

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA-21P02

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“O processamento de pseudopalavras no Português Brasileiro”

**Giovanna Vendeiro Vilar
Mariana Almeida Costa
Renata Massami Hirota
Viviana Giampaoli**

São Paulo, julho de 2021

CENTRO DE ESTATÍSTICA APLICADA - CEA - USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “O processamento de pseudopalavras no Português Brasileiro”

PESQUISADORA: Aline Benevides

ORIENTADORA: Profa. Dra. Raquel Santana Santos

INSTITUIÇÃO: Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

FINALIDADE DO PROJETO: Doutorado

RESPONSÁVEIS PELA ANÁLISE: Giovanna Vendeiro Vilar

Mariana Almeida Costa

Renata Massami Hirota

Viviana Giampaoli

REFERÊNCIA DESTE TRABALHO: ALMEIDA, M.C.; GIAMPAOLI, V.; HIROTA, R.M.; VILAR, G.V. **Relatório de análise estatística sobre o projeto: “O**

processamento de pseudopalavras no Português Brasileiro". São Paulo, IME-USP, 2021. (RAE-CEA-21P02)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS

AGRESTI, A. (2013). **Categorical Data Analysis**. 3.ed. New Jersey: Wiley.

COLLISCHONN, G. (1999). Acento Em Português. In Bisol, L. (Ed.) **Introdução a Estudos de Fonologia Do Português Brasileiro**. 1ed. Porto Alegre: EDIPUCRS.

FIORIN, J. L. (2019). **Linguística? Que é Isso?** 1.ed ed. São Paulo: Contexto.

GWET, K. (2002). Kappa Statistic is not Satisfactory for Assessing the Extent of Agreement Between Raters. **Statistical Methods for Inter-rater Reliability Assessment**, 1–5.

GWET, K. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. **Brit J Math Stat Psy**. **61**, 29–48

PETTER, M. (2007). **Introdução à Linguística i: Objetos Teóricos**. 5.ed ed. São Paulo: Contexto.

PRIBERAN. (2021). **"pseudopalavra"**, in *Dicionário Priberam Da Língua Portuguesa*. <<https://dicionario.priberam.org/pseudopalavra>>. [Acesso em: 01-05-2021].

TAPIA, A; LEIVA, V; DIAZ, M.d.P; GIAMPAOLI, V (2018). Influence diagnostics in mixed effects logistic regression models. **TEST** **28**, 920–942

PROGRAMAS COMPUTACIONAIS UTILIZADOS

Software R (versão 4.0.5)

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise de Dados Categorizados (06:030)

Outros (07:990)

ÁREA DE APLICAÇÃO

Linguística (14:110)

Resumo

A literatura na área de linguística afirma, em geral, que o acento tônico na penúltima sílaba (paroxítona) é o padrão, enquanto acentos na última (oxítona) e na antepenúltima (proparoxítona) sílabas são desvios. Além disso, a estrutura da palavra é considerada o fator determinante na acentuação tônica.

O estudo analisado neste relatório estatístico busca entender como se dá o processo da acentuação tônica (oxítona, paroxítona e proparoxítona) em falantes nativos do português brasileiro, através de um experimento utilizando pseudopalavras e, assim, definir quais as variáveis linguísticas que exercem influência sobre a acentuação. A partir desse objetivo, foram construídos modelos de regressão multinomial com efeitos aleatórios.

Concluiu-se que, de modo geral, existem variáveis linguísticas além da estrutura da palavra que influenciam no processo de acentuação tônica no português brasileiro. Características relacionadas aos falantes não apresentaram significância estatística nos modelos ajustados.

Sumário

1. Introdução	8
2. Objetivo	9
3. Descrição do estudo	10
4. Descrição das variáveis	12
5. Análise descritiva	15
6. Análise inferencial	23
7. Conclusões	31
APÊNDICE A	32
APÊNDICE B	38

1. Introdução

De acordo com Petter (2007), o interesse pela linguagem é antigo e vem sendo expressado por meio de mitos, lendas, cantos, rituais e trabalhos eruditos que buscam conhecer essa capacidade humana como sistema de comunicação. A autora aponta que, a partir do século XX, os estudos linguísticos passaram a ter um caráter científico, ou seja, centrados na observação dos fatos a partir de pressupostos teóricos da linguagem, no estabelecimento de hipóteses e na examinação mediante experimentos.

