

Estado da publicação: O preprint não foi publicado em outro meio.

Do Varbrul ao R: uma reanálise das vogais médias pretônicas em Fortaleza-CE

Ronaldo Manguiera Lima Júnior, Pablo Arantes, Guilherme Garcia, Luciana Lucente, Renata Passetti

<https://doi.org/10.1590/SciELOPreprints.15222>

Submetido em: 2026-02-25

Postado em: 2026-08-10 (versão 3)
(AAAA-MM-DD)

Justificativa da versão: Acréscimo de detalhes metodológicos.

Categoria da Contribuição: Protocolo de Relato Registrado - Envio inicial

Do Varbrul ao R: uma reanálise das vogais médias pretônicas em Fortaleza-CE

From Varbrul to R: a reanalysis of pretonic mid vowels in Fortaleza-CE

Ronaldo Lima Jr.

Doutor em Linguística; Universidade de Brasília; Departamento de Linguística,
Português e Línguas Clássicas; Brasília-DF; ronaldo.junior@unb.br;
<https://orcid.org/0000-0002-8610-0306>

Pablo Arantes

Doutor em Linguística; Universidade Federal de São Carlos; Departamento de Letras;
São Carlos-SP; pabloarantes@ufscar.br; <https://orcid.org/0000-0001-9707-8493>

Guilherme Garcia

Doutor em Linguística; Université Laval; Département de Langues, Linguistique et
Traduction; Québec-QC, Canadá; guilherme.garcia@lli.ulaval.ca;
<https://orcid.org/0000-0003-1412-3856>

Luciana Lucente

Doutora em Linguística; Universidade Federal de Minas Gerais; Faculdade de Letras;
Belo Horizonte-MG; lucente@ufmg.br; <https://orcid.org/0000-0001-6325-0531>

Renata Passetti

Doutora em Linguística; Pesquisadora independente; Sumaré-SP;
re.passetti@gmail.com; <https://orcid.org/0000-0002-1547-2831>

Resumo

Propomos reanalisar os dados da tese *As vogais médias pretônicas no falar popular de Fortaleza: uma abordagem variacionista* (Araújo, 2007), que investigou a realização das vogais pretônicas sob a perspectiva da Sociolinguística Variacionista. O estudo original utilizou o VARBRUL para modelar o efeito de variáveis linguísticas e sociais sobre os processos de alteamento, abaixamento e manutenção das vogais. Nossa proposta é utilizar modelos de regressão logística de efeitos mistos no R, incorporando avanços metodológicos da análise estatística, como a inclusão de efeitos aleatórios para informante e item lexical. Pretendemos verificar a robustez dos resultados originais frente a técnicas mais recentes, bem como discutir as implicações da modelagem mista para a interpretação de fenômenos variáveis. A reanálise busca contribuir para a promoção de práticas de reprodutibilidade na Linguística, oferecendo um exemplo de replicação com dados reais e disponibilização transparente do código e dos resultados.

Palavras-chave: fonética, R, sociolinguística, vogais

Abstract

We propose to reanalyze the data from the thesis "Pretonic Mid Vowels in the Popular Speech of Fortaleza: A Variationist Approach" (Araújo, 2007), which investigated the

realization of pretonic vowels from the perspective of Variationist Sociolinguistics. The original study used VARBRUL to model the effect of linguistic and social variables on the processes of raising, lowering, and maintaining vowels. Our proposal is to use mixed effects logistic regression models in R, incorporating methodological advances in statistical analysis, such as the inclusion of random effects for informants and lexical items. We intend to verify the robustness of the original results against more recent techniques, as well as discuss the implications of mixed modeling for the interpretation of variable phenomena. The reanalysis seeks to contribute to the promotion of reproducibility practices in Linguistics, offering an example of replication with real data and transparent availability of the code and results.

Keywords: phonetics, R, sociolinguistics, vowels

Introdução

Este trabalho propõe reanalisar os dados originais da tese de Araújo (2007), que investigou a realização das vogais médias pretônicas no falar popular de Fortaleza, Ceará, à luz da Sociolinguística Variacionista. O estudo original utilizou o programa VARBRUL (*Variable Rule Program*), modelo clássico da Sociolinguística Quantitativa desenvolvido por David Sankoff no início dos anos 1970, destinado à modelagem estatística de fenômenos linguísticos variáveis (Cedergren; Sankoff, 1974).

