

HIPERCORREÇÃO DA CODA FINAL (R) EM TEXTOS DIGITAIS: UM FENÔMENO VARIÁVEL

HYPERCORRECTION OF FINAL CODA (R) ON DIGITAL TEXTS:
A VARIABLE PHENOMENON

Caio Mieiromendonça | [Lattes](#) | mieiromendonca@letras.ufrj.br
Universidade Federal do Rio de Janeiro

Marcelo Alexandre Silva Lopes de Melo | [Lattes](#) | malmelo.lopes@letras.ufrj.br
Universidade Federal do Rio de Janeiro

Resumo: Este artigo tem como objetivo a análise da hipercorreção da coda final (r) em verbos cujas formas hipercorretas coincidem com o infinitivo padrão escrito (i. “o mundo não gira, ele *capotar*”; ii. “a maldade está nos olhos de quem *ver*”; iii. “eu *dormir* muito ontem”). O *corpus* da pesquisa é composto de dados de textos escritos em plataformas digitais, como as redes sociais Twitter e Instagram e o aplicativo de conversa WhatsApp. Como arcabouço teórico, elegeram-se a Teoria da Variação e Mudança (Weinreich, Labov; Herzog, 2018 [1968]) e os Modelos Baseados no Uso (Bybee, 2016 [2002]; Pierrehumbert, 2003, 2016; Cristófaros-Silva; Gomes, 2020). As variáveis de efeitos fixos analisadas indicam a relevância da *dimensão do vocábulo* e da *vogal antecedente* como favorecedoras da hipercorreção. A análise com efeitos aleatórios aponta para um possível condicionamento lexical do fenômeno. Os resultados sugerem impactos da representação escrita das palavras no conhecimento linguístico dos falantes.

Palavras-chave: Coda final (r). Variação fonológica. Hipercorreção. Escrita digital.

Abstract: This article aims to analyze the hypercorrection of final coda (r) on verbs whose hypercorrected forms match the standard infinitive written forms i. “o mundo não gira, ele *capotar*”; ii. “a maldade está nos olhos de quem *ver*”; iii. “eu *dormir* muito ontem”). The research *corpus* is composed by data of texts posted on digital platforms, such as the social medias Twitter and Instagram, as well as the chat app WhatsApp. The main theories of this study are Theory of Variation and Linguistic Change (Weinreich, Labov; Herzog, 2006 [1968]) and Usage-based Models (Bybee, 2016 [2002]; Pierrehumbert, 2003, 2016; Cristófaros-Silva; Gomes, 2020). The fixed effects analyses indicate word dimension and preceding vowel as significant factors affecting hypercorrection. The random effects analyses point to a possible lexical conditioning of the phenomenon, thus suggesting impacts of word’s written representations on the speakers’ linguistic knowledge.

Keywords: Final coda (r). Phonological variation. Hypercorrection. Digital writing.

1 Considerações iniciais

O presente estudo tem por objeto de análise a marcação hipercorretiva da coda final (r) em verbos, tais quais os exemplos i. “o mundo não gira, ele *capotar*”; ii. “a maldade está nos olhos de quem *ver*”; iii. “eu *dormir* muito ontem”, nos quais ocorre coincidência formal (homonímia gráfica) entre verbos hipercorretos e as formas infinitivas. Esta análise tem por orientações os princípios teóricos da Teoria da Variação e Mudança (Weinreich, Labov; Herzog, 2018 [1968]) e as postulações dos Modelos Baseados no Uso (Bybee, 2016 [2002]; Pierrehumbert, 2003, 2016; Cristófaros-Silva; Gomes, 2020). O *corpus* da pesquisa foi elaborado a partir de dados extraídos de textos digitais publicados nas redes sociais Twitter e Instagram e no aplicativo de conversa WhatsApp. O objetivo do estudo é investigar a relação entre representação mental e forma ortográfica, haja vista que se entende que todas as experiências dos usuários com a língua – incluindo a forma escrita das palavras – impactam o conhecimento linguístico dos falantes. O trabalho propõe-se a: i. traçar relações entre o uso hipercorreto da coda final (r) e a marca morfológica de infinitivo; ii. investigar como ocorre o espriamento da marca hipercorretiva dentro da classe dos verbos; e iii. discutir os contextos favorecedores da marcação. Para tanto, foram mapeados os condicionamentos linguísticos (estruturais e lexicais) para a realização do fenômeno.

As hipóteses do estudo são que a marcação hipercorretiva, inicialmente, ocorreria nos verbos oxítonos, por uma analogia entre a tonicidade final e o padrão sonoro do infinitivo. Os casos de sílaba final átona seriam motivados por analogia à representação gráfica do infinitivo. Adicionalmente, tomam-se como hipóteses que a dimensão do vocábulo é inversamente proporcional à marcação hipercorretiva (monossílabos > dissílabos > mais de três sílabas); que a vogal <i> favoreceria a hipercorreção, pois, na classe dos verbos, representa tanto o morfema de P1 do pretérito perfeito quanto a vogal temática da terceira conjugação; e que, em relação aos sujeitos, posições vazias e elementos funcionais favoreceriam a hipercorreção, enquanto sujeitos com núcleos lexicais desfavoreceriam a marcação.

Este artigo está organizado em cinco seções, além desta introdução. Na seção 2, aborda-se a modelagem do conhecimento linguístico, tendo em vista os construtos teóricos da Teoria da Variação e Mudança Linguística e dos Modelos Baseados no Uso, a relação fala x escrita e o tratamento da coda (r) pelos Modelos Baseados no Uso. A seção 3 versa sobre o fenômeno da hipercorreção, visitando tanto a literatura recente sobre o fenômeno quanto estudos progressos. O conteúdo da seção 4 diz respeito à literatura

sobre a coda final (r) nas modalidades falada e escrita do português brasileiro, bem como às menções ao seu uso hipercorreto. Os procedimentos metodológicos empregados na pesquisa são descritos na seção 5. A seção 6 descreve os resultados das análises estatísticas realizadas. Por fim, a seção 7 discute os resultados, a partir dos Modelos Baseados no Uso.

2 Variação e modelagem do conhecimento linguístico

Em oposição à tradição firmada nos estudos linguísticos desde o Estruturalismo e assumidas pelo Gerativismo de se conceber a língua como uma realidade abstrata, desvinculada de fatores históricos e sociais, a Teoria da Variação e Mudança Linguística (doravante TVM) postula que o conhecimento linguístico dos falantes é heterogêneo (Weinreich, Labov, Herzog, 2018 [1968]). Assim, para a TVM, a variação é inerente ao sistema linguístico, sem causar prejuízos à intercompreensão ou ao funcionamento do sistema. Além disso, a variabilidade observada no sistema não é aleatória, mas sim condicionada por fatores estruturais, sociais e cognitivos, sendo o sistema não autônomo e abstrato. Desta forma, a concepção de língua defendida pela TVM rompe com a ideia de que o sistema linguístico seria homogêneo, o que, conseqüentemente, levou à superação das dicotomias língua x fala e sincronia x diacronia, visto que a mudança linguística é observável tanto através do tempo quanto em tendências de uso verificáveis em uma sincronia. Defende-se, portanto, que variação é inerente ao sistema linguístico, lançando-se mão do conceito de heterogeneidade ordenada para mostrar que, mesmo sendo dinâmica, a variação linguística é passível de tratamento científico.

A TVM se propõe a descrever as relações entre língua e sociedade, com vistas a observar de quais maneiras os fatores externos à língua impactam o sistema e como tais fatores se coadunam com outros de ordem cognitiva e linguística. Identifica-se que as línguas não são estruturas estanques, mas construtos dinâmicos que admitem variação e são suscetíveis a mudanças. Ainda de acordo com WLH (2018 [1968], p. 108), os sistemas linguísticos são estruturados a partir de regras categóricas, que resultam sempre em uma mesma forma invariante, e variáveis, as quais resultam em formas alternantes. A variação linguística é definida como o processo pelo qual dois ou mais elementos do sistema linguístico (fones, morfemas, palavras, construções sintáticas etc.) podem ocorrer, mantendo o mesmo valor referencial (Labov, 1978) ou equivalência funcional (Lavandera, 1978), em um mesmo contexto.