Conforme descrito por Fiorin (2019), a linguística é uma ciência da linguagem porque, ao contrário da gramática, ela tem como objetivo estabelecer o que uma língua é e por que é de uma determinada maneira. Logo, a área estuda os aspectos fonéticos, morfológicos, sintáticos, semânticos, sociais e psicológicos de uma língua, neste caso, o português brasileiro. Dentro deste contexto, existe o conceito de *pseudopalavra*, que, de acordo com o dicionário Priberam (2021), é uma “*Sequência regular e pronunciável de caracteres que não tem um significado numa língua, apesar de obedecer às regras ortográficas, morfológicas ou de pronúncia*”.

No português brasileiro existem três classes de palavras segundo sua tonicidade: oxítona, paroxítona e proparoxítona. Essas denominações estão relacionadas à intensidade dada a determinadas sílabas na pronúncia das palavras. Aquela que é pronunciada de forma mais acentuada é a sílaba tônica. Assim, oxítona é a palavra cuja sílaba tônica é a última; paroxítona é a palavra cuja sílaba tônica é a penúltima; e proparoxítona é a palavra cuja sílaba tônica é a antepenúltima.

O intuito do trabalho é investigar a maneira como os falantes nativos do português atribuem a tonicidade em pseudopalavras parcialmente relacionadas a vocábulos existentes no idioma. Em outras palavras, busca-se compreender como o indivíduo, ao se deparar com uma palavra nova, nesse caso, uma pseudopalavra, define a sílaba tônica. Além disso, busca-se entender quais são os outros fatores, tais como os conhecimentos linguísticos do falante e as

estruturas linguísticas das palavras, que podem influenciar nesse processo de classificação e, portanto, na determinação da entonação.

2. Objetivos

O objetivo principal do trabalho é entender o processo de acentuação no Português Brasileiro, com base em um experimento utilizando pseudopalavras. Ou seja, buscamos compreender quais as características da palavra e quais as características do falante que influenciam nesse processo de acentuação tônica quando o indivíduo se depara com uma palavra nova.

Algumas perguntas a serem respondidas pela análise estatística são:

1. A classificação tônica das pseudopalavras pode recuperar o acento das palavras-alvo de referência? Em outras palavras, queremos entender se a tonicidade da palavra-alvo tem papel na predição do acento da pseudopalavra.
2. As características estruturais da pseudopalavra têm influência na acentuação tônica? Essas são as únicas características que realmente têm impacto nesse processo?
3. A similaridade entre palavra-alvo e pseudopalavra influencia na associação acentual? Buscamos entender se quanto mais similar a pseudopalavra for da palavra-alvo, maiores são as chances de atribuição do mesmo padrão acentual da palavra-alvo.
4. Há associação entre o participante estar na graduação em Letras e a classificação tônica da pseudopalavra? Há associação entre conhecimento em música e a classificação tônica da pseudopalavra? Espera-se que indivíduos com conhecimento em música ou que entraram recentemente no curso de Letras tenham um comportamento de classificação das pseudopalavras distinto dos demais.

3. Descrição do estudo

O estudo foi realizado de maneira remota com reuniões individuais entre a pesquisadora e cada um dos participantes via Google Meet. Consistiu em apresentar aos participantes, através do software Psychopy, 372 pseudopalavras agrupadas nos denominados grupos de classificação e registrar a forma como eles reproduziam verbalmente tais palavras criadas. A seguir, as respostas dos participantes foram classificadas de acordo com as três classes de acentuação tônica: oxítona, paroxítona e proparoxítona.

A coleta dos dados foi realizada no início de 2020 com 34 indivíduos que, por meio de divulgações em redes sociais e de colegas, se voluntariaram a participar do experimento. Os voluntários tiveram como pré-requisitos: ser maior de 18 anos, ser falante nativo do português brasileiro e não ter estudado linguística.

Entre os participantes da pesquisa encontram-se estudantes do primeiro semestre da faculdade de Letras da Universidade de São Paulo, músicos, alguns residentes de fora do estado de São Paulo, entre outros. Supõe-se que os alunos do primeiro semestre do curso de Letras ainda não têm conhecimento na área.

