lorenz, who wanted to account for the huge impact small local effects may have on global weather. Related to this is the notion of non-linearity, which means that there is a non-linear relation between the size of an initial perturbation of a system and the effects it may have in the long run. Some minor changes may lead to huge effects, while major perturbations may be absorbed by the system without much change (BOT, LOWIE, VESPOR, 2007, p.8).

que já haviam sido desenvolvidas. Em uma análise sobre a acentuação gráfica em textos de alunos brasileiros do 6º ao 9º anos do Ensino Fundamental II (EFII), Couto e Oliveira-Guimarães (2020) verificaram que os discentes do 8º ano erraram mais o acento gráfico do que os alunos do 7º ano. Se o aprendizado ocorresse linearmente, deveria haver uma correlação positiva entre o nível de escolaridade e o índice de acertos ortográficos; ou seja, haveria aumento das habilidades ortográficas à medida que se aumentasse o nível escolar. No entanto, como demonstrado por Couto e Oliveira-Guimarães (2020), o aprendizado é não linear e os índices de erros e acertos podem aumentar e diminuir ao longo dos anos escolares, já que há fatores variados envolvidos nesse percurso.

Nesta pesquisa, defendemos que o sistema oral é aberto e, por isso, interage com o sistema ortográfico de forma dinâmica, complexa e não linear. Especificamente, argumentamos que o aprendizado da ortografia motiva a instabilidade e, conseqüentemente, a auto-organização e a adaptação do sistema linguístico. Em síntese, a tese a ser defendida nesta pesquisa é a seguinte: *as letras são símbolos discretos que se associam a esquemas múltiplos, que emergem da interação complexa entre palavras, seqüências fonéticas, morfológicas e ortográficas.*

### **3.3 Resumo do Capítulo 3**

Este capítulo apresentou o quadro teórico que fundamenta a análise realizada nesta pesquisa. Na primeira seção, foram apresentados os pressupostos dos Modelos de Exemplares. Esperamos ter demonstrado que, diferentemente das abordagens tradicionais, os Modelos de Exemplares assumem que os padrões gerais (esquemas) emergem de associações complexas entre múltiplas informações linguísticas e não linguísticas. Nessa concepção, as letras têm representação discreta, mas os sons são parte de categorias múltiplas e detalhadas, que armazenam a variação linguística. Assim, uma mesma letra pode estar relacionada a diferentes esquemas múltiplos. Na segunda seção, foram abordadas as premissas da Teoria dos Sistemas Dinâmicos e Complexos (TSDC). Essa teoria propõe que linguagem é um sistema aberto, dinâmico, complexo e não linear. Apoiando-se na TSDC, argumentamos que o aprendizado do sistema ortográfico motiva a instabilidade e a conseqüente auto-organização e adaptação do sistema linguístico. Com a auto-organização do sistema, pode haver alterações na fala.



## 4. METODOLOGIA

Neste capítulo, são descritos os procedimentos utilizados antes, durante e após a coleta dos dados desta pesquisa. Em primeiro lugar, são apresentadas as decisões metodológicas que antecederam a coleta: a seleção das palavras, o perfil dos participantes e o desenho experimental. Em segundo lugar, é explicado como a coleta foi realizada. Por fim, apresentam-se a organização dos dados e os princípios estatísticos que orientaram a análise.

É relevante ressaltar que a análise foi feita em duas amostras de dados, sendo uma de dados de escrita e outra de dados de fala. O *corpus* de escrita foi utilizado para avaliar a influência da fala na escrita, a partir da análise dos índices dos erros ortográficos em cada ano escolar. O *corpus* de fala foi utilizado para investigar a influência da escrita na fala, a partir da análise da variação sonora na fala das crianças de cada ano escolar.

### 4.1 Procedimentos pré-coleta

Esta seção descreve como se deu a elaboração do desenho experimental. Os critérios utilizados na seleção das palavras da pesquisa, o campo de coleta, o perfil dos participantes e os experimentos utilizados na coleta de dados são apresentados.

#### 4.1.1 Seleção de palavras

As palavras desta pesquisa foram selecionadas no Dicionário Ilustrado de Português (BIDERMAN, 2004). Esse dicionário foi escolhido porque, além de ser uma ferramenta voltada ao público infantil, as palavras que o constituem são de alta frequência e estão presentes em livros didáticos do Ensino Fundamental (BIDERMAN, 2004, p. 8). O dicionário conta com cerca de 5.900 itens lexicais, dentre os quais foram selecionadas as 64 palavras-teste deste trabalho. Consideramos os seguintes critérios para filtrar a seleção:

**Quadro 11:** Critérios utilizados na seleção das palavras-teste

categoria	letras	Acento	classe gramatical
vogais postônicas finais	e, o	paroxítona	substantivo concreto
vogais pretônicas	e, o	paroxítona	substantivo concreto
verbos de infinitivo	R	Oxítona	verbo de ação
nomes com CVR- final	R	oxítona	substantivo concreto

**Fonte:** elaboração própria.

Foram selecionadas palavras paroxítonas terminadas com as letras <e> e <o> postônicas finais; palavras paroxítonas com <e> e <o> em posição pretônica; verbos e nomes oxítonos terminados em <r>. Além disso, controlamos a classe gramatical das palavras, em busca de facilitar a escolha das figuras que constituíram os experimentos. Foram selecionados apenas

substantivos concretos e verbos de ação que podem ser facilmente representados por meio de figuras. As palavras escolhidas estão listadas no Quadro 12:

**Quadro 12:** Palavras-teste

(1) postônicas finais	(2) pretônicas	(3) R-final verbos	(4) R-final nomes
chave	tesoura	lavar	celular
alface	bezerro	jogar	mar
sorvete	besouro	cortar	jantar
lustre	vestido	pescar	bar
cabide	peruca	varrer	calcanhar
peixe	bexiga	comer	colar
garfo	tomate	beber	professor
laço	morcego	morder	elevador
cavalo	boneca	abrir	cantor
caderno	formiga	dirigir	tambor
gato	coruja	imprimir	flor
suco	borracha	corrigir	pintor

**Fonte:** elaboração própria.

Primeiramente, foram escolhidas 12 palavras-teste para cada categoria linguística estudada. No caso das vogais postônicas finais (1), foram escolhidas 6 palavras paroxítonas terminadas em <e> e 6 palavras paroxítonas terminadas em <o>. Na categoria das vogais pretônicas (2), escolhemos 6 palavras paroxítonas com a letra <e> pretônica e 6 palavras paroxítonas com a letra <o> pretônica. Foram selecionadas apenas palavras em que o acento da vogal pretônica já foi atestado e documentado na literatura (OLIVEIRA, 1992; VIEGAS, 1995; CRISTÓFARO-SILVA; GRECO, 2010), tendo em vista que o acento vocálico é o fenômeno de interesse desta pesquisa. Além disso, escolhemos apenas palavras trissílabas em que a sílaba pretônica é a primeira da palavra e a única antes da tônica, em busca de mantermos o mesmo padrão silábico e acentual. Para a categoria em (3), foram escolhidos 4 verbos de cada conjugação. Para a análise de nomes terminados em <r> (4), foram escolhidas 6 palavras oxítonas terminadas em -ar e 6 palavras oxítonas terminadas em -or.

Em um segundo momento, foram selecionadas mais 16 palavras-teste, sendo 4 substantivos grafados com <i, u> pretônico, 4 substantivos grafados com <i> postônicos finais, 4 verbos terminados em vogal tônica e 4 nomes também terminados em vogal tônica. Vejamos o Quadro 13:

**Quadro 13:** Palavras-teste para análise da hipercorreção

(1) pretônicas	(2) postônicas finais	(3) R-final verbos	(4) R-final nomes
pipoca	táxi	sorri	Sofá
tucano	origami	lê	robô
biquini	safári	dormi	maracujá
muleta	júri	dá	dominó

**Fonte:** elaboração própria.

As palavras listadas no Quadro 13 foram selecionadas no intuito de analisar erros ortográficos, tradicionalmente, denominados de hipercorreções. Como já foi visto na revisão da literatura (Capítulo 2), as crianças em fase de alfabetização podem trocar <i, u> por <e, o> em casos como < \*origame, \*tocano>, em vez de < ‘origami’ e ‘tucano’> (TENANI; REIS, 2011; BOHN; SOUZA, 2017). Além disso, é possível encontrar nos textos infantis a inserção do <r> em final de verbos flexionados. Exemplo: <eu \*sorrir> em vez de < ‘eu sorri’> (CESAR, 2017; CASTELO BRANCO, 2020). Uma vez que esse tipo de erro está relacionado às categorias analisadas nesta pesquisa, é relevante investigá-lo. A descrição e análise desse tipo de erro ortográfico podem trazer evidências adicionais sobre o percurso de aprendizado da ortografia, além de indícios sobre a natureza dos esquemas múltiplos que constituem a representação mental.

#### *4.1.2 O campo de coleta e os participantes*

Os participantes da pesquisa foram crianças matriculadas no 1º, 3º, 5º, 7º e 9º anos escolares de uma escola municipal da região de Venda Nova em Belo Horizonte – MG. A escola municipal foi escolhida por trabalhar com todos os anos escolares do Ensino Fundamental, além de ser uma escola pública parceira da UFMG, que acolhe estagiários de diferentes licenciaturas e do Programa Institucional de Bolsas de Iniciação à Docência (PIBID). A escolha pela escola pública se deu porque, de acordo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (BRASIL,2020)<sup>14</sup>, a maioria dos alunos brasileiros se encontra matriculada no ensino público. Além disso, esta pesquisa é fomentada com verba pública, por meio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e, por isso, consideramos necessário devolver o investimento ao coletivo.

Na coleta de dados piloto desta pesquisa – realizada em março de 2021 – analisamos a fala e a escrita de alunos dos cinco primeiros anos do Ensino Fundamental I (1º a 5º anos do EFI). A análise comparativa dos dados do piloto apontou que as amostras de dados de escrita do 1º e do 2º anos, assim como as do 3º e do 4º não diferiram significativamente. Por essa razão, optamos por substituir o 2º e o 4º anos pelo 7º e 9º, visando ampliar a pesquisa para o EFII e, conseqüentemente, analisar a relação entre a fala e a escrita ao longo de todo o Ensino Fundamental.

As turmas de cada ano escolar foram escolhidas aleatoriamente pela coordenação da escola. Todos os alunos de cada turma escolhida receberam duas vias do Termo de Consentimento Livre Esclarecido (disponível no Apêndice A). Apenas os alunos que tiveram autorização dos responsáveis entraram na lista de participantes da pesquisa. Dessa lista, foram selecionados,

---

<sup>14</sup> Disponível em: <http://portal.inep.gov.br/sinopses-estatisticas-da-educacao-basica>

aleatoriamente, 10 alunos de cada ano escolar. Ao todo, 50 alunos participaram da pesquisa. Todos os participantes assinaram o Termo de Assentimento Livre e Esclarecido – TALE (Apêndice B). Todos os participantes são naturais de Belo Horizonte – MG. A idade média dos participantes do 1º, 3º, 5º, 7º e 9º é, respectivamente, 6, 8, 10, 12 e 14 anos.

Além dos alunos, 5 professoras(es), sendo 3 pedagogas e 2 professores especialistas em língua portuguesa, participaram da pesquisa. A participação dos docentes nos ajuda a entender a motivação dos erros ortográficos pela perspectiva do professor do Ensino Fundamental. Para isso, cada professor respondeu um questionário, que se encontra disponível no Apêndice C deste texto.

#### 4.1.3 Experimentos

O experimento<sup>15</sup> desenvolvido para a coleta de dados envolveu perguntas-guias associadas a figuras e foi dividido em três etapas subsequentes. A primeira etapa – chamada de **etapa de treinamento** – visou a familiarizar os participantes aos estímulos da pesquisa. Nessa etapa, os participantes foram orientados a memorizar os nomes de algumas figuras. A etapa de treinamento foi pensada com o objetivo de diminuir o índice de respostas não esperadas. A pesquisadora alertou as crianças de que as figuras e seus respectivos nomes seriam importantes para atividades futuras. A Figura 6 ilustra um estímulo utilizado na etapa 1.

**Figura 6:** Exemplo de estímulo experimental



**Fonte:** elaboração própria.

Cada estímulo apresentado às crianças continha uma figura e uma palavra gravada em áudio. As palavras-teste relacionadas às vogais pretônicas e postônicas finais foram todas pronunciadas com vogais altas. As palavras-teste relacionadas ao rótico foram faladas com cancelamento do R-final. A gravação das palavras foi feita para que todos os participantes tivessem acesso ao mesmo estímulo. A transcrição fonética foi colocada na Figura 6 apenas para

---

<sup>15</sup> O experimento foi realizado como parte do projeto “PAL: Percursos de aquisição da linguagem: oralidade e escrita em L1 e L2”, aprovado no Comitê de ética e pesquisa, em 2021. Número do CAAE: 48791021.4.0000.5149

exemplificar o estímulo que as crianças ouviram. A etapa de treinamento teve o intuito de treinar os participantes para as etapas posteriores, visando a diminuir, ao máximo, o índice de respostas diferentes do esperado.

A segunda etapa – chamada de **etapa oral** – teve a finalidade de coletar dados de fala. Cada participante foi orientado a observar as figuras, ouvir uma pergunta-guia e respondê-la em voz alta. As figuras e as perguntas foram organizadas no *Power Point* e houve aleatorização automática dos *slides* antes de cada coleta. A aleatorização diminui o risco de haver efeito (não controlado) da ordem das figuras nos dados. Não foram utilizados distratores, pois há grande variabilidade entre os padrões sonoros e ortográficos das palavras-teste. A pesquisadora apresentou uma figura por vez aos alunos juntamente às perguntas que guiam as suas respostas. As perguntas-guia foram gravadas em áudio para que todos os alunos ouvissem o mesmo estímulo. A seguir, encontram-se exemplos:

**Figura 7:** Exemplos de estímulos experimentais com perguntas-guia



**Fonte:** elaboração própria.

A figura (A) foi utilizada para coletar dados referentes às vogais postônicas finais. Junto à figura tem-se um áudio com a pergunta “*que verdura é essa?*”. Esperávamos obter a palavra ‘*alface*’ como resposta. A figura (B) foi utilizada para coletar dados referentes às vogais pretônicas. A pergunta “*que objeto é esse?*” guiou o falante a responder ‘*tesoura*’. A figura (C) foi utilizada para a coleta de dados referentes aos verbos no infinitivo. Escolhemos figuras que representem ações. Observe a pergunta: *O que a moça vai fazer com o cabelo?* Esperávamos a seguinte resposta: ‘*cortar*’. A figura (D) foi utilizada para coletar dados referentes aos nomes terminados com CVR. A pergunta “*O que é isso?*” orienta o falante a responder ‘*celular*’. As perguntas-guia estão escritas na Figura 7 apenas com a finalidade de exemplificar o experimento. Durante a coleta, apenas o áudio era tocado e não houve quaisquer estímulos ortográficos.

A terceira etapa – chamada de **etapa escrita** – envolveu a escrita das palavras-teste da pesquisa. Utilizamos as mesmas figuras e perguntas-guia das duas etapas anteriores. Os

participantes foram instruídos a escrever a resposta da pergunta-guia em uma folha de papel. A utilização de figuras e perguntas-guia para a coleta de dados foi pensada com a intenção de evitar efeitos da forma oral na forma ortográfica. A intenção do experimento é que os alunos grafem as palavras com os seus conhecimentos sobre a ortografia. As instruções, as figuras e as perguntas-guias se encontram no Apêndice D deste texto.

## 4.2 Coleta dos dados

A coleta de dados foi realizada de modo presencial em uma escola municipal de Belo Horizonte – MG. A direção da escola disponibilizou o laboratório de ciências, que contém mesas, cadeiras e tomadas. Os demais materiais – computador, fone de ouvido, microfone, gravador de áudio, folha de papel e caneta – foram providenciados pela pesquisadora. A coleta de dados foi feita individualmente com cada participante, em horários autorizados pelos professores dos alunos. A duração do experimento foi, em média, 25 minutos com cada participante. O tempo total de realização da coleta foi de 12 dias. Todos os participantes receberam a mesma instrução e o mesmo método foi aplicado aos participantes do 1º, 3º, 5º, 7º e 9º anos.

Primeiramente, o aluno foi instruído a colocar um fone para ouvir o nome das figuras que apareceriam na tela do computador. A pesquisadora enfatizou que a memorização desses nomes era importante para as etapas posteriores. Cumpria-se, neste momento, a etapa de treinamento. Esta etapa durou, em média, 4 minutos.

Em segundo lugar, foi aplicada a etapa oral. Nesta etapa, o participante foi instruído a observar novamente as figuras e a falar o nome delas. Mostramos ao aluno que cada figura estaria acompanhada de um áudio com uma pergunta-guia que o ajudaria a se lembrar do nome que ouviu na etapa anterior. Ressaltamos a importância de se falar o nome das figuras em voz alta. Antes de cada gravação de voz, a pesquisadora colocou um microfone de lapela na camiseta do participante. O microfone foi essencial para amenizar os ruídos típicos de um ambiente escolar. A voz dos alunos foi gravada por meio do gravador *Easy Voice Recorder*. Este gravador pode ser baixado gratuitamente nas lojas de aplicativos dos smartphones. Entre as vantagens desse aplicativo estão as seguintes: (i) há a possibilidade de configurar o formato da gravação para *.wav*; (ii) há a possibilidade de configurar a taxa de amostragem da gravação – configuramos em 44 kHz, pois esta taxa permite uma captura precisa de toda faixa de frequência da fala; (iii) há a possibilidade de armazenar o áudio gravado diretamente em nuvem, sem que este seja comprimido. Esta etapa durou, em média 4 minutos.

Em terceiro lugar, foi aplicada a etapa escrita. Os participantes foram instruídos a observar novamente as figuras, escutar novamente as perguntas-guia e respondê-las por escrito. A pesquisadora disponibilizou folha de papel e caneta ao aluno e enfatizou que a atividade não era

avaliativa. Ao final da atividade, as folhas de papel com as palavras escritas foram recolhidas e guardadas pela pesquisadora. Em um momento futuro, as folhas de papel foram escaneadas – por meio do aplicativo *CamScanner* – e foram armazenadas em um computador. A duração da atividade de escrita variou entre 5 e 30 minutos, dependendo do participante. Os participantes dos anos escolares iniciais precisaram de mais tempo para escrever, tendo em vista que ainda estão no início do processo de aprendizado.

### 4.3 Procedimentos pós - coleta

Os *corpora* contam com uma grande quantidade de dados – total de 5.964 dados – e, por isso, foi necessário estabelecer critérios para a organização do material. Primeiramente, descrevemos como foi feita a organização dos dados e, na sequência, explicamos os procedimentos realizados na análise.

#### 4.3.1 Organização dos dados

Os dados de escrita foram anotados quanto à letra grafada e quanto ao número de erros e acertos relacionados às vogais médias e aos róticos. Os dados de fala foram categorizados quanto ao som produzido. Os códigos e as variáveis criadas em cada amostra de dado estão descritos, separadamente, a seguir.

##### 4.3.1.1 Organização da amostra dados de escrita

Cada participante escreveu 16 palavras referentes a cada uma das 4 categorias linguísticas (12 palavras-teste + 8 estímulos para hipercorreção). No total, cada participante produziu 64 dados, o que resultou em 3.200 dados de escrita (64 palavras \* 50 participantes). Após a filtragem de dados descartados por escrita ilegível ou por escolha de palavras diferentes dos estímulos, a amostra de dados de escrita contou com 3.057 dados. A Tabela 1 resume a quantidade total de dados de escrita por categoria e por ano escolar.

**Tabela 1:** Quantidade total de dados de escrita

categoria ↓   ano →	1°	3°	5°	7°	9°	Total por categoria
Vogais pretônicas	157	154	153	156	154	774
Vogais postônicas finais	153	153	148	158	157	769
R-final em verbos	146	150	141	149	155	741
R-final em nomes	150	155	153	156	159	773
<b>Total por ano</b>	<i>606</i>	<i>612</i>	<i>595</i>	<i>619</i>	<i>625</i>	<b>3.057</b>

Fonte: elaboração própria.

Os dados de escrita foram categorizados quanto aos erros e aos acertos relacionados às vogais e aos róticos. Erros ortográficos que saíram do escopo da pesquisa não foram quantificados.<sup>16</sup> A categorização dos erros e dos acertos possibilitou a medição do índice de erros ortográficos por ano escolar e por categoria linguística.

#### 4.3.1.2 Organização da amostra dados de fala

As mesmas 16 palavras-teste coletadas para a amostra de escrita foram utilizadas na amostra de dados de fala. Assim sendo, cada participante produziu também 64 dados de fala, o que resultou em 3.200 dados de fala (64 palavras \* 50 participantes). Após a filtragem de dados descartados por falha no áudio ou por produção de palavras diferentes do esperado, a amostra de dados de fala contou com 2.907 dados de fala. A Tabela 2 resume a quantidade total de dados de fala por categoria e por ano escolar.

**Tabela 2:** Quantidade total de dados de fala

categoria↓  ano →	1°	3°	5°	7°	9°	Total por categoria
Vogais pretônicas	153	146	152	156	155	762
Vogais postônicas finais	135	147	140	154	148	724
R-final em verbos	128	140	143	138	140	689
R-final em nomes	134	145	152	148	153	732
<b>Total por ano</b>	<b>550</b>	<b>578</b>	<b>587</b>	<b>596</b>	<b>596</b>	<b>2.907</b>

Fonte: elaboração própria.

Cada dado de fala foi codificado pelo som produzido nos contextos pretônico, postônico final e em final de verbos e nomes. As variantes identificadas na fala, em cada categoria, estão descritas no Quadro 14:

**Quadro 14:** Variantes orais identificadas no conjunto de dados

Vogal postônica final	Vogal pretônica	Verbos	Nomes
vogal média anterior	vogal média anterior	R-final	R-final
vogal média posterior	vogal média posterior	Cancelamento	cancelamento
vogal alta anterior	vogal alta anterior		
vogal alta posterior	vogal alta posterior		
Apagamento			

Fonte: elaboração própria.

As variantes orais descritas no Quadro 14 foram agrupadas em duas classes: (a) variantes

<sup>16</sup> Vários tipos de erros, que fogem do escopo da pesquisa, foram encontrados nos dados. Por exemplo: <\*abi> (*abrir*); <\*aufasi> (*alface*); <\*bb> (*beber*); <\*bizoro, bizourro, bezolro,> (*besouro*); <\*mexiga> (*bexiga*); <\*bisero> (*bezerro*); <\*bonca> (*boneca*); <\*boraxa, \*borasa> (*borracha*); <\*cabite> (*cabide*); <\*cadeno> (*caderno*); <\*cacona; \*caukanha> (*calcanhar*); <\*cato> (*cantor*); <\*xavi> (*chave*); <\*coriji> (*corrigir*); <\*mucgo> (*morcego*), etc.



diferentes da forma ortográfica (em cinza claro); b) variantes próximas à forma ortográfica (em cinza escuro). Vejamos alguns exemplos que compõem essas categorias:

**Quadro 15:** Categorias dos dados de fala

Tipos de variante	Postônica fin.	Pretônica	Nome	Verbo
(a)proxescrita	[ˈlaso]	[peˈrukə]	[seluˈlah]	[pesˈkah]
(b)difescrita	[ˈlasø]; [ˈlasø];	[piˈrukə]	[seluˈlaø]	[pesˈkaø]

Fonte: elaboração própria.

A codificação ilustrada no Quadro 15 nos permitiu investigar os índices de variantes próximas à forma ortográfica na fala dos alunos do Ensino Fundamental. Pudemos verificar se o ano escolar – e conseqüentemente o tempo de estudo da escrita – influencia na forma de falar dos alunos do Ensino Fundamental. Dito de outra forma, a codificação nos possibilitou visualizar se, ao longo dos anos escolares, há aumento de variantes como laç[o], p[e]ruca celula[h], pesca[h], em detrimento de laç[u], p[i]ruca, celula[ø], pesca[ø].

As amostras de dados de fala e de escrita foram organizadas em uma mesma planilha do Excel. A planilha conta com as seguintes informações, distribuídas em colunas específicas:

**Quadro 16:** Planilha de organização dos dados

Nome da coluna	Códigos das variáveis
participante	P1, P2, P3, P4...
ano	1A, 3A, 5A, 7A, 9ª
palavra	cabide, boneca, pescar...
tipo de palavra	erro, hipercorrecao
fenômeno	postonica, pretonica, verbo, nome
categoria	vogalantterior, vogalposterior, rótico
forma oral	vogalalta, vogalmedia, apagamento, cancelamento, rfinal
tipo de variante	difescrita, proxescrita
forma escrita	xavi, auface, joga...
erro ou acerto	erro ou acerto

Fonte: elaboração própria.

#### 4.3.2 Análise dos dados

A análise dos dados foi feita por meio do Software R (R CORE, 2022). O R é um ambiente projetado em torno de uma linguagem computacional que auxilia os pesquisadores na análise estatística e na plotagem de gráficos. Uma das vantagens na utilização do R está na capacidade do software em fornecer *scripts* de programação para a análise de dados. Os *scripts* podem ser lidos por qualquer computador e podem ser compartilhados rapidamente via e-mail, usbs, drives, etc.. A facilidade de compartilhamento dos *scripts* permite que outros pesquisadores possam reproduzir, revisar ou auditar uma análise de dados (o *script* desenvolvido para esta pesquisa se encontra no Apêndice E deste texto).

A análise dos dados foi desenvolvida de modo a responder a seguinte pergunta de pesquisa: *categorias linguísticas, aparentemente semelhantes, se realizam de forma análoga ou diferente na fala e na escrita de alunos do Ensino Fundamental?* Para isso, utilizamos métodos estatísticos descritivos e inferenciais, em busca de descrever e explicar a amostra de dados, para, assim, tentar encontrar generalizações que expliquem as relações entre a fala e a escrita ao longo da escolarização. A seguir, esses métodos estatísticos serão explicados em mais detalhes.

#### 4.3.2.1 *Análise estatística*

A análise estatística de dados é realizada quando um pesquisador está interessado em encontrar relações entre determinadas variáveis de um conjunto de dados. Uma variável pode ser definida como uma característica a ser controlada e observada em um estudo experimental (GODOY, NUNES, em preparação). Por exemplo: o índice de erros e acertos ortográficos e o ano escolar são algumas das variáveis observadas nesta pesquisa. As variáveis podem ser subdivididas em dois grandes grupos: (1) variáveis quantitativas; (2) variáveis qualitativas. As variáveis quantitativas são de natureza numérica e podem ser discretas ou contínuas. As variáveis qualitativas, por sua vez, identificam uma categoria e, normalmente, possuem um número finito e determinado de elementos (GODOY; NUNES, em preparação; GRIES, 2019). Por exemplo: a ortografia correta ou incorreta (erro x acerto) constitui uma variável qualitativa. Esta pesquisa avalia apenas variáveis qualitativas – também chamadas de categóricas.

Quando o estudo está sendo planejado, deve-se esclarecer as relações entre as variáveis. Em outras palavras, deve-se identificar quais são as variáveis dependentes e independentes a serem analisadas. As variáveis dependentes são objeto de observação; ou seja, o pesquisador deve observar o comportamento de determinada variável em busca de informações. Esta variável não pode ser controlada; ela depende diretamente do comportamento dos participantes da pesquisa (GODOY; NUNES, em preparação). Neste estudo, as variáveis dependentes são: (1) a forma ortográfica (erro x acerto); (2) a variante oral (próxima à ortografia x diferente da ortografia). Essas duas variáveis não foram controladas no experimento. Ao contrário, apenas observamos a escrita e a fala dos participantes. As variáveis independentes, por outro lado, devem, em hipótese, explicar o comportamento da variável dependente (GODOY; NUNES, em preparação). Por exemplo: nesta pesquisa, partimos da hipótese de que a categoria linguística (vogal pretônica, vogal postônica final, R-final em verbo, R-final em nome) explica a persistência de determinados erros ortográficos. Assim sendo, devemos assumir que a categoria linguística é uma variável independente deste estudo. O Quadro 17 resume quais são as variáveis da presente pesquisa.

**Quadro 17:** Variáveis dependentes e independentes

	<b>Dados de fala</b>	<b>Dados de escrita</b>
<b>Variável dependente</b>	tipo de variante oral (dif. escrita x prox. escrita)	forma ortográfica (erros x acertos)
<b>Variável independente I</b>	categorias linguísticas vocálicas (vogal pretônica x vogal postônica final)	
<b>Variável independente II</b>	categorias linguísticas róticas (R-final em verbo x R-final em nome)	
<b>Variável independente III</b>	ano escolar (1°, 3°, 5°, 7°, 9°)	

**Fonte:** elaboração própria.

Uma vez que as variáveis dependentes e independentes foram identificadas e resumidas, pode-se propor as hipóteses da pesquisa (GRIES, 2019). Segundo Gries (2019, p. 23), as hipóteses científicas são uma afirmação sobre a relação entre duas ou mais variáveis. Vejamos, novamente, as hipóteses formuladas para este estudo:

*(1a) as trocas de <e, o> por <i, u> na escrita persistem, em índices significativamente diferentes, até estágios distintos da escolarização, a depender do contexto acentual: pretônico ou postônico final.*

*(1b) a omissão da letra <r> na escrita persiste, em índices significativamente diferentes, até estágios distintos da escolarização, a depender da classe gramatical: verbo ou nome.*

*(2a) o aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônica ou postônica final*

*(2b) o aumento do rótico final na fala ao longo dos anos escolares depende da classe gramatical: verbo ou nome.*

As hipóteses (1a) e (1b) foram testadas a partir da análise dos índices de erros ortográficos em cada categoria linguística e em cada ano escolar. Para testar a hipótese (1a), compararam-se os índices de erros ortográficos relacionados às vogais pretônicas e postônicas em cada ano escolar. Para testar a hipótese (1b), compararam-se os índices de erros ortográficos relacionados aos róticos em coda final dos verbos e dos nomes em cada ano escolar.

As hipóteses (2a) e (2b) foram avaliadas a partir da análise das variantes próximas à escrita – vogais médias e R-final – na fala dos alunos EF. Para testar a hipótese (2a), verificamos os efeitos do ano escolar e do contexto acentual – pretônico ou postônico final – na produção de vogais médias na fala. Para testar a hipótese (2b), comparamos os índices de produção oral do R-final em verbos e em nomes, em cada ano escolar.

A análise dos dados foi realizada por meio de métodos estatísticos descritivos e

inferenciais. A análise estatística descritiva auxiliou na exploração e na visualização gráfica dos dados. Como as variáveis da pesquisa são categóricas, as descrições envolveram contagens e percentagens. Foram contabilizados os índices percentuais de variantes orais e de erros e acertos na escrita. Esses índices estão representados em tabelas e em gráficos de colunas, pontos e linhas, nos Capítulos 5 e 6.

Além da descrição estatística, foram aplicados alguns testes de hipóteses às amostras, visando fazer uma análise exploratória para entender a distribuição dos dados e a relação entre as variáveis. Especificamente, aplicamos o teste Qui-quadrado de Person para medir a associação ou independência entre as variáveis categóricas dependentes e independentes (FRANKE, HO, CHRISTE, 2012). Para interpretar as relações entre duas variáveis, utiliza-se o p-valor. O valor de p é definido pelo pesquisador. Em ciências humanas, utiliza-se, geralmente, o valor de 0.05. Assim sendo, se as estatísticas entre as duas amostras não diferirem significativamente, o resultado do p-valor é maior que 0.05. Ao contrário, se houver diferenças estatísticas entre duas amostras o p-valor é menor que 0.05.

Por fim, ajustamos modelos lineares generalizados mistos ao conjunto de dados. Modelos lineares avaliam, matematicamente, a relação entre duas ou mais variáveis (GODOY, 2019). No caso desta pesquisa, os modelos lineares serão utilizados para avaliar se os índices de erros ortográficos e a variação oral estão relacionados à categoria linguística e ao ano escolar. Como as variáveis dependentes desta pesquisa são categóricas, o tipo de modelo linear ajustado deve ser generalizado. Neste tipo de modelo, calculam-se as probabilidades de se obter sucesso em um evento que tenha apenas dois resultados possíveis (GODOY, 2019). Por exemplo: ao escrever um texto, temos dois resultados possíveis: acertar ou errar a ortografia. O modelo linear generalizado nos permite medir se a probabilidade de errar a ortografia em um contexto A é diferente da probabilidade de errar a ortografia em um contexto B. Este tipo de análise estatística possibilita prever os resultados para a população de alunos do EF. Isto é, aplicando modelos estatísticos ao conjunto de dados de fala e de escrita de alguns participantes, é possível prever o comportamento de outros alunos belorizontinos do EF.

Vale lembrar, todavia, que para que os resultados possam ser modelados, é preciso considerar a variabilidade de cada dado. Isto é, é necessário incluir no modelo as variáveis de efeito aleatório: no caso, o indivíduo e a palavra (GODOY, 2019). Para entender por que o modelo deve considerar os efeitos aleatórios, pensemos na seguinte situação hipotética: os resultados de um modelo linear confirmam que há maior probabilidade de ocorrer erros ortográficos em verbos do que em nomes. No entanto, um participante, em específico, produziu mais da metade dos erros ortográficos relacionados aos verbos. Nesse caso, não há como dizer que os erros ortográficos estão relacionados apenas à classe gramatical; o indivíduo específico também deve ser considerado

como uma variável capaz de explicar os índices de erros ortográficos.

Além disso, a palavra também pode interferir no conjunto de dados. Suponhamos, por exemplo, que a maioria dos erros ortográficos ocorra em uma palavra específica. Se esta palavra for um verbo, o modelo poderá prever, erroneamente, que o erro ortográfico é explicado pela classe verbal, quando na verdade é explicado por uma só palavra. Por isso, é necessário incluir o item lexical no modelo. O tipo de modelo que considera a variabilidade do item lexical e do indivíduo é chamado de modelo misto. Neste tipo de modelo, calcula-se a relação entre as variáveis dependentes (forma ortográfica e variantes orais), os efeitos fixos (ano escolar e categoria linguística) e os efeitos aleatórios (participante e item lexical).

#### **4.4 Resumo do Capítulo 4**

O Capítulo 4 apresentou a metodologia utilizada para coleta, organização e análise dos dados desta pesquisa. A coleta dos dados foi realizada por meio de um experimento realizado em três etapas: treinamento, dados de fala e dados de escrita. As atividades experimentais envolveram perguntas-guia e figuras. O experimento foi realizado com alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental de uma escola municipal de Belo Horizonte – MG. A coleta foi realizada presencialmente. A análise dos dados foi desenvolvida com auxílio da Linguagem R. Para análise exploratória, aplicaram-se testes qui-quadrados à duas amostras: uma de dados de fala e outra de dados de escrita. Em busca de generalizações que expliquem os erros ortográficos e a variação oral, modelos lineares generalizados mistos foram ajustados aos dados.

## 5. ANÁLISE DOS DADOS: VOGAIS MÉDIAS

As vogais médias pretônicas e postônicas finais são descritas, tradicionalmente, como padrões similares. Na fala, essas vogais podem ser produzidas como vogais altas. Por exemplo: m[i]nino, b[u]neca, pent[i], pat[u]. Na escrita, ambas as vogais devem ser grafadas com as letras <e, o>, mas aprendizes grafam-nas, equivocadamente, com <i, u>. Exemplo: <\*minino>, <\*buneca>, <\*penti>, <\*patu>.

Se, por um lado, há semelhanças entre as vogais, por outro, há algumas particularidades em cada contexto acentual. As vogais médias pretônicas estão envolvidas no fenômeno fonológico variável de alçamento vocálico (BISOL, 1989; BATTISTI, 1993; SCHWINDT, 1995; VIEGAS, 2001; CHAVES, 2013; MAGALHÃES; CAMPOS, 2022). Já as vogais postônicas finais estão envolvidas no fenômeno fonológico variável denominado de redução vocálica (DIAS; SEARA, 2013). Ou seja, as vogais médias pretônicas e postônicas finais são pronunciadas de forma distinta na oralidade. Na escrita, por sua vez, há evidências de que os aprendizes cometem diferentes índices de erros ortográficos em contextos pretônico e postônico final (LEMOS, 2001; MIRANDA, 2010; CRISTÓFARO-SILVA; GUIMARÃES, 2013).

Considerando-se as semelhanças e as diferenças entre as vogais pretônicas e postônicas finais, esta tese se propôs a compará-las em busca de responder à seguinte questão: *categorias linguísticas, aparentemente semelhantes, se manifestam de forma análoga ou diferente na fala e na escrita dos alunos do EF?* Para isso, analisaremos, ao longo desta seção, os dados de fala e de escrita relacionados às vogais médias pretônicas e postônicas finais.

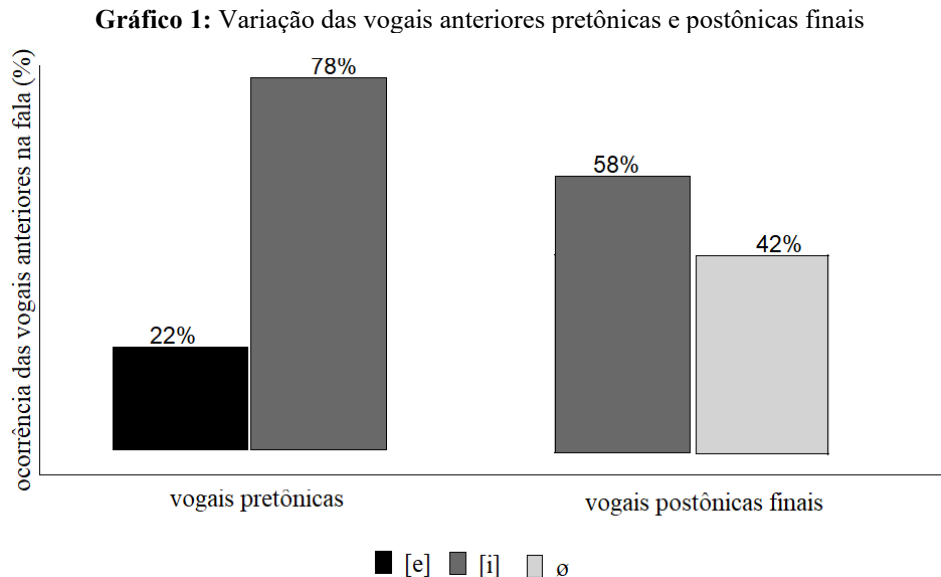
Este capítulo está organizado em quatro partes. A parte I apresenta a análise descritiva dos dados. A parte II testa as hipóteses (1a) e (2a). A parte III discute os dados à luz dos Modelos de Exemplos (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e da Teoria dos Sistemas Dinâmicos e Complexos (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). A parte IV aborda a visão dos professores sobre os erros ortográficos referentes às vogais médias pretônicas e postônicas finais.

### 5.1 PARTE I: Análise descritiva dos dados

A primeira parte da análise descreve as amostras de dados de fala e de escrita relacionadas às vogais médias pretônicas e postônicas finais. A descrição está subdividida da seguinte maneira: (i) vogais anteriores na fala; (ii) vogais posteriores na fala; (iii) vogais anteriores na escrita; (iv) vogais posteriores na escrita; (v) hipercorreções. As vogais anteriores e posteriores serão descritas separadamente, porque os dados mostraram que a variação sonora e o aprendizado da escrita ocorrem de forma diferente nas vogais [e] e [o].

### 5.1.1 Vogais anteriores na fala

Esta seção apresenta os dados de fala referentes às *vogais anteriores* pretônicas e postônicas finais. Busca-se, sobretudo, descrever a variação oral entre [e] ~ [i] ~ ø em ambos os contextos acentuais. Vejamos o Gráfico 1.



Fonte: elaboração própria.

As colunas à esquerda do Gráfico 1 se referem às vogais anteriores pretônicas. As colunas à direita se referem às vogais anteriores postônicas finais. A coluna preta mostra os índices de ocorrência da vogal [e]. As colunas cinza-escuras mostram os índices de ocorrência da vogal [i]. A coluna cinza-claro mostra os índices de cancelamento vocálico.

A vogal [e] ocorreu apenas em contexto pretônico, em 22% (N=58) dos dados. Já a vogal [i] foi produzida em contexto pretônico (78%, N =210) e em contexto postônico final (58%, N=165). Vale ressaltar ainda que, no contexto postônico final, houve 42% (N=118) de cancelamento da vogal anterior.

Comparamos as amostras do Gráfico 1 por meio do teste Qui-Quadrado. A análise estatística acusou diferenças significativas entre as vogais anteriores pretônicas e postônicas finais ( $\chi^2 = 180.22$ ,  $df = 2$ ,  $p\text{-value} < 2.2e-16$ ). Ou seja, confirmou-se, estatisticamente, que as amostras de dados de fala referentes às vogais anteriores pretônicas e postônicas finais são diferentes. As seguintes afirmações podem resumir os dados:

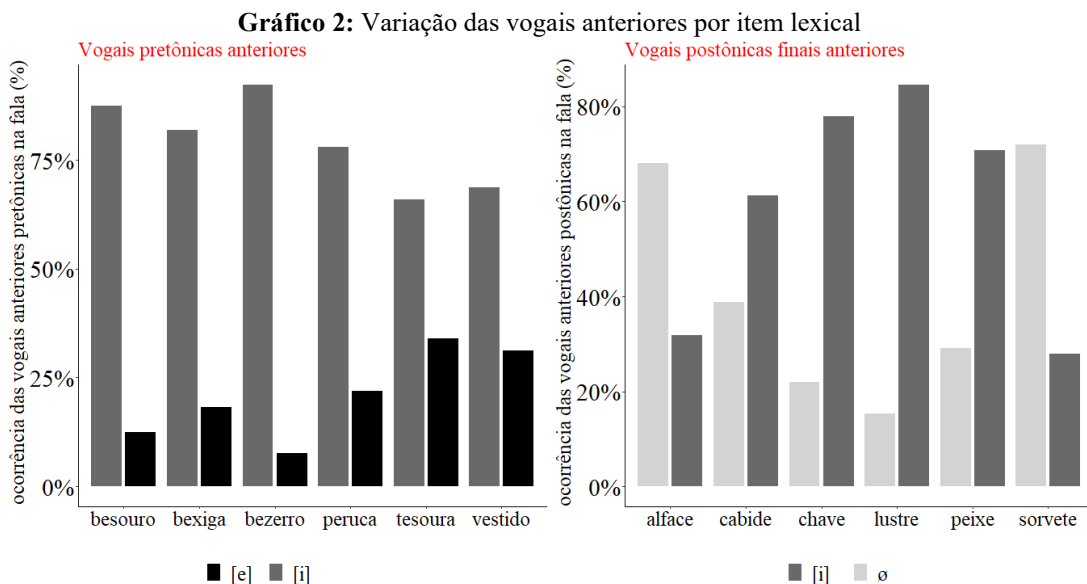
- As vogais médias anteriores – [e] – ocorreram somente em posição pretônica;
- As vogais altas anteriores – [i] – ocorreram nas posições pretônica e postônica final;
- O cancelamento de vogais anteriores foi restrito à posição postônica final.

Os resultados encontrados nos dados estão em consonância às seguintes afirmações já apresentadas na literatura:

- Há variação entre as vogais [e] ~ [i] em contexto pretônico (CHAVES, 2013).
- Há preferência pela produção da vogal [i] em contexto postônico final (BISOL, 2003).
- A vogal anterior postônica final pode ser cancelada (VIEGAS; OLIVEIRA, 2008).

De forma geral, os dados indicam que as vogais anteriores pretônicas e postônicas finais estão em diferentes estágios evolutivos da língua portuguesa. Para que fique claro, consideremos a trajetória proposta por Cristófar-Silva e Vieira (2015): [e] > [i] > ø. Lê-se a trajetória da seguinte maneira: a vogal [e] é o padrão mais antigo na língua portuguesa; a vogal [i] é o padrão atual; o cancelamento é o padrão inovador. Como os dados mostram que o cancelamento vocálico ocorreu apenas nas vogais postônicas finais, pode-se dizer que essas vogais se encontram em estágios evolutivos mais inovadores do que as vogais pretônicas. Esses resultados evidenciam que a língua é um sistema dinâmico que evolui de forma não linear (BOT; LOWIE; VERSPOOR, 2007).