Embora não descarte a variabilidade do indivíduo, Labov (2008 [1972]) sustenta que o falante não é autônomo em relação à sociedade em que se insere. Apesar de cada

indivíduo ter, em sua mente/cérebro, os conhecimentos linguísticos internalizados, configurando suas gramáticas, todo falante está inserido em contextos sócio-históricos específicos que permitiram a aquisição da(s) variedade(s) de língua que dominam. Mesmo tendo conhecimento da organização e da variação de sua língua, o sistema é um conhecimento compartilhado entre falantes, e a língua é influenciada pelas realidades de seus usuários, de modo que falar e avaliar a fala de outros é fruto das organizações sociais a que os falantes estão sujeitos. Esse compartilhamento forma um construto abstrato que regula os usos da língua, constituindo comunidades de fala específicas que apresentam comportamentos linguísticos distintos.

WLH (2018 [1968], p. 61-63) ressaltam a possibilidade de articulação de métodos e pressupostos de teorias distintas, a fim de que se possam atingir refinamentos teóricos que melhor auxiliem na análise dos fenômenos investigados. Os próprios estudos de Weinreich, Labov e Herzog incorporam métodos de outras teorias, tais quais fatores de orientação neogramática e difusionista para avaliar os processos de mudança, além das análises da fonologia a partir da perspectiva da Teoria Gerativa Clássica, admitindo a validade dos universais linguísticos (cf. WLH, 2018 [1968], p. 74-5). Tendo em vista que a própria TVM admite a incorporação de olhares diversos advindos de outras perspectivas teóricas, esta pesquisa conjuga os pressupostos teóricos da TVM, no que diz respeito à definição do conhecimento linguístico como contendo a heterogeneidade estruturada e da relação entre língua e sociedade, bem como os pressupostos de Modelos Baseados no Uso (Bybee, 2016 [2002]); Pierrehumbert, 2003, 2016), no que diz respeito à relação entre conhecimento abstraído e uso, à organização cognitiva do conhecimento linguístico do falante e à representação da variação em exemplares.

Os Modelos Baseados no Uso ou Modelos de Exemplares (doravante MBU) concebem a língua como um sistema que comporta em sua constituição variação e gradiência. O conhecimento linguístico do falante reúne as informações perceptíveis pela experiência linguística, articulando tanto aspectos estruturais, semânticos e contextuais quanto aqueles relacionados aos contextos sociais e físicos das interações e dos falantes (Bybee, 2016, p. 35-6).

Tal perspectiva teórica descreve a gramática como uma organização cognitiva de experiências com a língua (Bybee, 2016, p. 28). A língua é um sistema dinâmico cujas estruturas não são únicas, estáticas e homogêneas, no qual a gradiência exerce papel fundamental para a mobilidade do sistema. Além disso, para os MBU, a linguagem humana não é resultado de um processo cognitivo específico, mas sim da articulação de vários processos que fazem emergir o sistema linguístico, processos esses que perpassam os ní-

veis abstratos do sistema. Com isso, um dos aspectos mais marcantes de distinção entre os MBU e aqueles que o precederam é a emergência do sistema, ou seja, a ideia de que a língua emerge do uso, não preexistindo a ele.

Dentre os processos de domínio geral, Bybee (2016, p. 26-7) apresenta alguns que são mais relevantes para o estudo da linguagem, quais sejam i. *categorização*, que se refere a um emparelhamento de identidades de palavras, sintagmas ou estruturas menores a representações estocadas na memória; ii. *chunking*, que seria o agrupamento de sequências de unidades para a formação de estruturas complexas; iii. *memória enriquecida*, relativa à estocagem mental de todos os detalhes da experiência com a língua, e iv. *analogia*, um processo de criação de enunciados com base em estruturas enunciadas previamente.

Ainda de acordo com os MBU, toda a gramática é afetada pelas experiências extralinguísticas, razão pela qual “as circunstâncias de uso impactam a representação cognitiva da língua” (Bybee, 2016, p. 35). Como se entende que, a partir dos processos cognitivos de domínio geral, o falante seja capaz de estocar na memória tudo o que percebe e fala, as formas sonoras dos itens lexicais estão representadas no léxico e constituem generalizações a partir da fala; por meio de tais generalizações, as categorias fonológicas são abstraídas. Desta forma, segundo os MBU, a estocagem é redundante, sendo todas as ocorrências concretas de uso percebidas armazenadas no léxico. Em outras palavras, os MBU sustentam que “as representações das palavras no léxico são tanto discretas, contendo informação abstrata que compõe o item lexical, quanto empíricas, ou foneticamente detalhadas, contendo a gradualidade, isto é, o detalhe fonético presente na fala” (Cristófarosilva; Gomes, 2020, p. 13). Consequentemente, todas as possibilidades fonéticas de um item lexical se encontram representadas no léxico dos falantes, e a frequência das variantes com que esses itens são representados atua no processamento (Connine; Randbom; Patterson, 2018).

Da mesma forma, como as categorias fonológicas são formadas a partir das formas armazenadas no léxico, Pierrehumbert (2003) assume que o armazenamento de informação fonética detalhada e a representação fonológica envolvem padrões fonológicos recorrentes no léxico, relações fonotáticas e relações morfossintáticas entre os itens lexicais. Ademais, por não fazerem distinção entre léxico e gramática, os MBU concebem a ideia de o léxico ser organizado em uma rede de similaridades fonéticas e semânticas, em torno da qual as palavras se encontram relacionadas. Assim, Bybee (2001) e Pierrehumbert (2003) adotam a hipótese de os itens lexicais fazerem parte de uma rede de conexões por meio da qual regularidades e similaridades entre os itens podem ser observadas. Tal hi-

pótese permite que existam representações estáveis e, ao mesmo tempo, dinâmicas, haja vista que a estrutura emerge da própria experiência de ouvir e falar.

Como a experiência de ouvir e falar impacta as representações abstratas dos falantes, a frequência com que os itens lexicais são produzidos em uma determinada língua desempenha um papel fundamental na representação abstrata das palavras (Bybee, 2001), motivo pelo qual a frequência é assumida como parâmetro fundamental para o armazenamento do conhecimento linguístico. De acordo com Bybee (2001, p. 10), existem dois tipos de frequência a serem considerados: frequência de ocorrência (token frequency) e frequência de tipo (type frequency). A frequência de ocorrência é determinada pela quantidade de vezes em que uma unidade (palavra, estrutura silábica, unidade sonora) é encontrada em um determinado corpus. Já a frequência de tipo é determinada pela frequência de um padrão específico no léxico (ou dicionário).

Embora operem em harmonia, as frequências de tipo e de ocorrência produzem diferentes efeitos sobre as mudanças sonoras. Relativamente à frequência de ocorrência, Bybee (2001) sustenta que os itens lexicais com frequência de ocorrência mais alta, por questões articulatórias (automação de produção), tendem a ser afetados mais rapidamente pela mudança sonora, ou seja, mudanças foneticamente motivadas afetam mais rapidamente os itens lexicais que têm alta frequência de ocorrência. Por outro lado, nos casos em que não se observa motivação fonética na mudança sonora: (a) a alta frequência de ocorrência de um determinado item pode garantir que este item não seja afetado pela mudança, uma vez que a sua alta ocorrência oferece ao falante fonte suficiente de aprendizado em sua experiência com a língua; (b) a baixa frequência de ocorrência de um determinado item pode fazer com que este item seja mais rapidamente afetado pela mudança, uma vez que sua baixa ocorrência não oferece ao falante fonte suficiente de aprendizado, fazendo com que o falante recorra a formas mais frequentes e similares em sua experiência com a língua.

Além dos efeitos da frequência de ocorrência, deve ser observado o efeito da frequência de tipo (type frequency) na determinação da produtividade. Cristófaros-Silva e Gomes (2007, p. 171) sustentam que, apesar de a produtividade de um padrão não ser determinada apenas pela frequência de um determinado tipo de estrutura, quanto mais itens forem abarcados por um sistema, mais forte ele se torna e, conseqüentemente, mais disponível esse mesmo item se torna para ser aplicado a outros itens (Cristófaros-Silva; Gomes, 2007, p. 171).