3.1 Limitações do estudo

Identificamos dois eventuais problemas -um de caráter técnico e outro de aleatorização- na coleta de dados que tentaremos contornar nas análises. O primeiro é descrito a seguir, enquanto o outro será mencionado na Seção 4.4.

Destaca-se a perda de algumas respostas durante o processo de coleta de dados, visto que ruídos externos impediram que algumas entonações fossem captadas e registradas na gravação. Logo, na base de dados não temos 372 registros de pseudopalavras para todos os participantes. Portanto, um total de 12.511 dados serão utilizados na análise, em vez dos 12.648 esperados, o qual não representa uma perda substancial.

3.2 Conceitos de similaridade e validação

O conceito de similaridade entre palavra-alvo e pseudopalavra foi construído com base nas mudanças feitas na palavra-alvo até a obtenção da pseudopalavra. Essas alterações estão relacionadas a mudanças de ponto, modo e/ou vozeamento.

De acordo com o tipo e a quantidade de alterações foi estabelecido um valor de 1 a 10 -chamado de taxa de similaridade- onde, para mudanças de consoantes, valores menores do que 4 determinam pseudopalavras similares à sua palavra-alvo e valores maiores ou iguais a 5 determinam pseudopalavras dissimilares à sua palavra-alvo. Já para mudanças de vogais, valores acima de 1 determinam pseudopalavras dissimilares à sua palavra-alvo.

Diante disso, foi necessário definir um modo de validar essa classificação em similar e dissimilar, ou seja, verificar se a pseudopalavra classificada como similar -ou dissimilar- é, de fato, similar -ou dissimilar- à palavra da qual ela se originou (palavra-alvo). Nessa etapa -chamada de validação- pediu-se para 10 falantes do português, que não faziam parte do estudo final, listarem a palavra do português à qual eles associavam cada uma das pseudopalavras.

Dentre as pseudopalavras consideradas similares a palavras-alvo, considerou-se validadas como “similar à palavra-alvo” as pseudopalavras cuja associação foi a palavra-alvo na resposta de, no mínimo, oito indivíduos. Porém, pseudopalavras nas quais sete falantes apresentaram a associação correta foram classificadas como quase validadas e quando menos de sete falantes apresentaram a associação “correta” entendeu-se que a classificação em similar não foi validada (ela não está, de fato, parecida com sua palavra-alvo).

Da mesma forma, dentre as pseudopalavras consideradas dissimilares à palavra-alvo, foram consideradas validadas como “dissimilar à palavra-alvo” pseudopalavras não associadas a uma mesma resposta por até dois indivíduos. Em outras palavras, se até dois falantes associaram uma mesma palavra à pseudopalavra, ela foi considerada dissimilar à sua palavra-alvo. Pseudopalavras nas quais três falantes apresentaram a mesma associação

foram classificadas como quase validadas e, quando mais de três falantes lembraram de uma mesma palavra do português ao ler a pseudopalavra, entendeu-se que a classificação em dissimilar não foi validada (há indícios de que ela está parecida com uma palavra real).

Ressalta-se que, nesse caso, a associação não precisava ser exclusivamente com a palavra-alvo. Por exemplo, suponha que pesquisadora criou a pseudopalavra *Fanama* através de modificações na palavra *Paraná* (palavra-alvo) e definiu que essa pseudopalavra era dissimilar a sua palavra-alvo. Se, na etapa de validação, 5 dos 10 falantes lembraram da palavra *Canadá* ao ler *Fanama*, então considerou-se que a dissimilaridade não foi validada, pois muitos falantes associaram a pseudopalavra a uma mesma palavra do português (ela parece ser similar a uma palavra real).

4. Descrição das variáveis

Foram coletadas variáveis linguísticas -relacionadas às pseudopalavras-, extralinguísticas -relacionadas aos participantes- e experimentais -relacionadas ao estudo-, que podem, segundo a literatura da área, influenciar o comportamento acentual no português.

4.1 Variável dependente

A variável resposta de interesse é **Tonicidade de produção**, ou seja, a classificação tônica da pseudopalavra (oxítone, paroxítone ou proparoxítone).