A escolha de Araújo (2007) pelo modelamento estatístico no software VARBRUL reflete a prática majoritária dos estudos sociolinguísticos variacionistas da época, cuja literatura evidencia uma tradição de análises quantitativas conduzidas com essa ferramenta até o final da primeira década dos anos 2000 (Oushiro, 2022; Tagliamonte, 2011). A combinação entre recursos que permitiam a condução de análises inferenciais de dados variáveis e uma interface amigável e de uso intuitivo contribuiu para a ampla adoção do VARBRUL nesses estudos (Oushiro, 2022).

No entanto, a partir do final dos anos 2000, consolida-se uma mudança de paradigma nos métodos quantitativos da pesquisa variacionista, motivada por críticas às limitações das análises realizadas pelo VARBRUL e pela adoção de novas ferramentas estatísticas consideradas mais robustas para a modelagem de fenômenos linguísticos variáveis (Torres Vieira, 2022). O modelo mais recente do VARBRUL permite a condução de um único tipo de análise, a regressão logística, adequado para variáveis de resposta binárias (Lima Jr; Garcia, 2021). Além disso, as variáveis preditoras devem ser necessariamente nominais. Por isso, nos estudos sociolinguísticos que utilizam essa ferramenta para modelagem estatística dos dados, variáveis originalmente contínuas, como idade, são tratadas de maneira discreta, isto é, através de sua redução a faixas etárias (Oushiro, 2022). Ademais, o VARBRUL trabalha com um processo de seleção de variáveis, mantendo no modelo apenas aquelas que atingem significância estatística. Essa prática já não é mais aconselhada, uma vez que o modelo estatístico deve ser definido antes da análise, com base no modelo científico da área, e mesmo as variáveis sem resultados significativos devem ser mantidas no modelo a ser relatado, pois essa informação também é relevante para a área de conhecimento (McElreath, 2020).

Outra crítica às análises conduzidas pelo VARBRUL, e que está diretamente relacionada aos objetivos da presente pesquisa, está relacionada à condução de um modelo estatístico de efeitos fixos. Modelos de efeitos fixos trabalham com o pressuposto de independência das observações, ou seja, pressupõem que não há correlação entre os dados provenientes de diferentes indivíduos ou unidades de análise. Esse, porém, raramente é o caso de dados linguísticos (Lima Jr; Garcia, 2021). Por essa razão, fatores relacionados aos indivíduos que compõem a amostra e às suas escolhas lexicais, por exemplo, são denominados efeitos aleatórios.

Modelos estatísticos que consideram a existência de efeitos aleatórios, comumente referidos pela expressão “modelos de efeitos mistos”, são considerados mais robustos em comparação aos modelos de efeitos fixos, pois levam em conta a não independência na coleta de dados, o que, por sua vez, reduz a chance de erro do tipo I (Lima Jr; Garcia, 2021). Além disso, por calcularem tanto estimativas globais para o grupo como estimativas individuais por efeito aleatório, neste caso para cada indivíduo e para cada item lexical, os modelos de efeitos mistos permitem observar, reportar e prever comportamentos linguísticos intra e interindividual, aspecto tão caro para estudos sociolinguísticos.

As vantagens de modelos estatísticos de efeitos mistos sobre o modelo de efeito fixo conduzido pelo programa VARBRUL motivam o questionamento da replicação proposta: os resultados sobre a variação das vogais médias pretônicas no falar popular de Fortaleza, Ceará, obtidos por Araújo (2007), serão mantidos quando reanalisados com modelos de regressão logística de efeitos mistos? Nossa hipótese é que a inclusão de efeitos aleatórios para informante e para item lexical não alterará substancialmente as conclusões principais do estudo original, mas fornecerá estimativas mais robustas e generalizáveis, como aquelas advindas dos participantes e dos itens lexicais como efeitos aleatórios.

Consequentemente, a contribuição principal deste estudo é metodológica, mas com implicações interpretativas relevantes para a Sociolinguística Variacionista. Ao reanalisar os dados de Araújo (2007) por meio de modelos de regressão logística de efeitos mistos, pretendemos avaliar a estabilidade dos condicionamentos linguísticos e sociais identificados no estudo original, verificando em que medida esses efeitos permanecem quando se controla explicitamente a variabilidade associada aos informantes e aos itens lexicais por meio da modelagem de efeitos aleatórios. A reanálise permitirá examinar o papel do falante e do léxico na modelagem dos processos de alçamento e abaixamento das vogais médias pretônicas, fatores que não são tratados da mesma forma pelos modelos VARBRUL. Além disso, a comparação entre os resultados obtidos pelos dois métodos possibilitará discutir diferenças de interpretação decorrentes das escolhas analíticas, bem como os ganhos e limitações envolvidos na reanálise de dados previamente coletados. Dessa forma, o estudo busca contribuir para o debate sobre a robustez de generalizações sociolinguísticas e sobre as implicações metodológicas da adoção de técnicas estatísticas contemporâneas na investigação da variação linguística.