O caráter não linear da evolução linguística é evidenciado também pela variação lexical. Vejamos:



Fonte: elaboração própria.

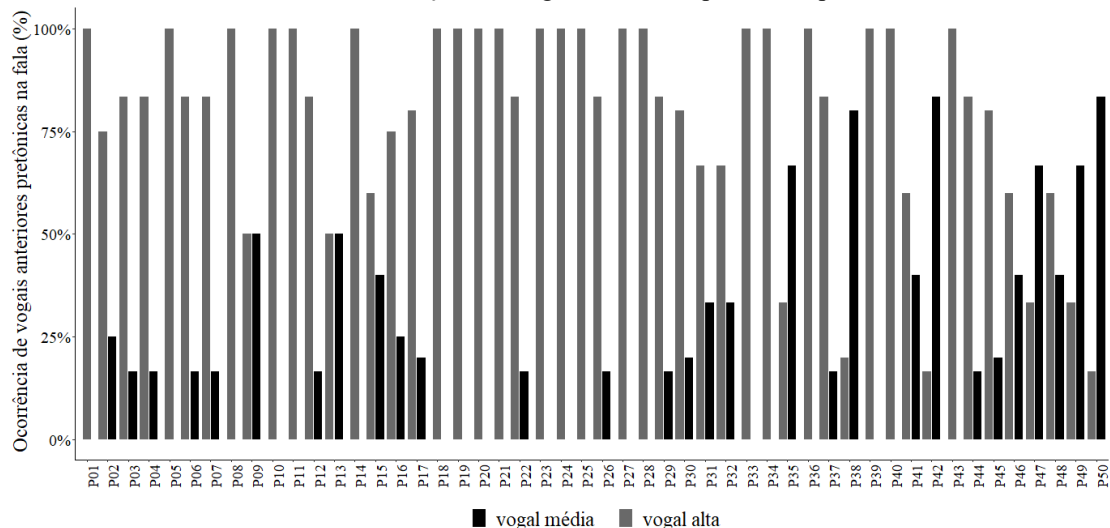
À esquerda do Gráfico 2, encontram-se as palavras-teste referentes às vogais anteriores pretônicas. À direita, estão as palavras-teste referentes às vogais anteriores postônicas finais. As colunas pretas do Gráfico 2 ilustram as palavras produzidas com [e]. As colunas cinza-escuras mostram as palavras pronunciadas com [i]. As colunas cinza-claras mostram os índices de cancelamento da vogal anterior em cada palavra. Notemos que cada item lexical apresentou intervalos específicos de variação. Na palavra 'besouro', por exemplo, os índices de [i] pretônicos



(87%, N=42) foram bem maiores do que os de [e] (12%, N=6). Já na palavra *'vestido'*, houve altos índices de [i] (69%, N=33), mas também houve índices consideráveis de [e] pretônicos (31%, N=15). No contexto postônico final, por sua vez, a vogal [i] foi cancelada em altos índices nas palavras *'alface'* (68%, N=32) e *'sorvete'* (72%, N=36), mas em baixos índices na palavra *'lustre'* (15%, N=6). Em resumo, os dados indicam que cada palavra se encontra em um estágio específico da trajetória evolutiva da língua portuguesa. Ou seja, os fenômenos de variação sonora operam de maneira não linear no léxico do PB.

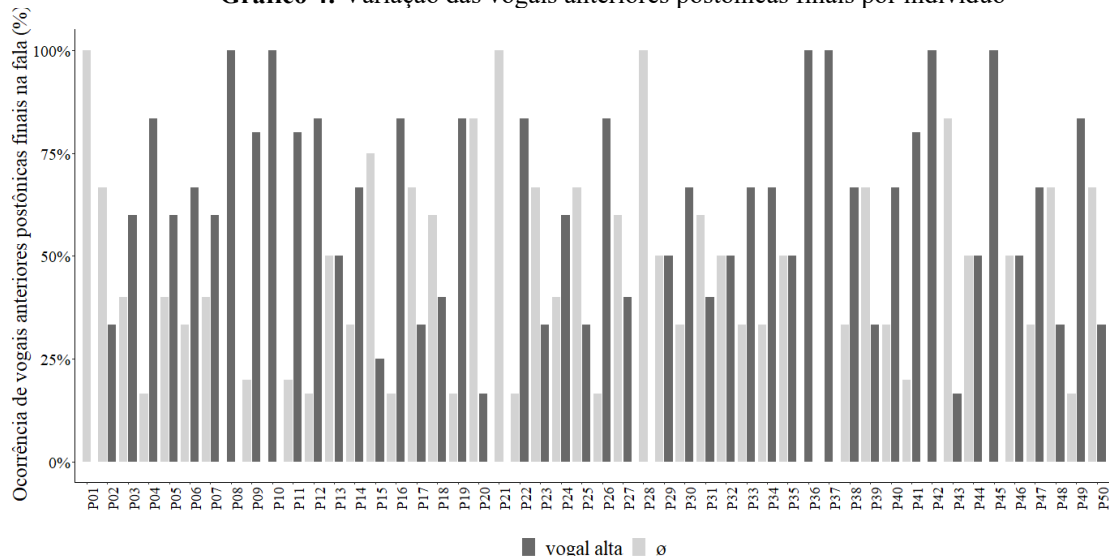
Além do léxico, há um outro fator importante contribuindo com a dinamicidade do sistema linguístico: o indivíduo. Cada participante da pesquisa produziu índices diferentes de vogais médias, altas e canceladas, a depender do contexto acentual. Vejamos os Gráficos 3 e 4:

**Gráfico 3:** Variação das vogais anteriores pretônicas por indivíduo



Fonte: elaboração própria.

**Gráfico 4:** Variação das vogais anteriores postônicas finais por indivíduo



Fonte: elaboração própria.

O Gráfico 3 mostra a produção oral de vogais anteriores *pretônicas* por indivíduo. O Gráfico 4 mostra a produção oral de vogais anteriores *postônicas finais* por indivíduos. Como se pode ver, a produção das vogais pretônicas e postônicas finais varia de indivíduo para indivíduo. Três resultados merecem destaque:

- Nenhum indivíduo produziu apenas a vogal [e] em contextos pretônico e postônico final.
- Todos os indivíduos produziram vogais [i] pretônicas e postônicas finais.
- A maioria dos indivíduos cancelou a vogal anterior postônica final.

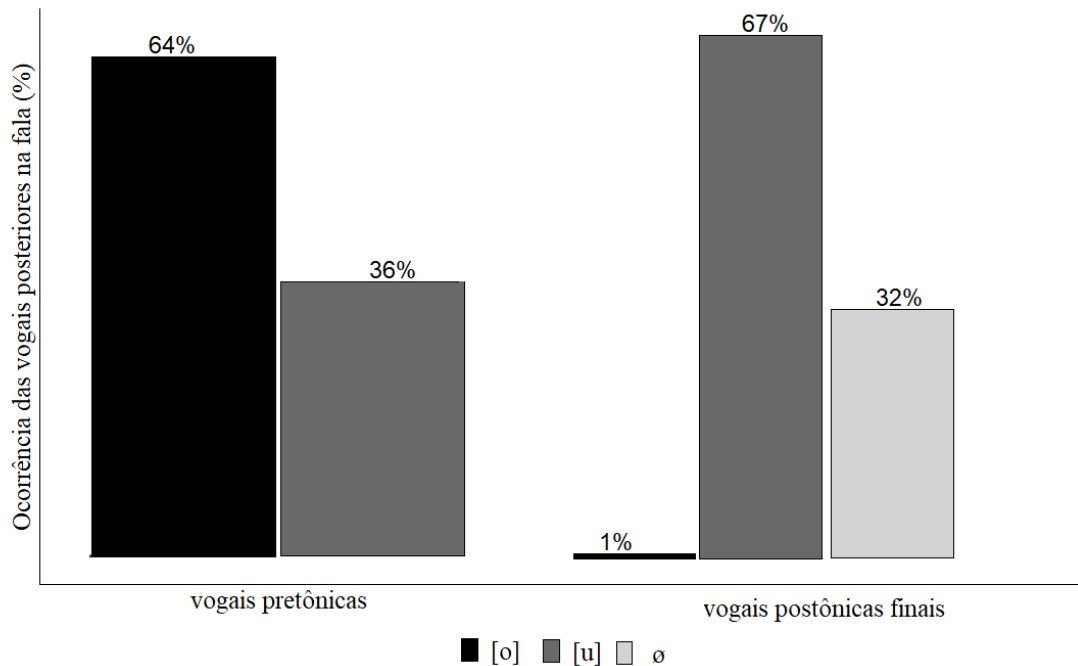
Os dados indicam, portanto, que a produção das vogais anteriores varia de maneira não linear também no nível individual. Alguns indivíduos foram mais conservadores do que outros ao produzirem as vogais pretônicas. Por exemplo: o indivíduo P50 produziu mais vogais [e] do que o indivíduo P01. No contexto postônico final, alguns indivíduos foram mais inovadores. Exemplo: os indivíduos P01, P20 e P28 cancelaram as vogais anteriores postônicas finais em 100%.

Vale ressaltar ainda que houve variação sonora ocorrendo na fala de um mesmo indivíduo. Vejamos, por exemplo, os dados do aluno P01 (1º ano). Em contexto pretônico, esse aluno produziu somente vogais [i]. Já em contexto postônico final, o mesmo aluno cancelou todas as vogais [i]. Ou seja, as vogais pretônicas e postônicas finais variam de forma não linear até mesmo na fala de único indivíduo.

Além das vogais anteriores, há, no português brasileiro, palavras pronunciadas com vogais posteriores – [o, u] – pretônicas e postônicas finais. As vogais posteriores também estão envolvidas nos fenômenos fonológicos de alçamento e redução vocálica. No entanto, os dados mostraram que tais fenômenos fonológicos operaram nas vogais [o, u] de forma diferente do que operaram nas vogais [e, i]. Por essa razão, optamos por analisar as vogais anteriores e posteriores separadamente.

### 5.1.2 Vogais posteriores na fala

Seguindo a mesma organização da seção anterior, descrevemos, nesta seção, a variação das vogais posteriores a nível segmental, lexical e individual. O Gráfico 5 ilustra a variação sonora das vogais posteriores no nível segmental.

**Gráfico 5:** Variação das vogais posteriores pretônicas e postônicas finais

Fonte: elaboração própria.

As colunas à esquerda do Gráfico 5 se referem às vogais pretônicas. As colunas à direita se referem às vogais postônicas finais. As colunas pretas mostram os índices de ocorrência da vogal [o]. As colunas cinza-escuras indicam o percentual de ocorrência da vogal [u]. As colunas cinza-claras se referem às taxas de cancelamento da vogal posterior. A vogal [o] ocorreu em 64% (N= 193) dos dados referentes às vogais pretônicas e em apenas 1% (N=2) dos dados referentes às vogais postônicas finais. Já a vogal [u] ocorreu em 36% (N= 107) dos dados referentes à posição pretônica e em 67% (N =198) dos referentes à posição postônica final. O cancelamento de vogais posteriores ocorreu em 32% (N=94) das vogais postônicas finais, mas não ocorreu em contexto pretônico. As amostras dos dados referentes às vogais posteriores se mostraram significativamente diferentes ( $\chi^2= 309.95$ ,  $df = 2$ ,  $p\text{-value} < 2.2e\text{-}16$ ). Podemos resumir os dados da seguinte maneira:

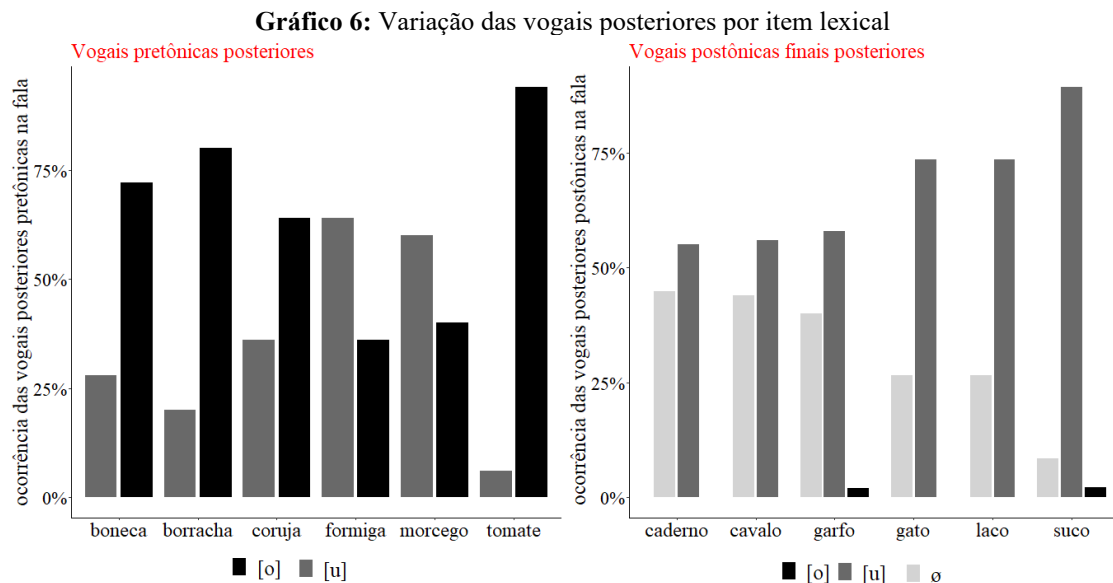
- As vogais médias posteriores – [o] – ocorreram majoritariamente em posição pretônica;
- As vogais altas posteriores – [u] – ocorreram nas posições pretônica e postônica final;
- O cancelamento de vogais posteriores foi restrito à posição postônica final.

Os resultados do Gráfico 5 estão de acordo com as seguintes afirmações já documentadas na literatura:

- Há variação entre as vogais [o] ~ [u] em contexto pretônico (CHAVES, 2013).
- Há preferência pela produção da vogal [u] em contexto postônico final (BISOL, 2003).
- A vogal anterior postônica final pode ser cancelada (SEARA; DIAS, 2013).

Assim como no caso das vogais anteriores, os dados indicam que as vogais posteriores pretônicas e postônicas finais estão em diferentes estágios evolutivos da língua portuguesa. Em diálogo com a Cristóforo-Silva e Vieira (2015), sugerimos a seguinte trajetória evolutiva:  $[o] > [u] > \emptyset$ . A vogal  $[o]$  é o padrão mais antigo no PB; a vogal  $[u]$  é o padrão atual; e  $\emptyset$  é o padrão inovador. Os dados indicam, portanto, que as vogais posteriores postônicas finais estão em estágios mais inovadores da língua do que as vogais posteriores pretônicas. Esses resultados expressam a dinamicidade e a não linearidade evolutiva do sistema linguístico.

As vogais posteriores pretônicas e postônicas finais foram produzidas com algum índice de variação em todas as palavras-teste avaliadas na pesquisa. Vejamos o Gráfico 6.



Fonte: elaboração própria.

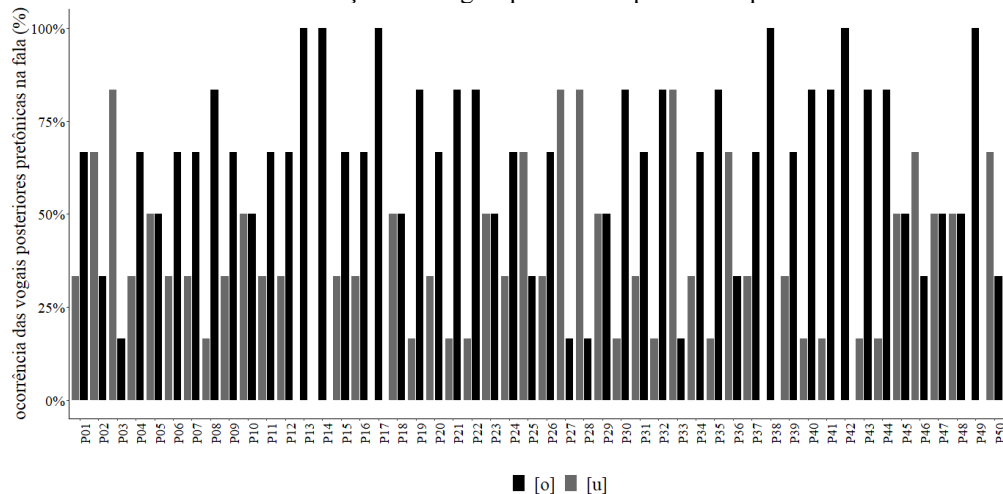
O Gráfico 6 mostra a produção oral das vogais posteriores por palavra. À esquerda, encontram-se as colunas referentes às vogais pretônicas posteriores. À direita, estão as colunas referentes às vogais postônicas finais. As colunas pretas, no Gráfico 6, se referem aos índices de vogal  $[o]$  em cada palavra. As colunas cinza-escuras mostram os índices de vogal  $[u]$  por palavra. As colunas cinza-claras indicam as taxas de cancelamento das vogais posteriores. Observe que todas as palavras estão envolvidas em algum índice de variação sonora. No entanto, a variação se manifestou de forma lexicalmente distinta. A palavra *'formiga'*, por exemplo, foi pronunciada com  $[u]$  pretônico em 64% (N=32) dos dados, enquanto a palavra *'tomate'* foi preferencialmente produzida com  $[o]$  (94%, N=47). Os dados indicam, portanto, que a palavra *'tomate'* está em estágios mais conservadores da trajetória  $[o] > [u] > \emptyset$  do que a palavra *'formiga'* na variedade de Belo Horizonte – MG.

No contexto postônico final, todas as palavras foram produzidas majoritariamente com  $[u]$ . Porém, todas elas também tiveram algum índice de cancelamento vocálico. As palavras

'cavalo' e 'caderno' foram as que tiveram maiores índices de cancelamento (44 e 45%, N= 22). Ou seja, essas palavras estão avançadas na trajetória rumo ao padrão inovador (cancelamento).

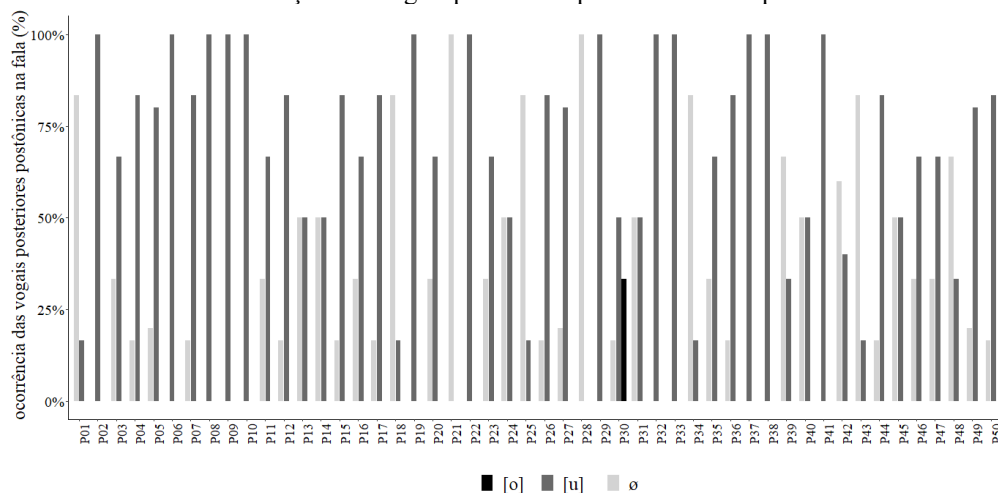
Um resultado que merece destaque diz respeito às palavras 'garfo' e 'suco'. Ambas destoaram das demais palavras por terem sido produzidas com [o] em 2% (N=1) dos dados. Uma análise tradicional da variação poderia considerar a ocorrência da vogal [o] postônica final como exceção devido ao baixo percentual. Ao descartar as "exceções", contudo, deixamos de interpretar resultados importantes. Os baixos índices de vogais médias nas palavras 'garfo' e 'suco' evidenciam a trajetória  $[o] > [u] > \emptyset$  em curso na língua. Vogais médias podem ocorrer em contexto postônico final devido ao fato de que a língua evolui de maneira não linear. Algumas palavras podem estar em estágios mais conservadores de evolução da língua porque os fenômenos de variação sonora afetam as palavras gradualmente. Além disso, os indivíduos – que também são sistemas dinâmicos – produzem diferentes intervalos de variação linguística. Vejamos.

**Gráfico 7:** Variação das vogais posteriores pretônicas por indivíduo



Fonte: elaboração própria.

**Gráfico 8:** Variação das vogais posteriores postônicas finais por indivíduo



Fonte: elaboração própria.

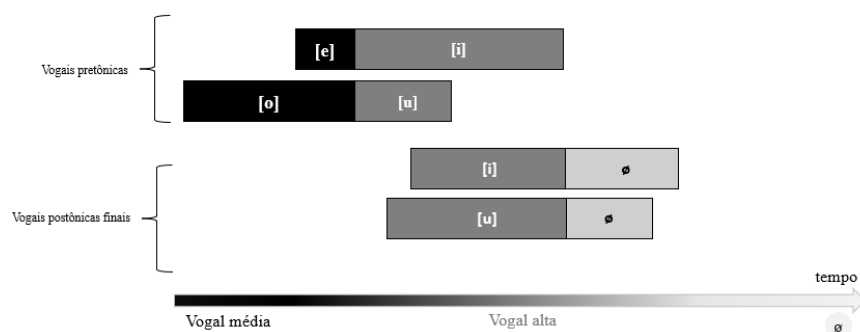
O Gráfico 7 mostra a produção oral de vogais posteriores pretônicas por indivíduo. O Gráfico 8 mostra a produção oral de vogais posteriores postônicas finais por indivíduo. Notemos, em primeiro lugar, as diferenças na pronúncia das vogais na fala de um mesmo indivíduo. Por exemplo: o participante P01 produziu altos índices (67%, N=4) de vogal [o] em contexto pretônico, mas, por outro lado, cancelou a vogal posterior postônica final na maioria dos dados (83%, N=5). Já o indivíduo P42 produziu apenas vogais [o] pretônicas, mas variou entre [u] (40%, N=2) e  $\emptyset$  (60%, N=3) em contexto postônico final. Em síntese, os dados mostram que cada pessoa se comporta de forma específica diante da produção oral de vogais pretônicas e postônicas finais.

Em segundo lugar, observemos a variação entre os indivíduos. No contexto pretônico, a maioria dos participantes produziu tanto vogal alta quanto vogal média posterior. Seis alunos se mostraram em estágios mais conservadores da trajetória  $[o] > [u] > \emptyset$  e produziram apenas vogais médias posteriores pretônicas na oralidade: P13, P14, P17, P38, P42, P49. No contexto postônico final, a maioria dos indivíduos variou entre [u] e  $\emptyset$ . Além disso, o indivíduo P30 foi o único que produziu a vogal [o] postônica final. De forma geral, os dados descritos nesta seção evidenciam que os fenômenos de alçamento e cancelamento vocálico evoluem em diferentes rotas na língua, podendo estar em estágios mais avançados ou mais conservadores a depender do contexto acentual, da palavra e do indivíduo.

### 5.1.3 Diferenças entre vogais anteriores e posteriores na fala

Ao compararmos as amostras de dados de fala das vogais anteriores e posteriores notamos diferenças quanto à produção oral. Em contexto pretônico, a vogal [i] ocorre com mais recorrência do que a vogal [u]. Em contexto postônico final, a vogal [i] foi mais cancelada do que a vogal [u]. Considerando-se que o cancelamento vocálico é o estágio final de um percurso evolutivo observado hoje (CRISTÓFARO-SILVA; VIEIRA, 2015), pode-se dizer que as vogais anteriores postônicas finais se encontram em estágios mais avançados do percurso evolutivo do que as vogais posteriores postônicas finais. Para ilustrar, elaboramos a Figura 8:

**Figura 8:** Percursos evolutivos das vogais pretônicas e postônicas finais



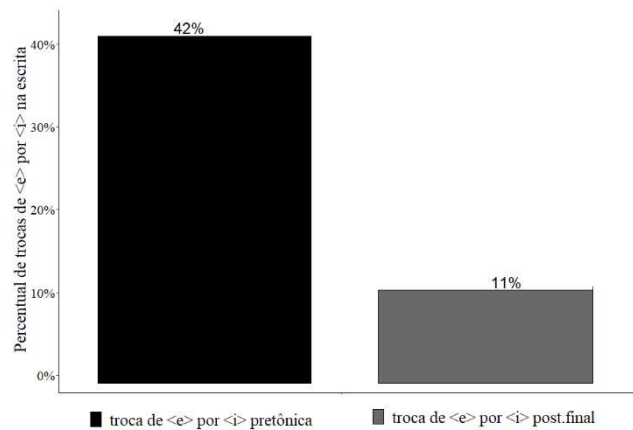
Fonte: elaboração própria.

A Figura 8 ilustra os percursos evolutivos que operam nas vogais pretônicas e postônicas finais. A seta representa a trajetória rumo ao cancelamento da vogal. A ponta esquerda da seta representa o estágio mais antigo da trajetória. O meio da seta representa o estágio atual do percurso evolutivo. A ponta direita da seta representa estágios avançados do percurso, que podem ser observados no hoje. As vogais [e, o] estão presentes no PB desde tempos mais antigos (FONTES, 2014) e, até hoje, ocorrem em contexto pretônico. As vogais [i, u] são padrões atuais e ocorrem tanto em contexto pretônico quanto em contexto postônico final. A vogal cancelada, por sua vez, é um padrão inovador na língua portuguesa. Os retângulos da Figura 8 foram projetados a partir dos dados de fala desta tese. Por isso, o retângulo que representa a vogal [o] é maior do que o que representa a vogal [e] na Figura 8. A Figura 8 ilustra, grosso modo, os diferentes estágios evolutivos os quais se encontram as vogais pretônicas e postônicas finais. As vogais postônicas finais se encontram em estágios mais avançados ( $\emptyset$ ) da trajetória do que as vogais pretônicas na Figura 8. Além disso, a Figura 8 mostra que as vogais [i, u] são recorrentes tanto em contexto pretônico quanto em contexto postônico final. Apoiando-se nos Modelos de Exemplos (BYBEE, 2001), pode-se dizer que as vogais [i, u] são padrões robustos no sistema linguístico atual da língua portuguesa.

De modo geral, os dados de fala relacionados às vogais pretônicas e postônicas finais desafiam a ideia tradicional de que processos fonológicos transformam vogais médias em altas em contexto átono. Se assim o fosse, as vogais médias, sejam elas anteriores ou posteriores, deveriam se manifestar como altas sempre que o contexto estivesse presente. Os dados de fala, contudo, mostram que os fenômenos de alçamento e redução vocálica estão mais avançados nas vogais anteriores do que nas vogais posteriores. Sabendo-se disso, pode-se perguntar: *Podemos os diferentes percursos de evolução das vogais pretônicas e postônicas finais refletirem na escrita infantojuvenil?* Vejamos, na próxima seção, o que nos dizem os dados.

#### 5.1.4 As trocas de <e> por <i> na escrita

Esta seção descreve os dados de escrita referentes às vogais anteriores pretônicas e postônicas finais. Especificamente, descrevemos os erros ortográficos envolvendo a troca da letra <e> pela letra <i> em contextos pretônico e postônico final. Contabilizamos, primeiro, o número total de vezes que os alunos trocaram a letra <e> pela letra <i> nos diferentes contextos acentuais. Vejamos, portanto, o Gráfico 9.

**Gráfico 9:** Trocas de <e> por <i> em contextos pretônico e postônico final

Fonte: elaboração própria.

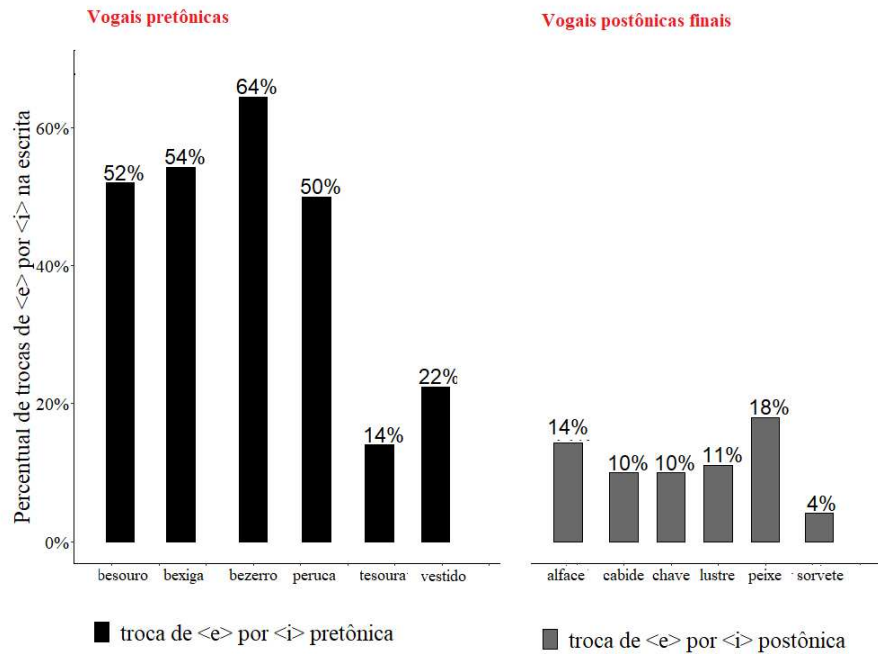
O Gráfico 9 ilustra os índices de erros ortográficos envolvendo a letra <e> pretônica e postônica final. A coluna preta indica o número percentual de vezes em que os alunos trocaram a letra <e> por <i> em contexto pretônico. A coluna cinza indica as trocas de <e> por <i> em contexto postônico final. Os alunos do Ensino Fundamental erraram mais a escrita da vogal <e> pretônica (42% N = 117) do que da vogal <e> postônica final (11% N = 33). A comparação estatística confirmou que as amostras de dados de escrita relacionadas às vogais anteriores pretônicas e postônicas finais apresentam diferenças significativas ( $\chi^2 = 67.921$ ,  $df = 1$ ,  $p\text{-value} < 2.2e-16$ ). Ou seja, os dados indicam que o contexto acentual interfere na grafia da letra <e>.

Esses resultados confrontam a ideia de que as trocas de <e> por <i> em contextos pretônico e postônico final ocorrem por uma única motivação: o alçamento vocálico (ALVARENGA, *et al.*, 1989). Se o fenômeno de alçamento vocálico fosse o único responsável pelos erros ortográficos, estes deveriam ocorrer em índices semelhantes em ambos os contextos acentuais analisados. No entanto, os dados indicam que as trocas de <e> por <i> na escrita ocorrem em diferentes índices a depender do contexto acentual. Assim, é coerente pensar que existem fatores, para além da ocorrência da vogal [i] na fala, que motivam os erros ortográficos. Argumentamos, nesta tese, que os diferentes percursos evolutivos da língua portuguesa podem explicar os diferentes índices de erros ortográficos.

Como foi observado nos dados de fala, os fenômenos de alçamento e redução vocálica evoluem de forma distinta nos contextos pretônico e postônicos final. As diferentes trajetórias evolutivas dos fenômenos fonológicos podem ter refletido na escrita infanto-juvenil. Vejamos o Gráfico 10.



Gráfico 10: Trocas de &lt;e&gt; por &lt;i&gt; por item lexical

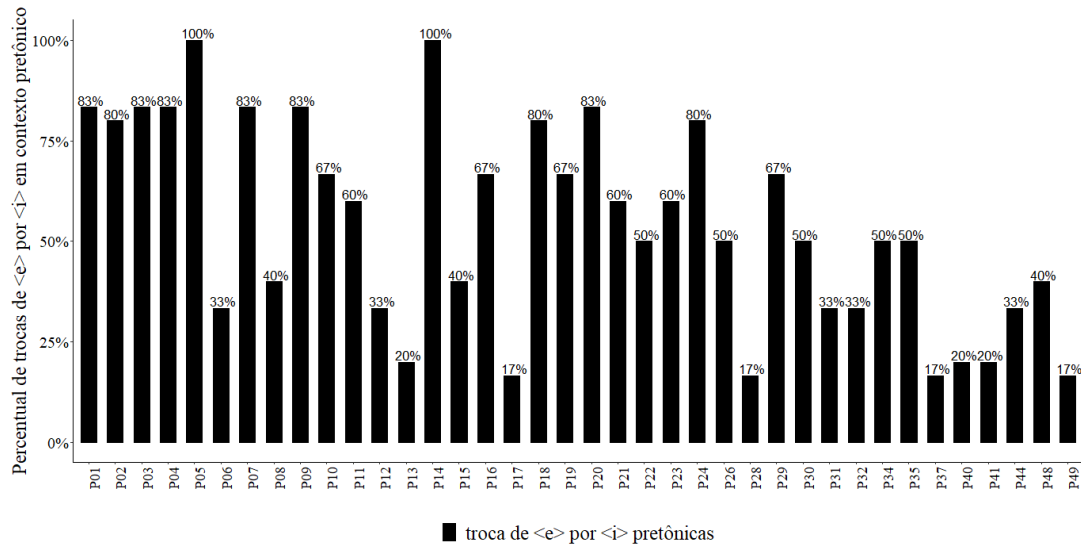


Fonte: elaboração própria.

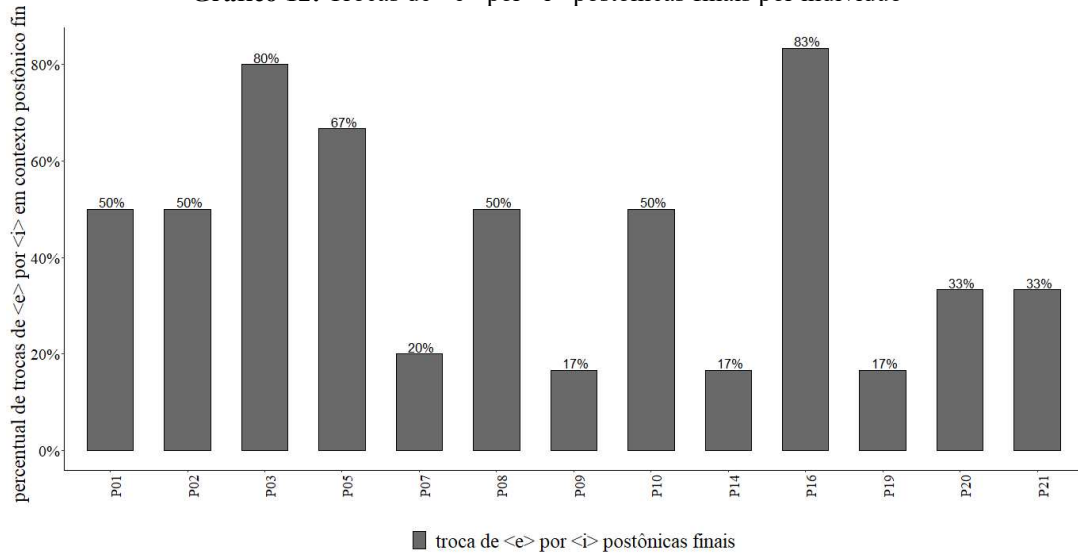
As colunas pretas no Gráfico 10 indicam as taxas percentuais de trocas das letras <e> por <i> em contexto pretônico por item lexical. As colunas cinzas mostram as trocas das letras <e> por <i> em contexto postônico final por item lexical. O primeiro resultado a se destacar é o seguinte: todas as 12 palavras-teste escolhidas para análise das vogais anteriores foram grafadas com algum índice de erro ortográfico. Ou seja, a escrita da letra <e> em contexto pretônico e postônico final causa dúvidas nos aprendizes do EF em diferentes palavras da língua portuguesa.

Outro resultado interessante diz respeito aos diferentes índices de erros ortográficos em cada palavra. Análises tradicionais tendem a propor que as trocas das letras <e> por <i> em cada palavra do Gráfico 10 são motivadas pela oralidade (ALVARENGA, *et al.*, 1989). Porém, há de se perguntar: *por que os alunos cometeram diferentes índices de erros em cada palavra?* Argumentamos, nesta tese, que os diferentes índices de erros ortográficos são resultado dos diferentes percursos de evolução dos fenômenos de alçamento de e redução vocálica na língua portuguesa. Cada palavra segue sua própria trajetória evolutiva na língua oral e, conseqüentemente, há diferentes trajetórias de aprendizado da ortografia em cada item lexical.

Além das trajetórias lexicais, há também percursos individuais envolvidos no percurso de aprendizagem da ortografia. Vejamos:

**Gráfico 11:** Trocas de <e> por <i> pretônicas por indivíduo

Fonte: elaboração própria.

**Gráfico 12:** Trocas de <e> por <i> postônicas finais por indivíduo

Fonte: elaboração própria.

O Gráfico 11 mostra as trocas de <e> por <i>, por indivíduo, em contexto pretônico. O Gráfico 12 mostra as trocas de <e> por <i>, por indivíduo, em contexto postônico final. Em primeiro lugar, vale destacar que 38 participantes (78%) erraram a grafia das vogais pretônicas, mas apenas 13 (26%) erraram as vogais postônicas finais. Esse resultado mostra que a maior parte dos alunos do EF têm mais dificuldade em escrever as vogais pretônicas do que as postônicas finais.

Outro resultado interessante diz respeito ao comportamento de um mesmo indivíduo diante da escrita das vogais. O indivíduo P07 (1º ano), por exemplo, cometeu 83% (N=5) de erros ortográficos em contexto pretônico e apenas 20% (N=1) em contexto postônico final. O indivíduo P29 (3º ano), por sua vez, errou a grafia da letra <e> pretônica em 67% (N=4) dos dados, mas não

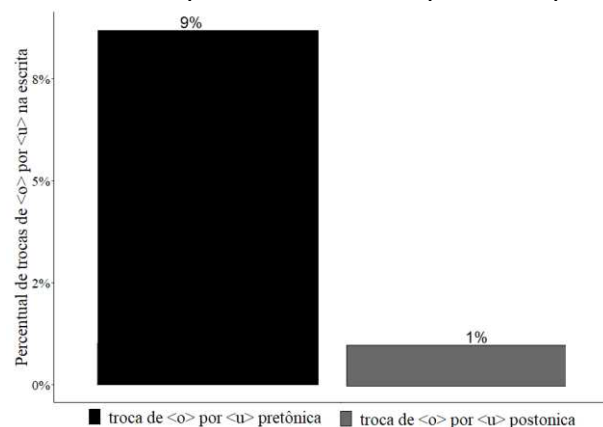
cometeu erros relacionados à letra <e> postônica final. Esses resultados indicam que um mesmo aprendiz apresenta diferentes índices de erros ortográficos a depender do contexto acentual. Considerando-se que cada indivíduo é um sistema dinâmico que evolui de maneira particular, é explicável que haja diferentes trajetórias de aprendizado da ortografia para cada aprendiz.

Além das trocas entre as letras <e> e <i>, os aprendizes de escrita também tendem a trocar a letra <o> pela letra <u> em contextos pretônico e postônico final. Vimos que, na fala, a vogal [o] se manifesta de forma diferente da vogal [e]. Vejamos, a seguir, se essas diferenças se mantêm nos dados de escrita.

### 5.1.5 As trocas de <o> por <u> na escrita

Esta seção descreve os dados de escrita dos alunos do EF referentes às vogais posteriores pretônicas e postônicas finais. Descrevemos, sobretudo, os erros ortográficos envolvendo as letras <o> pretônicas e postônicas finais. Primeiro, mostramos o número total de vezes que os alunos trocaram a letra <o> pela letra <u> nos diferentes contextos acentuais. Vejamos:

**Gráfico 13:** Trocas de <o> por <u> em contextos pretônico e postônico final

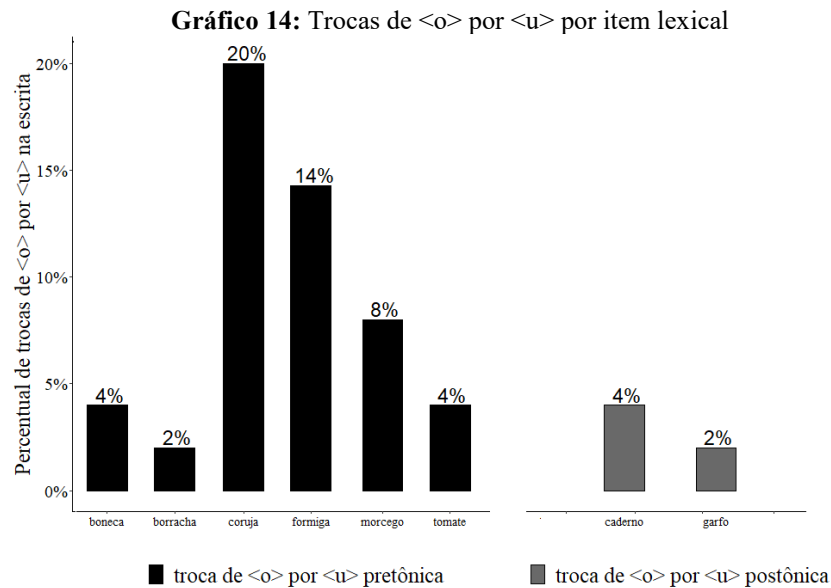


Fonte: elaboração própria.

A coluna preta do Gráfico 13 indica o percentual de trocas de <o> por <u> em contexto pretônico. A coluna cinza indica o percentual de trocas de <o> por <u> em contexto postônico final. Como se pode ver, os alunos cometeram mais erros ortográficos em contexto pretônico (9%, N=26) do que em contexto postônico final (1%, N=3). A comparação estatística confirmou que as amostras de dados de escrita relacionadas às vogais anteriores pretônicas e postônicas finais apresentam diferenças significativas ( $X^2 = 17.464$ ,  $df = 1$ ,  $p\text{-value} = 2.927e-05$ ). Ou seja, os dados indicam que o contexto acentual interfere na grafia da letra <o>.

Mais uma vez, os resultados questionam a ideia de que os erros ortográficos em contextos pretônico e postônico final são consequência de uma mesma motivação. Se assim o fosse, os erros ortográficos deveriam ocorrer em índices semelhantes. As diferenças nos índices de erros nos

levam a pensar que há outros fatores, como as diferentes trajetórias evolutivas da língua, motivando os erros ortográficos ao longo do percurso escolar. Vejamos o Gráfico 14:

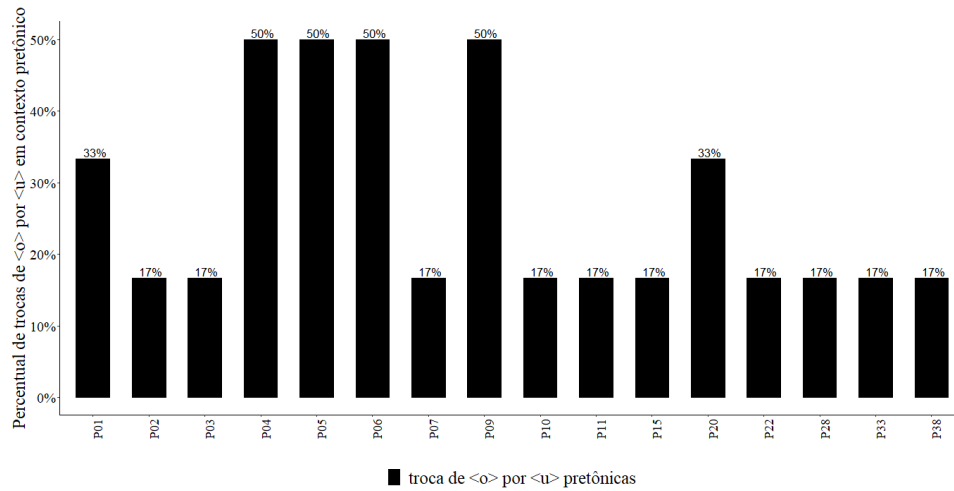


Fonte: elaboração própria.