4.2 Variáveis linguísticas

As pseudopalavras foram construídas com três sílabas de extensão para que os três padrões acentuais do português brasileiro pudessem ser produzidos. As palavras que deram origem às pseudopalavras, definidas como palavras-alvo,

foram classificadas em dois níveis de acordo com a sua ocorrência no Corpus brasileiro, corpus linguístico coordenado pelo pesquisador Antonio Paulo Berber Sardinha. Se a palavra possui mais de 100 mil ocorrências no corpus, ela é classificada como de alta frequência e se possui menos de 2 mil ocorrências, ela é classificada como de baixa frequência. A junção da ideia de frequência e similaridade entre a palavra alvo e a pseudopalavra resultou na criação de uma variável com 4 categorias chamada Grupo de classificação.

- **Validação:** s = sim, n = não validada e q = quase validada
- **Palavra-alvo:** palavra real que deu origem à pseudopalavra
- **Tonicidade da palavra-alvo:** oxítone, paroxítone e proparoxítone
- **Estrutura da palavra:** indica qual é a estrutura da pseudopalavra (CV-CV ou CV-CV-CVC), sendo que C indica Consoante e V indica Vogal
- **Pseudopalavra:** refere-se a cada uma das palavras criadas
- **Segmento modificado:** indica qual letra foi modificada na criação da pseudopalavra a partir da palavra real (consoante ou vogal)
- **Grupo de classificação:** indica o efeito da similaridade (entre a pseudopalavra e a palavra real) e da frequência (alta e baixa) na produção acentual
 - 1 = pseudopalavras similares de alta frequência
 - 2 = pseudopalavras dissimilares de alta frequência
 - 3 = pseudopalavras similares de baixa frequência
 - 4 = pseudopalavras dissimilares de baixa frequência

Ressaltamos que essa variável não foi controlada durante a coleta de dados, ou seja, não foi pré-definida uma quantidade de palavras de cada categoria em cada conjunto apresentado aos participantes.

4.3 Variáveis extralinguísticas

- **Falante:** identifica os 34 participantes do experimento
- **Idade:** de 18 a 60 (anos)
- **Gênero:** feminino e masculino
- **Naturalidade:** São Paulo SP, Outros municípios de SP e Outras UF
- **Escolaridade:** categorizada em Fundamental Completo, Superior Completo, Superior Incompleto e Pós-Graduação (Completo ou Incompleta)
- **Área de formação:** 0 = outros e 1 = Letras
- **Línguas:** 0 = não tem conhecimento em línguas e 1 = tem conhecimento em línguas
- **Música:** 0 = não tem conhecimento em música e 1 = tem conhecimento em música

4.4 Variáveis experimentais

As palavras foram aleatorizadas no Excel e divididas em 4 conjuntos -variável Bloco de apresentação- a serem apresentados aos participantes com um intervalo de tempo entre cada conjunto. Por limitações do software, a ordem de apresentação desses conjuntos não pôde ser aleatorizada, apenas a ordem das palavras dentro de cada conjunto.

- **Bloco de apresentação:** indica em qual bloco (ou conjunto) a pseudopalavra foi inserida (1, 2, 3 ou 4)

Problemas de aleatorização

Para alguns participantes o software Psychopy apresentou problemas e eles tiveram que continuar o experimento a partir de slides com uma ordem aleatória pré-estabelecida. Em outras palavras, todos os indivíduos que em algum

momento acompanharam o experimento pelos slides seguiram com palavras apresentadas na mesma ordem (a primeira aleatorização retirada do Excel).

- **Aleatorização:** codifica se o bloco de apresentação foi aleatorizado para o indivíduo ou não e, portanto, foi considerada a aleatorização prévia (s = o estímulo foi aleatorizado e n = o estímulo não foi aleatorizado).

5. Análise descritiva

A análise descritiva do projeto foi dividida em duas partes. A seguir, apresentamos algumas interpretações retiradas das volumetrias vistas. Ao final, discutimos uma ideia de análise de concordância envolvendo os falantes e a classificação tônica das pseudopalavras.

5.1 Perfil dos participantes

5.1.1 Gênero e idade

Os 34 participantes do estudo estão divididos entre 21 mulheres e 13 homens, com idades que variam entre 18 e 60 anos. A Figura B.1 mostra a distribuição dos respondentes segundo a faixa etária e gênero. A faixa etária foi dividida a partir dos quartis da variável idade, para resumir as informações da amostra. Com exceção da faixa etária de 31 a 38 anos, há mais mulheres do que homens em cada quartil.