Métodos

Trata-se da reanálise de um corpus oral de fala espontânea, constituído por entrevistas sociolinguísticas, utilizados na tese *As vogais médias pretônicas no falar popular de Fortaleza* (Araújo, 2007) e provenientes do Projeto Norma Oral do Português Popular de Fortaleza (NORPORFOR). O corpus é composto por gravações de 72 informantes, conforme a amostra original, totalizando 5.848 dados analisados (3.337 de /e/ e 2.511 de /o/), estratificados por sexo, faixa etária e escolaridade, conforme descrito na tese original. Não serão incluídos novos participantes nem haverá exclusão adicional de informantes. As entrevistas sociolinguísticas foram realizadas em Fortaleza (CE), representando o falar popular da região.

A escolha da tese de Araújo (2007) como objeto desta reanálise se justifica por três razões principais: (i) o estudo investiga não apenas o alçamento, mas também o abaixamento das vogais médias pretônicas; (ii) trata-se de um dialeto de região historicamente com menor cobertura nas investigações linguísticas de larga escala no Brasil; e (iii) a autora concordou em ceder os dados originais para fins desta reanálise.

Serão seguidas as seguinte etapas metodológicas:

1. Organização e formatação dos dados originais em planilha compatível com o R;
2. Codificação das variáveis preditoras e de resposta conforme a tese original;
3. Ajuste de modelos de regressão logística de efeitos mistos, incluindo informante e item lexical como efeitos aleatórios;
4. Comparação dos resultados com os obtidos pelo VARBRUL na tese original;
5. Divulgação do código e dos resultados em repositório aberto.

Os dados serão analisados no ambiente R, utilizando modelos de regressão logística de efeitos mistos. As variáveis de resposta serão a realização das vogais pretônicas, e as variáveis preditoras serão as mesmas testadas na tese original (9 fatores linguísticos: tipo de vogal acentuada, tipo de vogal átona seguinte, nasalidade, contexto fonológico precedente, contexto fonológico seguinte, tipo de sílaba, distância em relação à tônica, tipo de atonicidade da pretônica, estrutura morfológica da palavra; e 3 fatores sociais: sexo, faixa etária e escolaridade). Assim como no trabalho original, será primeiramente analisado o alçamento das vogais e depois o abaixamento, em modelos separados. Serão incluídos efeitos aleatórios para informante e item lexical, visando incorporar a variação intrafalantes e intrapalavras nas estimativas do modelo. É possível que o modelo não convirja ou apresente aviso de singularidade (*singular fit*) com a inclusão de item lexical como efeito aleatório por aumentar muito a quantidade de parâmetros do modelo. Caso isso ocorra, serão adotados dois caminhos, seguindo esta hierarquia: i) será ajustado um modelo apenas com participante como efeito aleatório, por ser o efeito aleatório mais importante para se observar a variação intra e interindividual e por demandar menos o modelo; e ii) será ajustado um modelo bayesiano com participante e item lexical como efeitos aleatórios, visto que essa abordagem pode mitigar limitações de convergência frequentemente observadas em modelos frequentistas mais complexos (Garcia; Lima Jr., 2021). Nesse caso, serão especificados priors adequados e realizados procedimentos de diagnóstico da convergência e de avaliação do ajuste do modelo. Além disso, os dois modelos serão reportados e comparados aos resultados originais, discutindo as limitações e vantagens de cada.

O modelo bayesiano, se utilizado como alternativa confirmatória, terá os seguintes elementos pré-especificados: priors levemente informativos, Normal(0; 2,5) na escala logística para os coeficientes fixos, e Normal(0; 1) truncada em valores positivos (half-Normal) para os desvios-padrão dos efeitos aleatórios; diagnóstico de convergência por meio do estatístico R-hat ($< 1,01$) e do tamanho efetivo da amostra (effective sample size > 400 por parâmetro); e verificação do ajuste do modelo por meio de checagem preditiva a posteriori (posterior predictive check).