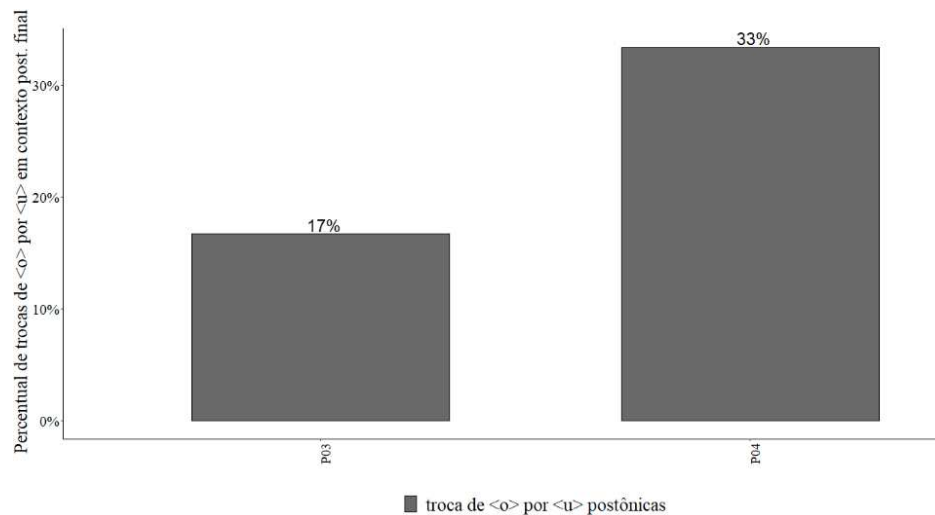
As colunas pretas no Gráfico 14 indicam as trocas de <o> por <u> em contexto pretônico por item lexical. As colunas cinzas mostram as trocas de <o> por <u> em contexto postônico final por item lexical. O primeiro resultado a se destacar é o seguinte: todas as palavras-teste escritas com a letra <o> pretônica foram grafadas com algum índice de erro ortográfico. Por outro lado, apenas duas, das seis palavras terminadas pela letra <o> postônica final, foram grafadas com <u>. Ou seja, as palavras escritas com <o> pretônico causaram mais dúvidas nos aprendizes do EF do que as palavras grafadas com <o> postônico final.

Um resultado merece destaque: cada palavra teve índices diferentes de erros ortográficos envolvendo a troca das letras <o> por <u>. Argumentamos, nesta tese, que as trajetórias evolutivas específicas de cada palavra refletem na escrita infantojuvenil. Nos dados de fala, que cada palavra apresentou um intervalo de variação sonora. Esse fato pode ter motivado os índices de erros ortográficos em cada palavra da língua.

Outro resultado interessante diz respeito às palavras '*caderno*' e '*garfo*'. Apenas essas duas palavras foram escritas com erros ortográficos em contexto postônico final. Análises tradicionais poderiam explicar os erros <\*cadernu> e <\*garfu> como consequência da oralidade. No entanto, tal explicação não consegue acomodar as particularidades lexicais. Argumentamos, nesta tese, que os diferentes percursos evolutivos de cada palavra da língua podem refletir de diferentes maneiras na escrita. Há ainda outro fator, para além da palavra, que pode ter motivado os erros ortográficos em contextos pretônico e postônico final: o indivíduo. Vejamos os Gráficos 15 e 16:

**Gráfico 15:** Trocas de <o> por <u> pretônicas por indivíduo

Fonte: elaboração própria.

**Gráfico 16:** Trocas de <o> por <u> postônicas finais por indivíduo

Fonte: elaboração própria.

O Gráfico 15 mostra as trocas de <o> por <u> em contexto pretônico por indivíduo. O Gráfico 16 mostra as trocas de <o> por <u> em contexto postônico final por indivíduo. Vale destacar, primeiro, que 16 participantes (32%) – dos 50 analisados - erraram a grafia da vogal <o> pretônica, mas apenas 2 (4%) erraram a grafia da vogal <o> postônica final. Esse resultado mostra, mais uma vez, que a maior parte dos alunos do EF comete mais erros em contexto pretônico do que em contexto postônico final.

Outro resultado interessante diz respeito aos índices de erros ortográficos cometidos por cada indivíduo em contexto pretônico. Como se pode ver, 10 participantes cometeram a mesma quantidade de erros ortográficos em contexto pretônico (17%, N=1). A análise pouco acurada deste resultado traz a sensação de que os alunos seguem um mesmo padrão de escrita ortográfica. No entanto, é válido pontuar que os alunos erraram a ortografia de palavras diferentes. O aluno P02

errou somente a palavra ‘*morcego*’ (<\*murcego>). Já o aluno P03 errou somente a palavra ‘*boneca*’ (\*<buneca>). Os alunos P07 e P10 erraram somente a palavra ‘*formiga*’ (<\*furmiga). Os alunos P11, P15, P22, P28, P33, P38 erraram somente a palavra ‘*coruja*’ (<\*curuja). Os dados mostram, portanto, que cada aluno parece enfrentar desafios específicos no processo de aprendizado da ortografia. Ou seja, os indivíduos traçam percursos diferentes para grafar a letra <o> em diferentes contextos acentuais e lexicais.

### 5.1.6 Diferenças entre as letras <e> e <o>

Os alunos do EF erraram mais a escrita das letras <e> do que das letras <o> pretônicas e postônicas finais. Vejamos a Tabela 3.

**Tabela 3:** Erros ortográficos por vogal anterior e posterior

	Troca de <e> por <i>	Troca de <o> por <u>
<b>v. pretônicas</b>	42% (N=117)	8% (N=26)
<b>v. postônicas finais</b>	11% (N=33)	1% (N=3)

**Fonte:** elaboração própria.

Considerando-se que as trocas de <e, o> por <i, u> são reflexos da oralidade, argumentamos que os diferentes índices de erros ortográficos vistos na Tabela 3 são consequência dos diferentes índices de variação sonora na produção de vogais anteriores e posteriores.

Os dados descritos nas seções anteriores evidenciaram que as vogais anteriores estão mais propícias ao alçamento e à redução vocálica (Gráfico 1) do que as vogais posteriores (Gráfico 5) na variedade de Belo Horizonte – MG. Este fato pode explicar os diferentes índices de erros ortográficos. Como os fenômenos de alçamento e redução vocálica operam em diferentes trajetórias evolutivas na língua portuguesa, é explicável que haja diferentes índices de erros ortográficos na escrita. Especificamente, os dados mostraram que as vogais anteriores se encontram em estágios mais avançados dos fenômenos de alçamento e redução vocálica. Esse fato pode ter refletido na escrita infantil e motivado os maiores índices de erros ortográficos na letra <e> se comparada à letra <o>.

### 5.1.7 Hipercorreções

Observamos, nas seções anteriores, que os alunos do EF cometeram erros ortográficos nas palavras grafadas com <e, o> pretônicos e postônicos finais. Nesta seção, veremos que os mesmos alunos também erraram a grafia das letras <i, u> nos mesmos contextos acentuais. Vejamos a Tabela 4:

**Tabela 4:** Hipercorreções em contextos pretônico e postônico final

Vogais pretônicas		
erro ortográfico	N	%
*moleta	8	17%
*tocano	3	6%
Vogais postônicas finais		
erro ortográfico	N	%
*origame	17	38%
*safare	11	24%
*jure	5	12%
*taxe	4	8%

Fonte: elaboração própria.

As palavras ‘muleta’ e ‘tucano’ devem ser escritas com a letra <u> pretônica conforme exige a norma ortográfica da língua portuguesa. Porém, alguns alunos do EF (22%, N=11 alunos) escrevem essas palavras com a letra <o>: <\*moleta> e <\*tocano>. As palavras ‘origami’, ‘safári’, ‘júri’ e ‘táxi’ são grafadas com a letra <i> de acordo com a ortografia do PB. No entanto, alguns aprendizes (76%, N=38 alunos) escreveram a letra <e> e grafaram <\*origame>, <\*safare>, <\*jure> e <\*taxe>.

Os erros ortográficos descritos na Tabela 4 são, tradicionalmente, denominados de hipercorreção. Foi visto, no Capítulo 2, que a literatura descreve as hipercorreções como tentativas de correções excessivas (BORTONE; ALVES, 2014). Pontuamos, no entanto, que a expressão ‘correção excessiva’ só pode fazer sentido em um *continuum* que parte do menos correto até alcançar o excessivamente correto. Em um sistema discreto como o ortográfico, porém, não existem formas mais corretas do que outras. Há apenas duas possibilidades: ou se escreve corretamente, ou se escreve errado. Assim é controverso dizer que determinadas formas de escrever são tentativas de corrigir para além do necessário. Argumentamos, nesta tese, que os aprendizes estão sempre em busca de escrever corretamente. Assim, os erros ortográficos – sejam ou não ‘hipercorreções’ – decorrem, na maioria das vezes, das tentativas criativas dos alunos de adequar as palavras à norma ortográfica.

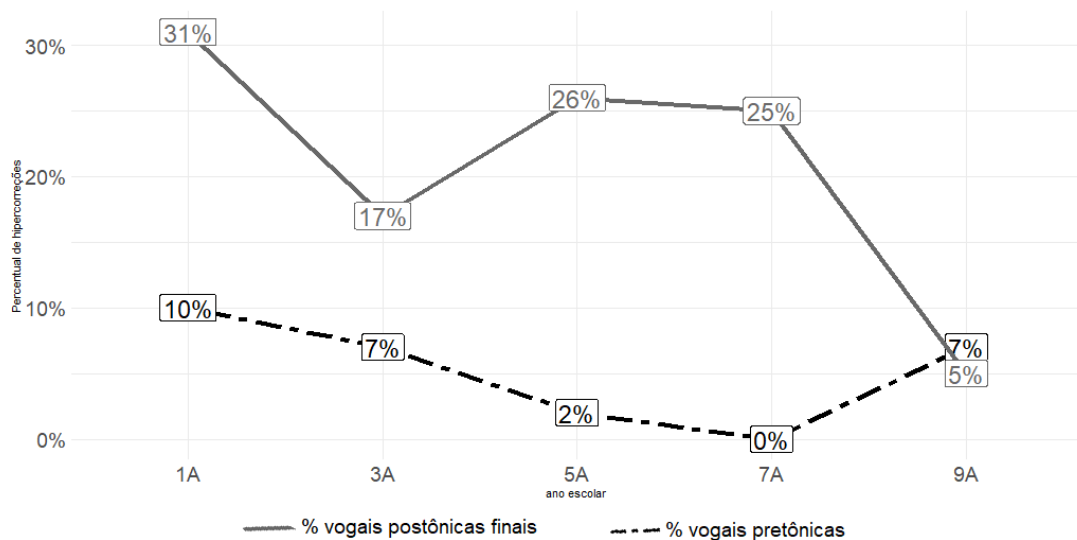
No caso das palavras listadas na Tabela 4, os erros ortográficos podem ser consequência da tentativa de relacionar, na escrita, palavras que têm alguma semelhança na fala. Por exemplo: as palavras ‘muleta’, ‘tucano’, ‘moleque’ e ‘formiga’ podem ser pronunciadas com vogal [u] pretônica na fala: [mu'letə], [tu'kãno], [mu'leki], [fu'fimiɡə]. Uma vez que as palavras têm semelhanças fonéticas, é compreensível que os alunos tentem aproximá-las também na escrita, grafando ambas as palavras com a letra <o>: <\*moleta> e <\*tocano> = <moleque> <formiga>

Seguindo o mesmo raciocínio, os alunos também podem errar a ortografia das vogais postônicas finais. As palavras ‘origami’, ‘safári’, ‘júri’ e ‘táxi’ são pronunciadas com [ɪ] final como as palavras ‘fome’ e ‘time’, por exemplo – [ori'ɡami], [sa'fari], [ʒuri], [takisi], [fõmi],

[ʃĩmɪ]. Já que essas palavras terminam por um som semelhante, é plausível pensar que elas sejam grafadas com a mesma letra: <e>, no caso.

Assim sendo, pode-se dizer que, indiretamente, os erros ortográficos da Tabela 4 são influência do conhecimento oral. Isso porque os alunos associam, na escrita, palavras que possuem certa similaridade na fala. Não se pode dizer, portanto, que o aprendiz faz uma correção excessiva; muito pelo contrário, o aprendiz está, necessariamente, realizando uma complexa reflexão dos padrões linguísticos da língua. Esse tipo de reflexão foi feito por alunos de todos os anos escolares em ambos os contextos acentuais – pretônico e postônico final – e em todos os anos escolares analisados. Consideremos o Gráfico 17:

**Gráfico 17:** Hipercorreções vocálicas por ano escolar



**Fonte:** elaboração própria.

O Gráfico 17 mostra os índices de ‘hipercorreções’ por ano escolar. A linha cinza indica as hipercorreções em contexto postônico final. A linha preta pontilhada ilustra as hipercorreções em contexto pretônico. Uma possível explicação para esses dados diz respeito à frequência dos padrões ortográficos. As letras <e, o> são muito frequentes em contexto postônico final, haja vista que ocorrem na maioria das palavras terminadas por vogais altas átonas. Já as letras <i, u> são pouco frequentes, posto que ocorrem em raras palavras<sup>17</sup>. Considerando-se que um padrão muito produtivo é também muito disponível para ser aplicado a novos itens (BYBEE, 2001), é

<sup>17</sup> Uma pesquisa feita no Portal da Língua Portuguesa identificou 12 palavras proparoxítonas ou paroxítonas escritas com a letra i átona final: *álibi, beribéri, cáqui, dândi, dóri, júri, trípoli, táxi, somali, pêni, maori, macúti*. Não foram encontradas palavras proparoxítonas e paroxítonas terminadas pela letra u. Há apenas empréstimos, como *shih-tzu, jiu-jitsu*.



justificável que os aprendizes escolham as letras <e, o> para escreverem as vogais postônicas finais na maioria das palavras que aprenderem na ortografia, mesmo que isso infrinja a norma ortográfica. Observe que os alunos do 5º ano foram os que mais cometeram hipercorreções. Curiosamente, também foram os alunos do 5º ano que mais acertaram a ortografia das vogais postônicas finais, escrevendo <pató e pentó>. Isso nos mostra que, quando o aluno entende que deve escrever conforme à norma ortográfica ele pode replicar um mesmo padrão ortográfico às palavras da língua que julga estarem relacionadas.

## 5.2. PARTE II: Hipóteses

Esta seção retoma as hipóteses da pesquisa referentes às vogais pretônicas e postônicas finais. São elas:

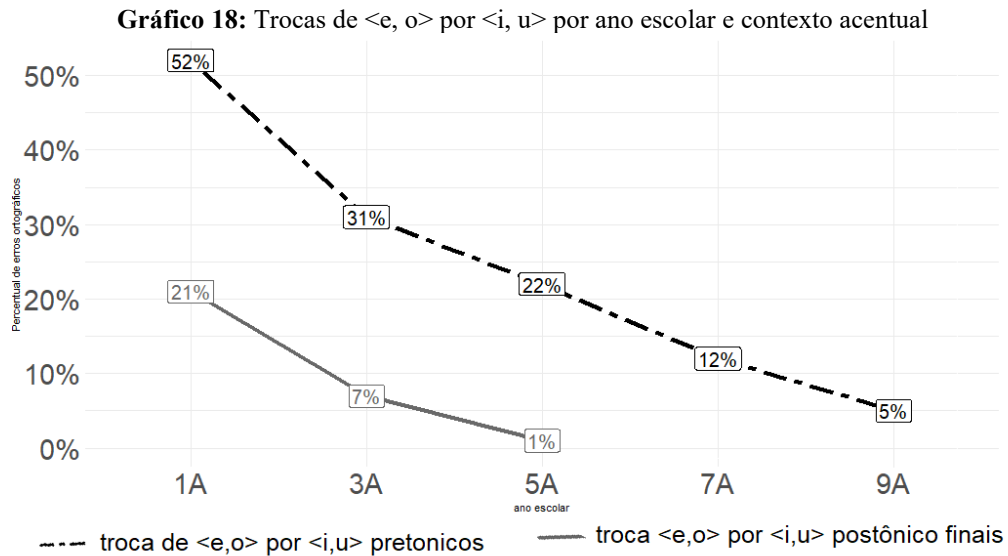
*(1a) as trocas de <e, o> por <i, u> na escrita persistem, em índices significativamente diferentes, até estágios distintos da escolarização, a depender do contexto acentual: pretônico ou postônico final.*

*(2a) O aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônico ou postônica final.*

Para testar a hipótese (1a), compararam-se os índices de erros ortográficos relacionados às vogais pretônicas e postônicas finais em cada ano escolar. Para testar a hipótese (2a), verificaram-se os efeitos do ano escolar e do contexto acentual – pretônico ou postônico final – na produção de vogais médias na fala. As hipóteses foram testadas com auxílio de modelos lineares generalizados mistos. A seguir, apresentam-se os resultados.

### 5.2.1 Hipótese (1a)

A análise descritiva dos dados mostrou que os alunos cometem mais erros ortográficos em contexto pretônico do que em contexto postônico final. Esses resultados dialogam com outros já apresentados na literatura (LEMOS, 2001; MIRANDA, 2010; CRISTÓFARO-SILVA; GUIMARÃES, 2013). Para avançar em relação à bibliografia precedente, esta tese se propôs a testar se os erros ortográficos em contextos pretônico e postônico final persistem, em índices significativamente diferentes, até estágios distintos da escolarização. Vejamos:



Fonte: elaboração própria.

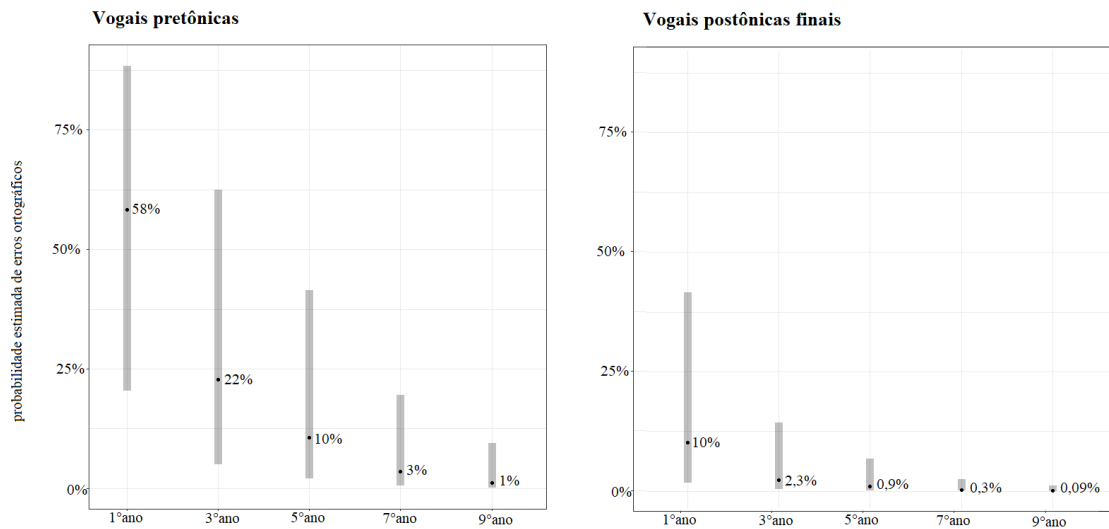
O Gráfico 18 apresenta os índices de erros ortográficos por ano escolar em contextos pretônico e postônico final. A linha preta pontilhada se refere ao contexto pretônico. A linha cinza se refere ao contexto postônico final. Como se pode ver, os índices de erros ortográficos decaem ao longo do tempo tanto em contexto pretônico quanto em contexto postônico final. A diminuição dos índices de erros mostra que os alunos passam a escrever as letras <e, o> corretamente em contextos pretônico e postônico final no decorrer dos anos escolares. Ou seja, os dados indicam que há desenvolvimento do conhecimento ortográfico com o aumento da escolaridade. No entanto, os dados também indicam que a trajetória de desenvolvimento ortográfico é diferente a depender do contexto acentual. Como se pode ver, os erros ortográficos são mais persistentes em contexto pretônico do que em contexto postônico final. Enquanto em posição pretônica os erros perduraram até o 9º ano do EF, em posição postônica final, os erros cessaram no 5º ano. Esses resultados se opõem a ideia de que as trocas de <e, o> por <i, u> decorrem do mesmo motivo, pois se assim o fosse, esses erros deveriam perdurar até estágios semelhantes da escolarização.

Um modelo linear generalizado misto<sup>18</sup> foi ajustado na amostra de dados de escrita referente às vogais pretônicas e postônicas finais, a fim de testar a hipótese (i): as trocas de <e, o> por <i, u> na escrita persistem, em índices significativamente diferentes, até estágios distintos da escolarização, a depender do contexto acentual: pretônico ou postônico final. A variável dependente do modelo foi binária: erro/acerto. O contexto acentual (pretônico/postônica final) e o ano escolar (1º, 3º, 5º, 7º e 9º) foram as duas variáveis independentes de efeito fixo do modelo. O item lexical e o participante foram as duas variáveis aleatórias ajustadas. Uma comparação de

<sup>18</sup> `m01<- glmer (ERROACERTO~ANO+FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA), dados.vogais, family=binomial)`

modelos aninhados mostrou que o melhor modelo ajustado continha o ano e o contexto acentual como variáveis relevantes de efeito fixo ( $\chi^2= 8.7528$ , p-value= 0.003091). Ou seja, o ano escolar explica os índices de erros ortográficos; o contexto acentual também. Além disso, o modelo estimou probabilidades de ocorrência de erros ortográficos em contexto pretônico e postônico final em cada ano escolar. Vejamos o Gráfico 19:

**Gráfico 19:** Probabilidades de erros em contextos pretônico e postônico final



Fonte: elaboração própria.

O Gráfico 19 mostra as probabilidades de erros ortográficos estimadas pelo modelo estatístico e os intervalos de confiança associados à essas probabilidades. À esquerda estão as probabilidades de erros ortográficos em contexto pretônico. À direita, estão as probabilidades de erros ortográficos em contexto postônico final. Lê-se o Gráfico 19 da seguinte maneira: considerando que os dados se mantêm constantes, a probabilidade de que um aluno do 1º ano cometa um erro ortográfico em contexto pretônico é de 58%. Em contexto postônico final, por outro lado, a probabilidade cai para 10% no 1º ano. Além de se considerar a probabilidade estimada, deve-se analisar os intervalos de confiança (sombra cinza no Gráfico 19). Esses intervalos estabelecem um valor mínimo e um valor máximo em que se estima estar a verdadeira probabilidade de erros ortográficos da população<sup>19</sup> (OUSHIRO, 2017). Por exemplo: a probabilidade estimada de erros ortográficos em contexto postônico final no 1º ano do EF é de 10%. Porém, este valor pode variar para cima ou para baixo, dentro do intervalo de confiança, em 95% das vezes que o mesmo experimento for replicado. Sabendo-se disso, pode-se dizer que, em

<sup>19</sup> A população é o “conjunto de elementos com alguma característica em comum, sobre o qual se deseja obter informações. A amostra é um subconjunto desta população” (GODOY; NUNES, p.56 (em fase de elaboração)). No caso desta pesquisa, a população seria o conjunto de dados de escrita das vogais médias e dos róticos de todos os alunos do EF de Belo Horizonte- MG.

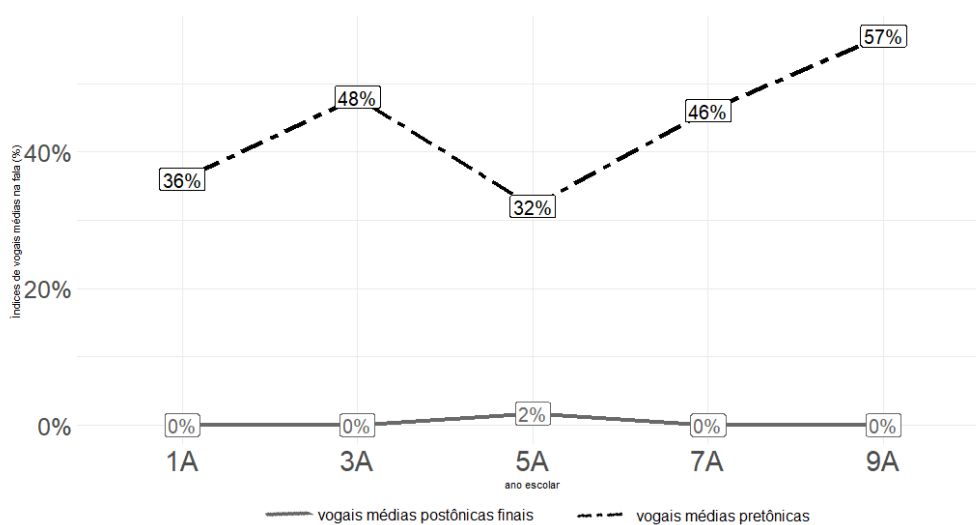
contexto pretônico, há maior variabilidade – e incerteza – da probabilidade de erros ortográficos do que em contexto postônico final. Mais do que isso, o valor máximo dos intervalos de confiança calculados para as vogais pretônicas é maior do que para as vogais postônicas finais em todos os anos escolares. Isso significa que a probabilidade de ocorrer erros ortográficos é maior em contexto pretônico do que em postônico final, mesmo com as incertezas.

Os resultados do modelo estatístico confirmam, portanto, a hipótese (1a): as trocas de <e, o> por <i, u> na escrita persistem, em índices significativamente diferentes, até estágios distintos da escolarização, a depender do contexto acentual: pretônico ou postônico final. Especificamente, os dados indicam que os erros ortográficos em contexto pretônico são mais recorrentes e persistentes do que os erros ortográficos em contexto postônico final.

### 5.2.2 Hipótese (2a)

Vimos, na seção anterior, que, ao longo dos anos escolares, diminuem-se as trocas de <e, o> por <i, u> na escrita infantojuvenil. Em outros termos, os alunos do EF passam a grafar com mais recorrência das letras <e, o> pretônicas e postônicas finais. Sabendo-se disso, essa seção se propõe a avaliar se os alunos do Ensino Fundamental aumentam o uso de vogais médias – [e, o] – na fala na medida em que aumentam o uso das letras <e, o> na escrita. Temos como hipótese que o contexto acentual – pretônico ou postônico final – interfere no aumento de vogais médias na fala ao longo dos anos escolares. Vejamos, então, o Gráfico 20.

**Gráfico 20:** Índices de vogais médias na fala por ano escolar



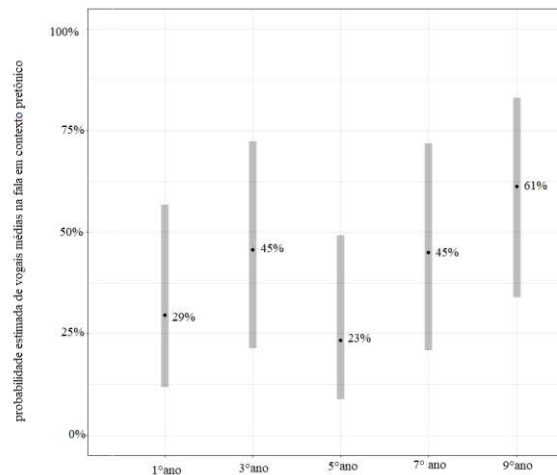
Fonte: elaboração própria.

O Gráfico 20 mostra os índices de [e, o] na fala por ano escolar. A linha preta pontilhada se refere aos índices de [e, o] pretônicas na fala por ano escolar. A linha cinza indica os índices de [e, o] postônicos finais na fala por ano escolar. Observa-se o aumento de vogais médias ao longo

dos anos escolares apenas em contexto pretônico. Em contexto postônico final, as vogais médias foram produzidas apenas no 5º ano e em índices extremamente baixos (2%).

Um modelo linear generalizado misto<sup>20</sup> foi ajustado na amostra de dados de fala referente às vogais pretônicas e postônicas finais, para testar a hipótese (2a): o aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônico ou postônica final. A variável dependente do modelo foi binária: variante diferente da escrita/ variante próxima à escrita. O contexto acentual (pretônico/postônico final) e o ano escolar (1º, 3º, 5º, 7º e 9º) foram as duas variáveis independentes de efeito fixo do modelo. O item lexical e o participante foram as duas variáveis aleatórias ajustadas. Uma comparação de modelos aninhados mostrou que o melhor modelo ajustado continha a interação entre as variáveis *ano* e *contexto acentual* ( $\chi^2=9.8634$ ,  $p\text{-value}=0.04279$ ). Ou seja, a produção de vogais médias na fala é explicada pelo ano escolar em interação com o contexto acentual. Além disso, o modelo estimou probabilidades de ocorrência de erros ortográficos em contexto pretônico e postônico final em cada ano escolar. A probabilidade de ocorrer vogais médias em contexto postônico final foi 0% no 1º, 3º, 7º e 9º anos. Por isso, reportaremos apenas as probabilidades de ocorrência de vogais médias pretônicas por ano escolar. Observemos o Gráfico 21:

**Gráfico 21:** Probabilidade de vogais médias pretônicas na fala por ano escolar



**Fonte:** elaboração própria.

Considerando-se que os dados se mantêm constantes, pode-se dizer que a probabilidade de ocorrer as vogais [e, o] pretônicas na fala oscila a cada ano escolar. Do 1º para o 3º ano, houve aumento da probabilidade de vogais médias na fala: de 29% para 45%. Do 3º para o 5º a probabilidade diminuiu: de 45% para 23%. Do 5º ano em diante a probabilidade voltou a aumentar:

<sup>20</sup> `m02 <- glmer(TIPODEVARIANTE~ANO*FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA), dados.fala.vogais, family=binomial)`

23% < 45% < 61%. Esses resultados indicam, a princípio, que os aprendizes de diferentes anos escolares produzem diferentes índices de vogais [e, o] na fala.

Há de se ressaltar, contudo, que a grande amplitude dos intervalos de confiança não nos permite garantir que a probabilidade de produção de vogais médias difere significativamente em todos os anos escolares. Fizemos uma análise *post-hoc* para comparar par a par os anos escolares. Os resultados dessa análise confirmaram que há diferenças significativas entre as amostras das vogais pretônicas do 5º e do 9º anos (p-value= 0.0390). Ou seja, podemos confirmar que os alunos do 5º ano falam menos vogais médias pretônicas do que os alunos do 9º ano. No caso das vogais postônicas finais, não houve diferenças significativas entre nenhum ano escolar.

De forma geral, os resultados confirmam a hipótese (2a): o aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônica ou postônica final. Especificamente, os dados indicam que o aumento da escolaridade – e conseqüentemente o aumento do contato com palavras grafadas com as letras <e, o> – motiva o aumento de vogais médias na fala em contexto pretônico, mas não em contexto postônico final. Por exemplo: os alunos passam a falar [me'nino] e [bo'neca], mas não falam ['pête] e ['pato].

### 5.3. PARTE III: Discussão teórica - vogais médias

A discussão feita nesta seção é ancorada nos pressupostos dos Modelos de Exemplos (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e da Teoria dos Sistemas Dinâmicos e Complexos (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Com o apoio das teorias e dos dados empíricos previamente descritos, tecem-se argumentos em busca de sustentar a tese de que as letras <e, o> se associam a esquemas múltiplos, que emergem da interação complexa entre palavras, seqüências fonéticas, morfológicas e ortográficas.

Como foi visto na descrição dos dados de fala, as vogais médias pretônicas e postônicas finais estão envolvidas em variação na oralidade. Tal variação já foi documentada por diversos trabalhos da área da fonética e da fonologia (BISOL, 1989; SILVA, 1989; BATTISTI, 1993; SCHWINDT, 1995; VIEGAS, 2001; CHAVES, 2013; MAGALHÃES; CAMPOS, 2022).

No entanto, embora haja um consenso sobre a variabilidade das vogais pretônicas e postônicas finais, há um debate no que diz respeito ao lugar da variação sonora nas teorias da linguagem. De um lado, os modelos tradicionais ((SAUSSURE, 2012 [1916]) postulam que a variação é externa ao sistema linguístico. De outro, abordagens multirrepresentacionais, como os Modelos de Exemplos, propõem que a variabilidade é parte das representações mentais (CRISTÓFARO-SILVA; GOMES, 2004).

Para os modelos tradicionais, apenas fonemas constituem a gramática fonológica; as variantes dos fonemas (fones e alofones) emergem apenas na superfície (BYBEE, 2001). Nesta perspectiva, as variantes das vogais anteriores pretônicas e postônicas finais, [e] ~ [i] ~ ø, são representadas por uma mesma unidade fonológica – /e/. Também as variantes de vogais posteriores pretônicas e postônicas finais – [o] ~ [u] ~ ø – são agrupadas como membros da mesma unidade subjacente: /o/. Assim sendo, a variação presente nos dados não é contabilizada em modelos tradicionais de linguagem, tendo em vista que apenas unidades discretas caracterizam os segmentos vocálicos.

Já nos Modelos de Exemplos, todas as variantes produzidas e percebidas na fala são estocadas na gramática, em categorias probabilísticas que representam diretamente a variação encontrada no uso linguístico (CRISTÓFARO-SILVA; GOMES, 2004). Ou seja, as categorias fonológicas armazenam não só os fonemas, mas também os alofones (BYBEE, 2001). Em outros termos, não há apenas uma unidade representacional, mas sim várias unidades estocadas em categorias múltiplas e dependentes do contexto. Nessa perspectiva, os diferentes intervalos de variação sonora encontrados nos dados de fala (Gráficos 1 a 8) estão estocados na gramática em categorias múltiplas e probabilisticamente organizadas.

Com apoio da premissa de multiplicidade das representações linguísticas, é possível argumentar que os aprendizes de escrita acessam categorias fonéticas múltiplas ao longo do aprendizado da escrita e, por isso, cometem diferentes índices de erros ortográficos ao grafarem palavras diferentes. Mais do que isso, argumentamos que as letras do sistema alfabético são associadas a categorias lexicais detalhadas. Ou seja, as letras se relacionam não apenas a um único fonema. Pelo contrário, uma mesma letra pode se associar a diversas categorias fonéticas. Assim sendo, o aprendizado das letras <e, o> em contexto pretônico pode envolver o gerenciamento de esquemas linguísticos distintos daqueles envolvidos no aprendizado das mesmas letras em contexto postônico final, o que explica os diferentes percursos de aprendizado da ortografia.

Trabalhos como o de Alvarenga *et al.* (1989) descrevem o alçamento vocálico como o responsável pelas trocas de <e, o> por <i, u> na escrita infantil. Nessa perspectiva, os erros ortográficos em contextos pretônico e postônico final teriam uma única motivação. Contudo, os dados desta tese desafiam essa ideia. Os aprendizes cometeram mais erros ortográficos nas letras <e> do que nas letras <o>. *Se os erros ortográficos são todos decorrentes do alçamento vocálico, por que as crianças têm mais dificuldade em grafar a letra <e> do que a letra <o>?* Uma possível resposta para essa pergunta está relacionada à não linearidade de evolução do sistema linguístico. De acordo com a TSDC, a língua é um sistema dinâmico que evolui de forma não linear. Isso significa que um mesmo fenômeno pode causar efeitos mais ou menos evidentes no sistema linguístico (BOT, LOWIE; VESPOR, 2007). Observamos, na análise descritiva dos dados de fala,

que os fenômenos de alçamento e redução vocálica se encontram em estágios mais avançados nas vogais postônicas finais do que nas vogais pretônicas. Esse fato expressa a não linearidade da evolução do sistema. Um mesmo fenômeno fonológico opera em diferentes trajetórias evolutivas na língua, a depender de vários fatores, como o contexto acentual (pretônico x postônico final), o tipo da vogal (anterior x posterior), a palavra e o indivíduo. Assim sendo, é totalmente possível que este fenômeno reflita na escrita de maneira distinta, causando diferentes índices de erros ortográficos.

Outro resultado que merece destaque diz respeito à persistência dos erros ortográficos. Os dados mostraram que os erros ortográficos em contexto pretônico persistem por mais anos escolares do que os erros ortográficos em contexto postônico final (ver Gráfico 18). Este fato também confronta a ideia de que há um mesmo motivo causando as trocas de <e, o> por <i, u> em diferentes contextos acentuais. Se o alçamento vocálico fosse o único motivador dos erros ortográficos, estes deveriam cessar assim que os aprendizes compreendessem tal motivação. No entanto, os alunos do EF tendem a aprender a ortografia das vogais postônicas finais mais cedo do que aprendem a ortografia das vogais pretônicas. Nos dados previamente descritos, os alunos do EF pararam de errar a escrita das vogais postônicas finais no 5º ano, mas continuaram errando as vogais pretônicas até o 9º ano. A pergunta que nos resta responder é: *por que as trocas de <e, o> por <i, u> na escrita são mais persistentes em contexto pretônico do que em contexto postônico final?*

Em primeiro lugar é válido pontuar a maior heterogeneidade contextual das vogais pretônicas em relação às vogais postônicas finais. As vogais pretônicas podem ocorrer em início e/ou meio da palavra, em sílabas próximas ou distantes da tônica, em mais de uma sílaba da mesma palavra, próximas de vogais e/ou consoantes (CRISTÓFARO-SILVA; GUIMARÃES, 2013). Por exemplo: ‘leão’, ‘inseguro’, ‘telefonema’. Por outro lado, as vogais médias postônicas finais ocorrem, necessariamente, em posição final de palavras. Exemplo: ‘dente’, ‘pato’. Considerando-se que as letras são associadas a esquemas múltiplos, podemos dizer que as letras <e, o> pretônicas se associam a esquemas mais complexos do que as letras <e, o> postônicas finais. Isso pode contribuir para a persistência dos erros ortográficos em contexto pretônico.

Na escrita, as vogais pretônicas também são mais complexas do que as vogais postônicas finais. As vogais postônicas finais [i] e [u] são representadas regularmente, pelas letras <e> e <o>, exceto em um conjunto restrito de palavras<sup>21</sup>. Já as vogais pretônicas têm o padrão ortográfico

---

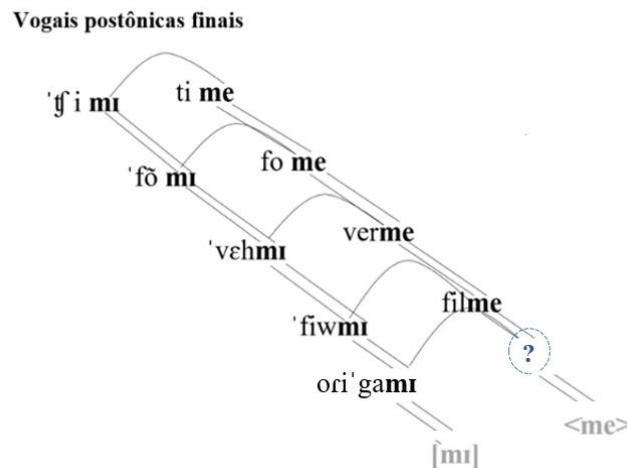
<sup>21</sup> Uma pesquisa feita no Portal da Língua Portuguesa identificou 12 palavras proparoxítonas ou paroxítonas escritas com a letra i átona final: *álibi, beribéri, cáqui, dândi, dóri, júri, tripoli, táxi, somali, pêni, maori, macúti*. Não foram encontradas palavras proparoxítonas e paroxítonas terminadas pela letra u. Há apenas empréstimos, como *shih-tzu, jiu-jitsu*.



lexicalmente determinado. A vogal [i] pretônica pode ser representada ortograficamente pelas letras <e> e <i>. Por exemplo: ‘perigo’ e ‘pirata’. A vogal [u] pretônica pode ser representada pelas letras <o> e <u>: ‘boneca’ e ‘buraco’. As variadas letras que podem ocorrer em contexto pretônico – <e, i, o, u> – podem contribuir para a persistência dos erros ortográficos nessa posição acentual.

Na perspectiva dos Modelos de Exemplos, as abstrações como morfemas, rimas, sílabas, sons emergem das relações entre itens lexicais (BYBEE, 2001). Ancorando-se neste pressuposto, sugerimos que abstrações relacionadas ao sistema de escrita, como padrões ortográficos, também podem surgir das relações do léxico. Consideremos a Figura 9.

**Figura 9:** Relações entre oralidade e escrita no contexto postônico final



Fonte: elaboração própria.

O diagrama da Figura 9 representa relações entre a sequência fonética [mi] e a sequência ortográfica <me> em contexto postônico final. Como se pode ver, todos os itens lexicais se associam ao mesmo esquema fonético-fonológico – [mi]. Além disso, quase todos os itens – com exceção de ‘origami’ – se associam ao mesmo esquema ortográfico – <me>. Isso porque a grande maioria das palavras faladas com [i] são grafadas com <e>. Assim sendo, pode-se dizer que o esquema fono-ortográfico relacionado às vogais postônicas finais tem alto grau de generalização. Isto é, a regularidade das vogais postônicas finais aumenta a probabilidade de padrões como [mi] e <me> se expandirem a novas palavras.

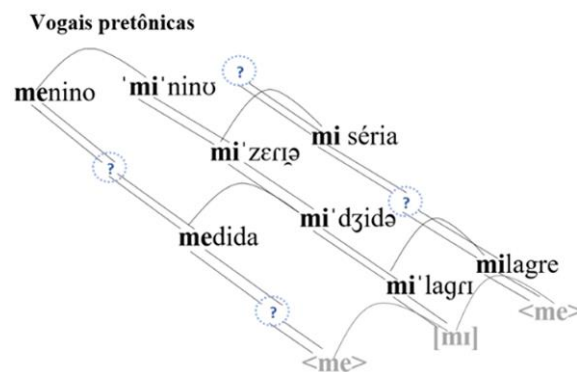
O esquema na Figura 9 explica ainda erros ortográficos chamados de ‘hipercorreções’. Observe o ponto de interrogação no esquema (I). Esse ponto indica uma lacuna presente no esquema fono-ortográfico das vogais postônicas finais. Essa lacuna pode ser preenchida pela

forma ortográfica <\*origame>. Tal forma não é prevista pela ortografia do PB, mas é completamente possível na língua portuguesa. Ou seja, as chamadas hipercorreções expressam as tentativas dos aprendizes de preencher as lacunas existentes no uso linguístico e nos esquemas, em busca de criar associações gráficas entre palavras que constituem o mesmo esquema oral.

Para que fique compreensível, imaginemos a seguinte situação hipotética: um aluno do EF deve escrever a seguinte pseudo-palavra ditada oralmente pela professora: [pa'ta.mi]. Mesmo sem nunca ter visto esta palavra escrita, a criança pode acessar o esquema da Figura 9 para escrevê-la. Feito isto, a criança, certamente, escreverá <patame>, haja vista que, conforme representado na Figura 9, palavras associadas ao padrão fonológico [mi] postônico final tendem a se associar também ao padrão ortográfico <me>.

No caso das vogais pretônicas, por outro lado, os aprendizes de escrita precisam gerenciar um outro esquema fono-ortográfico, um pouco mais complexo do que o esquema das vogais postônicas finais. Vejamos o diagrama na Figura 10:

**Figura 10:** Relações entre oralidade e escrita no contexto pretônico



Fonte: elaboração própria.

O diagrama da Figura 10 representa relações entre a sequência fonética [mi] e as sequências ortográficas <me, mi> em contexto pretônico. A variação sonora atrelada à heterogeneidade dos padrões ortográficos tende a diminuir o grau de generalização dos esquemas relacionados às vogais pretônicas. No diagrama da Figura 10 os itens lexicais estão associados ao mesmo esquema fonológico – [mi] – mas a diferentes esquemas ortográficos – <mi, me>. As palavras 'menino' e 'medida' estão associadas à sequência fonética [mi] e à sequência ortográfica <me>. Já as palavras 'miséria' e 'milagre' estão associadas às sequências [mi] e <mi>. O fato de existir mais de um esquema ortográfico associado à mesma sequência sonora pode explicar a persistência dos erros ortográficos em contexto pretônico na escrita infantojuvenil.

Consideremos outra situação hipotética: em um segundo momento, a professora dita oralmente para o aluno do EFI a seguinte pseudo-palavra: [mivi'datə]. A criança, que nunca viu

tal palavra escrita, poderá ter dúvidas quanto à ortografia, uma vez que a palavra nova pode ser associada a esquemas diferentes: <mivedata> <mividata>, <mevedata> ou <mevidata>, todos possíveis na língua portuguesa.