2.1 Retroalimentação da escrita na fala

Cristófaros Silva e Guimarães (2013) descrevem o item lexical como peça-chave para apropriação do conhecimento linguístico. O input fonológico a que o falante é exposto não é fragmentado e primariamente analisável, mas associado aos itens lexicais que reforçam e atualizam seus exemplares. A construção da representação mental dos sons, portanto, está ligada a itens inteiros e as generalizações que levam a criação de exemplares se dão via itens:

Na perspectiva multirrepresentacional, argumenta-se que o léxico atua como organizador de generalizações gramaticais e, em especial, de generalizações fonológicas. Sendo assim, as unidades representacionais de consciência fonológica seriam categorizadas via a palavra ou item léxico. (Cristófaros Silva; Guimarães, 2013, p. 319).

Os MBU, por adotarem a perspectiva teórica de que a gramática é uma complexa rede de associações emergente do uso, na qual informações linguísticas, contextuais e sociais se articulam produzindo complexos arquivos atualizáveis – exemplares –, admitem haver um pareamento entre fala e escrita que não é apenas unilateral (da fala para a escrita), mas retroalimentado.

A literatura, especificamente nos casos de falantes em processo de aquisição da escrita, discute a transposição de elementos de um sistema para outro, como a representação gráfica de vogais médias [e o] como altas [i u], a ausência de representação de segmentos praticamente ausentes da fala (fala por falar), a palatalização da consoante lateral (família por família), dentre outros fenômenos. O movimento inverso, entretanto, não tem tanta expressão, apesar de estudos já confirmarem que há retroalimentação da escrita na fala (Chevrot et al., 2000; Schwindt et al., 2007; Greco, 2009).

O trabalho de Cristófaros Silva e Guimarães (2013) defende que ocorre interferência da aquisição do sistema de escrita na fala, argumentando que tal impacto é fruto da natureza multimodal dos exemplares, que abarcam não apenas informações estruturais e sociais, mas também aquelas de ordem gráfica. A representação escrita das palavras é também mapeada no conhecimento linguístico do falante. Tal mapeamento, segundo as autoras, está intimamente relacionado com a consciência linguística dos falantes.

Se as abstrações gramaticais refletem generalizações de padrões diversos e o conhecimento linguístico é afetado também pelas representações gráficas, não é equívoco admitir que, à medida que o falante adquire conhecimentos sobre o sistema ortográfico de sua língua, haja o estabelecimento de novas generalizações que afetem todo o conhe-

cimento linguístico e, com isso, a fala. Teberosky e Colomer (2008 *apud* Cristófaró Silva; Guimarães, 2013) defendem que, sendo a escrita um sistema de representação, ao longo do processo de apropriação da escrita, tanto fala quanto escrita são afetadas.

Exemplos de retroalimentação da escrita na fala são apresentados por Greco (2009). A autora estudou o alçamento pretônico na fala e na escrita de crianças em idade escolar residentes em Belo Horizonte. Ao avaliar o impacto da apropriação das formas ortográficas de itens em que o alçamento era frequente, os resultados mostraram que, conforme os alunos avançavam na escolarização, menores eram os usos de alçamento na fala, o que indica que houve assimilação das formas gráficas das palavras e retroalimentação da escrita na fala.

Cristófaró-Silva e Guimarães (2013) discutem os resultados de Greco (2009), mostrando que os fenômenos de alçamento postônico e pretônico são manifestados na escrita de forma diferente, pois se apresentam de forma distinta na fala. A regularidade do alçamento postônico leva o escrevente a, com mais facilidade, generalizar a regra de pareamento entre o [ʊ] postônico e o grafema <o>, por exemplo, enquanto a irregularidade apresentada nos casos de alçamento pretônico gera maior dificuldade de assimilação das formas ortográficas e pode levar a um outro tipo de retroalimentação da escrita na fala, que produz hipercorreções:

Uma predição, decorrente da análise de Greco (2009), é que seria possível encontrarmos palavras que são grafadas com “i, u” pretônicos que passam a ser pronunciadas com [e,o] e, em decorrência disso, passam a ser escritas com “e, o”. Esse é exatamente o caso de palavras como “privilégio, meritíssimo, elucubração”. Em tais palavras, que entendemos ter conteúdo semântico relacionado ao prestígio, atestamos pronúncias como “pr[e]vilégio, mer[e]tíssimo” e grafias como “previlégio, meretíssimo”. (Cristófaró Silva; Guimarães, 2013, p. 32).

Por atravessamento de fatores externos, os falantes acabam produzindo pareamentos entre o conhecimento ortográfico e a organização da fala que levam à produção de formas hipercorretas, tais quais os casos de abaixamento vocálico apresentados pelas autoras, de outros abaixamentos como “orina” e “adevogado”, casos de ditongação, tais quais “tempeiro” e “pouvo”, e, dentre outros exemplos, casos que se assemelham aos estudados nesta pesquisa, que envolvem a inserção de segmentos róticos ao fim de palavras, como “maracujar”, “leite em por”¹.

¹ Ocorrências hipercorretas de ambas as palavras podem ser encontradas no seguinte vídeo: <https://youtu.be/nufG7IQXaqA>. Acesso em 08 set. 2021. Ainda não há na literatura trabalhos que abordem o fenômeno na fala.

3 Hipercorreção

Labov (2008 [1972]) postula que a hipercorreção é resultado da atribuição de significados sociais às formas variantes. O fenômeno decorre de uma hipersensibilidade ao prestígio, ou seja, a identificação por parte dos falantes tanto de traços linguísticos das normas utilizadas por grupos socialmente prestigiados quanto daqueles traços estigmatizados de suas próprias normas, sem que sejam sistematizados os contextos que geram estigma. Os usuários da língua, devido ao sentimento de insegurança linguística, hipermonitoram suas produções linguísticas corrigindo irrestritamente aqueles traços vistos como estigmatizados, a partir da imitação/reforço dos elementos das variedades prestigiadas tomados como régua. Devido à não sistematização, são produzidas formas hipercorretas que ultrapassam os contextos de uso dos traços reproduzidos das variedades modelares.

Trabalhos que abordam a escrita escolar se baseiam na leitura de Labov (2008 [1972]) sobre a hipercorreção, considerando que, no processo, ocorre reprodução de traços linguísticos tomados como modelares, mas postulando, adicionalmente, que são criadas formas linguísticas que se afastam das variedades prestigiadas e da norma-padrão (Bagno, 2012; Bortone; Alves, 2014). A literatura atual analisa o fenômeno da hipercorreção como associado à ideia de erro, fazendo menção ao conceito de hipercorreção que precedia a leitura de Labov (2008 [1972]), destacando que há reprodução de traços que a escola reforça com aporte na norma-padrão, mas cujos contextos de uso falha em sistematizar (Bortone; Alves, 2014). Bagno (2012, p. 83), por exemplo, se refere à hipercorreção como uma prática de “higiene verbal”, destacando que os falantes, tentando alcançar um ideal de “língua boa”, se afastam tanto de suas próprias gramáticas quanto da gramática normativa; Bortone e Alves (2014), seguindo pelo mesmo caminho, salientam que os falantes envolvidos em processos de mobilidade social ascendente que possuem algum conhecimento da norma-padrão, por vezes, se sentem inseguros quando do contato com modalidades novas de fala, e isso pode levar à hipercorreção se as regras da norma-padrão tiverem sido tardiamente interiorizadas por esses indivíduos (p. 130).

A forma pela qual a literatura vem reaplicando o conceito não é incoerente, haja vista que tais definições tomam por base a escola como instituição geradora da insegurança linguística. Mollica (1998) destaca que a avaliação das formas prestigiadas é oriunda das prescrições escolares e discorre acerca da inter-relação entre prestígio, mudança e escola, dizendo que a escola atua na perpetuação de comportamentos linguísticos e no

estabelecimento de estigmas sociais. A definição de Labov (2008 [1972]), entretanto, afirma que ocorre uma imitação exagerada das formas prestigiadas, mas o exagero não necessariamente resulta em formas inovadoras e que fujam ao padrão, podendo ser apenas o reforço de um traço específico, como a retroflexão dos róticos em coda abordada pelo autor. A relação que se estabelece com escolarização, portanto, é produtora de uma faceta do fenômeno da hipercorreção, e não do processo em si.