Sendo assim, os modelos a serem ajustados serão:

- Para vogal /e/:
 - p(alçamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1|participante) + (1|item lexical)
 - p(abaixamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1|participante) + (1|item lexical)
- Para vogal /o/:
 - p(alçamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1|participante) + (1|item lexical)
 - p(abaixamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1|participante) + (1|item lexical)

A variável de resposta de cada modelo será codificada como binária, assumindo o valor 1 para a variante alçada (nos modelos de alçamento) ou abaixada (nos modelos de abaixamento) e o valor 0 para a variante mantida (manutenção). O modelo de alçamento de cada vogal será ajustado apenas com as ocorrências de alçamento e de manutenção, excluindo-se as ocorrências de abaixamento; o modelo de abaixamento será ajustado apenas com as ocorrências de abaixamento e de manutenção, excluindo-se as ocorrências de alçamento. A fim de comparação exploratória, um modelo multinomial com a manutenção como referência poderá ser ajustado, mantendo-se todas as demais características dos modelos anteriores.

Os níveis de cada variável, bem como o nível de referência de cada uma, estão listados a seguir:

- vogal acentuada: [i], [e], [ɛ], [a], [ɔ], [o], [u], [ĩ], [ẽ], [ã], [õ], [ũ]

- Referência: [a], por ser a mais baixa e central
- vogal átona seguinte: [i], [e], [ɛ], [a], [ɔ], [o], [u], [ĩ], [ẽ], [ã], [õ], [ũ]
 - Referência: [a], por ser a mais baixa e central
- Nasalidade: nasal, oral
 - Referência: oral, por ser o menos marcado
- contexto fonológico precedente: alveolares, labiais, palatais, velares, glotais
 - Referência: alveolares, por serem as mais frequentes
- contexto fonológico seguinte: alveolares, labiais, palatais, velares, glotais
 - Referência: alveolares, por serem as mais frequentes
- tipo de sílaba: livre, fechada
 - Referência: livre, por ser a menos marcada
- distância da tônica: 1, 2, 3, 4
 - Referência: 1, a menos distante
- tipo de atonicidade da pretônica: permanente, casual
 - Referência: permanente, a menos marcada
- estrutura morfológica: sem sufixo, com sufixo
 - Referência: sem sufixo, a menos marcada
- sexo: F, M
 - Referência: F, ordem alfabética
- faixa etária: 1 (22-35 anos), 2 (36-49 anos), 3 (50+ anos)
 - Referência: 1, mais jovens
- escolaridade: A (0-4 anos de estudo), B (5-8 anos), C (9-11 anos)
 - Referência: A, menor escolaridade

Os modelos serão ajustados por meio da função `glmer()` do pacote `lme4`, com família binomial e função de ligação logit (e, no caso do modelo bayesiano, por meio da função `brm()` do pacote `brms`, também com família binomial e função de ligação logit). O critério inferencial adotado será o nível de significância de 5% ($p < 0,05$) para os modelos frequentistas, e a ausência de zero no intervalo de credibilidade de 95% para o modelo bayesiano. Não serão incluídas interações entre variáveis preditoras nos modelos confirmatórios; qualquer análise de interação eventualmente conduzida será tratada como exploratória.

Apesar da vantagem de modelos de regressão ajustados com efeitos mistos permitirem a inclusão de variáveis preditoras contínuas, as variáveis preditoras nesta reanálise continuarão sendo tratadas como categóricas por não haver as informações dos valores contínuos nos dados e para não adicionar mudanças demasiadas à análise original, possibilitando uma comparação mais direta. Os resultados serão comparados aos da análise original (VARBRUL), verificando se os efeitos principais (altura da vogal tônica, contexto fonológico, fatores sociais) se mantêm significativos e com direção semelhante. Será considerada reprodução bem-sucedida a obtenção dos mesmos padrões qualitativos e tendências gerais reportadas na tese, ainda que pequenas diferenças nos valores de probabilidade sejam esperadas pela inclusão de efeitos aleatórios. Optaremos por métodos mais recentes (modelos mistos) por serem estatisticamente mais robustos e amplamente aceitos atualmente, permitindo inferências mais generalizáveis que o VARBRUL, que não incorpora efeitos aleatórios.