Os pontos de interrogação, na Figura 10, representam lacunas existentes nos esquemas linguísticos. Essas lacunas podem ser preenchidas pelas seguintes formas ortográficas: <\*minino, \*meséria, \*midida, \*melagre>. Ou seja, os erros ortográficos em contexto pretônico (até os chamados de hipercorreções) podem ser explicados como resultado da tentativa dos aprendizes de preencherem as lacunas existentes no uso da língua e, conseqüentemente, nos esquemas da gramática.

Vale ressaltar que os diagramas das Figuras 9 e 10 expressam apenas algumas generalizações fono-ortográficas possíveis na língua. Muitas outras podem emergir da associação entre muitos elementos linguísticos e não linguísticos. O item lexical '*menino*', por exemplo, pode se associar à '*menina*', a '*meninho*', etc. Mesmo no caso das vogais postônicas finais, em que há maior regularidade, há múltiplos exemplares armazenados na categoria. O item lexical '*filme*', por exemplo, pode estar associado morfológicamente ao diminutivo '*filminho*', semanticamente a '*cinema*', etc.

Substantivos masculinos também estão associados entre si, dando origem a generalizações de gênero na categoria das vogais postônicas finais. A generalização de gênero morfológico pode, inclusive, ter sido um elemento auxiliar no aprendizado das vogais posteriores postônicas finais. Mesmo não sabendo a ortografia das palavras, os aprendizes têm grandes chances de acertá-las por via da morfologia. Como a maioria das palavras no masculino são associadas ao esquema ortográfico <o>, os aprendizes podem prever que esta letra é a melhor opção para ocorrer em final de palavras masculinas. Isso pode explicar os baixos índices de erros ortográficos envolvendo a letra <o> postônica final encontrados nos dados.

De forma geral, as Figuras 9 e 10 visaram ilustrar a proposta de que as vogais pretônicas e postônicas finais podem estar associadas a diferentes – e múltiplos – esquemas linguísticos probabilísticos. Queremos defender, em resumo, que, para grafarem as letras <e, o> em contexto pretônico, os aprendizes gerenciam esquemas múltiplos e diferentes daqueles gerenciados ao escreverem as mesmas letras em contexto postônico final. Esta tese é fundamentada nos Modelos de Exemplares e na TSDC (JOHNSON, 1997; BYBEE, 2001; PIERREHUMBERT, 2001; THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009) que postulam a multiplicidade e a complexidade do sistema linguístico.

Vale ressaltar ainda que, durante o aprendizado da escrita, os esquemas múltiplos podem ser reorganizados para que o conhecimento ortográfico se acomode no sistema linguístico. Partindo da premissa de que o sistema linguístico é aberto às influências de outros sistemas

(BECKNER *et al.*, 2009), argumentamos que o sistema de escrita pode influenciar nas representações linguísticas, fortalecendo algumas categorias. Isto é, as formas ortográficas recém aprendidas podem funcionar como atratores, moldando as representações linguísticas por um tempo e causando a adaptação do sistema linguístico. Por exemplo: as formas ortográficas <menino> e <tomate> podem fortalecer a representação das formas [me'nino] e [to'mate].

A reorganização da linguagem pode ser evidenciada pelas alterações na fala ao longo dos anos escolares. Os dados de fala evidenciaram que os alunos do 9º ano produziram mais vogais médias pretônicas na fala do que os alunos do 5º ano do EF. No caso das vogais postônicas finais, os dados mostraram que não houve diferenças significativas quanto a produção oral dos alunos do EF. Ou seja, os dados evidenciaram que, ao longo da escolarização, os alunos alteram apenas a pronúncia das vogais pretônicas. Sabendo-se disso, pode-se perguntar: *por que os aprendizes aumentam o uso de vogais médias pretônicas, mas não fazem o mesmo em contexto postônico final?* Com o apoio dos Modelos de Exemplares e da TSDC, argumentamos que o aprendizado da escrita causa diferentes efeitos na organização dos esquemas linguísticos e, conseqüentemente, há diferentes reflexos na oralidade.

A maior ou menor robustez dos exemplares pode influenciar na força de atração da escrita e, conseqüentemente, no aumento do uso de determinadas variantes fonéticas. Por exemplo: as vogais médias postônicas finais [e, o] têm representação muito menos robusta do que as vogais altas [ɪ, ʊ]. Isso porque, conforme evidenciam os dados, as vogais médias postônicas finais ocorrem em menor frequência do que as vogais altas na fala infantil. O aprendizado das formas ortográficas pode até fortalecer a representação das vogais médias postônicas finais. Por exemplo: a forma ortográfica <pente> pode motivar o fortalecimento da representação fonológica ['pête]. Porém, as vogais médias são tão raras em posição postônica final que, dificilmente, as formas ortográficas terão força suficiente para moldar os esquemas a ponto de [e, o] se tornarem tão robustas quanto [ɪ, ʊ]. Em outras palavras, as letras <e, o> podem até fortalecer a representação de [e, o], mas não de forma suficiente para que os índices de vogais médias átonas se igualem ou ultrapassem os índices de vogais altas átonas na fala dos aprendizes.

Já as vogais médias pretônicas [e, o] são muito recorrentes na fala. Assim sendo, as letras <e, o> podem promover ainda mais a produção oral dessas vogais. Isto é, a representação de palavras como [me'ninʊ] e [bo'nekə] pode se fortalecer quando associada às formas ortográficas que as representam: <menino> e <boneca>. Com o fortalecimento da representação pode haver aumento dessas variantes no uso linguístico. Em suma, queremos argumentar que as letras <e, o> fortalecem a representação e o uso de padrões sonoros que são produtivos na língua: como [e, o] pretônicos. Em caso de padrões não produtivos – como [e, o] postônico final – a escrita não tem força suficiente para alterar as representações e o uso linguístico.

De forma geral, esta seção buscou argumentar que a fala e a escrita podem estabelecer múltiplas relações. Em vez de se limitar à ideia unirepresentacional de que letras representam fonemas, propomos que uma mesma letra – <e, o> – está relacionada a esquemas linguísticos múltiplos e probabilísticos, os quais emergem da interação complexa entre palavras, sequências fonéticas, morfológicas e ortográficas.

#### 5.4. PARTE IV: A visão dos professores sobre as trocas de <e, o, i, u>

Nesta seção, analisam-se os questionários respondidos pelos professores do EF. O objetivo desta análise é avaliar a percepção docente sobre os erros ortográficos relacionados às vogais pretônicas e postônicas finais. As repostas de três perguntas serão avaliadas:

- 1) Atualmente, você leciona em qual ano escolar?
- 2) Você leciona há quantos anos?
- 3) Marque os erros ortográficos que você, professor, já observou nos textos dos seus alunos.

A terceira pergunta do questionário solicita que os professores marquem os erros ortográficos já observados nos textos dos seus alunos. Uma lista de erros foi disponibilizada para os docentes. Nesta lista, estavam as palavras-teste da pesquisa (ver Apêndice C). Cinco professores responderam às perguntas. Vejamos as respostas:

**Quadro 18:** Respostas dos professores – vogais médias

Prof.	Atualmente, você leciona em qual ano escolar?	Você leciona há quantos anos?	Marque os erros ortográficos que você já observou nos textos dos seus alunos	
			Contexto pretônico	Contexto postônico final
1	1° as 5° anos	10 anos	*bisouro, *burracha,	*laçu, *peixi
2	1° ano	20 anos	*bixiga, *bisouro, *bizerro, *tumate, *vistido	*cavalu, *chavi, *gatu, *peixi, *sorveti, *sucu,
3	4° e 5° anos	19 anos	*bisouro, *burracha, *curuja, furniga, *murcego, *piruca, *tumate, *vistido	*cavalu, *chavi, *gatu, *garfu, *origame, *sorveti,
4	6° ao 9° anos	32 anos	*bixiga, *bisouro, *bizerro, *moleta.	*bequini
5	9° ano	11 anos	*tumate	*cavalu, *chavi

**Fonte:** elaboração própria.

Dos cinco professores participantes da pesquisa, três atuam no primeiro ciclo e dois no segundo ciclo do EF. Pode-se perceber, na quarta coluna do Quadro 18, que todos os professores observam erros ortográficos em contextos pretônico e postônico final. Ou seja, os professores

confirmam que os alunos erram a ortografia das letras <e, o> pretônicas e postônicas finais ao longo do EF.

Contabilizamos os erros ortográficos marcados pelos professores. Os resultados da contagem mostram que as observações dos professores estão alinhadas com a análise quantitativa feita nesta pesquisa. Consideremos a Tabela 5:

**Tabela 5:** Quantidade de erros vocálicos observados pelos professores

	<b>Pretônica</b>	<b>Postônica fin.</b>	<b>Total</b>
N de erros marcados pelos profs. do 1º ciclo do EF	18	12	30
N de erros marcados pelos profs. o 2º ciclo do EF	6	2	8
Total	24	14	38

**Fonte:** elaboração própria.

Em primeiro lugar, é válido pontuar que os professores do 1º ciclo marcaram mais tipos de erros ortográficos do que os professores do 2º ciclo. Ou seja, os dados dos professores trazem evidências de que os alunos do 1º ao 5º anos cometem mais erros ortográficos em contextos pretônico e postônico final do que alunos do 6º ao 9º ano. Os dados de escrita dos alunos do EF mostraram que há mais trocas de <e, o> por <i, u> nos anos iniciais do que nos anos finais do EF (ver Gráfico 18). As observações dos docentes estão, portanto, de acordo com os dados de escrita avaliados nesta pesquisa.

Outro ponto que merece destaque na Tabela 5 é o número de erros ortográficos em cada categoria linguística avaliada. Em conjunto, os professores marcaram mais erros ortográficos relacionados às vogais pretônicas (24 tipos de erros) do que às vogais postônicas finais (14 tipos de erros). Curiosamente, esse resultado confirma o que vimos nos dados de escrita dos alunos do EF. Os erros ortográficos em contexto pretônico são mais recorrentes e mais persistentes ao longo do EF do que os erros em contexto postônico final. Mais uma vez, a percepção dos professores concorda com os dados de escrita avaliados nesta pesquisa.

De maneira geral, as observações dos professores seguem a mesma tendência encontrada nos dados de escrita avaliados nesta pesquisa. Os professores estão atentos aos erros ortográficos relacionados às vogais pretônicas e às vogais postônicas finais. Mais do que isso, as observações dos professores indicam que há diferenças na recorrência dos erros ortográficos. Tanto na análise quantitativa dos dados, quanto na percepção dos professores, os erros mais comuns nos textos dos alunos do EF são relacionados às vogais pretônicas, em comparação com as vogais postônicas finais.

Análises teóricas sobre a recorrência e a persistência de erros ortográficos – como a proposta neste capítulo – podem contribuir com a compreensão sobre os percursos de aprendizado

da ortografia e, mais do que isso, podem auxiliar os professores a se conscientizarem sobre aquilo que eles já observam na escrita de seus alunos.

### 5.5 Resumo do capítulo

Este capítulo descreveu e analisou dados de fala e de escrita referentes às vogais médias e altas pretônicas e postônicas finais. Na primeira parte do capítulo, foi feita a análise descritiva dos dados. Especificamente, avaliou-se a variação sonora em contextos pretônicos e postônicos finais e os índices de erros ortográficos envolvendo as trocas de <e, o> por <i, u>. De modo geral, os dados de fala indicaram que as vogais variam de forma distinta a depender do contexto acentual – pretônico ou postônico final. Os dados de escrita, por sua vez, indicaram que os alunos do EF cometem mais erros ortográficos em contexto pretônico do que em contexto postônico final.

Na segunda parte do capítulo, as hipóteses (1a) e (2a) foram testadas e confirmadas. No que se refere à hipótese (1a), os dados mostraram que a recorrência e a persistência das trocas de <e, o> por <i, u> dependem do contexto acentual. Os erros ortográficos em contexto pretônico persistiram por mais tempo e em índices maiores do que os erros ortográficos em contexto postônico final. No que se refere à hipótese (2a), os dados mostraram os alunos do 9º ano aumentam a pronúncia de vogais médias pretônicas. Por outro lado, não houve aumento de vogais médias postônicas finais na fala ao longo dos anos escolares. Em suma, os dados indicam que as alterações na fala ao longo da escolarização dependem do contexto acentual.

Na terceira parte do capítulo, os dados foram discutidos à luz dos Modelos de Exemplos e da TSDC. De forma geral, argumentamos que os diferentes percursos de aprendizado das letras <e, o> pretônicas e postônicas finais expressam o gerenciamento de esquemas múltiplos ao longo da escolarização. Além disso, propomos que o aprendizado das letras <e, o> causa a reorganização dos esquemas linguísticos. No entanto, como esses esquemas são probabilísticos, a escrita exerce diferentes forças na representação e, conseqüentemente, no uso linguístico.

Por fim, a parte IV do capítulo descreveu a visão dos professores sobre os erros ortográficos relacionados às vogais pretônicas e postônicas finais. Os dados indicaram que os professores estão sensíveis aos diferentes índices de erros ortográficos em contextos pretônico e postônico final. Ou seja, os professores confirmam os resultados apresentados nesta tese.

## 6. ANÁLISE DOS DADOS: RÓTICOS

Os verbos e os nomes terminados em róticos são descritos como padrões similares na fala e na escrita. Na fala, as palavras de ambas as classes gramaticais estão envolvidas no fenômeno fonológico de cancelamento do R-final. Por exemplo: [kã'tə] e [selu'la]. Na escrita, as formas verbais e nominais devem terminar com a letra <r>, mas aprendizes tendem a omitir, equivocadamente, o <r> final. Exemplo: <\*cantá>, <\*celulá>.

Embora haja similaridades entre as categorias, há também particularidades no uso dos róticos finais. O cancelamento do R-final opera sistematicamente em verbos (OLIVEIRA, 1981; 1997) e variavelmente em nomes (HUBACK, 2003). Já a omissão da letra <r> na escrita é mais recorrente em verbos do que em nomes (COSTA, 2009; TORRES; OLIVEIRA, 2015; FREIRE; HORA, 2019).

Considerando-se as semelhanças e as diferenças entre os verbos e os nomes terminados por R-final, esta tese se propôs a compará-los visando verificar se essas categorias se manifestam de forma análoga ou diferente na fala e na escrita. Para isso, serão analisados, ao longo desta seção, os dados de fala e de escrita relacionados aos róticos finais de verbos e de nomes.

A organização do capítulo é semelhante à do capítulo anterior. Primeiro, os dados de fala e de escrita são descritos. Em segundo lugar, as hipóteses são testadas. Na Parte III, os dados são discutidos à luz dos Modelos de Exemplares (JOHNSON, 1997; BYBEE, 2001; PIERREHUMBERT, 2001) e da TSDC (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Por fim, na última parte, apresentam-se as observações dos professores sobre os erros ortográficos envolvendo o <r> em final de verbos e de nomes.

### 6.1 PARTE I: Análise descritiva dos dados

A primeira parte da análise descreve as amostras de dados relacionados aos róticos finais de verbos e de nomes. A descrição está subdividida da seguinte maneira: (i) róticos finais na fala; (ii) róticos finais na escrita.

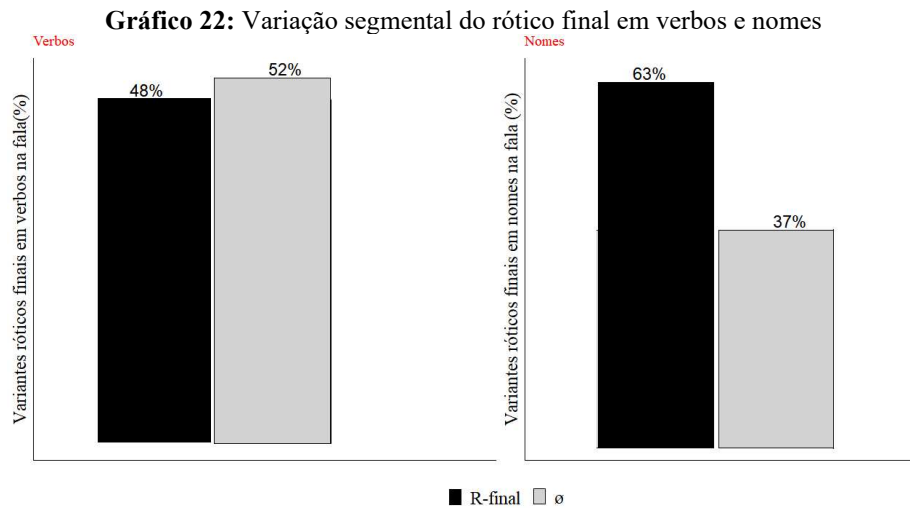
#### 6.1.1 Róticos finais na fala

A produção oral dos róticos finais será descrita em três etapas subsequentes: primeiro, apresentam-se os índices gerais de variantes fonéticas em verbos e em nomes. Em seguida, descrevem-se os dados de fala por item lexical. Por fim, mostra-se como cada indivíduo pronunciou o R-final em verbos e em nomes.

O R-final pode ser realizado com inúmeras variantes fonéticas, como vibrantes, tepes, fricativas, aproximantes e aproximantes aspiradas, etc. (RENNICKE, 2015). A análise de todas as



variantes, contudo, ultrapassa o escopo deste trabalho. As variantes fonéticas relacionadas aos róticos serão agrupadas em duas categorias: (1) R-final; (2) cancelamento. Vejamos o Gráfico 22:

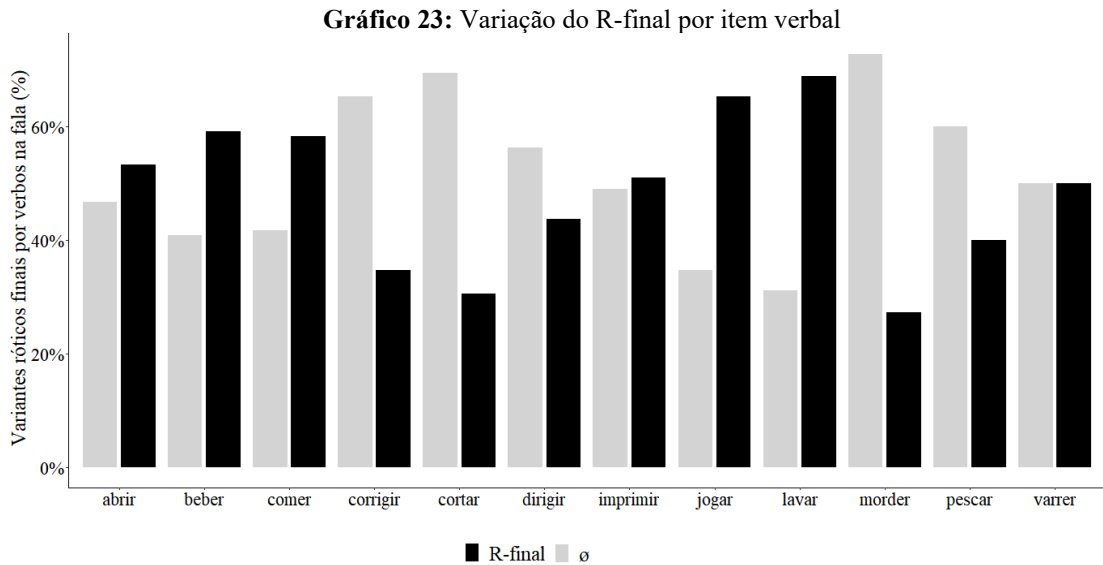


Fonte: elaboração própria.

O Gráfico 22 ilustra a variação sonora do rótico final em verbos e em nomes. À esquerda do Gráfico, encontram-se as colunas referentes aos verbos. À direita, as colunas referentes aos nomes. As colunas pretas indicam os índices de produção do R-final. As colunas cinza indicam os índices de cancelamento dos róticos em final de palavras. O cancelamento do R-final foi mais frequente em verbos (52%, N=281) do que em nomes (37%, N= 202). A análise estatística confirmou que as amostras de dados de fala referentes aos verbos e aos nomes são significativamente diferentes ( $X^2 = 21.663$ ,  $df = 1$ ,  $p\text{-value} = 3.25e-06$ ).

Esse resultado está em consonância com outros já reportados na literatura (OLIVEIRA, 1981; 1997; HUBACK, 2003). Oliveira (1981) propôs que os casos de cancelamento do R-final em nomes e em verbos fossem tratados separadamente, haja vista que, nos nomes, o R-final é variavelmente cancelado, enquanto que, nos verbos, o cancelamento ocorre em índices mais altos do que a produção do R-final. O autor sugere que “o cancelamento de (r), [...], é um processo fonético que se iniciou em posição final de palavra, preferencialmente em verbos, passando daí a nomes” (OLIVEIRA, 1981, p. 27). Dialogando com Oliveira (1981), os dados reportados no Gráfico 22 evidenciam que o cancelamento do R-final se dá em diferentes índices em verbos e em nomes. Em outras palavras, os dados indicam que o cancelamento do R-final opera em percursos evolutivos diferentes na língua portuguesa.

A variação sonora dos róticos finais se mostrou presente não apenas no nível segmental, como também no nível lexical. Vejamos os Gráficos 23 e 24:



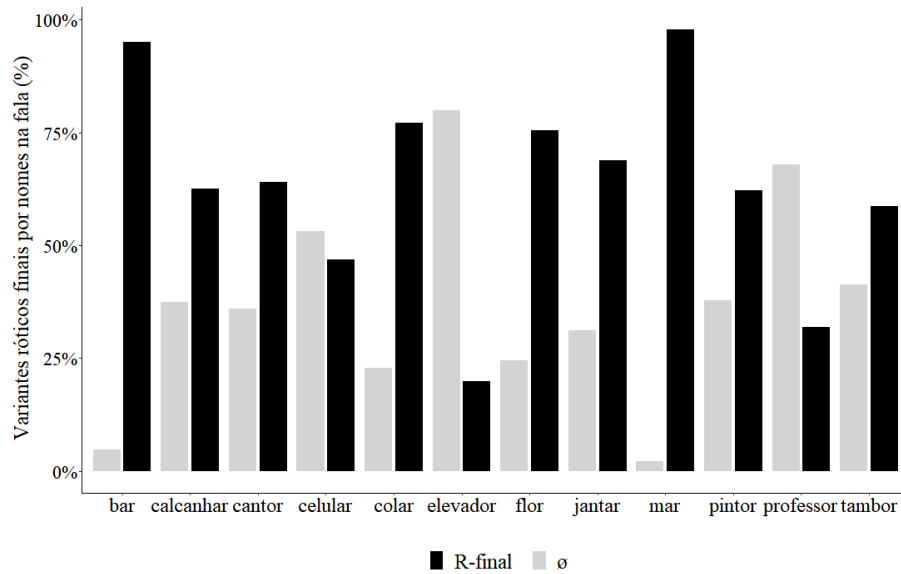
**Fonte:** elaboração própria.

O Gráfico 23 ilustra a variação do rótico final em cada verbo. As colunas pretas ilustram os verbos produzidos com R-final. As colunas cinza, por sua vez, ilustram os verbos produzidos com o cancelamento do rótico final. Observa-se que todas as palavras variaram. Ou seja, o rótico foi ora produzido, ora cancelado em todas as formas verbais. A variação se mostrou lexicalmente especificada, uma vez que cada verbo teve seu próprio intervalo de variação. Em algumas formas verbais, como “*abrir*”, “*beber*”, “*comer*”, “*imprimir*”, “*jogar*” e “*lavar*”, a produção do R-final prevaleceu. Em outros, como “*corrigir*”, “*cortar*”, “*dirigir*”, “*morder*” e “*pescar*”, houve maiores índices de cancelamento do R-final. Na palavra “*varrer*”, por sua vez, ambas as variantes ocorreram em índices semelhantes. Em conclusão, os dados de fala dos alunos do EF sugerem que o cancelamento do rótico final se implementa de forma lexicalmente gradual nos verbos.

Há de se pontuar, todavia, que a variação lexical encontrada nos verbos é um resultado curioso, pois vai de encontro à descrição documentada na literatura. É consensual na análise linguística do PB que o R-final em verbos é regularmente cancelado (OLIVEIRA, 1997). No entanto, nos dados aqui apresentados, o cancelamento do rótico ocorreu variavelmente. Uma possível explicação para os resultados divergentes à literatura pode estar relacionada ao fato de que os dados foram coletados em condições experimentais, fora do contexto de uso. O experimento, porém, permitiu confirmar o papel central do léxico em casos de variação e mudança sonora. Oliveira (1992) propôs que toda mudança sonora é lexicalmente implementada. Ou seja, até mesmo as mudanças regulares, que já atingiram todo o léxico, se implementariam de palavra à palavra na concepção do autor. Consensualmente, Bybee (2002) trouxe evidências de que a difusão lexical é muito comum em casos de mudanças sonoras. Tomando esses autores por referência, podemos dizer que os dados expostos no Gráfico 23 evidenciam que o cancelamento dos róticos se difunde lexicalmente nos verbos, em diferentes trajetórias evolutivas. Alguns

verbos, como *'lavar'* e *'jogar'* se encontram em estágios mais conservadores do percurso evolutivo rumo ao cancelamento do R-final; outros, como *'morder'* e *'cortar'* se encontram em estágios mais avançados dessa trajetória. Vejamos, a seguir, como se dá variação oral em nomes terminados por R-final.

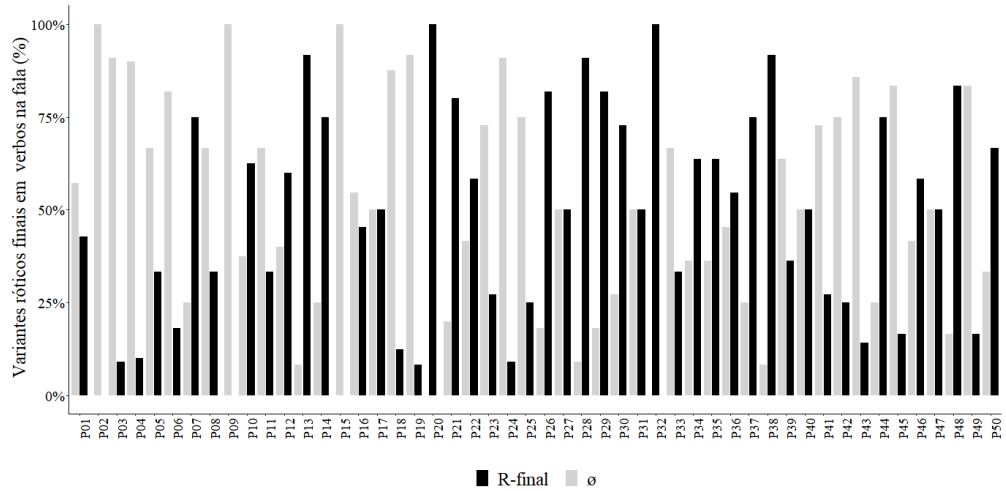
**Gráfico 24:** Variação do R-final por item nominal



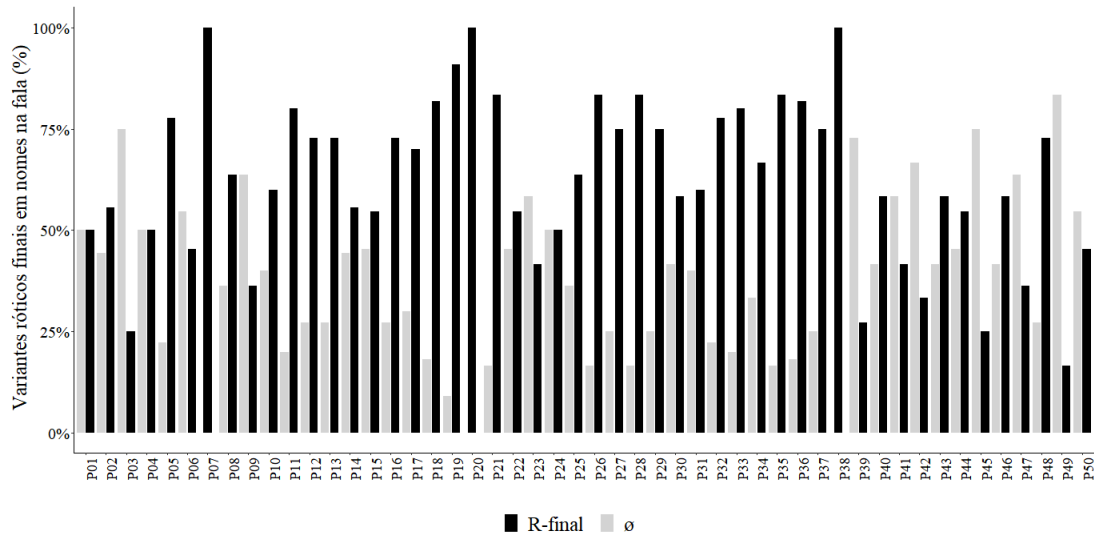
Fonte: elaboração própria

As colunas pretas, no Gráfico 24, ilustram nomes produzidos com o R-final. As colunas cinza se referem aos nomes produzidos com cancelamento do rótico final. Assim como nos verbos, todos os nomes terminados em R-final se mostraram envolvidos em variação. Contudo, os alunos do EF preferiram produzir do que cancelar o R-final na maioria dos nomes. Apenas as palavras *'celular'*, *'elevador'* e *'professor'* tiveram maiores índices de cancelamento do que de produção do rótico. As palavras *'bar'* e *'calcanhar'*, por exemplo, foram produzidas com R-final em quase 100% dos dados. Huback (2006) defendeu que o cancelamento do rótico em final de nominais se dá gradativamente no léxico, sendo que algumas palavras são mais suscetíveis ao fenômeno do que outras. Nos dados desta pesquisa, as palavras *'bar'* e *'calcanhar'* desfavoreceram o fenômeno de cancelamento do rótico, o que confirma que algumas palavras estão em estágio mais conservadores da evolução da língua.

Além do léxico, há um outro fator importante atuando na implementação do cancelamento dos róticos finais: o indivíduo. Cada participante cancelou o rótico em diferentes índices em cada classe gramatical. Vejamos os Gráficos 25 e 26:

**Gráfico 25:** Variação do R-final em verbos por indivíduo

Fonte: elaboração própria.

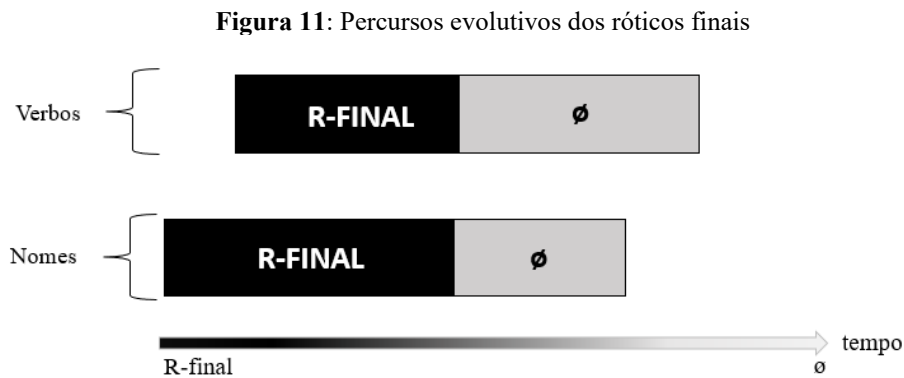
**Gráfico 26:** Variação do R-final em nomes por indivíduo

Fonte: elaboração própria.

O Gráficos 25 mostra a variação do R-final por indivíduo em verbos. O Gráfico 26 mostra a variação do R-final por indivíduo em nomes. O primeiro resultado a se destacar diz respeito à variação intra-individual (na fala de um mesmo indivíduo). Como se pode ver, os mesmos indivíduos produziram índices diferentes de cancelamento do R-final, a depender da classe gramatical. Por exemplo: o indivíduo P01 cancelou o rótico em 57% (N=4) dos verbos e em 50% (N=4) dos nomes. Já o participante P07 cancelou o R-final em 25% (N=3) dos verbos, mas não cancelou nenhum rótico final em nomes. Por fim, o P15 cancelou o R-final em 100% dos verbos, mas em apenas 45% (N=5) dos nomes. Esse resultado evidencia que, até mesmo no nível individual, o cancelamento do rótico final opera de forma específica a depender da classe gramatical.

Em segundo lugar, observemos a variação interindividual (entre os indivíduos). Nos verbos, todos os participantes cancelaram o rótico final em algum índice. Nos nomes, por outro lado, a produção do rótico final prevaleceu na fala da maioria dos alunos. Esse resultado nos mostra que os verbos estão em estágios mais avançados do cancelamento do rótico final do que os nomes também no nível do indivíduo. Vale pontuar, contudo, que alguns indivíduos produziram maiores índices de cancelamento do que outros, tanto em verbos quanto em nomes. Por exemplo: o participante P02 cancelou o R-final em 100% dos verbos. Já o P48 cancelou o R-final dos verbos em apenas 16% dos dados. Esses resultados evidenciam que o indivíduo, que também é um sistema dinâmico, evolui linguisticamente de forma não linear.

Os dados descritos nesta seção evidenciam, em resumo, que o cancelamento do R-final opera de forma específica na fala a depender da classe gramatical: os verbos tiveram o rótico mais cancelado do que os nomes. Além disso, a variação se mostrou determinada pelo léxico e pelo indivíduo. Ou seja, algumas palavras e alguns participantes favoreceram o cancelamento do R-final, enquanto outros se mostraram suscetíveis a produzir o rótico. Em busca de ilustrar as diferentes trajetórias evolutivas que operam em verbos e em nomes terminados com R-final, elaboramos a Figura 11:



**Fonte:** elaboração própria.

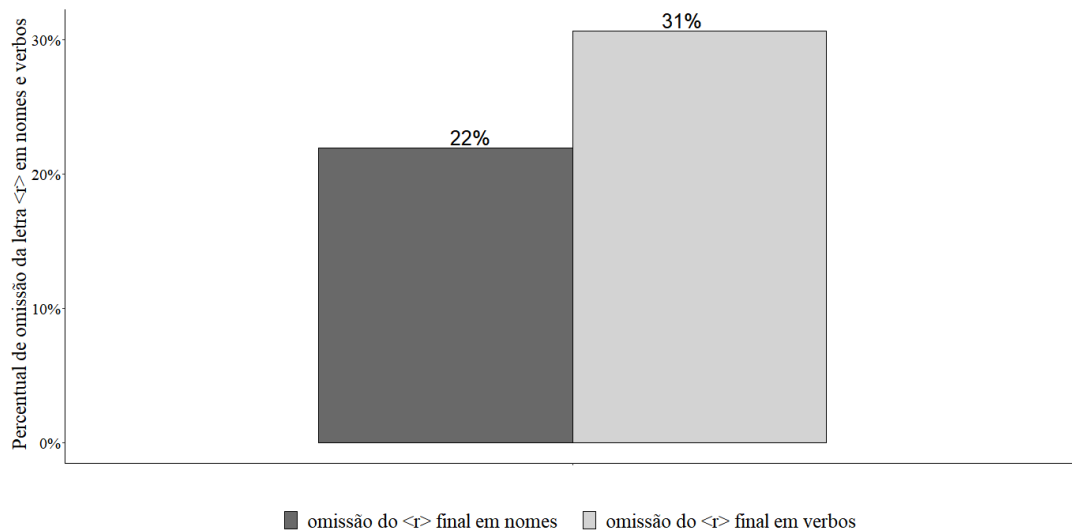
A seta da Figura 11 representa a trajetória evolutiva rumo ao cancelamento do R-final. A ponta esquerda da seta, parte preta, representa o estágio mais antigo da trajetória. A ponta direita da seta representa estágios avançados do percurso, que podem ser observados no hoje. O R-final ocorre no PB desde tempos mais antigos (CALLOU; MORAES; LEITE, 1996). O cancelamento do R-final é o padrão atual. Os retângulos da Figura 11 foram projetados a partir dos dados de fala desta tese. Por isso, o retângulo que representa o R-final é maior em nomes do que em verbos. A Figura 11 ilustra, portanto, que os diferentes estágios evolutivos em que se encontram os verbos e os nomes terminados com R-final. Como se pode ver, os verbos se encontram em estágios mais avançados do que os nomes na trajetória rumo ao cancelamento do R-final.

De modo geral, os dados de fala relacionados aos róticos finais desafiam a ideia tradicional de que um único processo fonológico motiva o cancelamento do R-final. Se assim o fosse, o rótico final de verbos e de nomes deveriam ser cancelados sempre que estivessem em final de palavras. Os dados analisados nesta tese mostram, contudo, que o fenômeno de cancelamento do R-final opera de diferentes formas na língua portuguesa, a depender da classe gramatical. Sabendo-se disso, podemos perguntar: *os diferentes percursos de evolução rumo ao cancelamento do R-final refletem na escrita infanto-juvenil?* Os dados descritos na próxima seção podem responder essa questão.

### 6.1.3 A omissão do <r> final na escrita

Esta seção descreve os dados de escrita dos alunos do EF referentes aos róticos finais de verbos e de nomes. Especificamente, descrevem-se os índices de erros ortográficos envolvendo a omissão da letra <r> final. Primeiro, o número total de vezes que os alunos omitiram a letra <r> final em cada classe gramatical foi contabilizado. Consideremos, portanto, o Gráfico 27.

**Gráfico 27:** Omissão do <r> final em verbos e em nomes

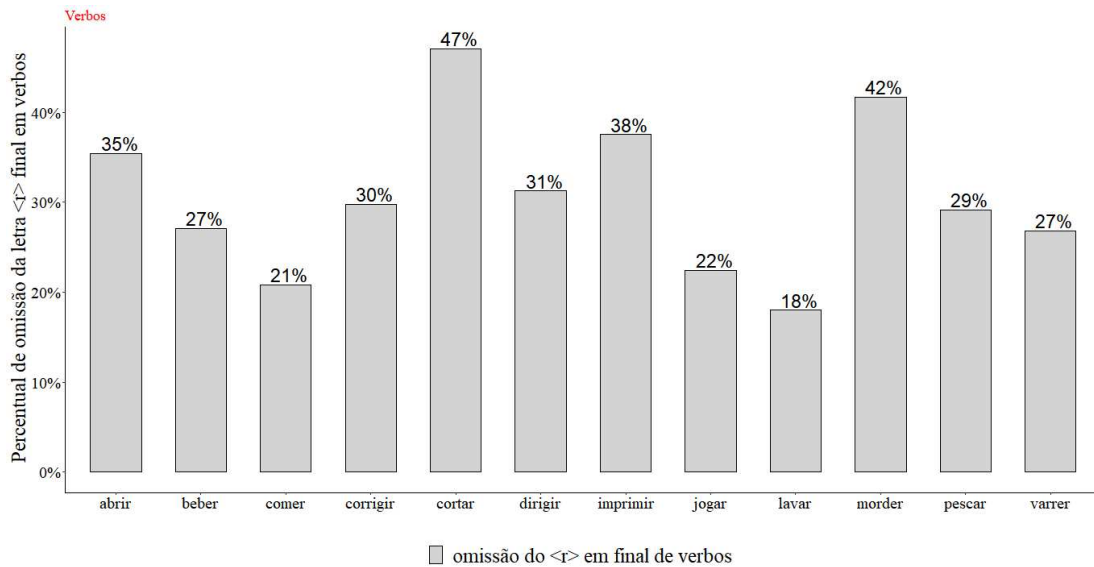


**Fonte:** elaboração própria.

O Gráfico 27 ilustra os índices de erros ortográficos envolvendo o R-final em verbos e em nomes. A coluna cinza-escuro do Gráfico 27 indica o número percentual de vezes em que os alunos omitiram a letra <r> no final de nomes. A coluna cinza-clara mostra os índices de omissão do <r> final em verbos. Os alunos do Ensino Fundamental erraram mais a escrita dos verbos (31%, N=176) do que dos nomes (22%, N=127). A análise estatística exploratória confirmou que as amostras de dados de escrita relacionadas aos verbos e aos nomes apresentam diferenças significativas ( $X^2 = 10.887$ ,  $df = 1$ ,  $p\text{-value} = 0.0009683$ ). Ou seja, os dados indicam que a classe gramatical interfere na grafia da letra <r>.

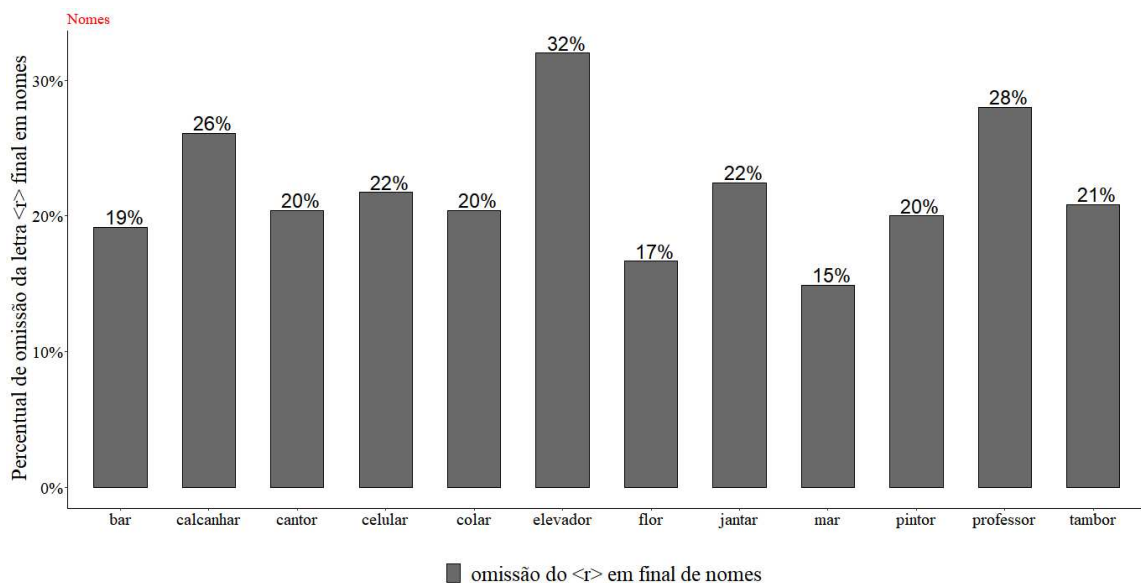
Esses resultados confrontam a ideia de que a omissão da letra <r> decorre de uma mesma motivação: o fenômeno fonológico de cancelamento do R-final. Se o fenômeno fosse o único responsável pelos erros ortográficos, estes deveriam ocorrer em índices semelhantes em ambas as classes gramaticais analisadas. Argumentamos, nesta tese, que a omissão do <r> final é resultado das múltiplas trajetórias evolutivas que operam em verbos e em nomes, no nível da palavra e do indivíduo. Vejamos os Gráficos 28 e 29.

**Gráfico 28:** Omissão do <r> final em itens verbais



Fonte: elaboração própria.

**Gráfico 29:** Omissão do <r> final em itens nominais



Fonte: elaboração própria.

O Gráfico 28 – superior – mostra os índices de erros ortográficos em verbos por item lexical. O Gráfico 29 – inferior – mostra os índices de erros ortográficos em nomes por item lexical.

Pode-se destacar, de início, que as 24 palavras-teste referentes aos róticos finais foram grafadas com algum índice de erro ortográfico. Esse resultado confirma que os aprendizes do EF têm dúvidas quanto à grafia da letra <r> final em diferentes palavras da língua portuguesa.