Outro aspecto relevante do tratamento da literatura sobre a hipercorreção é o fato de o fenômeno ser associado à insegurança linguística. Embora o fenômeno ocorra em situações que requerem maior monitoramento da produção linguística, como na produção de gêneros escritos com maior grau de formalidade, como artigos científicos, teses e dissertações (Belga, 2019), bem como na produção de gêneros orais tais quais conferências, entrevistas inquéritos, aconselhamentos etc., há também manifestação regular de formas hipercorretas, por exemplo, o caso do uso de vogais médias em itens como [o]rubu e d[e]lúvio, que são produzidas sistematicamente por alguns falantes, independentemente do grau de monitoramento da situação comunicativa e exibindo, com isso, um padrão previsível de hipercorreção.

4 Coda final (r): representação usual e hipercorreção

A realização da coda final (r) é variável no português brasileiro (doravante PB), sendo, conforme demonstram inúmeros trabalhos, quase que categoricamente ausente em final de verbos. A ausência da coda – variante [ø] – é analisada pelos estudos de orientação formalista como um apagamento, ao passo que investigações com base nos MBU a enxergam como uma ausência que está em alternância com outras realizações deste segmento: [h, h̃, x, ɣ, r, ɾ, ø]. A distinção entre as análises reside no fato de que estudos formalistas tomam por base uma concepção de gramática invariante, em que representações fonológicas são estanques, sendo as variações das formas fonéticas produzidas por aplicação de regras transformacionais, enquanto estudos com base nos MBU admitem variabilidade na gramática, postulando que são as formas fonológicas variantes e atualizáveis. A Teoria Gerativa Clássica dita, portanto, que o componente fonológico estoca apenas informação não redundante – ou seja, no sistema fonológico, não estariam representados todos os alofones de um mesmo fonema –, já nos MBU, as representações englobam não só informações categóricas, mas também informações redundantes e gradientes, tais como distinções de duração ou qualidade vocálica, reservando à variação *status* representacional. Tomando como base o item lexical “amar”, a visão formalista de-

fende a existência de uma unidade fonológica preexistente e invariante representativa do rótico, como na transcrição /amaR/, em que o arquifonema /R/ simula o local abstrato dessa coda na representação fonológica. A visão funcionalista dos MBU, entretanto, não admite a existência de uma entidade abstrata apriorística representante dessa coda (r), tendo em vista a postulação de que a gramática é modelada a partir do uso (Rennicke, 2016; Gomes, Carnaval, Melo, 2020). Logo, como não há *input* robusto o suficiente do segmento na fala para que se opere o reforço dos exemplares, na gramática interna, o item é majoritariamente representado sem o rótico, ao passo que as suas realizações com a coda (r) apresentam-se subpostas e vinculadas às experiências linguísticas a partir das quais os falantes as apreendem, podendo, com isso, apenas ser correlatas do maior monitoramento estilístico e/ou da expressão linguística escrita.

Conforme anteriormente assinalado, a literatura sobre a coda final (r) na fala mostra que em, itens verbais, sua realização é quase inexpressiva, o que parece refletir um processo de mudança já completamente implementado em diferentes variedades do PB (Oliveira, 1983; Callou; Moraes; Leite, 1998; Hora; Monaretto, 2003; Callou; Serra, 2012; Rennicke, 2016). Oliveira (1983) postula que ocorre, quanto ao uso do rótico em formas verbais, um processo de inserção motivado por variação estilística. O *input* que o falante de PB tem da coda final (r) em verbos, na fala, portanto, é mínimo e está restrito a contextos mais formais, de fala monitorada.

Dentre as variáveis linguísticas analisadas nos estudos de fala que são relevantes para a observação dos itens verbais, destacam-se *tonicidade silábica* e *dimensão do vocábulo*. Diversos estudos atestam que a tonicidade silábica favorece a não representação do rótico (Monaretto, 2000; Hora; Monaretto, 2003; Rennicke, 2016). Bortoni-Ricardo (2004; 2006) dá destaque à dimensão do vocábulo, mostrando que há maior incidência de uso da coda em monossílabos do que em itens com mais sílabas.

Apesar de a implementação da mudança ter levado à ausência do segmento na fala, o mesmo não se verifica na escrita, uma vez que o uso de <r> para sinalizar graficamente o morfema infinitivo é prescrito pela escola, tornando o paradigma verbal de infinitivo mais saliente à percepção e, com isso, à avaliação dos falantes. Os estudos que se debruçam sobre o registro da coda final (r) em itens verbais, na escrita atestam variabilidade de uso. A variação na representação do segmento mostra que a fala exerce impacto sobre a escrita, uma vez que a não representação da coda em verbos é categórica na fala; e que aprendizagens da escrita nem sempre sistematizam a regra de inserção do grafema <r>, o que se comprova pelo fato de a ausência de registro perdurar desde a escrita infantil até a fase adulta.

Os condicionantes internos mais relevantes para a manifestação da coda (r) em itens verbais, nos estudos sobre a escrita são *tonicidade silábica*, *contexto antecedente*, *dimensão do vocábulo*. A literatura aponta que a variável *tonicidade silábica* é relevante, apresentando palavras oxítonas como desfavorecedoras da marcação (Queiroz, 2016); a variável *contexto antecedente* indica a vogal [a] como desfavorecedora à representação da coda (Queiroz, 2016); por fim, a *dimensão do vocábulo* se mostra também significativa, tendo em vista o favorecimento da representação usual da coda em monossílabos e o desfavorecimento em vocábulos maiores, sendo dissílabos os itens que apresentam o menor índice de registro da coda, demarcando, portanto, resultados paralelos às análises da fala (Mollica, 2003; Queiroz, 2016; Martins, 2019; Branco, 2020).

O processo de hipercorreção da coda final (r) é tratado transversalmente em textos que avaliam erros escolares e em alguns estudos que abordam a posição de coda silábica. Trabalhos como Câmara Jr (2004 [1972]), Bagno (2012) e Bortone e Alves (2014) apontam esse caso de hipercorreção na escrita formal, em que se insere um rótico ao final de verbos, denominando-o como uma hipercorreção da marca de infinitivo. A literatura apresenta um consenso em relação à atuação da consciência linguística dos escreventes na manifestação dessas hipercorreções. Algumas pesquisas que investigam o apagamento do rótico na escrita escolar apenas relatam ocorrências do fenômeno, como Queiroz (2016), Almeida (2016) e Martins (2019). O estudo de Cesar (2017) se debruça especificamente sobre o fenômeno.

Queiroz (2016, p. 54) identifica a marcação hipercorretiva do rótico em verbos conjugados “mas isso não dar muito dinheiro”; em participios verbais “mais ele não estava lá devido ele ter idor preucurar”; em nomes, “Quero ter uma motor”; em preposições, “Não tinha linha nem pipa e foram ater uma barraca”, e conjunções, “por serem crianças sonhavam, uns tinha 12, 11, 10 er falavam que não”. A autora aponta a coincidência gráfica de alguns verbos conjugados e infinitivos (gosta x gostar) como fator de “confusão” para os escreventes. Os dados de Almeida (2016) mostram ocorrências de hipercorreções mais expressivas em tarefas espontâneas do que em atividades guiadas, o que é um indicativo de que a hipercorreção do rótico está relacionada às situações de menor monitoramento. O fenômeno, em itens não verbais, é analisado pela autora como uma hipercorreção formadora de verbos (Almeida, 2016, p. 48). Martins (2019) postula que, quando os escreventes realizam hipercorreções, demonstram capacidade para a assimilação, para a internalização e para a generalização de regras (p. 94) e avalia o fenômeno em itens verbais como aplicação da regra de infinitivo (p. 95).

Cesar (2017) estuda o acréscimo indevido do grafema <r> em posição de coda silábica, na escrita de estudantes do 9º ano do Ensino Fundamental, de uma escola da rede pública estadual do município de Mesquita, Região Metropolitana do Rio de Janeiro. Os dados da pesquisa foram analisados qualitativamente², considerando-se as variáveis: *extensão do vocábulo*, *contexto precedente*, *tonicidade do vocábulo* e *classe gramatical*. Os resultados de Cesar (2017), na variável dimensão do vocábulo, registram, por maior incidência de ocorrência, dissílabos, trissílabos, monossílabos e polissílabos. A *variável contexto precedente* considerou as vogais [a e i o], apontando a vogal anterior média alta como mais incidente, seguida pelas vogais anterior alta, central baixa e posterior média alta, respectivamente. A variável *tonicidade* não apresenta grandes distinções na amostra. A expressividade de verbos no *corpus* leva a autora a analisar o fenômeno como sendo originado nessa classe de palavras.