Para tornar operacional o critério de reprodução, o quadro a seguir especifica, para cada um dos quatro processos analisados, quais resultados de Araújo (2007) serão efetivamente comparados na reanálise, bem como o critério utilizado para classificar cada resultado como reproduzido, parcialmente reproduzido ou não reproduzido.

Processo	Variáveis/grupos de fatores selecionados por Araújo (2007)	Direção/ordenação dos efeitos considerada relevante	Critério de classificação
Alçamento de /e/	Ordem de seleção no VARBRUL (variáveis estatisticamente significativas, da mais para a menos relevante): natureza da vogal tônica; natureza da vogal átona contígua; consoante subsequente; atonicidade; consoante precedente; idade; sufixação; escolaridade (não selecionadas: distância em relação à tônica, estrutura silábica, sexo).	Alçamento favorecido por: vogal tônica/átona alta, oral ou nasal (harmonização vocálica); consoante precedente palatal ou velar; consoante subsequente velar, labial ou palatal; atonicidade casual variável ou permanente; sufixo verbal; informantes com 50 anos ou mais; menor escolaridade (nenhuma a 4 anos).	Reproduzido: mesma direção do efeito e significância estatística no modelo de efeitos mistos ($p < 0,05$ ou intervalo de confiança de 95% não contendo zero) para as variáveis selecionadas por Araújo (2007), mantendo a mesma ordenação relativa de magnitude dos efeitos (tolerância de uma posição). Parcialmente reproduzido: mesma direção do efeito, porém sem atingir o critério inferencial adotado, ou alteração na ordenação relativa entre as variáveis. Não reproduzido: direção do efeito oposta à original, ou variável originalmente selecionada pelo VARBRUL sem evidência de associação no modelo de efeitos mistos.
Abaixamento de /e/	Ordem de seleção no VARBRUL: natureza da vogal tônica; natureza da vogal átona contígua; atonicidade; consoante subsequente;	Abaixamento favorecido por: vogal tônica/átona baixa ou nasal não-alta (harmonização vocálica); atonicidade casual	Reproduzido: mesma direção do efeito e significância estatística no modelo de efeitos mistos ($p < 0,05$ ou intervalo de confiança de 95% não contendo zero) para as

	sufixação; consoante precedente; idade; tipo de sílaba (não selecionadas: distância em relação à tônica, sexo, escolaridade).	baixa; consoante subsequente velar ou aspirada; consoante precedente velar ou aspirada; sufixo nominal ou ausência de sufixo; informantes com 50 anos ou mais; sílaba travada (por /R/ ou por /S/).	variáveis selecionadas por Araújo (2007), mantendo a mesma ordenação relativa de magnitude dos efeitos (tolerância de uma posição). Parcialmente reproduzido: mesma direção do efeito, porém sem atingir o critério inferencial adotado, ou alteração na ordenação relativa entre as variáveis. Não reproduzido: direção do efeito oposta à original, ou variável originalmente selecionada pelo VARBRUL sem evidência de associação no modelo de efeitos mistos.
Alçamen to de /o/	Ordem de seleção no VARBRUL: natureza da vogal tônica; atonicidade; consoante subsequente; consoante precedente; natureza da vogal átona contígua; idade; sufixação; escolaridade (não selecionadas: distância em relação à tônica, estrutura silábica, sexo).	Alçamento favorecido por: vogal tônica/átona alta, oral ou nasal (harmonização vocálica); atonicidade casual variável ou permanente; consoante subsequente palatal ou labial; consoante precedente labial ou velar; informantes com 50 anos ou mais; sufixo verbal; menor escolaridade (nenhuma a 4 anos).	Reproduzido: mesma direção do efeito e significância estatística no modelo de efeitos mistos ($p < 0,05$ ou intervalo de confiança de 95% não contendo zero) para as variáveis selecionadas por Araújo (2007), mantendo a mesma ordenação relativa de magnitude dos efeitos (tolerância de uma posição). Parcialmente reproduzido: mesma direção do efeito, porém sem atingir o critério inferencial adotado, ou alteração na ordenação relativa entre as variáveis. Não reproduzido: direção do efeito oposta à original, ou variável originalmente selecionada pelo VARBRUL sem evidência