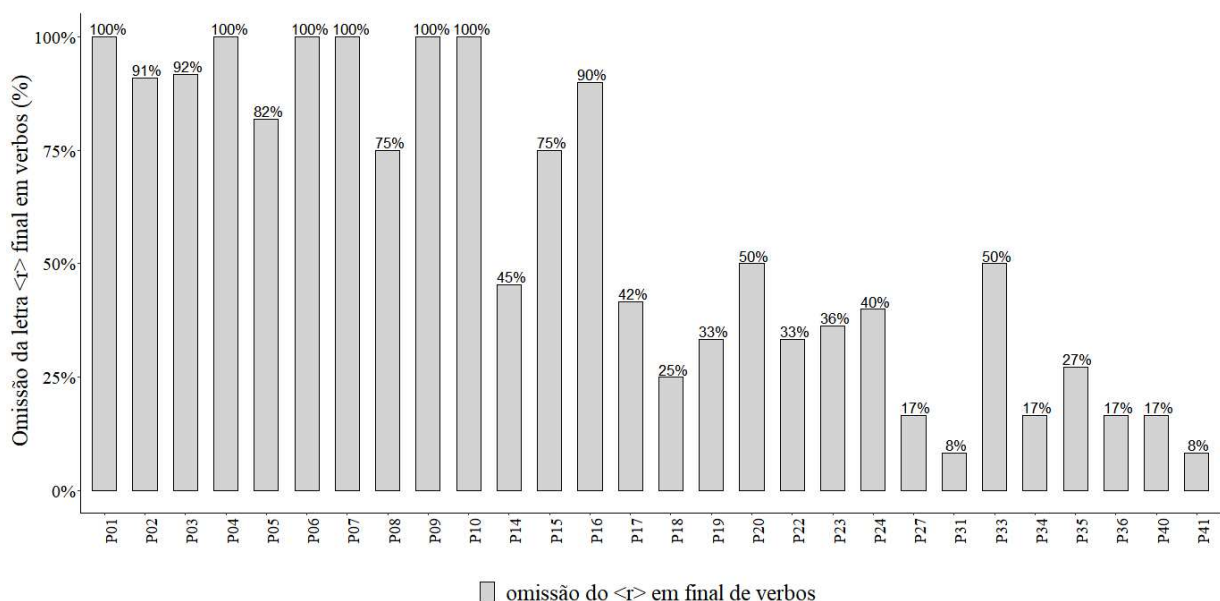
Os verbos grafados com maiores índices de erros ortográficos foram *'cortar'* e *'morder'*. Coincidentemente, foram essas as palavras com maiores índices de cancelamento na fala – *'cortar'*: 69%; *'morder'*: 72%. Além disso, o verbo *'lavar'* teve o menor índice de erros ortográficos (18%) e o menor índice de cancelamento na fala (31%) Em resumo, pode-se dizer que as diferentes trajetórias do cancelamento do R-final no léxico interferem na escrita infantil, de modo que alguns verbos são grafados com mais erros do que outros.

No que diz respeito aos nomes, as palavras *'elevador'* e *'professor'* foram escritas com a maior quantidade de erros ortográficos. Ambas as palavras também foram as que tiveram maiores índices de cancelamento na fala (elevador 80%; professor 68%). Já as palavras *'bar'*, *'flor'* e *'mar'* tiveram os menores índices de erros ortográficos e também baixos índices de cancelamento do rótico na fala (*'bar'*: 4%; *'flor'* 22%; *'mar'* 2%). Podemos pensar, então, que a variação lexicalmente especificada pode ter influenciado na escolha da forma ortográfica dos nomes.

Até este ponto, os resultados indicaram que os verbos e os nomes se manifestam de forma diferente na escrita, uma vez que a letra <r> foi mais omitida no final dos verbos do que dos nomes. Além disso, cada palavra foi escrita com índices específicos de erros ortográficos, o que destaca o papel do item lexical no percurso de aprendizado da ortografia.

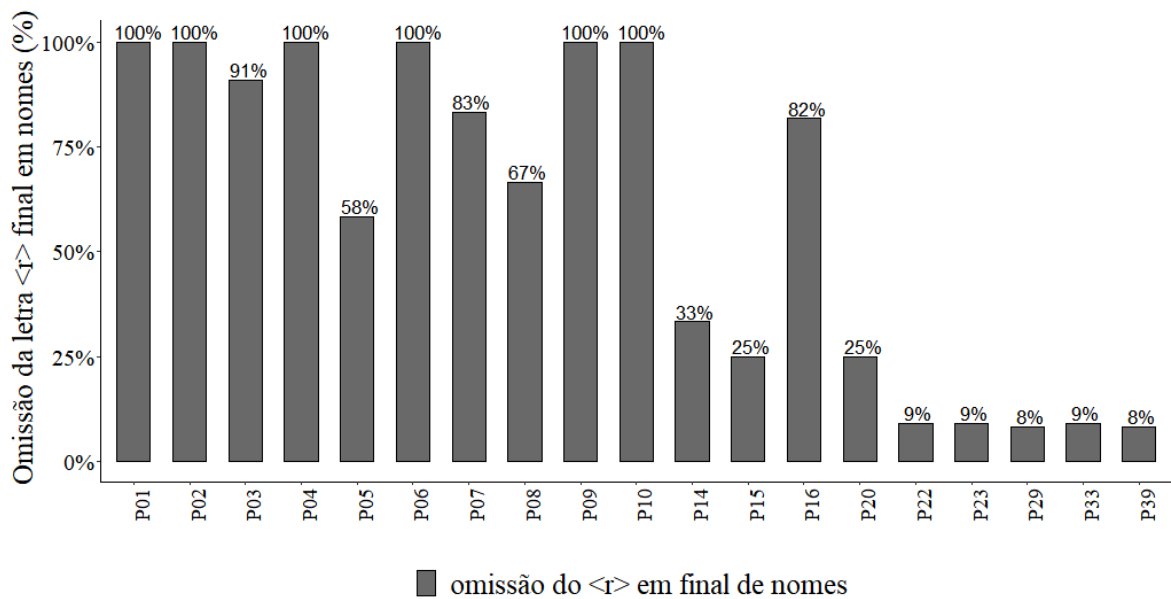
Vale ressaltar ainda que os índices de erros ortográficos variaram na escrita de cada indivíduo. Vejamos:

**Gráfico 30:** Omissão do <r> final em verbos por indivíduo



**Fonte:** elaboração própria.



**Gráfico 31:** Omissão do <r> final em nomes por indivíduo

Fonte: elaboração própria.

O Gráfico 30 mostra a omissão da letra <r> em final de verbos por indivíduo. O Gráfico 31 mostra a omissão da letra <r> em final de nomes por indivíduo. O primeiro resultado a se destacar é o seguinte: 28 alunos (56%) apresentaram erros ortográficos em verbos, mas apenas 19 (38%) apresentaram erros ortográficos em nomes. Esse resultado mostra que a maior parte dos alunos do EF erra mais as formas verbais do que as formas nominais.

Outro resultado interessante diz respeito à variabilidade nos índices de erros ortográficos na escrita de cada indivíduo. Consideremos, por exemplo, os participantes P14, P15, P16, P17, P18, P19, P20. Todos esses participantes estão matriculados no 3º ano do EF na mesma escola. Como as condições de aprendizado formais são parecidas, poderíamos ter como hipótese que os alunos evoluiriam linearmente no aprendizado do R-final. No entanto, como mostram os dados, os sete participantes do 3º ano cometeram diferentes índices de erros ortográficos em verbos e em nomes. Ou seja, os dados indicam que cada aluno percorre trajetórias particulares no decorrer do aprendizado da letra <r> final.

Por fim, é necessário destacar as diferenças na escrita dos verbos e dos nomes por um mesmo indivíduo. O participante P05, por exemplo, apresentou 82% (N= 9) de erros ortográficos em verbos e 58% (N=7) em nomes. Já o participante P39 não errou a ortografia dos verbos, mas apresentou 8% de erros em nomes. Por fim, o participante P40 errou apenas a escrita dos verbos. Esses resultados mostram que, ao grafar a mesma letra <r> em diferentes classes gramaticais, um mesmo aprendiz comete diferentes índices de erros ortográficos. Ou seja, a classe gramatical influencia na escrita ortográfica de um único indivíduo.

### 6.1.5 Hipercorreções

Vimos, nas seções anteriores, que os alunos do EF omitiram a letra <r> no final de verbos e de nomes. Nesta seção, veremos que os mesmos alunos inserem a letra <r> indevidamente no final de algumas palavras. Vejamos a Tabela 6.

**Tabela 6:** Hipercorreções em final de verbos e nomes

<b>Verbos</b>		
<b>erro ortográfico</b>	<b>N</b>	<b>%</b>
*a gente ler	17	61%
*a gente sorrir	11	35%
*cu dormir	5	19%
*a gente dar	8	26%
<b>Nomes</b>		
<b>erro ortográfico</b>	<b>N</b>	<b>%</b>
*dominor	3	6%
*robor	3	6%
*sofar	1	2%

Fonte: elaboração própria.

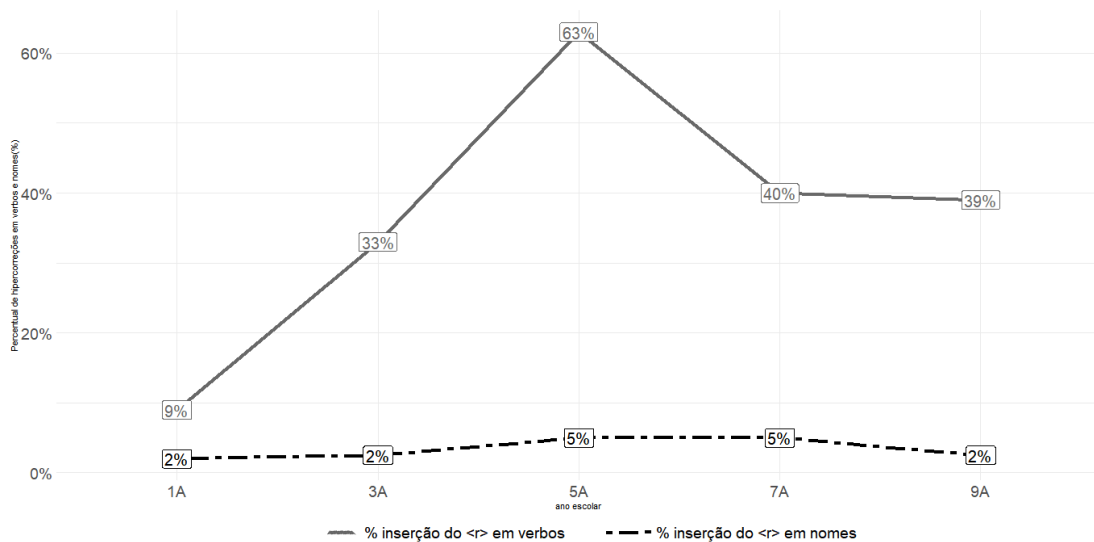
A Tabela 6 ilustra os índices de inserção do <r> em final de verbos e de nomes. Pode-se notar que o <r> final foi inserido, em algum índice, em todas as formas verbais flexionadas. Além disso, alguns alunos do EF inseriram a letra <r> no final dos nomes ‘dominó’, ‘robô’ e ‘sofá’. Análises tradicionais poderiam dizer que os dados da Tabela 6 são hipercorreções. No entanto, a hipercorreção ocorre quando o indivíduo desvia da norma na tentativa de se aproximar de um padrão ideal; de uma variante de prestígio (BAGNO, 2011, p. 949). Formas como <\*dominor, \*robor e \*sofar> sequer ocorrem na língua oral atual; então é difícil sugerir que elas representam variantes prestigiadas.

Argumentamos, nesta tese, que as tradicionais hipercorreções ortográficas são tentativas de se estabelecer relações ortográficas entre palavras que já possuem alguma similaridade linguística. Por exemplo: no caso dos verbos, os aprendizes podem relacionar as formas flexionadas “*lê*”, “*sorri*”, “*dormi*” e “*dá*” às suas formas de infinitivo “*ler*”, “*sorrir*”, “*dormir*” e “*dar*”. Assim, como os verbos se relacionam no nível morfológico, os aprendizes tentam relacioná-los também no nível ortográfico.

No caso dos nomes, os alunos podem estar tentando aplicar um mesmo padrão ortográfico em palavras que têm alguma similaridade fonética. Quando o R-final é apagado dos nomes, estes passam a terminar com vogal tônica na fala. Por exemplo: [selu’la]. A pronúncia [selu’la] é similar à pronúncia de outras palavras da língua, como [so’fa]. O aprendiz, portanto, pode criar a hipótese que as palavras terminadas em [a] tônica são grafadas da mesma forma. Por um lado, o aprendiz pode errar a palavra <\*celulá> para aproximá-la da palavra <sofá>. Por outro lado, o aprendiz pode colocar a letra <r> no final de <\*sofar> em busca de aproximá-la da palavra <celular>.

Vale ressaltar que há diferenças nos índices de ‘hipercorreções’ em nomes e em verbos. Vejamos o Gráfico 32:

**Gráfico 32:** Hipercorreções do <r> em final de verbos e de nomes



**Fonte:** elaboração própria.

As linhas cinza e preta, no Gráfico 32, mostram, respectivamente, os índices de ‘hipercorreção’ em verbos e em nomes. Como se pode ver, o <r> foi inserido em maiores índices nos verbos do que nos nomes. Uma possível explicação para este fato pode estar na morfologia verbal da língua portuguesa. Para conseguirem acertar a ortografia dos verbos, os indivíduos precisam se conscientizar das diferenças entre as formas verbais pertencentes a um mesmo esquema. Por exemplo: as formas flexionadas ‘sorri’ e ‘lê’ constituem o mesmo esquema morfológico das formas de infinitivo ‘sorrir’ e ‘ler’. Ademais, as formas flexionadas e no infinitivo podem ser pronunciadas com sequências fonéticas semelhantes na fala: vou [so.hi] = a gente [so.hi]. Partindo do pressuposto de que há conexão entre formas linguísticas semelhantes, é possível que as crianças estabeleçam representação ortográfica similar para palavras com sequências fonéticas semelhantes e cometam erros como <\*vou sorri ou \*a gente sorrir>.

## 6.2 PARTE II: Hipóteses

Esta seção retoma as hipóteses da pesquisa referentes aos róticos finais de verbos e nomes. São elas:

*(1b) a omissão da letra <r> na escrita persiste, em índices significativamente diferentes, até estágios distintos da escolarização, a depender da classe gramatical: verbo ou nome.*

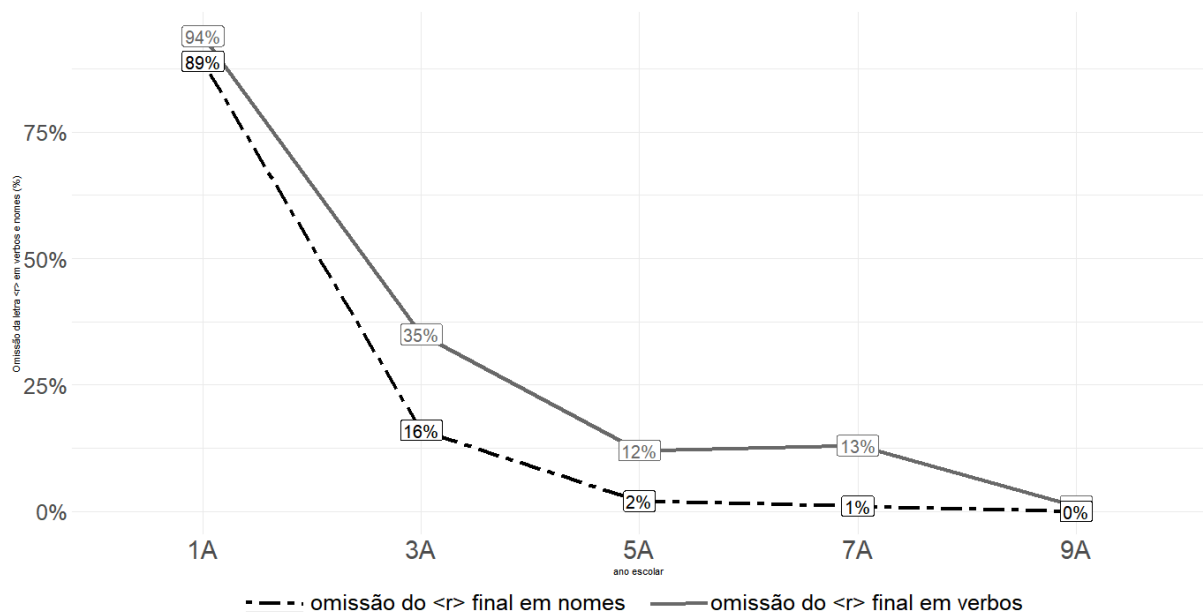
(2b) o aumento do rótico final na fala ao longo dos anos escolares depende da classe gramatical: verbo ou nome.

Para testar a hipótese (1b), compararam-se os índices de erros ortográficos relacionados ao <r> final em cada classe gramatical e em cada ano escolar. Para testar a hipótese (2b), verificaram-se os efeitos do ano escolar e da classe gramatical na produção oral de róticos finais. As hipóteses foram testadas com auxílio de modelos lineares generalizados mistos. A seguir, apresentam-se os resultados.

### 6.2.1 Hipótese (1b)

A análise descritiva dos dados mostrou que a omissão da letra <r> final é mais recorrente nos verbos do que nos nomes. Esses resultados dialogam com outros já documentados na literatura (COSTA, 2009; TORRES; OLIVEIRA, 2015; FREIRE; HORA, 2019). Em busca de avançar em relação à bibliografia precedente, esta tese se propôs a testar se os erros ortográficos envolvendo o R-final de verbos e de nomes persistem até estágios distintos da escolarização. Consideremos o Gráfico 33.

**Gráfico 33:** Omissão da letra <r> por ano escolar e por contexto acentual



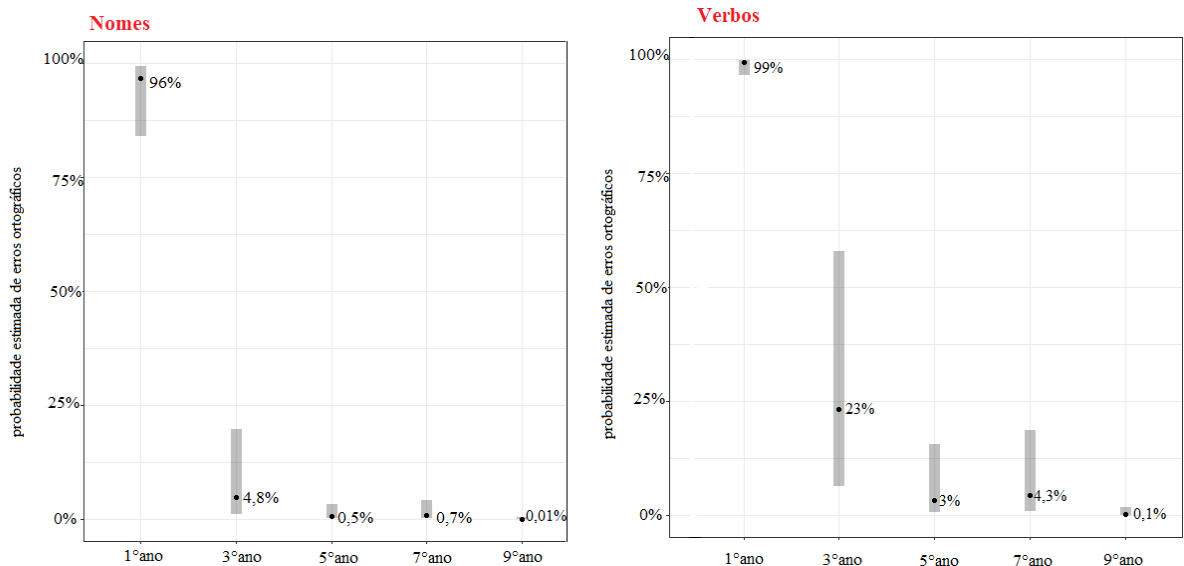
**Fonte:** elaboração própria.

O Gráfico 33 ilustra os índices de erros ortográficos em verbos e em nomes por ano escolar. A linha cinza indica as taxas de omissão da letra <r> no final de verbos. A linha preta pontilhada indica as taxas de omissão do <r> em final de nomes. Os erros ortográficos relacionados à letra <r> se concentraram, primordialmente, no 1º ano. Nos demais anos escolares, os índices de erros

decreceram progressivamente. No entanto, é necessário destacar dois resultados: (i) os índices de erros ortográficos em verbos são maiores do que em nomes em todos os anos escolares; (ii) os alunos erraram os verbos até o 9º ano; nos nomes, por sua vez, os erros ortográficos cessaram no 7º ano.

A fim de estimarmos a probabilidade de ocorrência de erros relacionados aos verbos e aos nomes em cada ano escolar, ajustamos um modelo linear generalizado misto<sup>22</sup> nas amostras de dados de escrita referentes aos róticos. Especificamente, o modelo visa testar a hipótese (1b): a omissão da letra <r> final persiste, em índices significativamente diferentes, até estágios distintos da escolarização, a depender da classe gramatical: verbo ou nome. A variável dependente do modelo foi binária: erro/acerto. A classe gramatical (verbo/nome) e o ano escolar (1º, 3º, 5º, 7º e 9º) foram as duas variáveis independentes de efeito fixo do modelo. O item lexical e o participante foram as duas variáveis aleatórias ajustadas. Uma comparação de modelos aninhados mostrou que o melhor modelo ajustado continha as variáveis ano escolar e classe gramatical como efeito fixo ( $\chi^2=966.474$ , p-value=1.259e-13). Ou seja, o ano escolar explica os índices de omissão da letra <r>; a classe gramatical também. Além disso, o modelo estimou probabilidades de ocorrência de erros ortográficos em verbos e em nomes em cada ano escolar. O Gráfico 34 mostra as probabilidades de erros ortográficos estimadas pelo modelo estatístico e os intervalos de confiança associados a essas probabilidades.

**Gráfico 34:** Probabilidade de omissão da letra <r> final



Fonte: elaboração própria.

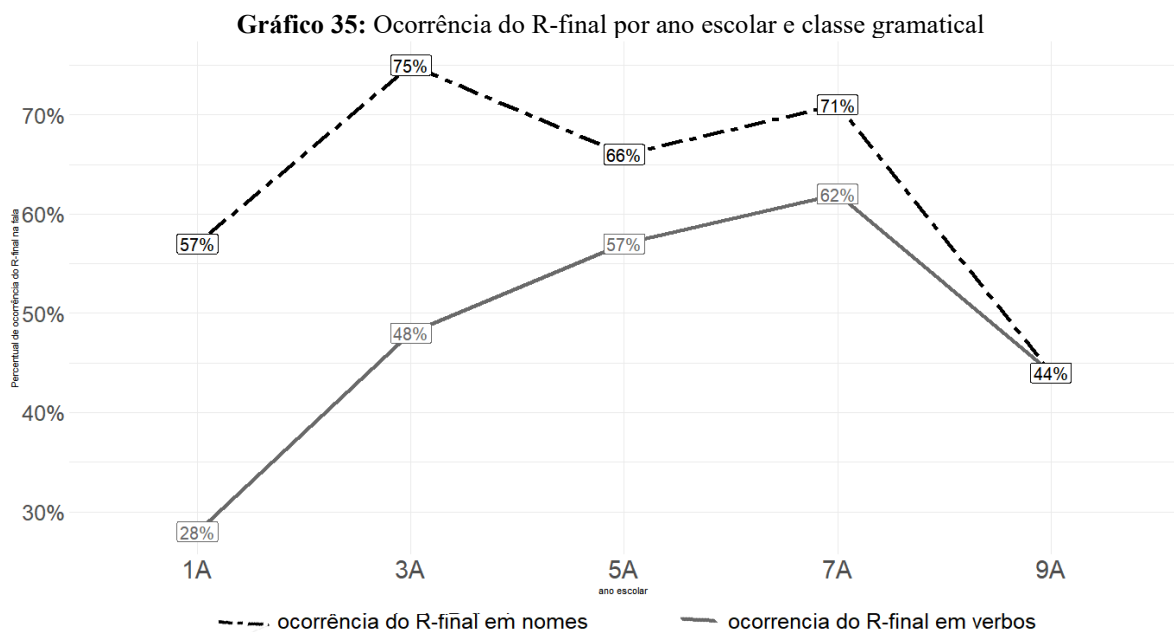
<sup>22</sup> m03<- glmer (ERROACERTO~FENOMENO+ANO+(1|PARTICIPANTE) +(1|PALAVRA), dados. roticos, family=binomial)

Considerando-se que os dados se mantêm constantes, pode-se dizer que a probabilidade de ocorrer a omissão da letra <r> na escrita é maior em verbos do que em nomes em todos os anos escolares. Além disso, os valores máximos dos intervalos de confiança (sombra cinza) são mais altos nos verbos do que nos nomes. Isso significa que, mesmo com as incertezas, a probabilidade de que um aprendiz omita a letra <r> no final de verbos é maior do que a probabilidade de que o aluno erre a escrita de nomes terminados em <r>.

De modo geral, os resultados demonstram que os índices de erros ortográficos em róticos finais ao longo dos anos parece depender da classe gramatical. Pode-se dizer, portanto, que os dados confirmam a hipótese (1b): a omissão da letra <r> na escrita persiste, em índices significativamente diferentes, até estágios distintos da escolarização, a depender da classe gramatical: nome ou verbo. Especificamente, os dados mostram que a omissão da letra <r> persiste mais na escrita de verbos do que na escrita de nomes.

### 6.2.2 Hipótese (2b)

Vimos, na seção anterior, que, com ao longo dos anos escolares, diminuem-se as omissões da letra <r> final na escrita infantojuvenil. Ou seja, os alunos passam a utilizar a letra <r> com mais frequência nos textos escritos. Sabendo-se disso, essa seção se propõe a avaliar se, com o aumento do uso da letra <r>, há também o aumento do uso do R-final na fala. Temos como hipótese que a classe gramatical – verbo ou nome – interfere no aumento do uso do rótico na fala ao longo dos anos escolares. Vejamos, então, o Gráfico 35:

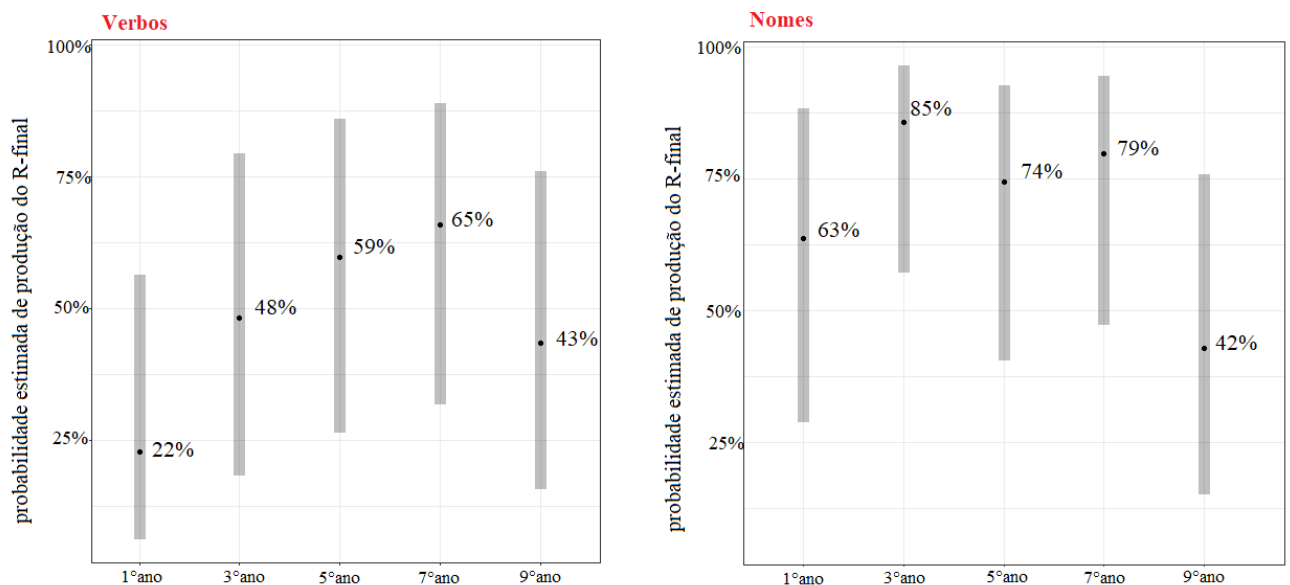


Fonte: elaboração própria.

O Gráfico 35 mostra os índices de produção do R-final na fala por ano escolar. A linha preta pontilhada ilustra a ocorrência de róticos finais em nomes. A linha cinza mostra os índices de ocorrência do R-final em verbos. No que diz respeito aos verbos, houve aumento gradativo da produção de róticos finais até o 7º ano. Nos nomes, a produção de róticos oscilou para cima e para baixo ao longo dos anos escolares.

Visando verificar o efeito do ano escolar e da classe gramatical na produção do R-final, foi ajustado um modelo linear generalizado misto<sup>23</sup> na amostra de dados de fala referente aos róticos finais. Especificamente, buscou-se testar a hipótese (2b): o aumento do rótico final na fala ao longo dos anos escolares depende da classe gramatical: verbo ou nome. A variável dependente do modelo foi binária: variante diferente da escrita/ variante próxima à escrita. A classe gramatical (verbo/nome) e o ano escolar (1º, 3º, 5º, 7º e 9º) foram as duas variáveis independentes de efeito fixo do modelo. O item lexical e o participante foram as duas variáveis aleatórias ajustadas. Uma comparação de modelos aninhados mostrou que o melhor modelo ajustado continha a interação entre as variáveis ano e classe gramatical como efeito fixo ( $\chi^2= 22.665$ , p-value= 0.0001477). Ou seja, a produção de róticos na fala é explicada pelo ano escolar em interação com a classe gramatical. Pode-se dizer, portanto, que os dados confirmam a hipótese (2b). Além disso, o modelo estimou probabilidades de ocorrência do R-final em verbos e em nomes e em cada ano escolar. Observemos o Gráfico 36:

**Gráfico 36:** Probabilidade de produção oral do rótico final



Fonte: elaboração própria.

<sup>23</sup> m.004.interaçao <- glmer(TIPODEVARIANTE~ANO\*FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA), dados.nomesverbos, family=binomial)

Considerando-se que os dados se mantêm constantes, pode-se dizer que a probabilidade de ocorrer róticos no final de verbos e de nomes é bastante variável se comparados os anos escolares. Como se pode observar, há aumento da probabilidade de ocorrência dos róticos tanto em nomes quanto em verbos. Contudo, a grande amplitude dos intervalos de confiança não nos permite garantir que tais probabilidades diferem significativamente em todos os anos escolares. Fizemos uma análise post-hoc para comparar par a par os anos escolares. Os resultados dessa análise confirmaram que os alunos do 3º ano falam mais róticos em final de nomes do que os alunos do 9º ano ( $p\text{-value}=0.0149$ ). Ademais, os alunos do 7º ano falam mais róticos em final de verbos do que os alunos do 1º ano ( $p\text{-value}= 0.0393$ ). Considerando-se que o aumento de R-final na fala pode refletir às influências da escrita na oralidade, há indícios de que a escrita influencia tanto a pronúncia de nomes quanto de verbos.

### **6.3 PARTE III: Discussão teórica - Róticos**

Nesta seção, os resultados previamente apresentados são discutidos à luz dos Modelos de Exemplares (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e da Teoria dos Sistemas Dinâmicos e Complexos (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Com o apoio das teorias e dos dados empíricos previamente descritos, busca- argumentar em defesa da seguinte tese: a letra <r> se associa a esquemas múltiplos, que emergem da interação complexa entre palavras, sequências fonéticas, morfológicas e ortográficas.

Como foi visto na análise descritiva dos dados de fala, a variação fonética do R-final é especificada pela classe gramatical, pelo item lexical e pelos indivíduos. Embora haja evidências de que o cancelamento do rótico final opera de forma distinta em diferentes contextos, as análises tradicionais insistem em propor uma única unidade representacional para os róticos finais: o arquifonema /R/. O arquifonema é uma unidade fonológica tradicional que representa a perda de contraste fonêmico entre dois ou mais fonemas em determinado contexto. Ou seja, uma única unidade fonológica representa variantes sonoras não contrastivas. Mais do que isso, a mesma unidade fonológica representa tanto as variantes do R-final de verbos quanto as variantes do R-final de nomes. Nessa perspectiva, as especificidades fonéticas de cada classe gramatical, evidenciadas nos dados empíricos desta tese, são excluídas da representação fonológica.

Diferentemente dos modelos tradicionais, os Modelos de Exemplares sugerem que as unidades fonológicas com as quais os indivíduos operam emergem das experiências de uso real da língua (TOMASELLO, 2000). Pode-se dizer, portanto, que a variação específica de cada classe gramatical expressa parte do conteúdo armazenado na representação mental dos alunos do EF. Ou seja, na gramática fonológica infantojuvenil, há registros de palavras pronunciadas com R-final e



também há registros do cancelamento do rótico. Como os verbos e os nomes foram produzidos com diferentes intervalos de variação segmental e lexical, é coerente propor que os esquemas linguísticos relacionados aos róticos de verbos e de nomes são distintos. No entanto, é válido pontuar também que, para os Modelo de Exemplares (BYBEE, 2001; JOHNSON, 2007), as categorias linguísticas estão conectadas entre si por meio de redes de similaridade. Nesse viés, pode-se dizer que a estrutura silábica (CVR) compartilhada pelos nomes e verbos permite associações entre as duas classes gramaticais e, conseqüentemente, permite a emergência de generalizações. A generalização CVR pode explicar por que o cancelamento do rótico ocorre em várias palavras (seja verbo, seja nome) terminadas com padrão CVR.

A variação lexicalmente especificada encontrada nos dados apresenta um problema para as abordagens tradicionais em fonologia, tendo em vista que representações únicas limitam o fonólogo a considerar apenas um padrão para representar contextos claramente diferentes. As categorias discretas simplificam a análise, ao ponto de se deixar alguns dados sem explicação (BYBEE, 2001). Ao propor, por exemplo, uma única unidade cognitiva para representar o R-final em diferentes palavras, os modelos estruturalistas deixam de explicar a variabilidade lexical encontrada nos dados reais de fala. Contudo, pode-se perguntar: *qual é a relevância de se introduzir a variação lexical na gramática?* Para Bybee (2001), a incorporação da variação na representação mental permite incluir a dinamicidade, inerente à linguagem, aos modelos de gramática. Em outras palavras, os Modelos de Exemplares incluem os mecanismos pelos quais a mudança linguística ocorre como parte integrante da arquitetura da linguagem (BYBEE, 2001, p. 57). Para além disso, a inclusão da variação lexical em modelos de gramática faz com que o léxico seja visto como elemento central do conhecimento linguístico. Assim, é possível avaliar as influências do léxico no comportamento linguístico dos indivíduos. Na presente tese, vimos que o léxico interfere no aprendizado da ortografia, uma vez que cada palavra é grafada com índices específicos de erros ortográficos.

Partindo do pressuposto de que a variação fonética e a variação lexical são parte da representação mental, propomos que os aprendizes de escrita acessam representações linguísticas múltiplas ao grafarem os róticos finais de verbos e de nomes. A multiplicidade das representações pode motivar as diferentes trajetórias de aprendizado da ortografia indicadas nos dados. Como foi evidenciado, os erros ortográficos em verbos são mais recorrentes e mais persistentes do que os erros ortográficos em nomes. Resta-nos pensar nos fatores que motivam os diferentes percursos de aprendizado da letra <r>.

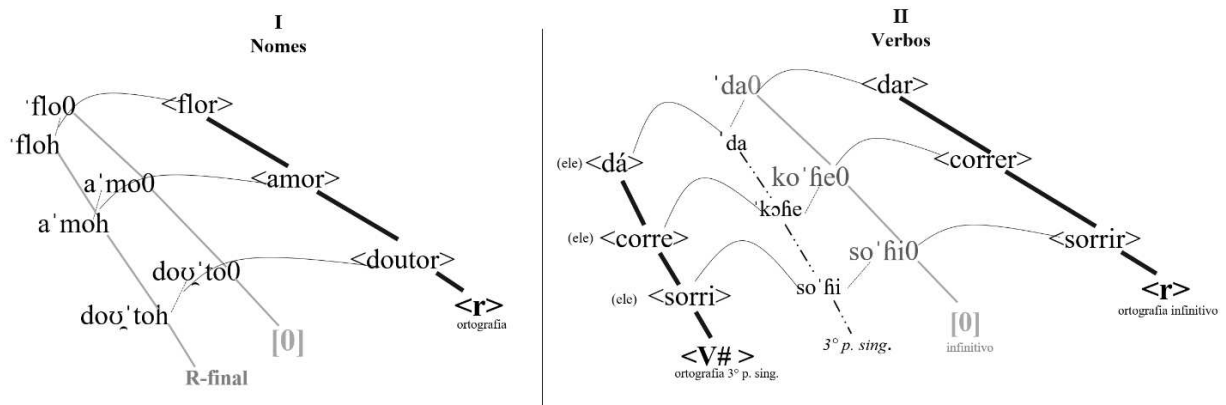
O primeiro motivo capaz de explicar os diferentes índices de erros ortográficos em verbos e em nomes diz respeito à variação sonora específica de cada classe gramatical. Vimos, ao longo desta seção, que o R-final é mais cancelado nas formas verbais do que nas formas nominais. Uma

vez que as categorias fonéticas têm exemplares mais e menos robustos (BYBEE, 2001), a representação do R-final cancelado seria mais robusta na categoria verbal do que na nominal. Considerando-se que, ao longo do aprendizado da escrita, os alunos relacionam as letras às categorias fonéticas, é explicável que eles escrevam os verbos de forma diferente dos nomes. Mais do que isso, é justificável que os verbos tenham maiores índices de omissão da letra <r> na escrita, tendo em vista que as formas canceladas são prototípicas na categoria verbal.

Em segundo lugar, a persistência dos erros ortográficos em verbos pode ter relação com a semântica e com a morfologia flexional do português. Formas de infinitivo como *'sorrir'* e *'ler'* estão relacionadas, foneticamente e semanticamente, às suas formas flexionadas *'sorri'* e *'lê'*. A relação fonética se dá pelo fato de que enunciados como *'eu sorri'* - *'eu vou sorrir'* podem ser pronunciados de forma similar como, por exemplo, [eʊso'hi], [eʊvosos'hi]. Já a relação semântica decorre do fato de que as formas infinitivas e flexionadas se referem à mesma ação. Paula (2007) mostrou que alunos mais avançados (5º e 7º ano) possuem mais sucesso em tarefas explícitas de consciência morfológica do que alunos dos anos iniciais do EF. Esse resultado evidencia que as habilidades relacionadas à morfologia da língua são aprimoradas em momentos tardios da escolarização. Isso pode explicar por que os erros relacionados aos verbos perduram por mais tempo na escrita dos aprendizes do que os erros relacionados aos nomes.

De forma geral, esta tese busca defender que os diferentes percursos de aprendizado da ortografia expressam o gerenciamento de esquemas linguísticos múltiplos ao longo da escolarização. Os diagramas I e II, a seguir, ilustram esses esquemas.

**Figura 12:** Relações múltiplas entre oralidade e escrita em verbos e nomes



Fonte: elaboração própria.

O diagrama I (à esquerda da Figura 12) representa as relações entre formas fonéticas e formas ortográficas na categoria dos nomes. As linhas cinzas representam a variação fonética – [ø] ~ R-final. A linha preta representa o padrão ortográfico – <r> – relacionado aos nomes. Como os róticos ocorrem variavelmente no final de nomes, a variação está representada no esquema.

Vale ressaltar que essas são algumas das generalizações possíveis. Muitas outras, como, por exemplo, o sufixo *-dor*, outras variantes posicionais, etc., estão armazenadas na categoria dos nomes. A representação de todas elas, no entanto, tornaria a ilustração extremamente complexa. P

O diagrama II (à direita da Figura 12) representa as relações fonéticas, morfológicas e ortográficas na categoria dos verbos. A linha cinza representa um padrão fonético-fonológico: o cancelamento do R-final. Esse padrão é produtivo no PB, haja vista que o rótico é frequentemente cancelado em final de verbos (OLIVEIRA, 1997; RENNICKE, 2016). A linha preta pontilhada representa um padrão morfofonológico: as formas de 3º pessoa do singular. A linha preta contínua representa padrões ortográficos das palavras. Observe que, na categoria dos verbos, há formas escritas com vogais finais – <sorri> – e formas escritas com o <r> final – <sorrir>. Ambos os esquemas foram representados para demonstrar que formas como ‘vou [so'fi]’ e ‘ele [so'fi]’ podem estar conectadas na representação, por meio de redes de similaridade fonética e morfológica. Essa conexão explica casos em que os aprendizes ora omitem o <r> final, como em <\*vou sorri'>, ora inserem-no: <\*ele sorrir> (CESAR, 2017, CASTELO BRANCO, 2020). Para conseguirem entender as especificidades ortográficas de cada item verbal, os aprendizes podem precisar ser instruídos quanto à morfologia da língua, de modo a desenvolver habilidades voltadas às flexões de modo e de tempo verbal.

Até o momento, argumentos foram tecidos em busca de defender que os diferentes percursos de aprendizado da ortografia dos verbos e dos nomes terminados em <r> são resultado do acesso aos esquemas linguísticos múltiplos no decorrer da escolarização. Vale ressaltar que, além de acessarem as categorias múltiplas, os aprendizes reorganizam os esquemas linguísticos ao longo dos anos escolares. Os dados indicaram que, durante o EF, há oscilação na produção oral dos róticos finais. Essa oscilação pode ser interpretada, com o apoio da TSCD (THELEN; SMITH, 2003), como evidência da instabilidade das categorias linguísticas róticas ao longo do aprendizado da ortografia. Quando os alunos aprendem a forma ortográfica das palavras, o novo conhecimento precisa ser acomodado na representação linguística dos aprendizes. Durante este processo, o sistema linguístico, que é dinâmico, pode se reorganizar. Assim, é explicável que ora haja aumento da produção do R-final, ora haja diminuição e ora haja estabilidade. Considerando-se que o aprendizado da ortografia pode fortalecer esquemas já existentes na gramática, pode-se dizer que o aprendizado da letra <r> pode fortalecer a representação do R-final e, conseqüentemente, aumentar a produtividade deste padrão. O aumento da produtividade do padrão pode ser evidenciado pelo aumento do seu uso na fala em determinados anos escolares.

De forma geral, sugerimos que as formas ortográficas das palavras, associadas aos esquemas múltiplos, podem interferir positivamente na produtividade de padrões; isto é, padrões sonoros já produtivos na língua – como os róticos finais – podem se tornar ainda mais produtivos

no uso linguístico após o aprendizado da escrita. Essa ideia só faz sentido se, com apoio dos Modelos de Exemplos e da TSDC, considerarmos que as letras se associam à esquemas múltiplos e dinâmicos, que emergem da interação complexa entre palavras, sequências fonéticas, morfológicas e ortográficas.

#### 6.4 PARTE IV: A visão dos professores sobre a omissão do <r> final

Além da análise quantitativa dos dados, foram analisadas, qualitativamente, algumas observações de professores a respeito da omissão da letra <r> em final de verbos e de nomes. Foi pedido para os docentes que marcassem em uma tabela (Apêndice C) os erros já observados na escrita de seus alunos. Vejamos as respostas no Quadro 19:

**Quadro 19:** Observações dos professores sobre a omissão da letra <r>

Prof.	Atualmente, você leciona em qual ano escolar?	Você leciona há quantos anos?	Marque os erros ortográficos que você já observou nos textos dos seus alunos	
			<i>verbos</i>	<i>nomes</i>
1	1° as 5° anos	10 anos		*maracujar
2	1° ano	20 anos	*abri	*ma
3	4° e 5° anos	19 anos	*bebe, *come, *corta, *ele dar, *dirigi, *imprimir, *lava*morde, *pesca, *varre	*calcanha, *canto, *celula, *cola, *elevado, *janta, *tambo
4	6° ao 9° anos	32 anos	*bebe, *come, *corrigi, *corta, *ele dar, dirigi, *eu dormir, *ele ler, *eu sorrir, *imprimir, *lava, *morde, *pesca, *varre	
5	9° ano	11 anos	*abri, *bebe, *come, *dirigi, *joga, *lava	*celula*janta

**Fonte:** elaboração própria.

Os cinco professores participantes da pesquisa já observaram erros ortográficos envolvendo a letra <r> final. Contabilizamos os erros ortográficos relacionados aos verbos e aos nomes. Os resultados da contagem mostram que as observações dos professores estão alinhadas com a análise quantitativa feita nesta pesquisa. Consideremos a Tabela 7:

**Tabela 7:** Quantidade de erros ortográficos observados pelos professores

	Verbos	Nomes	Total
N de erros marcados pelos profs. do 1° ciclo do EF	10	6	46
N de erros marcados pelos profs. o 2° ciclo do EF	16	2	23
Total	26	8	69

**Fonte:** elaboração própria.

A Tabela 7 mostra a quantidade de erros ortográficos que os professores do 1º e do 2º ciclos do EF marcaram em verbos e em nomes. Em primeiro lugar, é válido pontuar que os professores do 1º ciclo do EF marcaram mais tipos de erros ortográficos do que os professores do 2º ciclo. Ou seja, na visão dos professores, os alunos do 1º ao 5º anos cometem mais erros ortográficos do que alunos do 6º ao 9º ano. De fato, os dados de escrita dos alunos do EF mostraram que há mais erros ortográficos nos anos iniciais do que nos anos finais do EF. As observações dos professores estão, portanto, de acordo com os dados de escrita avaliados nesta pesquisa.