O reforço contínuo à inserção do grafema <r> em verbos imposto no contexto escolar faz com que os escreventes monitorem suas próprias escritas e também as dos demais para além da sala de aula, adotando posturas prescritivas em relação ao uso do rótico em itens verbais mesmo em ambientes nos quais o monitoramento estilístico é gradiente, como em plataformas de comunicação digital abertas e aplicativos de conversa. Nesses ambientes digitais, há recorrência de comportamentos prescritivos em relação ao registro gráfico da coda final (r). O trabalho de Mendonça (2022) observou prescrições na rede social Twitter quanto ao uso do grafema <r> em verbos, tanto no que diz respeito à ausência de registro em verbos no infinitivo quanto no que se refere à inserção hipercorretiva em formas verbais não infinitivas. O trabalho demonstra, em linhas gerais, o cunho majoritariamente assistemático e frequentemente incoerente das prescrições. Os problemas de prescrição quanto à ausência da coda (r) identificados pelo autor são: i. alusão aos termos “palavra” e “letra”; ii. menção a “palavras no infinitivo”, sem associá-las à classe dos verbos; iii. referência à classe dos verbos de maneira geral; iv. nomeação da categoria “infinitivo”, sem diferenciá das demais categorias verbais. Quanto à hipercorreção, os problemas de prescrição identificados foram: i. referência ao termo “palavras”; ii. associação do uso hipercorretivo da marca de infinitivo com “não conjugação”; iii. apontamento de algumas categorias verbais, sem apresentar qualquer sistematização. Tais problemas, além de atestarem a variação no registro escrito da coda (r) em ambientes digitais, apontam o desencontro entre o que está na gramática intuitiva da língua e o que se aprende via prescrição, bem como o fato de a regra ortográfica não ser aprendida significativamente mesmo por falantes com acesso ao ensino formal da língua.

² Não se apresentaram os dados em valores percentuais.

5 Procedimentos metodológicos

O fenômeno aqui investigado tem delimitado o seu *locus* de ocorrência nos ambientes digitais, tais quais as redes sociais Twitter e Instagram e o aplicativo de conversa WhatsApp. Em relação aos textos postados nas redes sociais públicas, dada a ausência de informações específicas sobre os usuários, como sexo, faixa etária, escolaridade, classe social etc., não há como identificar os produtores dos dados quanto a características sociais. Isso se verifica tanto em perfis pessoais, em que há apenas um usuário controlando as atividades da conta, quanto em perfis administrados por mais de um usuário, dentre os quais se destacam os de figuras públicas, os profissionais e institucionais, as páginas informativas e de humor, e os fãs clubes. Além disso, há que se considerar a existência de perfis anônimos. Assim, tendo em vista a impossibilidade de precisar características sociais dos produtores dos dados, as variáveis consideradas neste estudo são apenas de ordem linguística (estruturais e lexicais).

Quanto à recolha de dados em ambientes digitais, a coleta ocorreu tanto de maneira assistemática, à medida que as hipercorreções surgiam em textos lidos, quanto de forma orientada, e realizada na rede social Twitter. A busca por dados no Twitter foi realizada por meio da ferramenta “pesquisar” (lupa) da própria rede social, visando a paradigmas verbais específicos (itens de conjugações, dimensões e com tonicidades das sílabas finais distintas). Em relação aos verbos pesquisados, sua escolha foi orientada de duas maneiras: i. visando ao levantamento das ocorrências hipercorretas de determinados itens lexicais de maior aparição nas interações rotineiras do pesquisador – verbos como **buscar**, **ver**, **conseguir**, dentre outros –, e ii. a fim de reunir ocorrências de itens de grafia distinta do infinitivo padrão – tais quais **entender**, **dirar** e **comor**³. Quando da pesquisa por formas em que há coincidência gráfica entre verbos hipercorretos e infinitivos padrões (“estar”, “poder”, “assistir”), obtêm-se resultados tanto de hipercorreções quanto de verbos no infinitivo, então, como estratégia para diminuir a quantidade de ocorrências, fez-se necessário pesquisar sintagmas ou frases (*essa **estar**; o rapaz **poder** fazer; a primeira vez que **assistir***). Tendo em vista que o Twitter não se restringe à busca linear de dados, mesmo com tais combinações, as postagens em que ocorrem hipercorreções aparecem na busca em meio àquelas em que não há casos de hipercorreção. Feitas as pesquisas, são observadas as postagens apresentadas na guia, e, dentre todos os resultados obtidos, são selecionados apenas aqueles em que há pelo menos um caso de hipercorreção. O rastreamento de verbos cuja grafia não coincide com a do infinitivo padrão é mais simples, visto que os

³ Destaque-se que não se pretendia, no início do levantamento da amostra, observar a hipercorreção estatisticamente, porém, os dados apontaram o caráter variável do fenômeno.

resultados apresentam sempre formas hipercorretas (**vendir, irar, encontrour**). Nesses casos, o esforço maior da pesquisa é ampliar os contextos de uso dos verbos (*eu vendir x ela vendir*), para variar a amostra.

A pesquisa desenvolvida conta com um *corpus* de 419 ocorrências de verbos (170 sem a marca hipercorretiva e 249 com a marca), que faz parte de uma amostra com 1.635 dados, da qual foram selecionados apenas os casos em que ocorriam homonímias com os infinitivos padrões escritos (**falar** por *fala*; **poder** por *pode*; **mentir** por *menti*). A escolha do recorte se justifica por dois motivos, quais sejam, o objetivo de traçar uma investigação sobre o impacto da representação gráfica das palavras no conhecimento dos falantes e a menção na literatura sobre a interferência da coincidência formal nos casos de hipercorreção. A variável dependente selecionada foi a realização da coda final (r). Por se tratar de um fenômeno variável, os casos de não marcação foram considerados (i. *você sabe que hoje não dar*). Descartaram-se verbos com sílaba travada por nasal (“amam”) ou rótico (“quiser”), verbos no gerúndio (“mentindo”) e no particípio (“vendido”) e os dados em que o uso do -r é ambíguo, podendo ser o verbo interpretado como hipercorreção, infinitivo ou forma flexionada nos modos subjuntivo ou imperativo (i. *deixar as coisas para depois* [infinitivo ou imperativo]; ii. *quem crer, curte* [subjuntivo ou presente]). As variáveis independentes analisadas foram as seguintes:

Quadro 1 – variáveis independentes analisadas

VARIÁVEL	VARIANTES		
<i>Tonicidade da sílaba final</i>	tônica (<i>essa estar boa</i>)	átona (<i>ele só assistir tv</i>)	
<i>Vogal antecedente⁴</i>	<a> (estar)	<e> (saber)	<i> (mentir)
<i>Dimensão do vocábulo</i>	monossílabo (dar)	dissílabo (pedir)	mais de três sílabas (responder)
<i>Expressão do sujeito</i>	lexical (<i>mãe ver tudo</i>)	funcional (<i>você não engajar</i>)	não explícito (<i>não entendir</i>)

Fonte: elaboração própria.

⁴ Destaque-se que se optou pela representação grafemática dos segmentos na variável *vogal antecedente* (<a>, <e>, <i>), pelo fato de que a pesquisa buscou estabelecer um pareamento entre a forma escrita dos infinitivos padrões e a hipercorreção, de modo que se analisaram as representações gráficas dos vocábulos. Reforça-se, portanto, que o impacto investigado não se restringe à grafia padrão, visto que a variação entre vogais anteriores médias e altas na fala leva à produção de formas escritas como “*Marcia pedir*” (Marcia pede).