			de associação no modelo de efeitos mistos.
Abaixamento de /o/	Ordem de seleção no VARBRUL: natureza da vogal tônica; natureza da vogal átona contígua; atonicidade; consoante precedente; sufixação; estrutura da sílaba; consoante subsequente (não selecionadas: distância em relação à tônica, sexo, faixa etária, escolaridade).	Abaixamento favorecido por: vogal tônica/átona baixa ou nasal não-alta (harmonização vocálica); atonicidade casual baixa ou mista; consoante precedente alveolar, palatal ou aspirada; ausência de sufixo; sílaba travada por /R/; consoante subsequente velar ou alveolar. Faixa etária e escolaridade não foram selecionadas como relevantes para esta regra.	Reproduzido: mesma direção do efeito e significância estatística no modelo de efeitos mistos ($p < 0,05$ ou intervalo de confiança de 95% não contendo zero) para as variáveis selecionadas por Araújo (2007), mantendo a mesma ordenação relativa de magnitude dos efeitos (tolerância de uma posição). Parcialmente reproduzido: mesma direção do efeito, porém sem atingir o critério inferencial adotado, ou alteração na ordenação relativa entre as variáveis. Não reproduzido: direção do efeito oposta à original, ou variável originalmente selecionada pelo VARBRUL sem evidência de associação no modelo de efeitos mistos.

Quando a direção do efeito for semelhante à original, mas o resultado não atingir o critério inferencial adotado no modelo de efeitos mistos, o resultado será classificado como parcialmente reproduzido, e essa situação será discutida individualmente na seção de resultados, reportando a magnitude e o intervalo de confiança da estimativa. Inversamente, quando uma variável não selecionada pelo VARBRUL na tese original apresentar evidência de associação ($p < 0,05$ ou intervalo de confiança de 95% não contendo zero) no modelo de efeitos mistos, esse resultado será relatado como um achado adicional decorrente da mudança de método, sem ser utilizado para contestar a reprodução dos efeitos originalmente selecionados. Não será estabelecida equivalência numérica entre os pesos relativos do VARBRUL e os coeficientes dos modelos de efeitos mistos; a comparação restringe-se à direção, à significância e à ordenação relativa dos efeitos.

Quadro-resumo do desenho do estudo

Dados a serem reanalisados	Dados da tese de Araújo (2007): corpus oral NORPORFOR, 72 informantes, 5.848 ocorrências de vogais médias pretônicas (3.337 de /e/ e 2.511 de /o/).
Variáveis de resposta	Realização das vogais pretônicas /e/ e /o/: alçamento vs. manutenção; abaixamento vs. manutenção.
Variáveis preditoras	9 variáveis linguísticas (tipo de vogal acentuada, vogal átona seguinte, nasalidade, contexto fonológico precedente e seguinte, tipo de sílaba, distância da tônica, tipo de atonicidade, estrutura morfológica) e 3 sociais (sexo, faixa etária, escolaridade).
Efeitos aleatórios	Participante (informante) e item lexical.
Modelos previstos	Modelos de regressão logística de efeitos mistos no R (lme4). Alternativa bayesiana (brms) caso haja problemas de convergência ou singularidade. Para vogal /e/: p(alçamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1 participante) + (1 item lexical) p(abaixamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1 participante) + (1 item lexical) Para vogal /o/: p(alçamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1 participante) + (1 item lexical) p(abaixamento) ~ vogal acentuada + vogal átona seguinte + nasalidade + contexto fonológico precedente + contexto fonológico seguinte + tipo de sílaba + distância da tônica + tipo de atonicidade da pretônica + estrutura morfológica + sexo + faixa etária + escolaridade + (1 participante) + (1 item lexical)

CrITÉrios de comparaÇ�o com o estudo original	Reprodu��o bem-sucedida: manuten��o dos mesmos padr�es qualitativos e tend�ncias gerais reportadas em Ara�jo (2007), com pequenas diferen�as num�ricas esperadas.
An�lises confirmat�rias	Ajuste dos modelos pr�-especificados com todas as vari�veis preditoras do estudo original e efeitos aleat�rios para participante e item lexical.
An�lises explorat�rias	Qualquer an�lise n�o especificada neste protocolo e inclu�da posteriormente ser� explicitamente identificada como explorat�ria no relato de pesquisa.
Reposit�rios previstos	Open Science Framework: https://doi.org/10.17605/OSF.IO/HYXNK (dados anonimizados, scripts e documenta��o da an�lise).

Cronograma

A conclus o do estudo est  prevista para 6 meses ap s a publica  o do relato registrado. As atividades ser o realizadas de acordo com o cronograma a seguir.