Outro ponto que merece destaque na Tabela 7 é o número de erros ortográficos em cada categoria linguística avaliada. No que diz respeito aos róticos finais, a visão dos professores também está em concordância aos dados. Pode-se notar que, juntos, os professores marcaram 26 tipos de erros ortográficos nos verbos e apenas 8 tipos nos nomes. Vimos, nas seções anteriores que, de fato, os alunos do EF cometem mais omissões do <r> final nas formas verbais do que nas nominais. Além disso, foi observado que os erros ortográficos nos verbos perduram por mais anos escolares do que nos nomes. Pode-se ver, na Tabela 7, que os professores do 2º ciclo marcaram 16 tipos de erros nos verbos e apenas 2 tipos nos nomes. Ou seja, os docentes do segundo ciclo avaliam, nos textos do 6º ao 9º anos, que a letra <r> é mais omitida nas formas verbais do que nas nominais.

Além de marcarem os erros ortográficos observados, os professores também opinaram sobre as motivações desses erros. O Quadro 20 resume o que escreveu cada professor.

**Quadro 20:** Motivação dos erros ortográficos pela visão do professor

<b>Professor</b>	<b>De acordo com a sua opinião, quais motivos explicam erros ortográficos, como os listados anteriormente, cometidos pelos alunos?</b>
1 (1º ao 5º anos)	No início, e até mesmo durante a consolidação da base alfabética, é normal as crianças trocarem as letras ou omiti-las. A partir do 3º ano do 1º ciclo, com a sistematização da ortografia em sala de aula, os erros se tornam objeto de reflexão coletiva, sendo trabalhados didaticamente pelo professor. A maioria dos erros acontece porque as crianças escrevem do mesmo jeito que falam. Aos poucos vão internalizando as convenções ortográficas.
2 (1º ano)	A língua portuguesa não é tão simples, muitas palavras falamos de um jeito e escrevemos de outro.
3 (4º e 5º anos)	1- Reprodução da oralidade na linguagem escrita, sem perceber as características e diferenciações entre ambas; 2 - Prática da escrita sem reflexão, sem a leitura do que escreve; 3- Hábitos de leituras de diversos gêneros textuais ainda não consolidados; 4- Desconhecimento da adequação ou inadequação da linguagem formal/informal mediante o contexto.
5 (6º ao 9º anos)	Observo que eles grafam as palavras registrando muito a oralidade que é marcante e está no cotidiano deles. Pouco a pouco, eles vão percebendo as variações nos padrões de linguagem e distinguindo linguagem coloquial da formal.
6 (9º ano)	Uma parte das palavras > questão fonética, aproximação com a fala (oralidade). Outra parte > hipercorreção

**Fonte:** elaboração própria.

Embora cada professor tenha respondido o questionário de acordo com a sua opinião, há pontos convergentes entre as respostas transcritas no Quadro 20. A fala é a principal motivadora dos erros ortográficos, na visão dos professores. Todos os docentes mencionaram a relação entre oralidade e escrita em suas respostas. Podemos dizer que os comentários dos professores estão bastante alinhados com textos teóricos sobre ortografia, já que, na maioria dos manuais sobre ortografia, explicações parecidas com as dos docentes podem ser encontradas.

Como já foi mencionado nesta tese, autores como Alvarenga et.al (1989), Cagliari (1989), Gomes de Moraes (1998), Nóbrega (2013) propõem categorias para os erros cometidos por influência da oralidade. Comumente, erros ortográficos são explicados como tentativas de transcrição da fala. Assim, é justificável que os professores, estudiosos dos manuais sobre ortografia, expliquem os erros ortográficos conforme à tradição teórica. No entanto, é preciso ressaltar a sensibilidade dos professores na percepção dos diferentes índices de erros ortográficos em cada categoria linguística. Por um lado, os professores, assim como os manuais, explicam os erros ortográficos por via da oralidade. Por outro, esses docentes observam os diferentes índices de erros ortográficos na escrita infantojuvenil. As observações dos professores, atreladas à discussão teórica realizada nesta tese, podem motivar práticas pedagógicas inovadoras que considerem as particularidades de aprendizado de padrões ortográficos aparentemente semelhantes.

## **6.5 Resumo do capítulo**

Este capítulo descreveu e analisou dados de fala e de escrita referentes aos róticos finais de verbos e de nomes. Na primeira parte do capítulo, foi feita a análise descritiva dos dados. Especificamente, avaliou-se o fenômeno fonológico de cancelamento do rótico final e os índices de erros ortográficos envolvendo a omissão da letra <r> em final e verbos e de nomes. De modo geral, os dados de fala indicaram que o cancelamento do rótico final opera em trajetórias evolutivas específicas a depender da classe gramatical. Os dados de escrita, por sua vez, indicaram que os alunos do EF apresentam índices diferentes de erros ortográficos em formas verbais e nominais.

Na segunda parte do capítulo, as hipóteses (1b) e (2b) foram testadas e confirmadas. No que se refere à hipótese (1b), os dados mostraram que a persistência da omissão da letra <r> final depende do contexto acentual. Os erros ortográficos em verbos são mais recorrentes e persistentes em verbos do que em nomes. No que se refere à hipótese (2b), os dados mostraram que alunos do 3º ano falam mais róticos em final de nomes do que os alunos do 9º. Ademais, os alunos do 7º ano falam mais róticos em final de verbos do que os alunos do 1º ano. Em suma, os dados indicam que as alterações na fala ao longo da escolarização dependem da classe gramatical.

Na terceira parte do capítulo, os dados foram discutidos à luz dos Modelos de Exemplos e da TSDC. De forma geral, argumentamos que os diferentes percursos de aprendizado da letra <r> expressam o gerenciamento de esquemas múltiplos ao longo da escolarização. Além disso, propomos que o aprendizado da letra <r> pode interferir positivamente na produtividade de padrões; isto é, padrões sonoros já produtivos na língua – como os róticos finais – podem se tornar ainda mais produtivos no uso linguístico após o aprendizado da escrita.

Por fim, a parte IV do capítulo descreveu a visão dos professores sobre os erros ortográficos relacionados aos róticos finais. Os dados indicaram que os professores estão atentos aos diferentes índices de erros ortográficos em verbos e em nomes.

## 7. CONSIDERAÇÕES FINAIS

O objetivo desta pesquisa foi investigar as relações múltiplas que se estabelecem entre a fala e a escrita e entre a escrita e a fala na variedade de Belo Horizonte – MG. Quatro categorias linguísticas foram consideradas: (1) vogais médias e altas pretônicas, (2) vogais médias e altas postônicas finais, (3) róticos finais em verbos, (4) róticos finais em nomes. A investigação ocorreu a partir da análise de erros ortográficos e de fenômenos fonológicos variáveis. Buscou-se defender a seguinte tese: *as letras são símbolos discretos que se associam a esquemas múltiplos, que emergem da interação complexa entre palavras, sequências fonéticas, morfológicas e ortográficas.*

A tese foi ancorada nos pressupostos de Modelos de Exemplares (JOHNSON, 1997, 2005; BYBEE, 2001, 2002, 2010; PIERREHUMBERT, 2001, 2003; FOULKES; DOCHERTY, 2006; PORT, 2007) e da Teoria dos Sistemas Dinâmicos e Complexos (THELEN; SMITH, 1994, 2003; ELLIS; LARSEN-FREEMAN, 2009). Em síntese, estas teorias pressupõem que o sistema linguístico é múltiplo, por armazenar a variação; é dinâmico, por alterar ao longo do tempo; é não-linear, pois uma mudança pode causar efeitos de forças variadas no sistema; é probabilístico, por conter padrões mais e menos produtivos. Assim sendo, esses modelos teóricos suportam o argumento de que o aprendizado da escrita envolve o gerenciamento e a reorganização de esquemas múltiplos, dinâmicos e probabilísticos. Isto é, visamos defender que os esquemas linguísticos múltiplos são acessados e reorganizados pelos aprendizes ao longo do aprendizado do sistema de escrita.

Neste ponto, são retomados os objetivos específicos propostos na Introdução para que se apontem os resultados alcançados:

- *Comparar os índices de erros ortográficos em contextos pretônico e postônico final na escrita de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.*

Os resultados alcançados confirmaram a hipótese de que as trocas de <e, o> por <i, u> na escrita persistem, em índices significativamente diferentes, até estágios distintos da escolarização, a depender do contexto acentual: pretônico ou postônico final. Os dados mostraram que os alunos do Ensino Fundamental aprendem a ortografia das vogais postônicas finais em estágios anteriores ao aprendizado da ortografia das vogais pretônicas. Além disso, os resultados mostraram que a probabilidade de se cometer erros ortográficos é maior em contexto pretônico do que em contexto postônico final. Os diferentes estágios de aprendizado da ortografia das vogais médias e altas foram interpretados como evidências de que os aprendizes gerenciam esquemas múltiplos – e diferentes – para grafarem as letras <e, o> em diferentes contextos acentuais.



- *Comparar os índices de erros ortográficos em verbos e em nomes terminados com R-final na escrita de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.*

Os resultados alcançados confirmaram a hipótese de que a omissão da letra <r> na escrita persiste, em índices significativamente diferentes, até estágios distintos da escolarização, a depender da classe gramatical: verbo ou nome. Os dados mostraram que os erros ortográficos relacionados aos verbos perduram por mais anos do EF do que os erros ortográficos relacionados aos nomes. Além disso, a análise mostrou que a probabilidade de se cometer erros ortográficos nos verbos é maior do que nos nomes. Esses resultados foram interpretados como indícios de que os aprendizes gerenciam esquemas linguísticos para grafarem a letra <r> final em diferentes classes gramaticais.

- *Avaliar a variação oral das vogais pretônicas e postônicas finais na fala de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.*

Tanto as vogais pretônicas quanto as vogais postônicas finais foram pronunciadas com algum índice de variação em todos os anos escolares avaliados. No contexto pretônico, a variação se deu entre as vogais médias e altas. Exemplo: [mi'ninʊ] ~ [me'ninʊ]. No contexto postônico final, a variação se deu entre vogais altas e canceladas. Exemplo: ['patʊ] ~ ['pat]. Houve aumento de vogais médias pretônicas, sobretudo, na fala de alunos do 9º ano. Por outro lado, não houve aumento de vogais médias postônicas finais ao longo dos anos escolares. Portanto, confirmou-se a hipótese de que o aumento de vogais médias na fala ao longo dos anos escolares depende do contexto acentual: pretônico ou postônico final. Apoiando-se nas premissas dos Modelos de Exemplares e da TSDC, esses resultados foram interpretados como evidências de que o aprendizado da ortografia causa diferentes efeitos na fala e no sistema linguístico. Especificamente, argumentamos que as formas ortográficas podem motivar a reorganização dos esquemas linguísticos. Porém, o efeito da escrita na representação mental pode ser mais ou menos evidente a depender da categoria linguística, uma vez que estas são múltiplas, dinâmicas e probabilísticas. Argumentamos que padrões sonoros já produtivos na língua – como as vogais médias pretônicas – podem se tornar ainda mais produtivos no uso linguístico após o aprendizado da escrita. Por outro lado, em esquemas de baixa produtividade – como [e, o] postônicos finais – a forma ortográfica não tem força suficiente para atrair o sistema e moldar o esquema a ponto de aumentar significativamente a produtividade do padrão.

- *Avaliar a variação oral dos róticos finais de verbos e de nomes na fala de alunos do 1º, 3º, 5º, 7º e 9º anos do Ensino Fundamental.*

Todos os verbos e nomes terminados em R-final foram produzidos com algum índice de variação. Os verbos tiveram maiores índices de cancelamento do rótico final do que os nomes. A análise estatística mostrou que a classe gramatical e o ano escolar interferem na produção oral dos róticos. Esse resultado confirma a hipótese de que o aumento do rótico final na fala, ao longo dos anos escolares, depende da classe gramatical: verbo ou nome. Os alunos do 3º ano falaram mais róticos em final de nomes do que os alunos do 9º ano. Ademais, os alunos do 7º ano falaram mais róticos em final de verbos do que os alunos do 1º ano. Considerando-se que o aumento de R-final na fala pode refletir às influências da escrita na fala, houve indícios de que a escrita influencia tanto a pronúncia de nomes quanto de verbos. Com apoio dos Modelos de Exemplos e da TSDC, argumentamos que as formas ortográficas podem fortalecer esquemas já existentes na gramática. Assim, a letra <r> pode fortalecer a representação do R-final e, conseqüentemente, aumentar a produtividade e o uso deste padrão.

- *Discutir a natureza abstrata das vogais médias e dos róticos finais à luz de Modelos múltiplos e dinâmicos.*

Este trabalho analisou as vogais pretônicas e postônicas finais, bem como os róticos finais em verbos e em nomes na fala e na escrita infantojuvenil em busca de evidências sobre a natureza abstrata das categorias linguísticas. Os resultados alcançados indicam que os erros ortográficos em contextos pretônico e postônico final, assim como em verbos e em nomes perduram até estágios diferentes do percurso escolar. Esses resultados permitem a interpretação de que as crianças acessam categorias linguísticas múltiplas durante a escolarização, o que motiva diferentes trajetórias de aprendizagem. Além disso, os dados sugerem que a escrita causa diferentes efeitos na fala dos indivíduos. Esses resultados trazem indícios de que o sistema alfabético-ortográfico interage de maneira não-linear com o sistema linguístico, causando efeitos variados na produtividade dos esquemas múltiplos.

- *Discutir as relações entre oralidade e escrita, à luz de Modelos múltiplos e dinâmicos.*

Este trabalho analisou as relações entre oralidade e escrita, visando encontrar evidências que sustentem a ideia de que uma mesma letra - <e, o, r> - pode estar associada a esquemas múltiplos. Os resultados encontrados confirmaram essa ideia. Os aprendizes gerenciam categorias linguísticas múltiplas para escrever uma mesma letra em diferentes contextos. Além disso, o aprendizado das letras parece causar diferentes efeitos na produção oral de diferentes categorias linguísticas. Assim sendo, os dados confirmam a tese de que *as letras são símbolos discretos que*

se associam a esquemas múltiplos, que emergem da interação complexa entre palavras, sequências fonéticas, morfológicas e ortográficas. A fim de ilustrar a proposta deste trabalho, desenhamos um diagrama para representar as relações múltiplas entre a oralidade e a escrita. Vejamos:

**Figura 13:** Relações múltiplas entre oralidade e escrita



**Fonte:** elaboração própria.

Os círculos ovais nos digramas ilustram o conjunto de exemplares relacionados às quatro categorias avaliadas neste trabalho. Cada conjunto de exemplares armazena informações morfológicas, sociais, fonéticas, semânticas específicas de cada padrão avaliado. Além disso, há exemplares ortográficos, que armazenam o conhecimento da escrita. Os exemplares relacionados às letras <e, o, i, u> são parte do esquema das vogais pretônicas e das vogais postônicas finais. Os exemplares relacionados à letra <r> se associam tanto ao esquema dos verbos de infinitivo, quanto ao esquema dos nomes terminados com R-final. Em suma, uma mesma letra pode estar relacionada a esquemas linguísticos múltiplos.

Esperamos, de modo geral, que a presente pesquisa possa promover reflexões acerca das diferentes relações estabelecidas entre a fala e a escrita. Além disso, esperamos que este trabalho consiga destacar a relevância em se considerar a multiplicidade e a dinamicidade da gramática para se expressar as relações entre oralidade e escrita. A tese de que as letras estão associadas à esquemas múltiplos é aplicável a vários contextos da língua portuguesa – e não apenas às vogais médias átonas e aos róticos. Por exemplo: a letra <s> pode se associar a esquemas distintos a depender do contexto silábico que ocupa. Em final de palavras, por exemplo, o <s> está relacionado à flexão de número; por vezes, a sibilante final pode ser cancelada na oralidade. A letra <l> também se relaciona a diferentes esquemas a depender do contexto. Em final de palavras, o <l> pode ser aprendido com auxílio da morfologia lexical (sal > saleiro). Essa estratégia não se aplica ao <l> em final de sílabas, como em <soldado>, por exemplo. Em suma, a ideia de que uma mesma letra pode se associar a categorias múltiplas é capaz de enriquecer as explicações sobre as estratégias construídas pelos aprendizes e sobre as dificuldades enfrentadas ao longo do percurso de aprendizado da escrita.

As limitações deste trabalho são apresentadas como sugestões para pesquisas futuras. A primeira sugestão diz respeito ao uso de metodologia longitudinal para a avaliação das relações entre fala e escrita. Esta pesquisa se limitou a investigar dados transversais de fala e de escrita. Esse tipo de análise nos permitiu comparar amostras de dados de diferentes alunos, mas nos impossibilitou de avaliar possíveis alterações no comportamento oral e gráfico de um mesmo indivíduo. Isto é, não nos foi possível afirmar que os alunos alteram a fala após o aprendizado da escrita. Apontamos apenas evidências de que alunos de anos escolares diferentes produzem índices diferentes de variantes orais. Análises longitudinais poderão investigar se uma criança pré-alfabetizada, por exemplo, altera sua fala ao longo da alfabetização. Esse tipo de análise poderá confirmar, com mais ênfase, os efeitos da escrita na organização do conhecimento linguístico. Além disso, as análises longitudinais poderão verificar se um único aluno traça diferentes trajetórias de aprendizado de um mesmo padrão ortográfico em diferentes contextos.

A segunda sugestão de trabalho futuro é a de ampliação do conjunto de dados deste estudo. Com o aumento do número de participantes, será possível validar melhor os resultados e aumentar o índice de confiabilidade nas generalizações encontradas. Para possibilitar a replicação deste estudo, disponibilizamos, nos Apêndices D e E, os materiais experimentais e o *script* elaborado para a análise dos dados desta pesquisa.

Por fim, a última sugestão tem relação com aplicação prática da proposta teórica desenvolvida nesta tese. As descrições teóricas podem dar suporte para o desenvolvimento de metodologias e de livros didáticos. Trabalhos futuros, no âmbito da Linguística Aplicada, podem propor atividades sobre ortografia que considerem a multiplicidade do conhecimento linguístico.

Alguns livros didáticos ainda priorizam o ensino mecânico e transmissivo dos aspectos ortográficos do português. Frequentemente, as atividades trabalham com memorização, ao invés de impulsionarem os alunos na construção reflexiva do conhecimento (COUTO, 2020). Para exemplificar, vejamos um exemplo de atividade encontrada no livro didático *Ápis*, do 3º ano do EF (BORGATTO; BERIN; MARCHEZI, 2017):

**Figura 14:** Atividade didática sobre a letra <r>


**Palavras em jogo**

**Letra R em final de sílaba**

1 Leia e observe a sílaba destacada.


ervilhas

As imagens não estão representadas em proporção.




Shutterstock.com

Outras palavras têm sílabas com R final. Preencha os espaços de acordo com as imagens e leia em voz alta as palavras que encontrou.



Draigo.com.br

BAR CO



Shutterstock.com

TAM BOR

O R em final de sílaba pode estar no início, no meio ou no fim das palavras.

**Fonte:** Borgatto; Berin; Marchezi (2017, p. 133)

A atividade na Figura 14 pede aos alunos que se atentem para a primeira sílaba da palavra “ervilha”. Em seguida, os alunos são orientados a preencher os espaços em branco com sílabas também terminadas por <r>. O grande benefício desse tipo de atividade é levar o aluno a refletir sobre as semelhanças ortográficas entre as palavras. Considerando-se que as crianças criam generalizações linguísticas a partir da associação entre itens similares (BYBEE, 2001), é interessante propor atividades que os incentivem a perceber os padrões ortográficos em comum. Vale ressaltar, contudo, que as palavras a serem completadas na atividade têm diferenças: na palavra “barco”, o <r> ocupa o final de sílaba interna; na palavra “tambor”, o <r> está em final de palavras. Como foi visto na análise dos dados desta pesquisa, o R-final de nomes pode ser cancelado na oralidade. Em contrapartida, o cancelamento do rótico em final de sílaba é incipiente em algumas variedades como a de Belo Horizonte – MG (HUBACK, 2006). Assim, é necessário

considerar que a letra <r> pode estar relacionada a variantes orais distintas e, conseqüentemente, pode haver desafios distintos no aprendizado. O olhar para as múltiplas relações entre oralidade e escrita pode enriquecer a atividade. Trabalhos futuros podem propor atividades que orientem os aprendizes a refletir sobre as diferenças orais das palavras com o mesmo padrão ortográfico e sobre as particularidades de cada tipo silábico.

Esperamos, enfim, que a discussão proposta nesta tese possa colaborar para o enriquecimento das análises teóricas sobre as relações entre oralidade e escrita. Mais do que isso, esperamos que esta tese possa impulsionar reflexões acerca de práticas pedagógicas que considerem a multiplicidade do conhecimento linguístico e as relações múltiplas entre oralidade e escrita.

## 8. REFERÊNCIAS

- ADAMOLI, M. A. **Um estudo sobre o estatuto fonológico dos ditongos variáveis [aj] e [ej] do PB a partir de dados orais e ortográficos produzidos por crianças de séries iniciais.** 2012. 196 p. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade Federal de Pelotas, 2012.
- ALVARENGA, D.; SOARES, M. B.; OLIVEIRA, M. A.; NASCIMENTO, M. **Da forma sonora da fala à forma gráfica da escrita uma análise linguística do processo de alfabetização.** Caderno de Estudos Linguísticos, Campinas, n. 16, p. 5-30, jan./jun. 1989.
- ALVARENGA, D.; OLIVEIRA, M.A. de. **Canonicidade silábica e aprendizagem da escrita.** Revista de Estudos da Linguagem. Belo Horizonte: Faculdade de Letras da UFMG, v. 6, n. 5, 1997, p. 127-158. <https://doi.org/10.17851/2237-2083.5.1.127-158>
- ASSIS, A. **A emergência de consoantes finais no português brasileiro na microrregião de Araguaína/Tocantins.** 2017. 162 f. (Tese de doutorado em Estudos Linguísticos) Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.
- AZEREDO, J. C. S. **Gramática Houaiss da Língua Portuguesa.** São Paulo: Publifolha, 2018.
- BAGNO, Marcos. **Preconceito linguístico: o que é, como se faz?** 49. ed. São Paulo: Loyola, 2007.
- BAGNO, Marcos. **Gramática pedagógica do português brasileiro.** São Paulo: Parábola Editorial, 2011.
- BARTLETT, F. C. **Remembering. A Study in Experimental and Social Psychology.** Cambridge: Cambridge University Press, 1932
- BATTISTI, E. **Elevação das vogais médias pretônicas em sílaba inicial de vocábulo na fala gaúcha.** 1993. 283 f. Dissertação (Mestrado em Letras), Universidade Federal do Rio Grande do Sul, Porto Alegre, 1993.
- BECHARA, E. **Moderna gramática portuguesa,** 37. ed. Rio de Janeiro: Lucerna, 2000.
- BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHRISTIANSEN, M. H.; CROFT, W.; ELLIS, N. C.; HOLLAND, J.; KE, J.; LARSEN-FREEMAN, D.; SCHOENEMANN, T.; “The Five Graces Group”. **Language Is a Complex Adaptive System: Position Paper.** Language Learning, v. 59, n. supl., p. 1-26, 2009.
- BELGA, J. S. **Hipercorreção na escrita acadêmica: uma análise de textos de alunos da graduação em direito.** 2019. 178 f. Dissertação (Mestrado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2019.
- BIDERMAN, M. T. C. **Dicionário Ilustrado de Português.** São Paulo: Ática; 2004. 344 p.
- BISOL, L. **Harmonização vocálica: uma regra variável.** 1981. 333 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.
- BISOL, L. **Vowel harmony: A variable rule in Brazilian Portuguese.** Language Variation and Change, v. 1, n. 2, p. 185-198, 1989.
- BISOL, L. **A Neutralização das átonas.** Revista Letras, n. 61, esp., p. 273-283, 2003. Disponível em: <http://revistas.ufpr.br/letras/article/view/2891>. Acesso em: 10 jul. 2021.
- BLOCH, B. **A set of postulates for phonemic analysis.** Language, v. 24, n. 1, p. 3-46, 1948.
- BLOOMFIELD, L. **Language.** New York: Holt, Rinehart and Winston, 1996[1933]. 564 p.
- BOECKX, C. **Language in Cognition: Uncovering Mental Structures and the Rules Behind Them.** Oxford: Wiley, 2010. p. 266.

- BOD, R. B.; HAY, J.; JANNEDY, S. **Introduction**. In: BOD, R.; HAY, J. B.; JANNEDY, S. (ed.). *Probability theory in Linguistics*. Cambridge, MA: MIT Press, 2003. p. 1-10.
- BOHN, G.; SOUZA, M. **Hipercorreções na escrita como evidência da gramática fonológica da criança**. Work. Pap. Linguíst., Florianópolis, v. 18, n. 2, p. 153-170, ago./dez., 2017.
- BORTONI, S.; GOMES, C.; MALVAR, E. **A variação das vogais médias pretônicas no português de Brasília: um fenômeno neogramático ou de difusão lexical?** Revista de Estudos da Linguagem, v. 1, n. 1, p. 9-30, 1992.
- BORTONE, Marcia Elizabeth; ALVES, Scheyla Brito. **O fenômeno da hipercorreção**. In: BORTONI-RICARDO, Stella Maris *et al.* (Org.). *Por que a escola não ensina gramática assim?* São Paulo: Parábola Editorial, 2014
- BOT DE, K.; LOWIE, W.; VERSPOOR, M. **A Dynamic Systems Theory approach to second language acquisition**. Bilingualism: Language and Cognition, v. 10, n. 1, p. 7-21, 2007.
- BROWN, G. **Teaching the spoken language**. In: ASSOCIATION INTERNATIONALE DE LINGUISTIC APLIQUÉE. Bruxelas, Proceedings II: Lecture. Encontro de 1981.
- BYBEE, J. **The emergent lexicon**. Chicago Linguistic Society, v. 34, p. 421-435, 1998.
- BYBEE, J. L. **Phonology and language use**. Cambridge, UK; New York: Cambridge University Press, 2001. (Cambridge studies in linguistics).
- BYBEE, J. **Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change**. Language Variation and Change, v. 14, n. 3, p. 261-290, 2002.
- BYBEE, J. **Language, Usage and Cognition**. Cambridge: Cambridge University Press, 2010. Disponível em: <http://ebooks.cambridge.org/ref/id/CBO9780511750526>. Acesso em: 10 jul. 2021.
- BYBEE, J. **Grammatical and lexical factors in sound change: A usage-based approach**. Language Variation and Change, 29 (2017), 273–300. Cambridge University Press, 2017, doi:10.1017/S0954394517000199
- BYBEE, J.; BECKENER, C. **Emergence at the cross-linguistic level attractor dynamics in language change**. In: MACWHINNEY, B.; O'GRADY, W. (ed.). *The handbook of language emergence*. Oxford: Wiley, 2015. p. 183-200.
- BYRNE, Brian. **The foundation of literacy: the child's acquisition of the alphabetic principle**. UK: Psychology Press, 1998.
- CAGLIARI, L. C. **Alfabetização e Linguística**. São Paulo: Scipione, 1989.
- CAGLIARI, L. C. **Análise fonológica: introdução à teoria e à prática, com especial destaque para o modelo fonêmico**. Campinas: Mercado de Letras, 2002.
- CHAVES, I. **Panorama dos estudos das vogais pretônicas no português do Brasil: meta - análise das pesquisas desenvolvidas de 1980 a 2012**. Tese de Doutorado. Universidade Federal de Minas Gerais, 2013.
- CALDAS, V; CALLOU, D. **O apagamento do rótico em posição de coda silábica: indicadores linguísticos e sociais**. In: GELNE, 2014. Disponível em: <http://www.gelne.com.br/arquivos/anais/gelne-2014/anexos/256.pdf>. Acesso em: 26 jan. 2022.
- CALLOU, D.; MORAES, J.; LEITE, Y. **Variação e diferenciação dialetal: a pronúncia do /r/ no português do Brasil**. In: KOCH, I. G. V. (org.). *Gramática do português falado: desenvolvimentos*. Campinas: Unicamp/FAPESP, 1996. p. 465-493. v. 6.
- CAMÂRA JR., M. **Estrutura da Língua Portuguesa**. 47. ed. Petrópolis, RJ: Vozes, 2015 [1970].
- CAMERON, L.; LARSEN-FREEMAN, D. **Complex Systems and Applied Linguistics**. International Journal of Applied Linguistics, v. 17, n. 2, p. 226-240, 2007.



CARAVOLAS, M. **Spelling Development in Alphabetic Writing Systems: A Cross-Linguistic Perspective.** *European Psychologist*, 9(1), 3–14, 2004. <https://doi.org/10.1027/1016-9040.9.1.3>

CARAVOLAS, M; LERVAG, A; MOUSIKOU, P; EFRIM, C; LITAVSKY, M; ONOCHIE-QUINTANILLA, E; SALAS, N; SCHÖFFELOVÁ, M; DEFIOR, S; MIKULAJOVÁ, M; SEIDLOVÁ-MÁLKOVÁ G; HULME, C. **Common patterns of prediction of literacy development in different alphabetic orthographies.** *Psychol Sci.* 2012 Jun;23(6):678-86. doi: 10.1177/0956797611434536. Epub 2012 May 3. PMID: 22555967; PMCID: PMC3724272.

CARAVOLAS, M; SAMARA, A. **Learning to read and spell words in different writing systems.** In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of Reading*, 2015, (pp. 326–343). Oxford University Press.

CARRERA-SABATÉ, J. **Some connections between linguistic change and the written language: The behavior of speakers aged 3 to 20.** *Language Variation and Change*, v. 18, p. 15-34, 2006.

CASALIS, S.; LOUIS-ALEXANDRE, M. F. **Morphological analysis, phonological analysis and learnig to read French: a longitudinal study.** *Reading and Writing*, v. 12, p. 303-335. 2000.

CARLISLE, J. F. **Morphological awareness and early reading achievement.** In: FELDMAN, L. B. (ed.). *Morphological aspects of language processing*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1995. p. 189-209.

CARVALHO, G. T. **Grafema.** In: FRADE, I. C. A. S.; VAL, M. G. C. G.; BREGUNCI, M. G. C. *Glossário Ceale de termos de Alfabetização, leitura e escrita para educadores*. Belo Horizonte: CEALE/Faculdade de Educação da UFMG, 2014. Disponível em: <http://ceale.fae.ufmg.br/app/webroot/glossarioceale/verbetes/grafema&gt>. Acesso em: 10 jul. 2021.

CARVALHO, G. T. **Sistemas de escrita.** In: FRADE, I. C. A. S.; VAL, M. G. C. G.; BREGUNCI, M. G. C. *Glossário Ceale de termos de Alfabetização, leitura e escrita para educadores*. Belo Horizonte: CEALE/Faculdade de Educação da UFMG, 2014. Disponível em: <http://ceale.fae.ufmg.br/app/webroot/glossarioceale/verbetes/grafema&gt>. Acesso em: 10 jul. 2021.

CASTELO BRANCO, A. **O apagamento do rótico em coda final em produções escritas no ensino fundamental II.** 2020. 113 p. Dissertação (Mestrado Profissional em Letras) – Universidade Federal de Uberlândia, 2020. Disponível em: <https://repositorio.ufu.br/handle/123456789/29928>. Acesso em: 10 jul. 2021.

CESAR, H. H. F. **Acréscimo do grafema <r> em coda silábica: intervenção para casos de hipercorreção.** 2017. 135 p. Dissertação (Mestrado Profissional em Letras). Instituto de Ciências Humanas e Sociais, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2017.

CHEVROT, J-P.; BEAUD, L.; VARGA, R. **Developmental data on a French sociolinguistic variable: Post-consonantal word-final /R/.** *Language Variation and Change*, v. 12, n. 3, p. 295-319, 2000.

CHOMSKY, N. **O conhecimento da língua: sua natureza, origem e uso.** Trad. Anabela Gonçalves e Ana Teresa Alves. Lisboa: Caminho, 1994.

COULMAS, F. **The Writing Systems of the World.** Blackwell, Oxford, 1989

CRISTÓFARO SILVA, T. **Modelos Multirrepresentacionais em Fonologia.** In: Renata Coelho Marchezan; Arnaldo Cortina (org). (Org.). *Os fatos da linguagem, esse conjunto heteróclito*. 1 ed. Araraquara: FCL-UNESP Laboratório Editorial. São Paulo: Cultura Acadêmica, p. 171-186. 2006.

CRISTÓFARO-SILVA, T. **Representações mentais na aquisição da linguagem oral e escrita.** Revista da ABRALIN, v. 10, n. 4, 2011. Disponível em: <http://revistas.ufpr.br/abralin/article/view/32427>. Acesso em: 10 jul. 2021.

CRISTÓFARO SILVA, T. **Fonética e Fonologia do Português: Roteiro de Estudos e Guia de Exercícios.** 10 ed. São Paulo: Contexto, 2013.

CRISTÓFARO-SILVA, T. **Trajatórias fonológicas: evolução e complexidade.** Revista Linguística/Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro, v. esp., p. 215-229, dez. 2016. Disponível em: <http://www.letras.ufrj.br/poslinguistica/revistalinguistica>. Acesso em: 10 jul. 2021.

CRISTÓFARO-SILVA, T. **Dicionário de fonética e fonologia.** 1ºed. São Paulo: Contexto, 2017

CRISTÓFARO SILVA, T.; GRECO, A. **Representações Fonológicas: contribuições da oralidade e da escrita.** Letras de Hoje, v. 45, p. 87-93, 2010.

CRISTÓFARO-SILVA, T.; GUIMARÃES, D. O. **A aquisição da linguagem falada e escrita: o papel da consciência linguística.** Letras de Hoje, v. 48, n. 3, p. 316-323, 2013.

CRISTÓFARO-SILVA, T.; GOMES, C. A. **Representações múltiplas e organização do componente fonológico.** Fórum Linguístico, Florianópolis, v. 4, n. 1, p. 147-177, jul. 2004.

CRYSTAL, D. **Dicionário de Linguística e Fonética.** Trad. e adapt. Maria Carmelita Pádua Dias. Rio de Janeiro: Jorge Zahar, 1988. v. 1.

COUTO, A. L. S.; OLIVEIRA-GUIMARÃES, D. M. L. O. **O acento gráfico no Ensino Fundamental II: reflexões sobre os textos dos alunos e práticas possíveis.** Veredas – Revista de Estudos Linguísticos, v. 24, n. 3, p. 178-194, 2020.

COSTA, G. B. **O apagamento do rótico em coda silábica na escrita de estudantes catuenses.** Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura, ano 5, n. 10, 2009.

COSTA, J. **Descrevendo línguas brasileiras: Yaathe, a língua dos índios fulni-ô.** Revista do GELNE, [S. l.], v. 17, n. 1/2, p. 93–111, 2016. Disponível em: <https://periodicos.ufrn.br/gelne/article/view/10176>. Acesso em: 5 jan. 2023.

DEACON, S.; BRYANT, P. **What young children do and do not know about the spelling of inflections and derivations.** Developmental Science, v. 8, n. 6, p. 583-594, 2005.

DIAS, E. SEARA, I; **Redução e Apagamento de Vogais Átonas Finais na fala de crianças e adultos de Florianópolis: uma análise acústica,** Letrônica, Porto Alegre, v. 6, n. 1, p. 71-93, jan./jun., 2013

DUBOIS, Mathée Giacomo *et al.* **Dicionário de linguística.** Trad. Frederico Pessoa de Barros *et al.* São Paulo: Cultrix, 1995

ELLIS, N. C.; LARSEN-FREEMAN, D. (ed.) **Language as a Complex Adaptive System.** Language Learning, n. 59, supl. 1, 275 p., Dec. 2009.

FARACO, C. A. **Linguística histórica: uma introdução ao estudo da história das línguas.** São Paulo: Ática, 1991.

FARACO, C. A. **Linguística histórica: uma introdução ao estudo da história das línguas.** São Paulo: Parábola Editorial, 2005.

FARACO, C. A. **Escrita e Alfabetização.** São Paulo: Contexto, 2015.

FERREIRO, E. **Literacy development: Psychogenesis.** In: GOODMAN, Y (Ed.) *How children construct literacy; Piagetian perspectives* (pp.12-25). Newark, DE: International Reading Association, 1990

- FISCHER, Steven R. **História da Escrita**; tradução Mirna Pinsky. São Paulo: Editora UNESP, 2009.
- FOULKES, P.; DOCHERTY, G. **The social life of phonetics and phonology**. *Journal of Phonetics*, v. 34, n. 4, p. 409-438, 2006.
- FONTES-MARTINS, R. M. **A organização do componente fonológico e o comportamento do indivíduo**. 2007. 275 p. Tese (Doutorado) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.
- FONTES, J.S. **As vogais pretônicas do português antigo a partir da variação gráfica no Cancioneiro Geral de Garcia de Resende**. *ESTUDOS LINGÜÍSTICOS*, São Paulo, 43 (1): p. 333-346, jan-abr 2014.
- FRANKE, T. M.; HO, T.; CHRISTIE, C. A. **The Chi-Square Test: Often Used and More Often Misinterpreted**. *American Journal of Evaluation*, v. 33, n. 3, p. 448-458, 2012.
- FREIRE, J. B.; HORA, D. **Apagamento do /r/ em final de palavras: o que dizem os textos escolares**. *Iniciação & Formação Docente*, v. 6, n. 1, p. 1, 2019.
- GRECO, A. **O alçamento de vogais médias pretônicas na fala de crianças de Belo Horizonte: uma investigação acerca da influência retroalimentadora da escrita na oralidade**. 2009. Monografia (Linguística) – Faculdade de letras, Universidade Federal de Minas Gerais, 2009.
- GOODMAN, K; GOODMAN, Y. **Learning to Read is Natural**. *In: RESCNIK, L. B.; WEAVER, P. A. (ed.). Theory and Practice of Early Reading*. Hillsdale: Lawrence Erlbaum, 1979. p. 137-54. v. 1.
- GODOY, M. C. **Introdução aos modelos lineares mistos para os estudos da linguagem**. PsyArXiv. 2019, <https://doi.org/10.17605/OSF.IO/9T8UR>
- GODOY, M.; NUNES, M. **Estatística para as Ciências da Linguagem**. (Em preparação)
- GOMBERT, J. E. **Metalinguistic development**. London, UK: Harvester Wheatsheaf, 1992.
- GOMBERT, J. **Atividades metalinguística e aquisição da leitura**. *In: MALUF, M. R. (org.). Metalinguagem e Aquisição da escrita*. São Paulo: Casa do Psicólogo, 2003. p. 19-64.
- GOMES DE MORAIS, A. **Ortografia: ensinar e aprender**. São Paulo: Editora Ática, 1998.
- GOMES DE MORAIS, A. **A apropriação do sistema de notação alfabética e o desenvolvimento de habilidades de reflexão fonológica**. *Letras de Hoje*, Porto Alegre, v. 39, n. 3, p. 175-192, 2004.S
- GOMES DE MORAIS, A. **O desenvolvimento da consciência fonológica e a apropriação da escrita alfabética entre crianças brasileiras**. *Revista Brasileira de Alfabetização*, Vitória, ES, v. 1, n. 1, p. 59-76, 2015.
- GOMES DE MORAIS, A; TEBEROSKY, A. **Erros e transgressões infantis na ortografia do português**. "Discursos [Em linha: estudos de língua e cultura portuguesa". ISSN 0872-0738. Nº 8, 1994, p. 15-51
- GRIES, S. **Estatística com R para linguística: uma introdução prática/ Stefan Th. Gries**. Organizadora: Heliana R. Mello; tradução: Heliana R. Mello... [et.al] – Belo Horizonte: FALE/UFMG, 2019
- HAN, D, ADOLPH, K.E. **The impact of errors in infant development: Falling like a baby**. *Dev Sci*. 2021 Sep; 24(5): e13069. doi: 10.1111/desc.13069. Epub 2021 Jan 4. PMID: 33278863; PMCID: PMC8178414.
- HORA, D. **Vocalização da lateral /l/: correlação entre restrições sociais e estruturais**. *Scripta*, v. 9, n. 18, p. 29-44, 2006.

- HOUAISS, Antônio. **Dicionário Houaiss Eletrônico**. Objetiva, 2009.
- HUBACK, A. P. S. **Cancelamento do (r) final em nominais na cidade de Belo Horizonte: uma abordagem difusionista**. 2003. Dissertação (Mestrado em Linguística) – Universidade Federal de Minas Gerais, Belo Horizonte, 2003.
- HUBACK, A. P. S. **Cancelamento do (R) final em nominais: uma abordagem difusionista**. Scripta, Belo Horizonte, v. 9, n. 18, p. 11-28, 2006.
- BRASIL. IBGE – *INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA*. Censo Brasileiro de 2019. Rio de Janeiro: IBGE, 2019.
- IRAN-NEJAD, Asghar. **The Schema: a structural or a functional pattern** (Center for the Study of Reading Technical Report No. 159). University of Illinois at Urbana-Champaign, 1984
- JOHNSON, K. **Speech perception without speaker normalization: An exemplar model**. In: JOHNSON, K.; MULLENNIX, J. W. (ed.). *Talker Variability in Speech Processing*. San Diego: Academic Press, 1997.
- JOHNSON, K. **Decisions and Mechanisms in Exemplar-based Phonology**. UC Berkeley Phonology Lab Annual Reports, v. 1, p. 289-311, 2005. Disponível em: <https://escholarship.org/uc/item/7m49b843>. Acesso em: 10 jul. 2021.
- JOTA, Zélio dos Santos. **Dicionário de Linguística**. Rio de Janeiro: Presença, 1976.
- KARMILOFF-SMITH, A.; GRANT, J.; SIMS, K.; JONES, M.-C.; CUCKLE, P. **Rethinking metalinguistic awareness: representing and accessing knowledge about what counts as a word**. *Cognition*, v. 58, n. 2, p. 197-219, 1996.
- KATO, M. **No mundo da escrita: uma perspectiva psicolinguística**. São Paulo: Ática, 1990, 3ª ed.
- KATZ, L; FROST, R. **The reading process is different for different orthographies: The orthographic depth hypothesis**. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). North-Holland, 1992. [https://doi.org/10.1016/S0166-4115\(08\)62789-2](https://doi.org/10.1016/S0166-4115(08)62789-2)
- KENT, R.; READ, C. **Análise acústica da fala**. São Paulo: Cortez, 2015
- LABOV, W. **A estratificação social do (r) nas lojas de departamentos na cidade de Nova York**. In: LABOV, W, *Padrões sociolinguísticos*. Tradução: Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues de Oliveira. São Paulo: Parábola Editorial, 2008.
- LABOV, W. **A hipercorreção pela classe média baixa como fator de mudança linguística**. In: LABOV, W, *Padrões sociolinguísticos*. Tradução: Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues de Oliveira. São Paulo: Parábola Editorial, 2008.
- LADEFOGED, P.; MADDIESON, I. **The Sounds of the World's Languages**. Oxford: Blackwell, 1996.
- LANGACKER, Ronald W. **Cognitive grammar: a basic introduction**. Oxford; Nova York: Oxford University Press, 2008.
- LEMLE, M. **Guia Teórico do alfabetizador**. São Paulo: Editora ática, 1987.
- LEMOS, F. A. P. **Interferência da oralidade na escrita: o caso do registro ortográfico de e, i, o, u átonos**. *Letras de Hoje*, Porto Alegre, v. 36, n. 3, p. 465-471, 2001.
- LIMA JÚNIOR, R. M. **A influência da idade na aquisição da fonologia do inglês como língua estrangeira**. 2012. 187 f. Tese (Doutorado em Estudos Linguísticos) – Programa de Pós-Graduação em Linguística, Departamento de Linguística, Português e Línguas Clássicas, Instituto de Letras, Universidade de Brasília, Brasília, 2012.