5.1.2 Naturalidade

A Tabela A.1 exibe o perfil dos informantes segundo sua naturalidade (variável agrupada em razão da baixa volumetria de indivíduos por UF). É possível perceber que a maior proporção (82%) dos participantes reside no estado de São Paulo, portanto, essa variável não será considerada na análise.

5.1.3 Área de formação

A Tabela A.2 mostra o agrupamento da formação dos voluntários. A influência do curso de Letras na categorização das pseudopalavras é um dos fatores de possível interesse na análise dos dados.

Analisando a distribuição por formação, 19 (56%) participantes são de outras áreas e 15 (44%) são da área de Letras; no entanto, cruzando essa variável com as informações de escolaridade (Tabela A.3), notamos que o grupo majoritário são estudantes de Letras com ensino superior incompleto (11 pessoas, o que representa cerca de 32% do total). Destacamos que a amostra é pequena e não representativa da população brasileira, pois apenas um indivíduo tem nível de escolaridade abaixo do universitário.

A distribuição das variáveis línguas e música pode ser observada nas Tabelas A.4 e Tabela A.5, respectivamente. Observamos na Figura B.2 que, dos participantes que têm conhecimento de línguas, mais de metade (58%) são do curso de Letras, e todas as pessoas sem nenhum conhecimento de outro idioma são de outras áreas. Isso pode levar a um confundimento do efeito dessas duas variáveis, pois não há nenhum indivíduo que curse Letras sem conhecimento de línguas no estudo.

5.2 Variáveis linguísticas

A tonicidade de produção (classificação da pseudopalavra) do tipo proparoxítona é a menos expressiva (4%), enquanto a categoria paroxítona aparece com maior frequência nas respostas dos participantes (57%), como visto na Tabela 1 a seguir. Portanto, o processo de acentuação das pseudopalavras nesse estudo parece ir de acordo com a afirmação de Collischonn (1999) de que *“Podemos considerar que o acento proparoxítono é marcado, no sentido de que é menos usual. É um acento especial, contrário à tendência geral de acentuar a penúltima sílaba”*.

Tabela 1: Distribuição da Tonicidade de produção

Tonicidade de produção	n
Oxítona	4.904 (39%)
Paroxítona	7.163 (57%)
Proparoxítona	444 (4%)
Total	12.511 (100%)

De forma geral, comparando a tonicidade da produção e a tonicidade das palavras-alvo (vide Tabela A.6), nota-se que aproximadamente 73% das pseudopalavras classificadas como proparoxítonas tinham uma palavra-alvo de tonicidade proparoxítona. Entretanto, cerca de 77% das pseudopalavras que têm como palavra-alvo uma proparoxítona foram categorizadas pelos participantes como paroxítonas (Tabela A.7).

Além disso, nota-se também que pseudopalavras originadas de palavras-alvo oxítonas e paroxítonas foram acentuadas de forma praticamente uniforme entre oxítonas e paroxítonas, o que não ocorre com as pseudopalavras criadas a partir de palavras-alvo proparoxítonas. Pode-se pensar então, que há influência da palavra-alvo proparoxítona na atribuição de tonicidade proparoxítona.

Observa-se na Tabela A.8 que os grupos de classificação 2 e 4 (ou seja, os grupos em que as palavras sofreram mais alterações em relação à referência original) apresentam uma proporção menor de proparoxítonas produzidas quando comparados aos grupos 1 e 3.

O fluxo entre tonicidade da palavra-alvo (lado esquerdo) e tonicidade da produção (lado direito) pode ser analisado graficamente por grupo de classificação na Figura 1.