Ao todo, foram coletados 145 verbos distintos; dentre eles, 33 apresentaram maior frequência, ocorrendo mais de três vezes no *corpus*⁵. A seguir, os verbos mais frequentes e os valores totais de ocorrência (variante culta/variante hipercorreta): acha (1/3); acorda (2/3); assisti (0/9); busca (0/7); chama (4/0); chega (1/3); consegui (2/7); continua (2/2); cre (1/8); curte (2/7); curti (1/3); dá (7/32); deixa (3/0); descobri (0/4); desculpa (2/1); deve (2/1); está (16/10); fala (7/4); fica (4/3); manda (3/1); mora (2/1); multiplica (0/3); ouvi (0/3); para (3/0); parece (6/3); pode (7/0); responde (0/3); ri (0/14); sabe (8/3); sai (5/5); segui (1/2); senti (2/9); vê (2/28). Dentre os 112 itens computados com menos de três dados, 45 apresentam apenas a variante hipercorreta; 54 somente a variante culta, e 13 ambas as variantes.

A pesquisa aqui desenvolvida tem orientação quantitativa. Após a coleta e a computação dos dados, o *corpus* foi submetido à regressão logística no programa Rbrul, que é uma função interativa do programa estatístico de regressão logística R⁶, que permite ao usuário analisar variáveis fixas e aleatórias (Gomes, 2012). Neste trabalho, as variáveis de efeito fixo analisadas são *tonicidade da sílaba final*, *vogal antecedente*, *dimensão do vocábulo* e *expressão do sujeito*, e a variável aleatória é *item lexical*.

Em textos digitais, ocorre a interferência de ferramentas de correção ortográfica. Não se consegue delimitar exatamente até onde atua essa interferência, uma vez que os corretores automáticos trazem, em suas configurações, informações prévias dos *softwares* e, além disso, são alimentados pelos usuários, fixando formas ortográficas específicas e assimilando novas palavras. Tendo em vista esses fatores, a coleta de dados por busca lexical foi empregada também como uma estratégia metodológica para minimizar os efeitos do corretor automático. A forma de coleta proporcionou a observação de itens inovadores, como “**encontrarar**”, “**perdir**”, “**queror**” e “**estour**”, atestando a ocorrência

⁵ A escolha pelo número de três aparições no *corpus* como valor limítrofe da frequência de ocorrência deve-se a dois fatos, quais sejam: i. a busca lexical de itens cujas formas hipercorretas geram homônimas com os infinitivos padrões escritos resultar em um número muito grande de postagens para rastreamento, dentre as quais há apenas uma ou duas ocorrências; ii. serem considerados todos os verbos passíveis de receber a marca hipercorretiva que se encontram nos textos em que há pelo menos um caso de hipercorreção (conforme descrito no envelope da variação). Observe-se, portanto, que nem todos os itens computados na amostra foram objetos de pesquisa da busca sistemática via Twitter.

⁶ Consideram-se três valores na análise estatística do programa Rbrul: *p*-valor, *log-odds* e peso relativo. O nível de significância pré-selecionado pelo *software* é 0.05 (5%). Se *p*-valor for maior que o nível de significância, a variável é menos significativa, ao passo que, quanto menor é *p*-valor, mais significativa é a variável. O programa indica *log-odds* positivos e negativos. Se positivo, a variante favorece a aplicação do fenômeno, enquanto, se negativo, há o seu desfavorecimento. O índice de peso relativo, adotado pelo Rbrul é 0.5 (50%), tal que os valores maiores do que o índice são favorecedores da aplicação da variante investigada e os menores desfavorecem sua aplicação.

e a variabilidade dessas hipercorreções⁷. Além disso, a busca por itens nos quais ocorre a coincidência gráfica com o infinitivo, tais como “falar”, “correr” e “rir” encontra como resultados dados em que há verbos iguais em mesmo contexto, mas que se comportam diferentemente em relação à hipercorreção, tal como em “*mas eu ri muito com o áudio dele, ah eu rir*”. Adicionalmente, a busca por sintagmas e frases auxilia no mapeamento de contextos estruturais específicos, por exemplo, a presença de verbos auxiliares nas sentenças, como “*quem decide como o dia **vair** ser é nós mesmo*”, e ainda a identificação de estruturas paralelísticas, como “*eu tô em triagem agora, **conseguir** fugir a tempo*”. A forma de coleta, portanto, atua na pesquisa ampliando o escopo de investigação, a fim de garantir a variabilidade do *corpus*.

A unidade a seguir descreve os resultados obtidos pela análise estatística.

6 Análises

As análises desenvolvidas tiveram por base a realização de duas rodadas de regressão logística no programa Rbrul. A primeira rodada considerou apenas variáveis de efeito fixo, enquanto a segunda utilizou os itens que ocorriam mais de três vezes no *corpus* como variáveis aleatórias, a fim de tentar capturar possíveis efeitos de natureza lexical.

A tabela a seguir demonstra os resultados da rodada com efeitos fixos:

Tabela 1 – p-valores das variáveis independentes analisadas

Variável	Dimensão do vocábulo	Vogal antecedente	Tonicidade da sílaba final	Expressão do sujeito
p-valor	5.11e-05 ⁸	0.00375	0.0953	0.295

Fonte: Mendonça (2021, p. 61).

Na análise com efeitos fixos, as variáveis selecionadas pelo programa foram *dimensão do vocábulo* e *vogal antecedente*, como se observa na tabela anterior. As variáveis *tonicidade da sílaba final* e *expressão do sujeito* não foram apontadas como favorecedoras do fenômeno.

A seguir, os resultados estatísticos da variável *dimensão do vocábulo*:

⁷ Como mencionado anteriormente, esses casos não são abordados neste trabalho por o recorte do *corpus* considera apenas os casos de coincidência gráfica com o infinitivo padrão escrito, haja vista o interesse inicial da pesquisa em observar como as representações gráficas interagem com o conhecimento linguístico dos falantes.

⁸ O símbolo e-número representa a quantidade de zeros antes do número indicado. Portanto, p-valor é igual a 0.00000511.

Tabela 2 – Resultados para *dimensão do vocábulo*

Variante	Log-odds	Tokens	%	Peso relativo
Monossílabo	0.753	102	84.3	0.680
Mais de três sílabas	0.009	122	61.5	0.502
Dissílabo	-0.762	195	45.1	0.318

Fonte: Mendonça (2021, p. 64)

A variável indicou como favoráveis à hipercorreção as variantes *monossílabo* (*tu crer?*) e *mais de três sílabas* (*aí adquirir um*), apontando a variável *dissílabo* (*isso se tornar mais difícil*) como desfavorável. Alguns trabalhos já haviam demonstrado índices semelhantes quanto à marcação da coda (r) na escrita, como Almeida (2016) e Queiroz (2016). Ambas as pesquisas indicam que monossílabos e dissílabos estão em polos opostos no que tange à marcação de róticos na escrita escolar, sendo os primeiros mais marcados, enquanto os últimos recebem menos marcações. Os resultados apontam que o padrão manifestado na representação usual da coda se repete nas hipercorreções (monossílabos > mais de três sílabas > dissílabos), indicando que o *input* de representação gráfica da coda (r) que o falante recebe impacta positivamente na marcação hipercorretiva do segmento. Uma vez que há mais ocorrências de verbos monossílabos com <r> do que verbos com mais de três sílabas e dissílabos, respectivamente (Branco, 2019; Martins, 2019), isso influencia para que, no momento da escrita, o falante identifique os contextos favorecedores para o uso da marca hipercorretiva com base no uso. Tal resultado evidencia o efeito da frequência de tipo para o condicionamento do fenômeno, haja vista que a inserção hipercorretiva do rótico, ao que indicam as análises, é um decalque de macro-contextos estruturais do *input* de representação usual da coda final (r) em verbos.