- LIGHTFOOT, D. **Language acquisition and language change**. WIREs Cognitive Science, v. 1, n. 5, p. 677-684, 2010.
- LUCCHESI, D. **Língua e Sociedade Partidas: a polarização sociolinguística do Brasil**. São Paulo: Contexto, 2015.
- MAGALHÃES, J.; AGRELI MELO CAMPOS, R. **As vogais médias pretônicas em Minas Gerais e em São Paulo: o Rio Grande como fronteira linguística**. Revista do GELNE, [S. l.], v. 24, n. 2, p. 152–167, 2022. DOI: 10.21680/1517-7874.2022v24n2ID30061. Disponível em: <https://periodicos.ufrn.br/gelne/article/view/30061>. Acesso em: 5 fev. 2023.
- MAN, J. **A história do alfabeto**. John Man; tradução Edith Zonenschain. Rio de Janeiro: Ediouro, 2002
- MASSINI-CAGLIARI, G.; CAGLIARI, L. C. **Diante das Letras: a escrita na alfabetização**. Campinas: Mercado de Letras/Associação de Leitura do Brasil/FAPESP, 1999.
- MASSINI-CAGLIARI, G.; CAGLIARI, L. C. **Categorização gráfica e funcional na aquisição da escrita e da leitura em língua materna**. Calidoscópico, v. 2, n. 1, p. 89-94, 2004.
- MATTINGLY, I.G. **Linguistic awareness and orthographic form**. In: R. Frost, L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning*, Elsevier Science Publishers, pp. 11-26, Amsterdam, 1992.
- MAZZAFERRO, G. T.; MATZENAUER, C. L. B. **O comportamento das vogais postônicas finais na fronteira do Brasil com Uruguai**. Linguística, v. 35, n. 1, p. 57-79, 2019.
- MENDOZA-DENTON, N., HAY, J.; JANNEDY, S. **Probabilistic sociolinguistics: Beyond variable rules**. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 99–138). Cambridge, MA: MIT Press, 2003.
- MIRANDA, A. R. **Um estudo sobre a aquisição ortográfica das vogais do português**. In: Anais da ANPEDSul (UFSM), Santa Maria, 2006.
- MIRANDA, A. R. M. **A Aquisição Ortográfica Das Vogais do Português: Relações com a Fonologia e a Morfologia**. Letras, n. 36, p. 151-168, 2008.
- MIRANDA, A. R. M. **Um estudo sobre o erro ortográfico**. In: HEINING, O. L.; FRONZA, C. A. (org.). *Diálogos entre linguística e educação*. Blumenau: EDIFURB, 2010. p. 141-16. v. 1.
- MIRANDA, Ana Ruth Moresco. **A FONOLOGIA EM DADOS DE ESCRITA INICIAL DE CRIANÇAS BRASILEIRAS**. Linguística, Montevideo, v. 30, n.2, p. 45-80, 2014.
- MIRANDA, A. R. M. **Um Estudo Sobre a Natureza Dos Erros (Orto)gráficos Produzidos por Crianças dos Anos Iniciais**. Educação em Revista, v. 36, e221615, 2020.
- MOLL, K; RAMUS, F; BARTLING, J; BRUDER, J; KUNZE, S; NEUHOFF, N; *et al.* **Cognitive mechanisms underlying reading and spelling development in five European orthographies: Is English an outlier orthography?** Learning and Instruction, 29, 2014, pp. 65-77, <https://doi.org/10.1016/j.learninstruc.2013.09.003>.
- MOLLICA, Maria Cecília; BRAGA, Maria L. **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto, 2004
- MONARETTO, V. N. O. **O Apagamento da Vibrante Pós-vocálica nas Capitais do Sul do Brasil**. Revista Letras de Hoje, Porto Alegre, v. 35, n. 1, 2000.
- MORAIS, J.; CARY, L.; ALEGRIA, J.; BERTELSON, P. **Does awareness of speech as a sequence of phones arise spontaneously?** Cognition, v. 7, n. 4, p. 323-331, 1979.
- MORAIS, José. **A arte de ler**. UESP, São Paulo, 1996

- MOTA, M. M. P. E. **Explorando a relação entre consciência morfológica, processamento cognitivo e escrita.** *Estudos de Psicologia*, Campinas, v. 29, n. 1, p. 89-94, 2012.
- NAGY, W. E.; ANDERSON, R. C. **Metalinguistic awareness and literacy acquisition in different languages.** In D. A. Wagner, R. C. Venezky, & B. V. Street (Eds.), *Literacy: An international handbook*, 1999, (pp. 155–160). Boulder, CO: Westview Press.
- NÓBREGA, M. J. **Ortografia.** São Paulo: Melhoramentos, 2013.
- NOGARO, A; GRANELLA, E. **O erro no processo de Ensino e Aprendizagem.** *Revista de Ciências Humanas*, v. 5., n.5 2004
- OLIVEIRA, M. A. **Aspectos da difusão lexical.** *Revista de Estudos da Linguagem*, v. 1, n. 1, p. 31-41, 1992.
- OLIVEIRA, M. A. Reanalizando o processo de cancelamento do (r) em final de sílaba. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 6, n. 2, p. 31-58, jul./dez. 1997.
- OLIVEIRA, M. A. **A variação fonológica na perspectiva da linguagem como um sistema adaptativo complexo.** In: MAGALHÃES, J. S. (org.). *Fonologia*. Uberlândia: EDUFU, 2014.
- OLIVEIRA, M. A. **A auto-organização como mecanismo para a resolução da variação linguística.** *Cadernos de Estudos Linguísticos*. v. 58, p. 383-399, 2016.
- OSLON, D. **Literacy and Language Development.** In: WAGNER, Daniel A.; Venezky, Richard L.; STREET, Brian V. *Literacy: an International Handbook*. Oxford: Westview Press, 1999, p.132-36
- OUSHIRO, L. **Introdução à estatística para linguística.** Instituto da Linguagem, UNICAMP, Tutorial Online. Disponível em: <https://rpubs.com/oushiro/iel>. v1.0.1, 2017.
- PAIVA, M. C. **O percurso da monotongação [ey]: observações no tempo real.** In: PAIVA, M.C.; DUARTE, E. (org.). *Mudança linguística em tempo real*. Rio de Janeiro: FAPERJ/Contra Capa, 2003. p. 31-46.
- PAULA, F. V. **Conhecimento morfológico implícito e explícito na linguagem escrita.** 2007. 200 p. Tese (Doutorado em Psicologia) – Instituto de Psicologia, Universidade de São Paulo, São Paulo, 2007.
- PIERREHUMBERT, J. B. **Exemplar dynamics: Word frequency, lenition and contrast.** In: BYBEE, J. L.; HOPPER, P. J. (org.). *Typological Studies in Language*. Amsterdam: John Benjamins Publishing Company, 2001. p. 137. v. 45. Disponível em: <https://benjamins.com/catalog/tsl.45.08pie>. Acesso em: 10 jul. 2021.
- PIERREHUMBERT, J. **Probabilistic theories of phonology.** In: BOD, R.; HAY, J. B.; JANNEDY, S. (ed.). *Probability theory in linguistics*. Cambridge, MA: MIT Press, 2003. p. 177-228.
- POLLO, T. C.; TREIMAN, R.; KESSLER, B. **Uma revisão crítica de três perspectivas sobre o desenvolvimento da escrita.** *Estudos de Psicologia*, Campinas, v. 32, n. 3, p. 449-459, 2015.
- PORT, R. **How are words stored in memory? Beyond phones and phonemes.** *New Ideas in Psychology*, v. 25, n. 2, p. 143-170, 2007.
- R Core Team. **R: A Language and Environment for Statistical Computing.** Vienna: R Foundation for Statistical Computing, 2022.
- READ, C; YUN-FEI, Z; HONG-YIN, N; BAO-QING, D; **The ability to manipulate speech sounds depends on knowing alphabetic writing,** *Cognition*, Volume 24, Issues 1–2, 1986, p. 31-44, [https://doi.org/10.1016/0010-0277\(86\)90003-X](https://doi.org/10.1016/0010-0277(86)90003-X).

RENNICKE, I. **Variation and Change in the Rhotics of Brazilian Portuguese.** 2015. 355 p. Tese (Doutorado em Linguística Teórica e Descritiva) Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

RENNICKE, I. **Representação fonológica dos róticos do Português Brasileiro: uma abordagem à base de exemplares.** Scripta, Belo Horizonte, v. 20, n. 38, p. 70-97, 2016.

RICHGELS, D. J. **Schema theory, linguistic theory, and representations of Reading comprehension.** *The Journal of Educational Research*, 1982 76(1), 54-62. <https://doi.org/10.1080/00220671.1982.10885424>

RUMELHART, D.E. **The Building Blocks of Cognition.** In: SPIRO, R.J.; BRUCE, B.C.; BREWER, W.F. (orgs) *Theoretical Issues in Reading Comprehension: Perspective from Cognitive Psychology, Artificial Intelligence and Education* Hillsdale, N.J.: Laurance Erlbaum Assoc, 1980, p. 33-58

RUSSAK, S; ZERETSKY, E. **Investigating spelling across typologically diverse orthographies: An introduction,** *Written Language & Literacy*, Volume 25, Issue 1, Jun 2022, p. 1 – 10, <https://doi.org/10.1075/wll.00058.rus>

SAUSSURE, F. **Curso de linguística geral.** Organização C. Bally e A. Sechehaye. Colaboração A. Riedlinger. Tradução A. Chelini, J. P. Paes e I. Blikstein. 28. ed. São Paulo: Cultrix, 2012[1916].

SANTOS, G. F. L. **Aquisição da ortografia: influências fonéticas, fonológicas, morfológicas ou lexicais?** 2000. 151 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte, 2000.

SCHWINDT, L. C. S. **A harmonia vocálica em dialetos do sul do país: uma análise variacionista.** 1995. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1995.

SCHWINDT, L. C.; QUADROS, E.; TOLEDO, E.; GONZALES, C. **A influência da variável escolaridade em fenômenos fonológicos variáveis: efeitos retroalimentares da escrita.** *RevEL: Revista Virtual de Estudos da Linguagem*, v. 5, n. 9, ago. 2007.

SEYMOUR, P. H. K. **Early Reading Development in European Orthographies.** In: M. J. SNOWLING, M. J.; HULME, C. (ed.). *The science of reading: A handbook.* Oxford: Blackwell Publishing, 2005. p. 296-315.

SEYMOUR, P. H. K., ARO, M.; ERSKINE, J. M. **Foundation literacy acquisition in European orthographies.** *British Journal of Psychology*, 94, 143-174. 2003 doi:10.1348/000712603321661859

SMITH, F. **Compreendendo a leitura: uma análise psicolinguística da leitura e do aprender a ler.** Porto Alegre: Artes Médicas, 1989.

STANOVICH, K. E. **Putting children first by putting science first: The politics of early reading instruction.** In K. E. Stanovich (Ed.), *Progress in understanding reading: Scientific foundations and new frontiers.* New York, NY: Guilford, 2000

SOARES, M. **Letramento e Alfabetização: as muitas facetas.** *Revista Brasileira de Educação.* Online, n.25, 2003, pp.05-17.

SOARES, M. **Alfabetização e literatura.** In: *Guia da Alfabetização*, n. 2, p. 17, 2010. SOARES, M. **Alfabetização: a questão dos métodos.** São Paulo: Contexto, 2018.

SOUZA, R. **Redução de vogais altas pretônicas no português de Belo Horizonte: uma abordagem baseada na gradiência.** 2012. 146 f. (Dissertação de Mestrado em Estudos Linguísticos) Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.

- SPRENGER-CHAROLLES, L. **Linguistic Processes in Reading and Spelling: The Case of Alphabetic Writing Systems: English, French, German and Spanish.** In: Nunes, T., Bryant, P. (eds) *Handbook of Children's Literacy*. Springer, Dordrecht, 2004 [https://doi.org/10.1007/978-94-017-1731-1\\_4](https://doi.org/10.1007/978-94-017-1731-1_4)
- TENANI, L.; REIS, M. C. “E vereram felizes para sempre”: análise de grafias não convencionais de vogais pretônicas. *Verba Volant*, v. 2, n. 1. Pelotas: Editora e Gráfica Universitária da UFPel, jan.-jun. 2011.
- THELEN, E.; SMITH, L. B. **A dynamic systems approach to the development of cognition and action.** Cambridge: MIT Press, 1994.
- THELEN, E.; SMITH, L. B. **Development as a dynamic system.** *Trends in Cognitive Sciences*, v. 7, n. 8, p. 343-348, 2003.
- TRAVAGLIA, L. C. **Gramática e interação: uma proposta para o ensino de gramática.** 14. ed. São Paulo: Cortez, 2009.
- TREIMAN, R.; CASSAR, M.; ZUKOWSKI, A. “What Types of Linguistic Information Do Children Use in Spelling? The Case of Flaps.” *Child Development*, vol. 65, no. 5, 1994, pp. 1318–37. *JSTOR*, <https://doi.org/10.2307/1131501>. Accessed 30 Dec. 2022.
- TREIMAN, R.; KESSLER, B. **How children learn to write words.** New York, NY: Oxford University Press, 2014.
- TREIMAN, R. **Learning to Spell Words: Findings, Theories, and Issues.** *Scientific Studies of Reading*, v. 21, n. 4, p. 265-276, 2017.
- TREIMAN, R.; ZHANG, L. Spelling. In: HUPP, S.; JEWELL, J. (org.). **The Encyclopedia of Child and Adolescent Development.** Oxford: Wiley, 2020.
- TOMASELLO, M. **First steps toward a usage-based theory of language acquisition.** *Cognitive Linguistics*, v. 11, n. 1-2, 2000. Disponível em: <https://www.degruyter.com/document/doi/10.1515/cogl.2001.012/html>. Acesso em: 10 jul. 2021.
- TOMASELLO, M. **Constructing a Language: A Usage-Based Theory of Language Acquisition.** Harvard University Press, 2005. <https://doi.org/10.2307/j.ctv26070v8>.
- TORRES, P. F. J.; OLIVEIRA, J. M. **O apagamento do -r no final de vocábulo em produções escolares na cidade de Feira de Santana – BA.** *Cadernos do CNLF*, v. XIX, n. 1 – Fonética, Fonologia, Ortografia e Política Linguística. Rio de Janeiro: CiFEFiL, 2015.
- TUNMER, W.E., HERRIMAN, M.L. **The Development of Metalinguistic Awareness: A Conceptual Overview.** In: Tunmer, W.E., Pratt, C., Herriman, M.L. (eds) *Metalinguistic Awareness in Children.* Springer Series in Language and Communication, vol 15. Springer, Berlin, Heidelberg, 1984 [https://doi.org/10.1007/978-3-642-69113-3\\_2](https://doi.org/10.1007/978-3-642-69113-3_2)
- VELOSO, J. **Da Influência do Conhecimento Ortográfico sobre o Conhecimento Fonológico.** 2003. 521 p. Dissertação (Doutorado em Linguística), Universidade do Porto, Porto, Portugal, 2003.
- VELOSO, J. **A Língua na Escrita e a Escrita da Língua. Algumas considerações gerais sobre transparência e opacidade fonêmicas na escrita do português e outras questões.** *Da Investigação às Práticas. Estudos de Natureza Educacional.* [Escola Superior de Educação de Lisboa, Centro Interdisciplinar de Estudos Educacionais]. Vol. VI, Nº 1, 2005, pp. 49-69
- VIEGAS, M. C. **O acento de vogais médias pretônicas e os itens lexicais.** *Revista de Estudos da Linguagem*, v. 3, n. 2, p. 101-122, 1995.



VIEGAS, M. C. **O alçamento das vogais médias pretônicas e os itens lexicais**. 2001. Tese (Doutorado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2001.

VIEIRA, M. J. **Neutralização das vogais médias postônicas**. 1994. 110 f. Dissertação (Mestrado em Letras: Língua Portuguesa) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 1994.

VIEIRA, M. J. B. As vogais médias postônicas: uma análise variacionista. *In*: BISOL, L.; BRESCANCINI, C. **Fonologia e Variação**: recortes do português brasileiro. Porto Alegre: EDIPUCRS, 2002.

## APÊNDICES

### Apêndice A: Termo de Consentimento Livre e Esclarecido

Prezado(a) responsável,

o(a) aluno(a) de sua responsabilidade está sendo convidado(a) como voluntário(a) a participar do projeto **Percursos de aquisição da Linguagem: Oralidade e Escrita em L1 e L2 (PAL)**, coordenado pela professora Dra. Thaís Cristóforo Silva. O objetivo do projeto é investigar a fala e a escrita de crianças do 1º ao 9º anos do Ensino Fundamental, da cidade de Belo Horizonte, visando ao amplo conhecimento da linguagem oral e escrita infantil em língua materna e estrangeira. Pedimos a sua autorização para a coleta, o armazenamento, a utilização e o descarte de dados de fala e de escrita do(a) educando(a) seu(sua) dependente.

**Esta pesquisa tem os seguintes objetivos específicos:** (1) mapeamento fonológico do vocabulário do português infantil; (2) mapeamento lexical do vocabulário do português infantil; (3) mapeamento ortográfico de material escrito do português infantil; (4) mapeamento fonológico e lexical em estágios iniciais da aquisição de língua estrangeira. A questão norteadora da pesquisa é: Como se dá o percurso de aquisição da linguagem oral e escrita de crianças brasileiras? Em busca de alcançar os objetivos e responder à questão, adotaremos os seguintes procedimentos: gravação da voz e coleta da escrita do(a) educando(a) pelo qual você é responsável. Tanto a gravação da voz quanto a coleta da escrita serão realizadas virtualmente, por meio da plataforma Google Meet. A gravação da voz será realizada em uma sala com mesa e cadeira, para proporcionar conforto durante o tempo de coleta. Os instrumentos utilizados na gravação são: um gravador de voz e um computador. A gravação terá tempo mínimo de 10 minutos e máximo de 20 minutos. Seu(sua) dependente poderá interromper a gravação de voz a qualquer momento, caso sinta necessidade. Os dados de escrita serão coletados também virtualmente, na mesma reunião online em que será realizada a coleta dos dados de fala. O seu(sua) dependente estará sentado em uma cadeira durante toda a coleta e poderá se hidratar, se alimentar e interromper a coleta a qualquer momento. Após a coleta, os dados de fala e de escrita serão armazenados em um servidor de uso exclusivo da pesquisadora. Em nenhum momento a sua identidade pessoal e/ou a identidade do estudante por quem você é responsável será exposta.

**Os riscos envolvidos na pesquisa** são: (1) possível cansaço ao participar da gravação de voz e da atividade de escrita; (2) possível constrangimento em relação à utilização da norma-padrão da língua portuguesa e das línguas estrangeiras e (3) possível problema na conexão da internet no dia da coleta. Para minimizar os riscos em (1), os participantes da pesquisa podem interromper a coleta em caso de cansaço ou de desconforto. Para minimizar os riscos em (2), os participantes poderão falar e escrever normalmente, sem qualquer julgamento por parte da pesquisadora. Para minimizar os riscos em (3), o encontro online será reagendado em caso de perda de conexão sem que haja quaisquer penalidades ao participante.

**A pesquisa contribuirá** com a descrição e análise da pronúncia e da escrita do Português Brasileiro, do Inglês e do Espanhol contemporâneos. Os resultados poderão beneficiar várias áreas do conhecimento, como a linguística, a educação e a saúde. Como desfecho primário deste projeto temos a ampliação do conhecimento sobre a sonoridade e a sobre escrita infantis contemporâneas de língua materna e estrangeiras. O desfecho secundário consiste em informar a população do percurso de aquisição de língua materna e estrangeira. Sendo assim, os participantes voluntários têm papel primordial no desenvolvimento desta e de outras investigações na área de aquisição da linguagem.

**Para participar deste estudo**, você não terá nenhum custo, nem receberá qualquer vantagem financeira. Apesar disso, caso sejam identificados e comprovados danos provenientes desta pesquisa, você tem assegurado o direito à indenização. Você terá o esclarecimento sobre a pesquisa em qualquer aspecto que desejar e estará livre para autorizar ou recusar a participação do estudante, por quem você é responsável, a qualquer tempo e sem quaisquer prejuízos, valendo a desistência a partir da data de formalização desta.

**A participação, portanto, é voluntária** e a recusa não acarretará qualquer penalidade ou modificação na rotina e na avaliação escolar. Os resultados obtidos pela pesquisa estarão à sua disposição e você poderá acessá-los a qualquer momento, obtendo uma cópia gratuita gravada em um CD. Como dito, não haverá identificação e nem divulgação dos dados de nenhum participante em nenhuma publicação que possa resultar da pesquisa. A utilização do material coletado está vinculada somente a esta pesquisa, mas, caso você autorize, os dados poderão ser utilizados em pesquisas futuras.

Este termo de consentimento terá duas vias originais, sendo que uma será arquivada pelos pesquisadores responsáveis e a outra será fornecida a você. Os dados de voz e de escrita, utilizados na pesquisa, ficarão arquivados com o pesquisador responsável por um período de 5 (cinco) anos até 10 (dez) anos, na sala 3104 da Faculdade de

Letras da UFMG e, após esse tempo, serão destruídos. Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, atendendo à legislação brasileira (Resolução 466/2012, do Conselho Nacional de Saúde, Resolução 510/2016 e suas complementares), utilizando as informações somente para fins acadêmicos e científicos.

Eu, \_\_\_\_\_, portador (a) do documento de Identidade \_\_\_\_\_, responsável pelo (a) aluno (a) \_\_\_\_\_, matriculado na escola \_\_\_\_\_, no ano \_\_\_\_\_, fui informado (a) dos objetivos, métodos, riscos e benefícios da pesquisa intitulada de “**Percursos de aquisição da linguagem: Oralidade e Escrita em L1 e L2 (PAL)**”, de maneira clara e detalhada, e esclareci minhas dúvidas. Sei que a qualquer momento poderei solicitar novas informações e modificar minha decisão de autorizar a participação do (a) educando (a), por quem sou responsável, se assim o desejar.

- Concordo que os **dados de voz** do educando, por quem sou responsável, sejam utilizados nesta pesquisa.
- Concordo que os **dados de escrita** do educando, por quem sou responsável, sejam utilizados nesta pesquisa.
- Concordo que os **dados de voz em língua estrangeira** do educando, por quem sou responsável, sejam utilizados nesta pesquisa.
- Concordo que os dados do educando, por quem sou responsável, possam ser utilizados em outras pesquisas, mas serei comunicado pelo pesquisador novamente e assinarei outro termo de consentimento livre e esclarecido que explique para que será utilizado o material

*Rubrica da pesquisadora responsável:* \_\_\_\_\_  
*Rubrica do (a) responsável pelo (a) aluno (a):* \_\_\_\_\_

Declaro que concordo e autorizo a participação do (a) estudante, por quem sou responsável, nesta pesquisa. Recebi uma via original deste Termo de Consentimento Livre e Esclarecido assinado por mim e pelos pesquisadores, que me deram a oportunidade de ler e esclarecer todas as minhas dúvidas.

\_\_\_\_\_  
 Local e data \_\_\_\_\_  
 Nome completo do responsável pelo (a) aluno (a) \_\_\_\_\_  
 Assinatura do responsável pelo (a) aluno (a)

**Nome completo da Pesquisadora Responsável:** Thaís Cristóvão Alves da Silva  
 Endereço: Av. Presidente Antônio Carlos, 6627, Bairro: Pampulha. CEP: 31270-901 /Belo Horizonte – MG  
 Telefones: (31) 3409 – 6025 E-mail:  
 thaiscristofaro@gmail.com

\_\_\_\_\_  
 Local e data \_\_\_\_\_  
 Assinatura da pesquisadora responsável

Em caso de dúvidas, com respeito aos aspectos éticos desta pesquisa, você poderá consultar:

**COEP-UFMG - Comissão de Ética em Pesquisa da UFMG**  
 Av. Antônio Carlos, 6627. Unidade Administrativa II - 2º andar - Sala 2005.  
 Campus Pampulha. Belo Horizonte, MG – Brasil. CEP: 31270-901.  
 E-mail: coep@prpq.ufmg.br. Tel: 34094592.

Apêndice B: TALE

**TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO PARA COLETA DE DADOS DE L1**

Prezado (a) aluno (a),

Você está sendo convidado (a) a participar como voluntário (a) do projeto **Percursos de aquisição da Linguagem: Oralidade e Escrita em L1 e L2**. Queremos saber como alunos de Belo Horizonte falam e escrevem textos e palavras. Precisamos de produções orais e escritas de alunos do 1º ao 9º anos. Se você concordar em participar da pesquisa, sua voz será gravada e você deverá escrever um texto e/ou algumas palavras. Seu nome e seus dados nunca serão expostos.

**A pesquisa contribuirá** com a ampliação do conhecimento sobre a fala e escrita de alunos do Ensino Fundamental. Os resultados ajudarão os pesquisadores e professores a entender como o português brasileiro é falado e quais são as dificuldades percurso de aprendizado de escrita. Sendo assim, você, aluno(a), têm um papel importante no desenvolvimento desta pesquisa.

**Os riscos envolvidos na pesquisa são:** (1) possível cansaço ao participar das atividades; (2) possível dúvida em relação à utilização da língua-padrão; (3) possível falha na conexão da internet. Se você se sentir cansado ou desconfortável, você pode pedir para encerrarmos a atividade. Não se preocupe com as regras gramaticais. Você pode escrever livremente, sem se preocupar com erros ou acertos. Se perdermos a conexão com a internet, faremos as atividades em outro momento. Seus responsáveis estão cientes dos riscos e benefícios da pesquisa e autorizaram sua participação.

Aceito participar da atividade de gravação de voz.

Aceito participar da atividade de escrita.

---

Local e data

---

Assinatura do (a) aluno (a) menor de idade

**Nome completo da Pesquisadora Responsável:** Thaís Cristófaros Alves da Silva

Endereço: Av. Presidente Antônio Carlos, 6627, Bairro: Pampulha. CEP: 31270-901 /Belo Horizonte – MG

Telefones: (31) 3409 – 6025

E-mail: [thaiscristofaro@gmail.com](mailto:thaiscristofaro@gmail.com)

Em caso de dúvidas, com respeito aos aspectos éticos desta pesquisa, você poderá consultar:

**COEP-UFMG - Comissão de Ética em Pesquisa da UFMG**

Av. Antônio Carlos, 6627. Unidade Administrativa II - 2º andar - Sala 2005.

Campus Pampulha. Belo Horizonte, MG – Brasil. CEP: 31270-901.

E-mail: [coep@prpq.ufmg.br](mailto:coep@prpq.ufmg.br). Tel: 34094592

Apêndice C: Questionário para Professores

Esta pesquisa tem como objetivo investigar erros ortográficos cometidos por alunos do Ensino Fundamental. Para que o estudo alcance o seu objetivo, contamos com a experiência de vocês, professoras(es). Pedimos, por favor, que vocês respondam as perguntas deste questionário levando em conta as observações de vocês sobre a escrita dos alunos da turma que lecionam atualmente na Escola Municipal XXXXX. A ajuda de vocês será muito importante para o trabalho. Qualquer dúvida, entrem em contato comigo. Obrigada! Atenciosamente, Cecília Toledo. Telefone para contato: 00000000

**Atualmente, você é professora de português de qual ano escolar?**

1° ano  2° ano  3° ano  4° ano  5° ano  7° ano  9° ano

**Para quais anos escolares (séries) você leciona ou já lecionou?** \_\_\_\_\_

**Você leciona há quantos anos?** \_\_\_\_\_

**Qual a sua formação?**

Graduação em Letras;  Graduação em Pedagogia;  Normal Superior;  Outro: \_\_\_\_\_







**As opções a seguir exemplificam alguns erros ortográficos possíveis na escrita de alunos do EF. Marque todas as opções que contêm um erro ortográfico que você já observou nos textos de seus atuais alunos.**

- |  |  |  |  |
|--|--|--|--|
| <input type="checkbox"/> abri ( <i>abrir</i> )         | <input type="checkbox"/> cola ( <i>colar</i> )         | <input type="checkbox"/> garfu ( <i>garfo</i> )        | <input type="checkbox"/> peixi ( <i>peixe</i> )        |
| <input type="checkbox"/> alfaci ( <i>alface</i> )      | <input type="checkbox"/> come ( <i>comer</i> )         | <input type="checkbox"/> imprimir ( <i>imprimir</i> )  | <input type="checkbox"/> piruca ( <i>peruca</i> )      |
| <input type="checkbox"/> ba ( <i>bar</i> )             | <input type="checkbox"/> corrigi ( <i>corrigir</i> )   | <input type="checkbox"/> janta ( <i>jantar</i> )       | <input type="checkbox"/> pintô ( <i>pintor</i> )       |
| <input type="checkbox"/> bebe ( <i>beber</i> )         | <input type="checkbox"/> corta ( <i>cortar</i> )       | <input type="checkbox"/> joga ( <i>jogar</i> )         | <input type="checkbox"/> professo ( <i>professor</i> ) |
| <input type="checkbox"/> bixiga ( <i>bexiga</i> )      | <input type="checkbox"/> curuja ( <i>coruja</i> )      | <input type="checkbox"/> júre ( <i>júri</i> )          | <input type="checkbox"/> robor ( <i>robô</i> )         |
| <input type="checkbox"/> bisouro ( <i>besouro</i> )    | <input type="checkbox"/> chavi ( <i>chave</i> )        | <input type="checkbox"/> lava ( <i>lavar</i> )         | <input type="checkbox"/> safare ( <i>safári</i> )      |
| <input type="checkbox"/> bizerro ( <i>bezerro</i> )    | <input type="checkbox"/> dar ( <i>ele dá</i> )         | <input type="checkbox"/> laçu ( <i>laço</i> )          | <input type="checkbox"/> sofar ( <i>sofá</i> )         |
| <input type="checkbox"/> bequini ( <i>biquíni</i> )    | <input type="checkbox"/> dirigi ( <i>dirigir</i> )     | <input type="checkbox"/> lustr ( <i>lustre</i> )       | <input type="checkbox"/> sorveti ( <i>sorvete</i> )    |
| <input type="checkbox"/> buneca ( <i>boneca</i> )      | <input type="checkbox"/> eu dormir ( <i>eu dormi</i> ) | <input type="checkbox"/> ma ( <i>mar</i> )             | <input type="checkbox"/> sucu ( <i>suco</i> )          |
| <input type="checkbox"/> burracha ( <i>borracha</i> )  | <input type="checkbox"/> dominor ( <i>dominó</i> )     | <input type="checkbox"/> maracujar ( <i>maracujá</i> ) | <input type="checkbox"/> tambo ( <i>tambor</i> )       |
| <input type="checkbox"/> cabidi ( <i>cabide</i> )      | <input type="checkbox"/> elevado ( <i>elevador</i> )   | <input type="checkbox"/> moleta ( <i>muleta</i> )      | <input type="checkbox"/> táxe ( <i>táxi</i> )          |
| <input type="checkbox"/> caderno ( <i>caderno</i> )    | <input type="checkbox"/> ele ler ( <i>ele lê</i> )     | <input type="checkbox"/> morde ( <i>morder</i> )       | <input type="checkbox"/> tisoura ( <i>tesoura</i> )    |
| <input type="checkbox"/> cavalu ( <i>cavalo</i> )      | <input type="checkbox"/> eu sorrir ( <i>eu sorri</i> ) | <input type="checkbox"/> murcego ( <i>morcego</i> )    | <input type="checkbox"/> tumate ( <i>tomate</i> )      |
| <input type="checkbox"/> calcanha ( <i>calcanhar</i> ) | <input type="checkbox"/> fúrmiga ( <i>formiga</i> )    | <input type="checkbox"/> origame ( <i>origami</i> )    | <input type="checkbox"/> tocano ( <i>tucano</i> )      |
| <input type="checkbox"/> canto ( <i>cantor</i> )       | <input type="checkbox"/> flo ( <i>flor</i> )           | <input type="checkbox"/> pepoca ( <i>pipoca</i> )      | <input type="checkbox"/> varre ( <i>varrer</i> )       |
| <input type="checkbox"/> celula ( <i>celular</i> )     | <input type="checkbox"/> gatu ( <i>gato</i> )          | <input type="checkbox"/> pesca ( <i>pescar</i> )       | <input type="checkbox"/> vistido ( <i>vestido</i> )    |

**De acordo com a sua opinião, quais motivos explicam os erros ortográficos, como os exemplificados acima, cometidos pelas crianças.**

Apêndice D: Instruções e Estímulos Experimentais














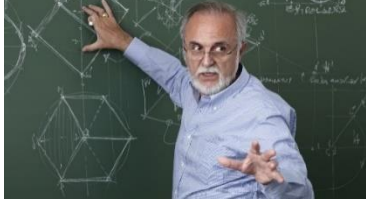

Etapas	Instruções – todas as instruções foram grafadas em áudio
Etapa 1- treinamento	Olá participante, A seguir você vai ver algumas figuras e vai escutar o nome dessas figuras. Tente gravar na memória esses nomes.
Etapa 2- dados de fala	Nesta etapa queremos saber se você se lembra das palavras que ouviu antes. A atividade é a seguinte: 1- Você verá as figuras acompanhadas de perguntas. 2- Você deverá observar as figuras, ouvir as perguntas e responder as perguntas <b>EM VOZ ALTA</b> 3- A resposta é <b>UMA PALAVRA</b> que você ouviu na atividade anterior.
Etapa 3 – dados de escrita	Nesta etapa queremos saber se você se lembra das respostas que deu antes. A atividade é a seguinte: 1- Você verá as figuras acompanhadas de perguntas. 2- Você deverá observar as figuras, ouvir as perguntas e responder as perguntas <b>POR ESCRITO</b> 3- a resposta é a palavra que você falou na atividade anterior.

<p align="center">Que inseto é esse?</p>  <p align="center"><i>besouro</i></p>	<p align="center">Que fruta é essa?</p>  <p align="center"><i>maracujá</i></p>	<p align="center">O que é isso?</p>  <p align="center"><i>borracha</i></p>
<p align="center">Qual o nome desse objeto?</p>  <p align="center"><i>muleta</i></p>	<p align="center">Como se chama essa peça de roupa?</p>  <p align="center"><i>biquíni</i></p>	<p align="center">Como se chamam as pessoas que julgam um crime?</p>  <p align="center"><i>júri</i></p>

<p>Complete a frase: em aniversários, a gente _____ presentes para o aniversariante</p>  <p><i>a gente dá</i></p>	<p>Como se chama esse tipo de passeio no meio da selva?</p>  <p><i>safári</i></p>	<p>O que é isso?</p>  <p><i>bexiga</i></p>
<p>Como se chama esse jogo?</p>  <p><i>dominó</i></p>	<p>O que a pessoa vai fazer com a lata?</p>  <p><i>abrir</i></p>	<p>Como se chama a refeição que fazemos a noite?</p>  <p><i>jantar</i></p>
<p>Que verdura é essa?</p>  <p><i>alface</i></p>	<p>O que é isso?</p>  <p><i>cabide</i></p>	<p>Que bebida é essa?</p>  <p><i>suco</i></p>
<p>O que a moça vai fazer com o cabelo?</p>  <p><i>cortar</i></p>	<p>O que a impressora vai fazer com o papel?</p>  <p><i>imprimir</i></p>	<p>Que objeto é esse?</p>  <p><i>caderno</i></p>
<p>Que sobremesa é essa?</p>  <p><i>sorvete</i></p>	<p>Como se chama essa obra de arte?</p>  <p><i>origami</i></p>	<p>O que a professora vai fazer com as provas?</p>  <p><i>corrigir</i></p>

<p>O que a criança vai fazer com a água?</p>  <p><i>beber</i></p>	<p>Que animal é esse?</p>  <p><i>gato</i></p>	<p>Como se chama esse tipo de carro?</p>  <p><i>táxi</i></p>
<p>Como se chama essa peça de roupa?</p>  <p><i>vestido</i></p>	<p>Que animal é esse?</p>  <p><i>tucano</i></p>	<p>Que objeto é este?</p>  <p><i>chave</i></p>
<p>Como se chama esse animal?</p>  <p><i>bezerro</i></p>	<p>Complete a frase: ontem eu estava muito cansada, então eu _____</p>  <p><i>dormi</i></p>	<p>Como se chama esse objeto na cabeça da criança?</p>  <p><i>peruca</i></p>
<p>O que é isso?</p>  <p><i>laço</i></p>	<p>Que animal é este?</p>  <p><i>peixe</i></p>	<p>O que o cachorro vai fazer com a mão?</p>  <p><i>morder</i></p>
<p>Como se chama esse alimento?</p>  <p><i>tomate</i></p>	<p>O que o menino vai fazer com o xadrez?</p>  <p><i>jogar</i></p>	<p>Como se chama a parte de trás do pé?</p>  <p><i>calcanhar</i></p>



<p>O que a moça vai fazer com a louça?</p>  <p>lavar</p>	<p>O que é isso?</p>  <p>flor</p>	<p>Qual é o nome desse animal?</p>  <p>morcego</p>
<p>Como se chama esse inseto?</p>  <p>formiga</p>	<p>O que o menino vai fazer com o arroz?</p>  <p>comer</p>	<p>Como se chama esse animal?</p>  <p>coruja</p>
<p>O que é isso?</p>  <p>robô</p>	<p>Que animal é esse?</p>  <p>cavalo</p>	<p>O que a vassoura vai fazer com a casa?</p>  <p>varrer</p>
<p>Os tubarões, as baleias e os peixes moram no fundo do?</p>  <p>mar</p>	<p>Que brinquedo é esse?</p>  <p>boneca</p>	<p>Como se chama o lugar onde os adultos vão comer, beber e se divertir?</p>  <p>bar</p>
<p>O que é isso no pescoço da mulher?</p>  <p>colar</p>	<p>Qual a profissão desse homem?</p>  <p>professor</p>	<p>Que comida é essa?</p>  <p>Pipoca</p>

O que o homem vai fazer com o peixe?



*pescar*

Que objeto é esse?



*tesoura*

Qual a profissão desse homem?



*cantor*

Complete a frase: ontem eu estava muito feliz, por isso eu \_\_\_



*eu sorri*

Que objeto é esse?



*celular*

Qual a profissão desse homem?



*pintor*

Que instrumento é esse?



*tambor*

O que o menino vai fazer com o carro?



*dirigir*

O que a gente faz com os livros?



*a gente lê*

O que é isso?



*elevador*

Que objeto é esse?



*garfo*

Que objeto é esse?



*sofá*

O que é isso?



*lustre*

## Apêndice E: Script Análise

```
#####
#Script Tese - Cecilia Toledo#
#####

#diretorio de trabalho
setwd("C:/Users/Usuario/Dropbox/Doutorado dropbox/TESE/CORPUS")

#exportando conjunto de dados
dados.brutos <- read.csv("dadosbrutostese.csv", stringsAsFactors = TRUE)

#carregando pacotes
library(tidyverse)
library(emmeans)
library(broom)
library(lme4)
library(forcats)
library(lmerTest)
library(patchwork)
library(effects)
library(rms)
library(car)
library(ggrepel)
library(ggpubr)
library(forcats)

#explorando conjunto de dados
view(dados.brutos)
head (dados.brutos)
str(dados.brutos)

#criando conjunto de dados (excluí NA e dados de hipercorreção)
dados.limpos <- dados.brutos %>%
select(PARTICIPANTE,ANO,PALAVRA,TIPODEPALAVRA,FENOMENO,CATEGORIA,FORMAORAL,TIPODE
VARIANTE,FORMAESCRITA,ERROACERTO) %>%
filter(TIPODEPALAVRA=="experimental")%>%
na.omit()%>%
droplevels()

#explorando novo conjunto de dados
view(dados.limpos)
head (dados.limpos)
str(dados.limpos)

##### CAPUTULO 5 #####
#####ANALISE DAS VOGAIS MEDIAS #####
#####
```

#### vogais anteriores na fala #####

**#GRÁFICO 1 - índice geral de vogais anteriores pretonicas e postonicas finais na fala**

```
graf1<-ggplot(dados.limpos %>%
select (ANO,FENOMENO,FORMAORAL,CATEGORIA) %>%
filter (FENOMENO == "postonica"| FENOMENO == "pretonica")%>%
filter (CATEGORIA == "vogalantterior")%>%
group_by(FENOMENO,CATEGORIA,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n)),
aes(x=FENOMENO, y=porcentagem*100,fill=FORMAORAL))+
  scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
  scale_x_discrete(" ",labels= c("postonica" = "vogais postônicas finais", "pretonica"="vogais pretônicas"))+
  geom_col(position="dodge",width=0.6,colour="black") +
  scale_fill_manual(labels= c("cancelamento", "[i]", "[e]"), values = c("lightgrey", "dimgray", "black"))+
  geom_text(aes(label= scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
  colour="black",position=position_dodge(.5), size=8)+
  labs(title="", x="", y="%",fill="")+
  theme_classic()+
  theme (legend.position = "bottom",plot.title = element_text(size = 20,color="red",vjust=1),legend.text=
  element_text(size=25),axis.text.x = element_text(size =
  25,color="black"),axis.text.y=element_blank(),axis.ticks=element_blank(),text = element_text (size=25,
  family="serif"))
```

**# X<sup>2</sup> diferenças entre vogais anteriores pretonicas e postonicas finais**

```
dados.limpos %>%
select (ANO, FENOMENO, FORMAORAL, CATEGORIA) %>%
filter (FENOMENO == "postonica"|FENOMENO=="pretonica")%>%
group_by(FENOMENO,CATEGORIA,FORMAORAL)%>%
summarise(n=n())%>%
mutate(porcentagem = n / sum(n))
```

**#tabela X<sup>2</sup>**

```
tab_vogaisanteriores_prepos<-matrix(c(117,0,163,208,0,58),byrow= T, nrow = 3)
tab_vogaisanteriores_prepos
chisq.test(tab_vogaisanteriores_prepos) #vogais anteriores pretônicas e postônicas finais são diferentes na fala
```

**## gráfico 2 - variação na fala por item lexical- anteriores**

```
palavras.pret.anteriores <-dados.limpos %>%
select (PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "pretonica")%>%
filter (CATEGORIA=="vogalantterior")%>%
group_by(FENOMENO,CATEGORIA, PALAVRA,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>% mutate(porcentagem = n / sum(n))
```

```

lexico.pret.anteriores <-ggplot(palavras.pret.anteriores,
  aes(x=PALAVRA,y=porcentagem,fill=FORMAORAL))+
  scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
  geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
  scale_fill_manual(values = c("dimgray","black"),labels=c("[i]","[e]"))+
  labs( y= " %", x = "",fill="", title= "Vogais pretônicas anteriores")+
  theme_classic()+
  theme (legend.position = "bottom",plot.title = element_text(size = 25,color="red"),axis.text.x =
  element_text(size = 25,color="black"),
  axis.text.y = element_text(size = 30,color="black"),text = element_text(size=25,
  family="serif"),legend.text= element_text(size=25))

  palavras.post.anteriores <- dados.limpos %>%
select (PALAVRA,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "postonica")%>%
filter (CATEGORIA=="vogalanterior")%>%
group_by(CATEGORIA,PALAVRA,FORMAORAL)%>%
na.omit(dados.limpos)%>%
summarise(n=n())%>%
mutate(porcentagem = n / sum(n))
lexico.post.anteriores<-ggplot(palavras.post.anteriores,
  aes(x=PALAVRA,y=porcentagem,fill=FORMAORAL))+
  scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
  geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
  scale_fill_manual(values = c("lightgray","dimgray"),labels=c("cancelamento","[i]"))+
  labs( y= " %", x = "",fill="", title= "Vogais postônicas finais anteriores")+
  theme_classic()+
  theme (legend.position = "bottom",plot.title = element_text(size = 25,color="red"),axis.text.x = element_text(size =
  25,color="black"),
  axis.text.y = element_text(size = 30,color="black"),text = element_text(size=25, family="serif"),legend.text=
  element_text(size=25))

graf2 <-lexico.pret.anteriores+lexico.post.anteriores

```

### **#grafico 3 - vogais anteriores pretônicas por individuos**

```

#vogais pretonicas anteriores
individuo.pre <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "pretonica")%>%
filter (CATEGORIA=="vogalanterior")%>%
group_by(PARTICIPANTE,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>% mutate(porcentagem = n / sum(n))

```

```
graf3 <-ggplot(individuo.pre,
  aes(x=PARTICIPANTE,y=porcentagem,fill=FORMAORAL))+
  scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
  geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
  scale_fill_manual(values = c("dimgray", "black"),labels=c("vogal alta","vogal média"))+
  labs( y= " %", x = "",fill="", title= "")+
  theme_classic()+
  theme (legend.position = "bottom",plot.title = element_text(size = 18,color="darkred"),axis.text.x = element_text(size =
  12,color="black"),axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25,
  family="serif"),legend.text= element_text(size=25))
```

#### **#Gráfico 4: vogais anteriores postônicas finais por indivíduo**

```
individuo.pos<-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "postonica")%>%
filter (CATEGORIA=="vogal anterior")%>%
group_by(PARTICIPANTE,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))
```

```
graf4 <-ggplot(individuo.pos,
  aes(x=PARTICIPANTE,y=porcentagem,fill=FORMAORAL))+
  scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
  geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
  scale_fill_manual(values = c("lightgray","dimgray"),labels=c("cancelamento","vogal alta"))+
  labs( y= " %", x = "",fill="", title= "")+
  theme_classic()+
  theme (legend.position = "bottom",plot.title = element_text(size = 18,color="darkred"),axis.text.x = element_text(size
  =12,color="black"),
  axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
  element_text(size=25))
```

#### **#### seção 6.1.2 - vogais posteriores na fala #####**

#### **#GRÁFICO 5 - índice geral de vogais posteriores preônicas e postônicas finais**

```
graf5<- ggplot(dados.limpos %>%
  select (ANO,FENOMENO,FORMAORAL,CATEGORIA) %>%
  filter (FENOMENO == "postonica"| FENOMENO == "pretonica")%>%
  filter (CATEGORIA == "vogal posterior")%>%
  group_by(FENOMENO,CATEGORIA,FORMAORAL)%>%
  summarise(n=n())%>%
  na.omit(dados.limpos)%>%
  mutate(porcentagem = n / sum(n)),
  aes(x=FENOMENO, y=porcentagem*100,fill=FORMAORAL))+
```

```

scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
scale_x_discrete (" ",labels= c("postonica" = "vogais postônicas finais", "pretonica"="vogais pretônicas"))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(labels= c("cancelamento","[u]", "[o]"), values = c("lightgrey","dimgray","black"))+
geom_text(aes(label = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
  colour="black",position=position_dodge(.5), size=8)+
labs(title="", x="", y="%",fill="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 20,color="red",vjust=1),legend.text=
  element_text(size=25),axis.text.x = element_text(size =
  25,color="black"),axis.text.y=element_blank(),axis.ticks=element_blank(),text = element_text(size=25, family="serif"))

```

### # X<sup>2</sup> diferenças entre vogais anteriores pretonicas e postonicas finais

```

dados.limpos %>%
select (ANO,FENOMENO,FORMAORAL,CATEGORIA) %>%
filter (FENOMENO == "postonica"|FENOMENO=="pretonica")%>%
group_by(FENOMENO,CATEGORIA,FORMAORAL)%>%
summarise(n=n())%>%
mutate(porcentagem = n / sum(n))

tab_vogaisposteriores_prepos<-matrix(c(0,94,106,200,193,2),byrow= T, nrow = 3)
tab_vogaisposteriores_prepos
chisq.test(tab_vogaisposteriores_prepos)#vogais posteriores pretonicas e postonicas finais sao diferentes na fala.