Tempo	Etapas
1� m�s ap�s publica��o do relato registrado	Adequa��o dos dados em planilha para ser utilizada em ambiente R.
2� m�s ap�s publica��o do relato registrado	An�lise descritiva dos dados. Confe��o de tabelas de propor��o e de gr�ficos e confer�ncia com os achados no estudo original.
3� ao 5� m�s ap�s publica��o do relato registrado	Ajuste e avalia��o dos modelos de regress�o log�stica com efeitos mistos; confer�ncia com os resultados do estudo original.
6� m�s ap�s publica��o do relato registrado	Reda��o e submiss�o do relato de pesquisa.

Link para Preprint

<https://doi.org/10.1590/SciELOPreprints.15222>

Declaração de disponibilidade de dados

Todos os arquivos e materiais necessários para replicar esta reanálise (planilha de dados, scripts de análise e documentação da análise) serão disponibilizados publicamente no Open Science Framework, sob DOI [10.17605/OSF.IO/HYXNK](https://doi.org/10.17605/OSF.IO/HYXNK). Os participantes serão anonimizados na planilha de dados.

Declaração de uso de inteligência artificial

Os autores declaram que nenhuma ferramenta de inteligência artificial foi utilizada na criação deste manuscrito.

Declaração ética

Os dados utilizados pertencem ao corpus Norma do Português Oral Popular de Fortaleza (NORPORFOR), cuja utilização não carece de nova aprovação por comitê de ética, conforme parecer 4.025.173 do CAAE 31405620.3.0000.5054, uma vez que não há forma de identificação dos falantes, cuja contribuição com o corpus cumpriu com os requisitos éticos. Além disso, esta reanálise não utilizará os áudios dos participantes, apenas os dados já extraídos e tabulados de maneira anonimizada.

Conflito de interesses

Os autores declaram ausência de conflitos de interesse.

Contribuição de autoria

Funções do CRediT: Ronaldo Lima Jr.: Administração do projeto, Conceitualização, Curadoria de dados, Análise formal, Metodologia, Software, Redação – rascunho original; Pablo Arantes: Análise formal, Investigação, Metodologia, Software, Validação, Visualização, Redação – rascunho original, Redação – revisão e edição; Guilherme Garcia: Conceitualização, Curadoria de dados, Análise formal, Investigação, Metodologia, Software, Visualização, Redação – rascunho original, Redação – revisão e edição; Luciana Lucente: Análise formal, Investigação, Validação, Visualização, Redação – rascunho original, Redação – revisão e edição; Renata Passetti: Análise formal, Investigação, Metodologia, Supervisão, Validação, Visualização, Redação rascunho original, Redação – revisão e edição.

Referências:

ARAÚJO, A. A. de. **As vogais médias pretônicas no falar popular de Fortaleza: uma abordagem variacionista**. [s. l.], 2007.

CEDERGREN, H. J.; SANKOFF, D. Variable rules: Performance as a statistical reflection of competence. **Language**, [s. l.], p. 333–355, 1974.

GARCIA, G. D.; LIMA JR., R. M. Introdução à estatística bayesiana aplicada à linguística. **Revista da ABRALIN**, [S. l.], v. 20, n. 2, p. 1–24, 2021. DOI: 10.25189/rabralin.v20i2.1914. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/1914>. Acesso em: 26 maio. 2026.

LIMA JR., R. M.; GARCIA, G. D. Diferentes análises estatísticas podem levar a conclusões categoricamente distintas. **Revista da ABRALIN**, [s. l.], p. 1, 2021.

MCELREATH, R. **Statistical rethinking**: A Bayesian course with examples in R and Stan. 2nd edition. Boca Raton: Chapman and Hall/CRC, 2020.

OUSHIRO, L. **Introdução à Estatística para Linguistas**. 1. ed. [S. l.]: Editora da Abralin, 2022. Disponível em: <https://editora.abralin.org/>. Acesso em: 27 out. 2025.

TAGLIAMONTE, S. A. **Variationist sociolinguistics: Change, observation, interpretation**. [S. l.]: John Wiley & Sons, 2011.

TORRES VIEIRA, N. M. **Monotongação de ditongos orais no português brasileiro: uma revisão sistemática da literatura**. 1. ed. [S. l.]: Editora da Abralin, 2022. Disponível em: <https://editora.abralin.org/>. Acesso em: 27 out. 2025.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.