Na sequência, os resultados da variável *vogal antecedente*:

Tabela 3 – Resultados para *vogal antecedente*

Variante	Log-odds	Tokens	%	Peso relativo
<i>	0.847	89	84.3	0.70
<a>	-0.352	236	50.8	0.413
<e>	-0.495	94	57.4	0.379

Fonte: Mendonça (2021, p. 66)

A variável indicou apenas a vogal <i> (*ainda bem que eu **dormir***) como favorável à hipercorreção, apontando as variantes <a> (*agora **soltar** no chão*) e <e> (***ver** os preços e me fala*) como desfavorecedoras. Essa manifestação mais expressiva das hipercorreções em verbos terminados em <i> é analisada como resultante da transversalidade do segmento na classe de palavras. O segmento <i>, na grafia padrão, representa dois morfemas distintos: tanto a marca de terceira conjugação quanto a de P1 do pretérito perfeito do indicativo. Esse fato é o possível desencadeador de o escrevente hipercorrigir com maior frequência nos verbos em que a vogal final é <i>. O resultado da variável *vogal antecedente* evidencia a relação que se estabelece entre o uso e o conhecimento linguístico, visto que há dois contextos significativos no paradigma de flexão verbal do PB que fazem uso do segmento, e tal fato é relevante para o conhecimento linguístico do falante. Ou seja, considera-se que o fato de a vogal <i> representar morfemas tipicamente verbais (enquanto <a> e <e> estão presentes também na morfologia nominal) é gatilho da consciência linguística quanto à classe de palavras, auxiliando o falante a identificar os verbos terminados em <i> como membros da classe.

Nenhuma variável independente foi selecionada pelo programa quando os *itens* foram incluídos como variável aleatória, o que pode indicar um possível condicionamento lexical. Assim, é possível que, por reconhecer a classe de palavras – no caso, a dos verbos –, o falante faça, de forma generalizada, uma analogia a todas as formas verbais. Tendo em vista que todos os dados analisados neste estudo formam homônimos com os infinitivos padrões escritos, é corroborada a hipótese de que haveria interferência expressiva das representações gráficas das formas verbais infinitivas no conhecimento linguístico. Tal fato reforça as constatações da literatura de que os verbos cujas formas conjugadas geram coincidência gráfica com o infinitivo sem a coda (r) favorecem a hipercorreção do segmento por serem motivo de confusão para os escreventes.

A recursividade do processo de hipercorreção quanto ao uso da coda final (r) em verbos se deve inicialmente a dois fatos: a) no português brasileiro, o segmento é praticamente ausente na fala, e b) seu uso é resgatado na escrita pela escola. A “interferência” da fala na escrita leva à produção de formas verbais infinitivas escritas sem a coda, como i. “vou **da** tudo de mim”, ii. “quero **sabe** quem é você”, iii. “eu já cansei de **assisti** isso”, e a escola, por sua vez, atua como entidade de propagação do estigma, repreendendo os usos que se afastam da norma-padrão ortográfica. Tal repreensão leva os escreventes a corrigirem inicialmente à insegurança linguística, que faz com que os falantes ultrapassem os contextos

de uso do segmento e registrem a coda final (r) em formas verbais conjugadas, como nos exemplos i. “*eu **conseguir** o brinquedo*”, ii. “*conhecer e ver o que **rolar***”, iii. “*aquela amiga que só **responder** com sim*”. O caso da coda final (r), por se tratar de uma hipercorreção resultante (em primeira ou segunda instância) de uma regra ortográfica reforçada pela escola, passa a ser manifestado regularmente nas produções escritas, e se enquadra nos casos de hipercorreção sistemática abordados na seção 2 deste artigo, pois ocorre tanto em textos mais monitorados quanto em ambientes digitais, como redes sociais e aplicativos de conversa, nos quais se apresentam textos com gradações de formalidade quanto à escrita. Os dados aqui analisados, portanto, são evidências de que as pressões que o falante sofre surtem efeito não apenas na escolha por uma ou outra variante, em processos de variação estilística, mas podem impactar o sistema, levando à produção regular de hipercorreções até em contextos de baixo monitoramento.

Na seção seguinte, os resultados apresentados para as variáveis estruturais serão explorados à luz dos Modelos Baseados no Uso, discutindo como tais modelos podem melhor explicar o fenômeno em análise, tendo em vista o impacto da representação escrita das palavras no conhecimento linguístico dos falantes.

7 Discussão geral

Os resultados do estudo evidenciam o efeito do atravessamento escolar nos comportamentos linguísticos do falante/escrevente. A supressão da coda (r) nos verbos, por ser um fenômeno natural da fala, transborda para a escrita, enquanto a sua representação é um fenômeno praticamente restrito à escrita, que é reforçado pela escola. Uma vez que algumas regras apreendidas a partir do ambiente escolar são propaladas *extra muros*, é provável que o ensino normativo afete não apenas aqueles que tiveram acesso à educação formal. O usuário da língua, por não ter a coda final (r) em itens verbais representada no seu conhecimento linguístico, precisa localizar os membros da classe de palavras para, a partir daí, identificar e sistematizar os contextos de uso do segmento, o que ocorre via escolarização. Analisa-se o uso hipercorretivo da coda final (r) em verbos como um indício do impacto da representação gráfica das palavras no conhecimento linguístico do falante. O falante opera generalizações e a não sistematização da regra ortográfica faz com que se identifique a classe dos verbos como o contexto de uso da coda (r), levando ao seu espraiamento para outros elementos além dos infinitivos.

Face aos resultados apresentados, faz-se necessário um modelo de organização do conhecimento linguístico que assuma uma modelagem capaz de melhor acomodar a va-

riação, a fim de que seja possível explicar como a experiência do falante com a língua, em diferentes modalidades (fala e escrita), impacta a organização e o processamento da linguagem. Neste sentido, os MBU, por assumirem que o uso impacta a gramática, permitem postular que em “pessoas que já adquiriram a escrita, os exemplares também contêm informações a respeito da forma gráfica de cada item léxico” (Cristófar-Silva; Guimarães, 2013, p. 318). Assim, como os falantes não ouvem a coda (r) em final de verbos, mas, em razão da experiência com a escrita, sabem que uma letra “r” deve ser colocada em itens verbais, estendem essa informação para todas as formas verbais, independentemente de ser um verbo na forma nominal do infinitivo ou não. Um modelo formal de gramática que preveja apenas a aplicação de regras variáveis não seria capaz de explicar a generalização sistemática observada nos dados desta pesquisa, uma vez que, conforme se observou, parece não haver condicionamentos estruturais atuando no fenômeno quando o item lexical é incluído nos modelos de análise como variável de efeito aleatório. Ademais, mesmo que se pense em processo de difusão lexical, não seria possível – ou pouco provável – explicar, a partir de um modelo formal de gramática, a interferência da fala na escrita e vice-versa.

Outra questão desta análise refere-se ao tratamento do fenômeno da hipercorreção. Os estudos sobre a hipercorreção descrevem o fenômeno como um processo estritamente relacionado ao monitoramento da produção linguística (Labov, 2008 [1972]; Bagno, 2012; Bortone; Alves, 2014). Observa-se, entretanto, nos dados uma supergeneralização, ou seja, mesmo em textos menos monitorados, aparecem formas hipercorretas. Isso parece indicar que não ocorre apenas hipermonitoramento quando do uso da coda (r), mas que o processo já se encontra computado na gramática desses escreventes.

Embora não se saibam quais as características sociais dos escreventes dada a natureza dos dados, há que se considerar padrões recorrentes de interação entre os redatores em ambientes digitais (tal como as posturas prescritivas em relação ao grafema <r> em itens verbais) e ainda como tais padrões podem moldar o uso da língua, estabelecendo práticas de letramento dos usuários de tais plataformas. Os dados extraídos da comunicação digital em plataformas abertas, portanto, mesmo que imponham certas restrições em relação ao mapeamento dos impactos de fatores sociais na língua, não impossibilitam a pesquisa linguística, dada a sua inserção em contextos reais de uso.

Apesar do fato de a hipótese da tonicidade não ter sido comprovada, as análises futuras da pesquisa, que considerarem os demais dados da amostra (que não geram homônimas com os infinitivos escritos) poderão revisitar tal hipótese, a fim de contrastar

os resultados estatísticos para os casos de hipercorreção com aqueles obtidos por estudos que analisam a produção usual da coda final (r) em textos falados e escritos.

Este estudo se mostra relevante para as pesquisas linguísticas por: i. ser o primeiro a analisar o uso hipercorretivo da coda (r) em ambientes digitais e a tratar o fenômeno como variável; ii. contribuir com as investigações sobre a manifestação da coda final (r) em textos digitais, área com pouca expressividade de trabalhos ainda; iii. investigar a relação entre representação mental e forma ortográfica e seus impactos na modelagem do conhecimento linguístico.

Referências

ALMEIDA, D. C. *Tratamento didático do apagamento e inserção da rótica em final de verbos*. Fortaleza: Universidade Federal do Ceará (Dissertação - Mestrado Profissional em Letras), 2016.