```

### ## gráfico 6- variação na fala por item lexical- posteriores

```

palavras.pret.posteriores <-dados.limpos %>%
select (PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "pretonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(FENOMENO,CATEGORIA, PALAVRA,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

lexico.pret.posteriores<-ggplot(palavras.pret.posteriores,
  aes(x=PALAVRA,y=porcentagem,fill=FORMAORAL))+
  scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
  geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
  scale_fill_manual(values = c("dimgray","black"),labels=c("vogal alta","vogal média"))+
  labs(y= "%", x = "",fill="", title = "Vogais pretônicas posteriores")+
  theme_classic()+
  theme (legend.position = "bottom",plot.title = element_text(size = 25,color="red"),axis.text.x = element_text(size =
  24,color="black"),
  axis.text.y = element_text(size = 23,color="black"),text = element_text(size=25, family="serif"),legend.text=
  element_text(size=25))

```

```

palavras.post.posteriores <- dados.limpos %>%
select (PALAVRA,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "postonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(CATEGORIA,PALAVRA,FORMAORAL)%>%
na.omit(dados.limpos)%>%
summarise(n=n())%>%
mutate(porcentagem = n / sum(n))

```

```

lexico.post.posteriores<-ggplot(palavras.post.posteriores,
aes(x=PALAVRA,y=porcentagem,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
scale_fill_manual(values = c("lightgray","dimgray","black"),labels=c("cancelamento","vogal alta","vogal
média"))+
labs( y= " %", x = "",fill="", title= "Vogais postônicas finais posteriores")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 25,color="red"),axis.text.x =
element_text(size = 24,color="black"),
axis.text.y = element_text(size = 23,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

```
graf6 <-lexico.pret.posteriores+lexico.post.posteriores
```

### ***#grafico 7 - vogais posteriores por individuos***

#### ***#vogais pretonicas posteriores***

```

individuo.pre <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "pretonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(PARTICIPANTE,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

```

```

graf7 <-ggplot(individuo.pre,
aes(x=PARTICIPANTE,y=porcentagem,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
scale_fill_manual(values = c("dimgray", "black"),labels=c("vogal alta", "vogal média"))+
labs( y= " %", x = "",fill="", title= "")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
12,color="black"),

```



```
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
  element_text(size=25))
```

### ***#vogais postônicas posteriores***

```
individuo.pos<-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "postonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(PARTICIPANTE,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

graf8 <-ggplot(individuo.pos,
  aes(x=PARTICIPANTE,y=porcentagem,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
scale_fill_manual(values = c("lightgray","dimgray","black"),labels=c("cancelamento","vogal alta","vogal média"))+
labs( y= "%", x = "",fill="", title= "")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size = 12,color="black"),
  axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
  element_text(size=25))
```

### **#### seção 6.2.1 - vogais anteriores na escrita #####**

#### **# GRÁFICO 9 índice total de erros ortográficos - vogais anteriores**

```
erro.total.anteriores <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO == "postonica"| FENOMENO=="pretonica")%>%
filter (CATEGORIA=="vogalanterior")%>%
group_by(FENOMENO,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n)) %>%
filter (ERROACERTO=="erro")

graf9 <-ggplot(erro.total.anteriores,
  aes(x=ERROACERTO,y=porcentagem,fill=FENOMENO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
scale_x_discrete (" ",labels= c("postonica" = "vogais postônicas finais", "pretonica"="vogais pretônicas"))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("dimgray","black"),labels=c("troca de <e> por <i> post.final","troca de <e> por <i> pretônica"))+
geom_text(aes(label =
  scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
  colour="black",position=position_dodge(.5), size=8)+
```

```

labs( y= " %", x = "",fill="", title="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
12,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

#testes qui-quadrado nas amostras de escrita vogais anteriores

dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO == "postonica"| FENOMENO=="pretonica")%>%
filter (CATEGORIA=="vogalanterior")%>%
group_by(FENOMENO,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(portentagem = n / sum(n))

tab_escritaanteriores_prepos<-matrix(c(260,33,162,117),byrow= T, nrow = 2)
tab_escritaanteriores_prepos
chisq.test(tab_escritaanteriores_prepos) #vogais anteriores pretônicas e postônicas finais são diferentes na escrita

# GRÁFICO 10 índice de erros ortográficos por palavra - vogais anteriores

erro.palavra.pret.anteriores <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="pretonica"| FENOMENO=="postonica")%>%
filter (CATEGORIA=="vogalanterior")%>%
group_by(FENOMENO,PALAVRA,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(portentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

graf10 <-ggplot(erro.palavra.pret.anteriores,
aes(x=PALAVRA,y=portentagem,fill=FENOMENO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
facet_wrap(~FENOMENO)+
scale_fill_manual(values = c("dimgray", "black"),labels=c("troca de <e> por <i> postônica", "troca de <e> por <i> pretônica"))+
geom_text(aes(label =
scales::percent(portentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=8)+
labs( y= " %", x = "",fill="", title="Vogais pretônicas")+
theme_classic()+

```

```

theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
18,color="black"),
  axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

# GRÁFICO 11 índice de erros ortográficos por individuo - vogais anteriores pretonicas

```

erro.individuo.pret.anteriores <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="pretonica")%>%
filter (CATEGORIA=="vogalantterior")%>%
group_by(PARTICIPANTE,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(percentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

graf11<-ggplot(erro.individuo.pret.anteriores,
  aes(x=PARTICIPANTE,y=percentagem,fill=ERROACERTO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("black"),labels=c("troca de <e> por <i> pretônicas"))+
geom_text(aes(label = scales::percent(percentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=6)+
labs( y= " %", x = "",fill="", title="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
14,color="black"),
  axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

# GRÁFICO 12 índice de erros ortográficos por individuo - vogais anteriores postonicas

```

erro.individuo.post.anteriores <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="postonica")%>%
filter (CATEGORIA=="vogalantterior")%>%
group_by(PARTICIPANTE,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(percentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

```

```

graf12<-ggplot(erro.individuo.post.anteriores,

```

```

aes(x=PARTICIPANTE,y=porcentagem,fill=ERROACERTO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("dimgray"),labels=c("troca de <e> por <i> postônicas finais"))+
geom_text(aes(label = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=6)+
labs( y= " %", x = "",fill="", title="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
15,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

#### seção 6.2.2 - vogais posteriores na escrita #####

# GRÁFICO 13- índice total de erros ortográficos - vogais posteriores

```

erro.total.posteriores <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO == "postonica" | FENOMENO=="pretonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(FENOMENO,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

graf13 <-ggplot(erro.total.posteriores,
aes(x=ERROACERTO,y=porcentagem,fill=FENOMENO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
scale_x_discrete(" ",labels= c("postonica" = "vogais postônicas finais", "pretonica"="vogais pretônicas"))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("dimgrey","black"),labels=c("troca de <o> por <u> postonica","troca de <o> por <u> pretônica"))+
geom_text(aes(label = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=8)+
labs( y= " %", x = "",fill="", title="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
12,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

#testes qui-quadrado nas amostras de escrita vogais posteriores

```

dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO == "postonica" | FENOMENO=="pretonica")%>%

```

```

filter (CATEGORIA=="vogalposterior")%>%
group_by(FENOMENO,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

```

```

tab_escritaposteriores_prepos<-matrix(c(295,3,273,26),byrow= T, nrow = 2)
tab_escritaposteriores_prepos
chisq.test(tab_escritaposteriores_prepos) #vogais anteriores pretônicas e postônicas finais são diferentes na escrita

```

# GRÁFICO 14 índice de erros ortográficos por palavra - vogais posteriores

```

erro.palavra.posterior<-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="pretonica"| FENOMENO=="postonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(FENOMENO,PALAVRA,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

```

```

graf14 <-ggplot(erro.palavra.posterior,
aes(x=PALAVRA,y=porcentagem,fill=FENOMENO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
facet_wrap(~FENOMENO)+
scale_fill_manual(values = c("dimgray", "black"),labels=c("troca de <o> por <u> postônica", "troca de <o> por <u> pretônica"))+
geom_text(aes(label = PALAVRA, y = porcentagem, fill = FENOMENO,
colour="black",position=position_dodge(.5), size=8)+
labs (y= "%", x = "",fill="", title="Vogais pretônicas")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
14,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

# GRÁFICO 15 índice de erros ortográficos por individuo - vogais posteriores

```

erro.individuo.pret.posterior<-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="pretonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(PARTICIPANTE,ERROACERTO)%>%

```

```

summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

graf15<-ggplot(erro.individuo.pret.posteriores,
  aes(x=PARTICIPANTE,y=porcentagem,fill=ERROACERTO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("black"),labels=c("troca de <o> por <u> pretônicas"))+
geom_text(aes(label = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=6)+
labs( y= "%", x = "",fill="", title="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
12,color="black"),
  axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

#GRÁFICO 16- índice de erros ortográficos por individuo - vogais posteriores

```

erro.individuo.post.posteriores <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="postonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(PARTICIPANTE,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

```

```

graf16<-ggplot(erro.individuo.post.posteriores,
  aes(x=PARTICIPANTE,y=porcentagem,fill=ERROACERTO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("dimgrey"),labels=c("troca de <o> por <u> postônicas"))+
geom_text(aes(label = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=6)+
labs( y= "%", x = "",fill="", title="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
12,color="black"),
  axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

```

#gráfico 17 - hipercorreções em contexto pretônico e psotônico
hiper.prepos<-dados.hiper %>%
select(PARTICIPANTE,ANO,PALAVRA,FENOMENO,ERROACERTO) %>%
group_by(FENOMENO,ANO,ERROACERTO)%>%
filter(FENOMENO=="pretonica"|FENOMENO=="postonica")%>%
summarise(n=n())%>%
mutate(porcentagemescrita = n / sum(n))%>%
filter(ERROACERTO=="hiper")

#dataframe com porcentagens de hipercorreção por contexto acentual

hiper.vogais <- tibble(ano = c("1A", "3A", "5A", "7A", "9A"),
  pretonica= c("hiperpre", "hiperpre", "hiperpre", "hiperpre", "hiperpre"),
  npre = c(4,3,1,0,3),
  pre= c(0.10,0.07,0.02,0.0,0.07),
  postonica=c("poshiper", "poshiper", "poshiper", "poshiper", "poshiper"),
  npos= c(11,6,8,10,2),
  pos=c(0.31,0.17,0.26,0.25,0.05))

graf17<-ggplot(hiper.vogais,
  aes(x=ano))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_point (aes(y=pre,colour = factor(pretonica)), size=7)+
geom_path(aes(y=pre,colour = factor(pretonica)),group=1,size=1.5,linetype = "twodash")+
geom_label(aes(y=pre,label = scales::percent(pre,accuracy=1)),colour="black",size=7)+
geom_point (aes(y=pos,colour = factor(postonica)), size=6)+
geom_path(aes(y=pos,colour = factor(postonica)),group=1,size=1.5)+
geom_label(aes(y=pos,label = scales::percent(pos,accuracy=1)),colour="dimgray",size=6)+
labs( y= "%", x = "ano escolar",title="", colour="",size=16)+
theme (text = element_text(size=19, family="serif"))+
scale_colour_manual(values= c("black", "dimgray"),labels= c("% vogais pretônicas", "% vogais postônicas finais"))+
theme_minimal()+
theme(legend.position = "bottom",axis.text.x=element_text(size=19), plot.title= element_text(size=19, colour=
"black"),axis.text.y=element_text(size=19),legend.text= element_text(size=19))

#### seção 6.2 hipóteses #####

# HIPOTESE 1A - grafico 18 erros ortograficos pretonicos e postonicos finais por ano escolar

hipotese.1A <- tibble(ano = c("1A", "3A", "5A", "7A", "9A"),
  pretonica= c("erropre", "erropre", "erropre", "erropre", "erropre"),
  npretonica =c(61,36,26,14,6),
  porcentagempre=c(0.52,0.31,0.22,0.12,0.05),
  postonica = c("erropos", "erropos", "erropos", "erropos", "erropos"),

```

```

npostonica= c(25,9,2,0,0),
percentagempos=c(0.21,0.07,0.01,0.00,0.00))

graf18 <- ggplot(hipotese.1A,
  aes(x=ano))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_point (aes(y=percentagempre,colour = factor(pretonica)), size=6)+
geom_path(aes(y=percentagempre,colour = factor(pretonica)),group=1,size=1.5,linetype = "twodash")+
geom_label(aes(y=percentagempre,label = scales::percent(percentagempre,accuracy=1)),colour="black",size=7)+
geom_point (aes(y=percentagempos,colour = factor(postonica)), size=6)+
geom_path(aes(y=percentagempos,colour = factor(postonica)),group=1,size=1.5)+
geom_label(aes(y=percentagempos,label = scales::percent(percentagempos,accuracy=1)),colour="dimgray",size=7)+
labs( y= "%", x = "ano escolar", colour="",size=30)+
theme (text = element_text(size=20, family="serif"))+
scale_colour_manual(values= c("dimgray","black"),labels= c("troca <e,o> por <i,u> postônico finais", "troca de <e,o> por <i,u>
pretonicos"))+
theme_minimal()+
theme(legend.position = "bottom",axis.text.x=element_text(size=29), plot.title= element_text(size=29, colour=
"black"),axis.text.y=element_text(size=25),legend.text= element_text(size=25))

### modelo generalizado 1

dados.vogais <- dados.brutos %>%
select(PARTICIPANTE,ANO,PALAVRA,TIPODEPALAVRA,FENOMENO,CATEGORIA,ERROACERTO) %>%
filter(TIPODEPALAVRA=="experimental")%>%
filter(FENOMENO=="pretonica"|FENOMENO=="postonica")%>%
#filter(ANO=="1A"|ANO=="3A"|ANO=="5A")%>%
na.omit()%>%
droplevels()

m.zero <-glmer(ERROACERTO ~ 1+(1|PARTICIPANTE) +(1|PALAVRA), dados.vogais,family=binomial)
m.ano <-glmer(ERROACERTO~ANO+(1|PARTICIPANTE) +(1|PALAVRA), dados.vogais,family=binomial)
m.fenomeno<-glmer(ERROACERTO~FENOMENO+(1|PARTICIPANTE) +(1|PALAVRA), dados.vogais,family=binomial)
m.anofenomeno<- glmer(ERROACERTO~ANO+FENOMENO+(1|PARTICIPANTE) +(1|PALAVRA),
dados.vogais,family=binomial)
m.anofenomenointeracao <-glmer(ERROACERTO~ANO*FENOMENO+(1|PARTICIPANTE) +(1|PALAVRA),
dados.vogais,family=binomial)

anova(m.zero,m.ano) #modelo com ano é melhor do que sem
anova(m.ano,m.fenomeno) #ano explica mais que fenomeno
anova(m.ano, m.anofenomeno) #ano+fenomeno é melhor que ano ***** esse é O MELHOR!
anova(m.anofenomeno,m.anofenomenointeracao) #interacao nao difere de sem interação

#o escolhido
m001<- glmer(ERROACERTO~ANO+FENOMENO+(1|PARTICIPANTE) +(1|PALAVRA), dados.vogais,family=binomial)

```



```

drop1(m001,test="Chisq")

plot(allEffects(m001),type = "response")
summary(m.anofenomeno)

post.hoc.m001<-emmeans(m001,~ANO+FENOMENO,type="response",adjust="tukey")
pairs(post.hoc.m001)
post.hoc.m001

graf19 <-plot(post.hoc.m001, color="black") +
  coord_flip()+
  labs(y= "ano + contexto acentual", x="probabilidade estimada de erros ortográficos")+
  theme_bw()

# HIPÓTESE 2a - grafico 14 Vogais médias na fala por ano escolar

hipotese.2a <-tibble(ano = c("1A", "3A", "5A","7A","9A"),
  pretonica =c("vpre", "vpre", "vpre", "vpre", "vpre"),
  npre = c(42,53,37,54,66),
  porcentagempre= c(0.36,0.48,0.32,0.46,0.57),
  postonica= c("vpos", "vpos", "vpos", "vpos", "vpos"),
  npos= c(0,0,2,0,0),
  porcentagempos= c(0.0,0.0,0.017,0.0,0.0))

#grafico 20 indices de vogais medias por ano escolar

graf20<- ggplot(hipotese.2a,
  aes(x=ano))+
  scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
  geom_point (aes(y=porcentagempre,colour = factor(pretonica)), size=6)+
  geom_path(aes(y=porcentagempre,colour = factor(pretonica)),group=1,size=1.5,linetype = "twodash")+
  geom_label(aes(y=porcentagempre,label = scales::percent(porcentagempre,accuracy=1)),colour="black",size=7)+
  geom_point (aes(y=porcentagempos,colour = factor(postonica)), size=6)+
  geom_path(aes(y=porcentagempos,colour = factor(postonica)),group=1,size=1.5)+
  geom_label(aes(y=porcentagempos,label = scales::percent(porcentagempos,accuracy=1)),colour="dimgray",size=7)+
  labs ( y= "%", x = "ano escolar", colour="",size=20)+
  theme (text = element_text(size=20, family="serif"))+
  scale_colour_manual(values= c("dimgray","black"),labels= c("vogais médias postônicas finais", "vogais médias pretônicas"))+
  theme_minimal()+
  theme(legend.position = "bottom",axis.text.x=element_text(size=29), plot.title= element_text(size=29, colour=
"black"),axis.text.y=element_text(size=25),legend.text= element_text(size=16))

#modelo estatístico 2: probabilidade de vogais médias na fala por ano escolar

```

```
#primeiro, faz-se uma tabela para ver as taxas de vogais médias
```

```
tabela.variantesproximasprepos<- dados.limpos %>%
select(PARTICIPANTE,ANO,PALAVRA,FENOMENO, TIPODEVARIANTE) %>%
group_by(ANO,TIPODEVARIANTE)%>%
filter(FENOMENO=="pretonica"|FENOMENO=="postonica")%>%
na.omit()%>%
summarise(n=n())%>%
mutate(porcentagemescrita = n / sum(n))
```

```
dados.fala.vogais<- dados.brutos %>%
```

```
select(PARTICIPANTE,ANO,PALAVRA,TIPODEPALAVRA,FENOMENO,CATEGORIA,FORMAORAL,TIPODEVARIAN
TE) %>%
filter(TIPODEPALAVRA=="experimental")%>%
filter(FENOMENO=="pretonica"|FENOMENO=="postonica")%>%
na.omit()%>%
droplevels()
```

```
levels(dados.fala.vogais$FENOMENO)
```

```
m.null<-glmer(TIPODEVARIANTE~1 +(1|PARTICIPANTE) +(1|PALAVRA), dados.fala.vogais,family=binomial)
m00 <- glmer(TIPODEVARIANTE~ANO+(1|PARTICIPANTE) +(1|PALAVRA), dados.fala.vogais,family=binomial)
m01 <- glmer(TIPODEVARIANTE~ANO+FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA),
dados.fala.vogais,family=binomial)
m02 <- glmer(TIPODEVARIANTE~ANO*FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA),
dados.fala.vogais,family=binomial)
```

```
anova(m.null, m02)# interação é importante
anova(m.null,m00)# ano é mais importante do que nada
anova(m.null,m01)#ano e fenomeno sao mais importantes do que nada
anova(m00,m02) #interação é mais importante do que ano sozinho
anova(m00,m01) #ano e fenomeno é mais importante do que ano sozinho
anova(m01,m02) # interaçao melhor do que só +
```

```
drop1(m02,test="Chisq")
```

```
summary(m02)
```

```
post.hoc.m02<-emmeans(m02,~ANO*FENOMENO,type="response")
post.hoc.m02
pairs(vpp)
```

```
graf.21<-plot(post.hoc.m02, color="black") +
```

```

coord_flip()+
labs(y= "contexto acentual+ano escolar", x="probabilidade estimada de variantes próximas à escrita em contexto pretonico")+
theme_bw()

##### CAPUTULO 6 #####
#ANALISE DOS ROTICOS #####
#####

#variação dos róticos na fala

# Gráfico 22: Variação segmental do rótico final em verbos e nomes

verbos.fala<-ggplot(dados.limpos %>%
  select (ANO,FENOMENO,FORMAORAL,CATEGORIA) %>%
  filter (FENOMENO == "verbo") %>%
  group_by(FENOMENO,CATEGORIA,FORMAORAL) %>%
  summarise(n=n()) %>%
  na.omit(dados.limpos) %>%
  mutate(percentagem = n / sum(n)),
  aes(x=CATEGORIA, y=percentagem*100,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
scale_x_discrete (" ",labels= c("rotico" = ""))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(labels= c("", ""), values = c("lightgrey", "black"))+
geom_text(aes(label = scales::percent(percentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=8)+
labs(title="Verbos", x="", y="%",fill="")+
theme_minimal()+
theme (legend.position = "bottom",plot.title = element_text(size = 20,color="red",vjust=3),legend.text=
element_text(size=25),axis.text.x = element_text(size =
25,color="black"),axis.text.y=element_blank(),axis.ticks=element_blank(),text = element_text(size=25, family="serif"))

nomes.fala<-ggplot(dados.limpos %>%
  select (ANO,FENOMENO,FORMAORAL,CATEGORIA) %>%
  filter (FENOMENO == "nome") %>%
  group_by(FENOMENO,CATEGORIA,FORMAORAL) %>%
  summarise(n=n()) %>%
  na.omit(dados.limpos) %>%
  mutate(percentagem = n / sum(n)),
  aes(x=CATEGORIA, y=percentagem*100,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
scale_x_discrete (" ",labels= c("rotico" = ""))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(labels= c("cancelamento do R-final", "produção do R-final"), values = c("lightgrey", "black"))+

```

```

geom_text(aes(label = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=8)+
labs(title="Nomes", x="", y="%",fill="")+
theme_minimal()+
theme (legend.position = "bottom",plot.title = element_text(size = 20,color="red",vjust=3),legend.text=
element_text(size=25),axis.text.x = element_text(size =
25,color="black"),axis.text.y=element_blank(),axis.ticks=element_blank(),text = element_text(size=25, family="serif"))

```

```
graf.22 <- verbos.fala+nomes.fala
```

```
# $\chi^2$  para ver diferenças entre verbos e nomes
```

```

dados.limpos %>%
select (ANO,FENOMENO,FORMAORAL,CATEGORIA) %>%
filter (FENOMENO == "nome"|FENOMENO=="verbo")%>%
group_by(FENOMENO,CATEGORIA,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

```

```
#análise estatística:comparação das amostras de escrita em contexto postonico
```

```

roticos.verbos.nomes.fala<-matrix(c(202,339,281,264),byrow= T, nrow = 2)
roticos.verbos.nomes.fala
chisq.test(roticos.verbos.nomes.fala) #há diferenças nas amostras de escrita do 1º e 3º ano

```

```
#Gráfico 23: Variação do R-final por item verbal
```

```

verbos <-dados.limpos %>%
select (PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "verbo")%>%
group_by(PALAVRA,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

```

```

graf.23 <-ggplot(verbos,
aes(x=PALAVRA,y=porcentagem,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
scale_fill_manual(values = c("lightgray","black"),labels=c("cancelamento do R-final","produção do R-final"))+
labs( y= "%", x = "",fill="", title= "")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 25,color="red"),axis.text.x = element_text(size =
22,color="black"),

```

```
axis.text.y = element_text(size = 23,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

# Gráfico 24: variação do R-final por item nominal

```
nomes <- dados.limpos %>%
select (PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "nome")%>%
group_by(PALAVRA,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

graf.24 <- ggplot(nomes,
aes(x=PALAVRA,y=porcentagem,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
scale_fill_manual(values = c("lightgray","black"),labels=c("cancelamento do R-final","produção do R-final"))+
labs( y= "%", x = "",fill="", title= "")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 25,color="red"),axis.text.x = element_text(size =
24,color="black"),
axis.text.y = element_text(size = 23,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

#Gráfico 25: variação do R-final em verbos por indivíduo

```
individuo.rotico.verbo <- dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "verbo")%>%
group_by(PARTICIPANTE,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

graf25 <- ggplot(individuo.rotico.verbo,
aes(x=PARTICIPANTE,y=porcentagem,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
scale_fill_manual(values = c("lightgray", "black"),labels=c("cancelamento","R-final"))+
labs( y= "%", x = "",fill="", title= "")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="darkred"),axis.text.x = element_text(size =
14,color="black"),
```

```
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

#Gráfico 26: variação do R-final em nomes por indivíduo

```
indivíduo.rotico.nome <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,FORMAORAL,CATEGORIA)%>%
filter (FENOMENO == "nome")%>%
group_by(PARTICIPANTE,FORMAORAL)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))

graf26 <-ggplot(indivíduo.rotico.nome,
aes(x=PARTICIPANTE,y=porcentagem,fill=FORMAORAL))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col (position = position_dodge2(width = 0.3, preserve = "single"))+
scale_fill_manual(values = c("lightgray", "black"),labels=c("cancelamento", "R-final"))+
labs( y= " %", x = "",fill="", title= "")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="darkred"),axis.text.x = element_text(size =
14,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

#variação dos róticos na escrita

#grafico 27- índice total de erros ortográficos - róticos

```
erro.total.roticos <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO == "verbo"| FENOMENO=="nome")%>%
group_by(FENOMENO,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

graf27 <-ggplot(erro.total.roticos,
aes(x=ERROACERTO,y=porcentagem,fill=FENOMENO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
scale_x_discrete (" ",labels= c("erro" = ""))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("black", "lightgray"),labels=c("omissão do <r> final em nomes", "omissão do <r> final em verbos"))+
```

```

geom_text(aes(label = scales::percent(portentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=8)+
labs( y= " %", x = "",fill="", title="")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
12,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))

```

```

#testes qui-quadrado nas amostras de escrita verbos e nomes
dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO == "verbo"| FENOMENO=="nome")%>%
group_by(FENOMENO,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(portentagem = n / sum(n))

```

```

tab_escritaroticos <-matrix(c(452,127,398,176),byrow= T, nrow = 2)
tab_escritaroticos
chisq.test(tab_escritaroticos) #róticos são diferentes na escrita

```

```

# GRÁFICO 28 índice de erros ortográficos por palavra - roticos verbos

```

```

erro.palavra.roticos.verbo <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="verbo")%>%
group_by(FENOMENO,PALAVRA,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(portentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

```

```

graf28 <- ggplot(erro.palavra.roticos.verbo,
aes(x=PALAVRA,y=portentagem,fill=FENOMENO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("lightgray"),labels=c("omissão do <r> em final de verbos"))+
geom_text(aes(label = scales::percent(portentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=8)+
labs( y= " %", x = "",fill="", title="Verbos")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
20,color="black"),

```

```
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

```
# GRÁFICO 29 índice de erros ortográficos por palavra - roticos nomes
```

```
erro.palavra.roticos.nome <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="nome")%>%
group_by(PALAVRA,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

graf29<-ggplot(erro.palavra.roticos.nome,
aes(x=PALAVRA,y=porcentagem,fill=ERROACERTO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("black"),labels=c("omissão do <r> em final de nomes"))+
geom_text(aes(label = "omissão do <r> em final de nomes",
               = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=8)+
labs( y= " %", x = "",fill="", title="Nomes")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size
=20,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

```
# GRÁFICO 30- índice de erros ortográficos por individuo - verbo
```

```
erro.individuo.roticos.verbo <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="verbo")%>%
group_by(PARTICIPANTE,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")

graf30<-ggplot(erro.individuo.roticos.verbo,
aes(x=PARTICIPANTE,y=porcentagem,fill=ERROACERTO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("lightgray"),labels=c("omissão do <r> em final de verbos"))+
geom_text(aes(label = "omissão do <r> em final de verbos",
               = scales::percent(porcentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=6)+
```



```
labs(y = "%", x = "", fill = "", title = "")+
theme_classic()+
theme (legend.position = "bottom", plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
17,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

```
erro.individuo.post.posteriores <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="postonica")%>%
filter (CATEGORIA=="vogalposterior")%>%
group_by(PARTICIPANTE,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(percentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")
```

# GRÁFICO 31- índice de erros ortográficos por individuo - nome

```
erro.individuo.roticos.nome <-dados.limpos %>%
select (PARTICIPANTE,PALAVRA,ANO,FENOMENO,CATEGORIA,ERROACERTO)%>%
filter (FENOMENO=="nome")%>%
group_by(PARTICIPANTE,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(percentagem = n / sum(n))%>%
filter (ERROACERTO=="erro")
```

```
graf31<-ggplot(erro.individuo.roticos.nome,
aes(x=PARTICIPANTE,y=percentagem,fill=ERROACERTO))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_col(position="dodge",width=0.6,colour="black") +
scale_fill_manual(values = c("black"),labels=c("omissão do <r> em final de nomes"))+
geom_text(aes(label = scales::percent(percentagem,accuracy=1)),vjust=-0.2,hjust=0.4,
colour="black",position=position_dodge(.5), size=6)+
labs(y = "%", x = "", fill = "", title = "")+
theme_classic()+
theme (legend.position = "bottom",plot.title = element_text(size = 18,color="red"),axis.text.x = element_text(size =
17,color="black"),
axis.text.y = element_text(size = 20,color="black"),text = element_text(size=25, family="serif"),legend.text=
element_text(size=25))
```

#### hipercorreções nos roticos ###

#Gráfico 32 - hipercorreções nos róticos

```
dados.hiper <- dados.brutos %>%
```

```
select(PARTICIPANTE,ANO,PALAVRA,TIPODEPALAVRA,FENOMENO,ERROACERTO) %>%
filter(TIPODEPALAVRA=="hipercorrecao")%>%
na.omit()%>%
droplevels()
```

```
hiper.roticos<-dados.hiper %>%
select(PARTICIPANTE,ANO,PALAVRA,FENOMENO,ERROACERTO) %>%
group_by(FENOMENO,ANO,ERROACERTO)%>%
filter(FENOMENO=="verbo"|FENOMENO=="nome")%>%
summarise(n=n())%>%
mutate(porcentagemescrita = n / sum(n))%>%
filter(ERROACERTO=="hiper")
```

#dataframe com porcentagens de hipercorreção por contexto acentual

```
hiper.verbosnomes <- tibble(ano = c("1A", "3A", "5A", "7A", "9A"),
  verbo = c("hiperverbo", "hiperverbo", "hiperverbo", "hiperverbo", "hiperverbo"),
  nverbo = c(3,12,19,13,14),
  pverbo = c(0.09,0.33,0.63,0.40,0.39),
  nome = c("nomehiper", "nomehiper", "nomehiper", "nomehiper", "nomehiper"),
  nnome = c(1,1,2,2,1),
  pnome = c(0.02,0.025,0.05,0.05,0.025))
```

```
graf32<-ggplot(hiper.verbosnomes,
  aes(x=ano))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_point(aes(y=pverbo,colour = factor(verbo)), size=7)+
geom_path(aes(y=pverbo,colour = factor(verbo)),group=1,size=1.5,linetype = "twodash")+
geom_label(aes(y=pverbo,label = scales::percent(pverbo,accuracy=1)),colour="dimgray",size=7)+
geom_point(aes(y=pnome,colour = factor(nome)), size=6)+
geom_path(aes(y=pnome,colour = factor(nome)),group=1,size=1.5)+
geom_label(aes(y=pnome,label = scales::percent(pnome,accuracy=1)),colour="black",size=7)+
labs(y = "%", x = "ano escolar",title="", colour="",size=16)+
theme(text = element_text(size=19, family="serif"))+
scale_colour_manual(values = c("dimgray", "black"),labels = c("% inserção do <r> em verbos", "% inserção do <r> em nomes"))+
theme_minimal()+
theme(legend.position = "bottom",axis.text.x=element_text(size=19), plot.title = element_text(size=19, colour = "black"),axis.text.y=element_text(size=19),legend.text = element_text(size=19))
```

#Hipotese 1B- persistencia dos erros em verbos e nomes

#Grafico 33

```
erros.indice<-dados.limpos %>%
select(PARTICIPANTE,ANO,PALAVRA,FENOMENO,TIPODEVARIANTE,ERROACERTO) %>%
```

```

filter(FENOMENO=="verbo"|FENOMENO=="nome")%>%
group_by(ANO,FENOMENO,ERROACERTO)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(percentagemescrita = n / sum(n))%>%
filter(ERROACERTO=="erro")

persistencia.erro.roticoo <-tibble(ano = c("1A", "3A", "5A","7A","9A"),
  verbo= c("verbo","verbo","verbo","verbo","verbo"),
  nverbo = c(106,40,14,15,1),
  percentagemverbo= c(0.94,0.35,0.12,0.13,0.01),
  nome=c("nome","nome","nome","nome","nome"),
  nnome= c(103,19,3,2,0),
  percentagemnome=c(0.89,0.16,0.02,0.01,0.0))

ggplot(persistencia.erro.roticoo,
  aes(x=ano))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_point (aes(y=percentagemverbo,colour = factor(verbo)), size=6)+
geom_path(aes(y=percentagemverbo,colour = factor(verbo)),group=1,size=1.5)+
geom_label(aes(y=percentagemverbo,label = scales::percent(percentagemverbo,accuracy=1)),colour="dimgray",size=7)+
geom_point (aes(y=percentagemnome,colour = factor(nome)), size=6)+
geom_path(aes(y=percentagemnome,colour = factor(nome)),group=1,size=1.5,linetype = "twodash")+
geom_label(aes(y=percentagemnome,label = scales::percent(percentagemnome,accuracy=1)),colour="black",size=7)+
labs( y = "%", x = "ano escolar", colour="",size=30)+
theme (text = element_text(size=20, family="serif"))+
scale_colour_manual(values= c("black","dimgray"),labels= c( "omissão do <r> final em nomes","omissão do <r> final em
verbos"))+
theme_minimal()+
theme(legend.position = "bottom",axis.text.x=element_text(size=29), plot.title= element_text(size=29, colour=
"black"),axis.text.y=element_text(size=25),legend.text= element_text(size=25))

#modelo estatístico para ver a probabilidade de erros ortográficos em verbos e em nomes e em cada ano escolar

dados.roticos <- dados.brutos %>%
select(PARTICIPANTE,ANO,PALAVRA,TIPODEPALAVRA,FENOMENO,CATEGORIA,ERROACERTO) %>%
filter(TIPODEPALAVRA=="experimental")%>%
filter(FENOMENO=="verbo"|FENOMENO=="nome")%>%
na.omit()%>%
droplevels()

m.zeroroticos<-glmer(ERROACERTO~1+(1|PARTICIPANTE) +(1|PALAVRA), dados.roticos,family=binomial)
m.anoroticos <- glmer(ERROACERTO~ANO+(1|PARTICIPANTE) +(1|PALAVRA), dados.roticos,family=binomial)
m.fenomenoroticos <- glmer(ERROACERTO~FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA),
dados.roticos,family=binomial)

```

```

m.fenomenoanoroticos <- glmer(ERROACERTO~FENOMENO+ANO+(1|PARTICIPANTE) +(1|PALAVRA),
dados.roticos,family=binomial)
m.intercao <- glmer(ERROACERTO~FENOMENO*ANO+(1|PARTICIPANTE) +(1|PALAVRA),
dados.roticos,family=binomial)

```

```

anova(m.zeroroticos,m.anoroticos) #ano é relevante
anova(m.anoroticos, m.fenomenoanoroticos) #fenomeno é mais relevante que ano
anova (m.fenomenoroticos, m.fenomenoanoroticos) #fenomeno + ano é mais relevante que só fenomeno
anova(m.fenomenoanoroticos, m.intercao) #intercao nao é rlevante

```

```
#MODELO ESCOLHIDO
```

```

m003 <- glmer(ERROACERTO~FENOMENO+ANO+(1|PARTICIPANTE) +(1|PALAVRA), dados.roticos,family=binomial)
drop1(m003,test="Chisq")

```

```

plot(allEffects(m003),type = "response")
summary(m003)

```

```

r<-emmeans(m003,~ANO+FENOMENO,type="response")
pairs(r)
r

```

```

graf34 <-plot(r, color="black") +
coord_flip()+
labs(y= "ano + contexto acentual", x="probabilidade estimada de erros ortográficos")+
theme_bw()

```

```
#hipótese 2 B - aumento de roticos na fala em verbos e nomes
```

```
#Grafico 35
```

```

roticos.indice<-dados.limpos %>%
select(PARTICIPANTE,ANO,PALAVRA,FENOMENO, TIPODEVARIANTE) %>%
filter(FENOMENO=="verbo"|FENOMENO=="nome")%>%
group_by(ANO,FENOMENO,TIPODEVARIANTE)%>%
summarise(n=n())%>%
na.omit(dados.limpos)%>%
mutate(porcentagemescrita = n / sum(n))%>%
filter(TIPODEVARIANTE=="proxescrita")

```

```

rotico.fala.aumento <-tibble(ano = c("1A", "3A", "5A","7A","9A"),
verbo= c("verbo","verbo","verbo","verbo","verbo"),
nverbo = c(28,54,64,70,49),
porcentagemverbo= c(0.28,0.48,0.57,0.62,0.44),
nome=c("nome","nome","nome","nome","nome"),
nnome= c(57,80,74,78,50),

```

```

porcentagemnome=c(0.57,0.75,0.66,0.71,0.44))

ggplot(rotico.fala.aumento,
  aes(x=ano))+
scale_y_continuous(labels=scales::percent_format(accuracy = 1))+
geom_point(aes(y=porcentagemverbo,colour = factor(verbo)), size=6)+
geom_path(aes(y=porcentagemverbo,colour = factor(verbo)),group=1,size=1.5)+
geom_label(aes(y=porcentagemverbo,label = scales::percent(porcentagemverbo,accuracy=1)),colour="dimgray",size=7)+
geom_point(aes(y=porcentagemnome,colour = factor(nome)), size=6)+
geom_path(aes(y=porcentagemnome,colour = factor(nome)),group=1,size=1.5,linetype = "twodash")+
geom_label(aes(y=porcentagemnome,label = scales::percent(porcentagemnome,accuracy=1)),colour="black",size=7)+
labs(y = "%)", x = "ano escolar", colour="",size=30)+
theme(text = element_text(size=20, family="serif"))+
scale_colour_manual(values= c("black","dimgray"),labels= c("ocorrência do R-final em nomes","ocorrencia do R-final em verbos"))+
theme_minimal()+
theme(legend.position = "bottom", axis.text.x=element_text(size=29), plot.title= element_text(size=29, colour="black"),axis.text.y=element_text(size=25),legend.text= element_text(size=25))

```

#### #MODELO 4 - NOMES VERBOS FALA AUMENTO ROTICO

```
dados.nomesverbos<- dados.brutos %>%
```

```
select(PARTICIPANTE,ANO,PALAVRA,TIPODEPALAVRA,FENOMENO,CATEGORIA,FORMAORAL,TIPODEVARIANTE) %>%
```

```
filter(TIPODEPALAVRA=="experimental")%>%
```

```
filter(FENOMENO=="nome"|FENOMENO=="verbo")%>%
```

```
na.omit()%>%
```

```
droplevels()
```

```
levels(dados.nomesverbos$FENOMENO)
```

```
m.004.null <- glmer(TIPODEVARIANTE~1 +(1|PARTICIPANTE) +(1|PALAVRA), dados.nomesverbos,family=binomial)
```

```
m.004.ano <- glmer(TIPODEVARIANTE~ANO+(1|PARTICIPANTE) +(1|PALAVRA), dados.nomesverbos,family=binomial)
```

```
m.004.classegramatical <- glmer(TIPODEVARIANTE ~ FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA),
dados.nomesverbos,family=binomial)
```

```
m.004.anoclasse<- glmer(TIPODEVARIANTE~ANO+FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA),
dados.nomesverbos,family=binomial)
```

```
m.004.interaçã <- glmer(TIPODEVARIANTE~ANO*FENOMENO +(1|PARTICIPANTE) +(1|PALAVRA),
dados.nomesverbos,family=binomial)
```

#### #comparação de modelos aninhados

```
anova(m.004.null,m.004.ano)# ano é melhor do que nada
```

```
anova(m.004.ano,m.004.classegramatical)#ano explica mais que classe
```

```
anova(m.004.ano,m.004.anoclasse)#ano e classe explica mais que ano sozinho
```

```
anova(m.004.anoclasse, m.004.interaçã)#interaçã explica mais
```

```
drop1(m.004.interação,test="Chisq")
```

```
summary(m.004.interação)
```

```
#gráfico 36
```

```
rvn<-emmeans(m.004.interação,~ANO*FENOMENO,type="response", adjust="tukey")
```

```
rvn
```

```
pairs(rvn)
```

```
graf36 <-plot(rvn, color="black") +
```

```
coord_flip()+
```

```
labs(y= "classe gramatical*ano", x="probabilidade estimada de r-final em nomes e verbos")+
```

```
theme_bw()
```