BAGNO, M. *Gramática pedagógica do português brasileiro*. São Paulo: Parábola Editorial, 2012.

BELGA, J. S. *Hipercorreção na escrita acadêmica: uma análise de textos de alunos da graduação em Direito*. Belo Horizonte: Faculdade de Letras da UFMG, Dissertação (mestrado), 2019.

BORTONE, M. E.; ALVES, S. B. O fenômeno da hipercorreção. In: BORTONI-RICARDO, S. M. *Por que a escola não ensina gramática assim?* São Paulo: Parábola Editorial, 2014.

BORTONI-RICARDO, S. M. *Educação em língua materna: a sociolinguística em sala de aula*. São Paulo: Parábola, 2004.

BORTONI-RICARDO, S. M. O estatuto do erro na língua oral e na língua escrita. In: GORSKI, E. M.; COELHO, I. L. (Org.). *Sociolinguística e ensino: contribuições para a formação do professor de língua*. Florianópolis: UFSC, p. 267-276, 2006.

BRANCO, A. A. T. C. *O apagamento do rótico em coda final em produções escritas no Ensino Fundamental II*. Instituto de Letras e Linguística da Universidade Federal de Uberlândia Mestrado Profissional em Letras (PROFLETRAS), 2020.

BYBEE, J. Usage-based Theory and Exemplar Representation. In: HOFMANN, T.; TROUSDALE, G. (Org.). *The handbook of Construction Grammar*. Oxford: Oxford University Press, p. 49-69, 2013.

- BYBEE, J. L. *Língua, uso e cognição*. São Paulo: Cortez, 2016 [2002].
- CALLOU, D.; MORAES, J.; LEITE, Y. Apagamento do R final no dialeto carioca: um estudo em tempo aparente e em tempo real. *DELTA*, São Paulo, v. 14, n. spe, 1998.
- CALLOU, D.; SERRA, C. Variação do rótico e estrutura prosódica. *Revista do GELNE (UFC)*, v. 1, p. 41-57, 2012.
- CAMARA Jr., J. M. Erros escolares como sintomas de tendências linguísticas no português do Rio de Janeiro [1972]. In: UCHÔA, C. E. F. (Org.). *Dispersos de Mattoso Camara Jr.* Rio de Janeiro: Lucerna, 2004.
- CESAR, H. H. F. *Acréscimo do grafema <r> em coda silábica: intervenção para casos de hipercorreção*. Dissertação (Mestrado Profissional em Letras). Instituto de Ciências Humanas e Sociais, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2017.
- CHEVROT, J.; BEAUD, L.; VARGA, R. Developmental data on French sociolinguistic variable: post-consonantal word-final. *Language Variation and Change*. Cambridge, v. 12, p. 295-319, 2000.
- CONNINE, C. M.; RANBOM, L. J.; PATTERSON, D. J. Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, v. 70, p. 403-411, 2008.
- 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. *Letras de Hoje*, Porto Alegre, v. 48, n. 3, p. 316-323, jul./set. 2013.
- CRISTÓFARO SILVA, T.; GOMES, C. A. Representações Múltiplas e Organização do componente linguístico. *Fórum Linguístico*, Universidade Federal de Santa Catarina. Centro de Comunicação e Expressão. Pós-Graduação em Linguística, n.1, v. 4, 147-177, 2007.
- CRISTÓFARO SILVA, T.; GOMES, C. A. Teoria de Exemplares. In: HORA, D.; MATZENAUER, C. L. *Fonologia, fonologias: uma introdução*. São Paulo: Contexto, p. 157-168, 2017.
- CRISTÓFARO SILVA, T.; GOMES, C. A. Fonologia na perspectiva dos Modelos de Exemplares. In: GOMES, C. A. (Org.). *Fonologia na perspectiva dos Modelos de Exemplares: para além do dualismo natureza/cultura na ciência linguística*. São Paulo: Contexto, 2020.

GOMES, C. A.; CARNAVAL, M.; MELO, M. A. S. L. . Variação da coda (r) em interior de palavra na comunidade de fala do Rio de Janeiro: aspectos gradientes. In: GOMES, C. A. (Org.). *Fonologia na perspectiva dos Modelos de Exemplares: para além da dicotomia natureza/cultura na ciência linguística*. 1ed. São Paulo: Contexto, 2020, v. 1, p. 157-175.

GOMES, C. A. Para além dos pacotes estatísticos VARBRUL/GOLDVARB e RBRUL: qual a concepção de gramática? *Revista do GELNE*, v. 14, n. 1/2, p. 259-272, 2012.

GOMES, C. A. (Org.). *Fonologia na perspectiva dos Modelos de Exemplares: para além do dualismo natureza/cultura na ciência linguística*. São Paulo: Contexto, 2020.

GRECO, A. *Alçamento de vogais médias pretônicas do português na oralidade de crianças de Belo Horizonte*. Monografia (Graduação) – Faculdade de Letras da UFMG, Belo Horizonte, 2009.

HORA, D.; MONARETTO, V. O. Enfraquecimento e cancelamento dos róticos. In: HORA, D.; Collischonn, G. (Org.). *Teoria Linguística: fonologia e outros temas*. João Pessoa: Editora Universitária, p. 114-143, 2003.

LABOV, W. Where does the linguistic variable stop? A response to Beatriz Lavandera. *Sociolinguistic Working Papers*, 44. Austin, Texas: Southwest Educational Development Laboratory, 1978.

LABOV, W. *Padrões sociolinguísticos*. São Paulo: Contexto (2008 [1972]).

LAVANDERA, B. R. Where does the sociolinguistic variable stop? *Language in Society*, v. 7, p. 171-82, 1978.

MARTINS, K. G. *Variação na escrita do /r/ final: uma análise em textos escritos e dados orais de alunos do Ensino Fundamental I*. Universidade Federal de Ouro Preto: Departamento de Letras. Programa de Letras: Estudos da Linguagem (Dissertação - Mestrado Acadêmico), 2019.

MENDONÇA, C. M. “O mundo não gira, ele capotar”: hipercorreção e variabilidade no uso da coda (r) ao final de verbos. Rio de Janeiro: Universidade Federal do Rio de Janeiro, Faculdade de Letras, 2021. Dissertação de Mestrado em Letras Vernáculas (Língua Portuguesa)

MENDONÇA, C. M. Da hipercorreção ao reforço pragmático: considerações sobre a marca de infinitivo. *Cadernos do NEMP*, v. 13, p. 35-47, 2022.

MOLLICA, M. C. *Influência da fala na alfabetização*. Rio de Janeiro: Tempo Brasileiro, 1998.

MONARETTO, V. N. O apagamento da vibrante posvocálica nas capitais do sul do Brasil. *Letras de Hoje*, Porto Alegre, v. 35, n. 1, p. 275-284, 2000.

OLIVEIRA, M. A. Reanálise de um problema de variação. *Português: Estudos Lingüísticos*, Uberaba, n. 7, p. 23-51, (Série Estudos), 1983.

PIERREHUMBERT, J. B. Knowledge of Variation. CLS 30, *Papers from Parasession on Variation and Linguistic Theory*, 1994.

PIERREHUMBERT, J. B. Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P. Hopper (Ed.). *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001.

PIERREHUMBERT, J. B. Probabilistic Phonology: discrimination and robustness. In: R. BOD, J. HAY, S. JANNEDY (Ed.). *Probabilistic Linguistics*. Cambridge: MIT Press, p. 177-228, 2003.

PIERREHUMBERT, J. B. Phonological representation: beyond abstract versus episodic. *Annual Review of Linguistics*, v. 2, p. 33-52, 2016.

QUEIROZ, V. T. *A ausência de registro da vibrante na escrita de alunos do Ensino Fundamental II em ambiente on-line e off-line*. Rio de Janeiro: PROFLETRAS UFRJ, 2016.

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.

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

TEBEROSKY, A.; COLOMER, T. *Aprender a ler e escrever: uma proposta construtivista*. Porto Alegre: Artemed, 2003.

WEINREICH, U.; LABOV, W.; HERZOG, M. I. *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo: Parábola Editorial, 2018. [1968].



Data de submissão: 21/12/2022

Data de aceite: 10/03/2023