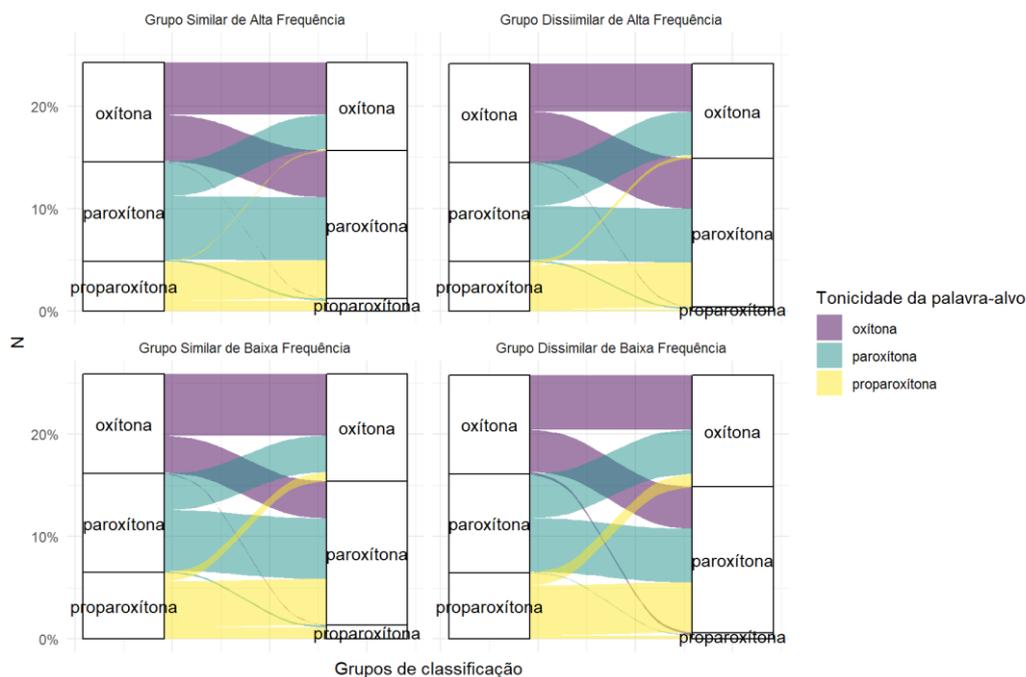


Figura 1: Tonicidade de palavra-alvo e pseudopalavra

No geral, muitas pseudopalavras cujas palavras de referência são paroxítonas foram classificadas pelos falantes como oxítonas, e vice-versa.

Consegue-se perceber que nos grupos de baixa frequência, a quantidade de pseudopalavras que foram classificadas como oxítona e cuja tonicidade da palavra-alvo era proparoxítona é consideravelmente maior do que nos grupos de alta frequência. Além disso, para os grupos dissimilares a quantidade de pseudopalavras acentuadas como proparoxítonas é ainda menor do que nos grupos similares. Uma hipótese é de que os grupos dissimilares (2 e 4) diminuem a probabilidade de atribuição proparoxítona. Os gráficos de fluxo podem ser vistos detalhadamente nas Figuras B.3 a B.6.

Na Tabela A.9, vemos que 85% das pseudopalavras cuja estrutura era CV-CV-CV foram classificadas como paroxítonas, e 79% das pseudopalavras cuja estrutura era CV-CV-CVC foram classificadas como oxítonas. Logo, parece existir forte influência da estrutura na atribuição da tonicidade em palavras do português brasileiro.

Por fim, explorando os dados no âmbito de validação, pode-se calcular pela Tabela A.2 que, aproximadamente, 73% das pseudopalavras foram validadas de acordo com o processo descrito anteriormente na Seção 3. A Tabela 2 indica, também, que cerca de 81% das pseudopalavras cuja similaridade/dissimilaridade não foi validada foram classificadas como paroxítonas.

Tabela 2: Frequência da Tonicidade de produção por nível da variável Validação.

Validação			
Tonicidade produção	Não Validada	Quase Validada	Validada
Oxítona	335 (16%)	529 (39%)	4.040 (45%)
Paroxítona	1.727 (81%)	771 (57%)	4.665 (51%)
proparoxítona	56 (3%)	48 (4%)	340 (4%)
Total	2.118 (100%)	1.348 (100%)	9.045 (100%)

Dentre as não validadas, aproximadamente 86% das pseudopalavras pertencem ao grupo das similares. Ou seja, percebe-se que nos grupos 1 (pseudopalavras similares de alta frequência) e 3 (pseudopalavras similares de baixa frequência) a porcentagem de palavras não validadas é superior aos demais grupos (vide Figura 2).

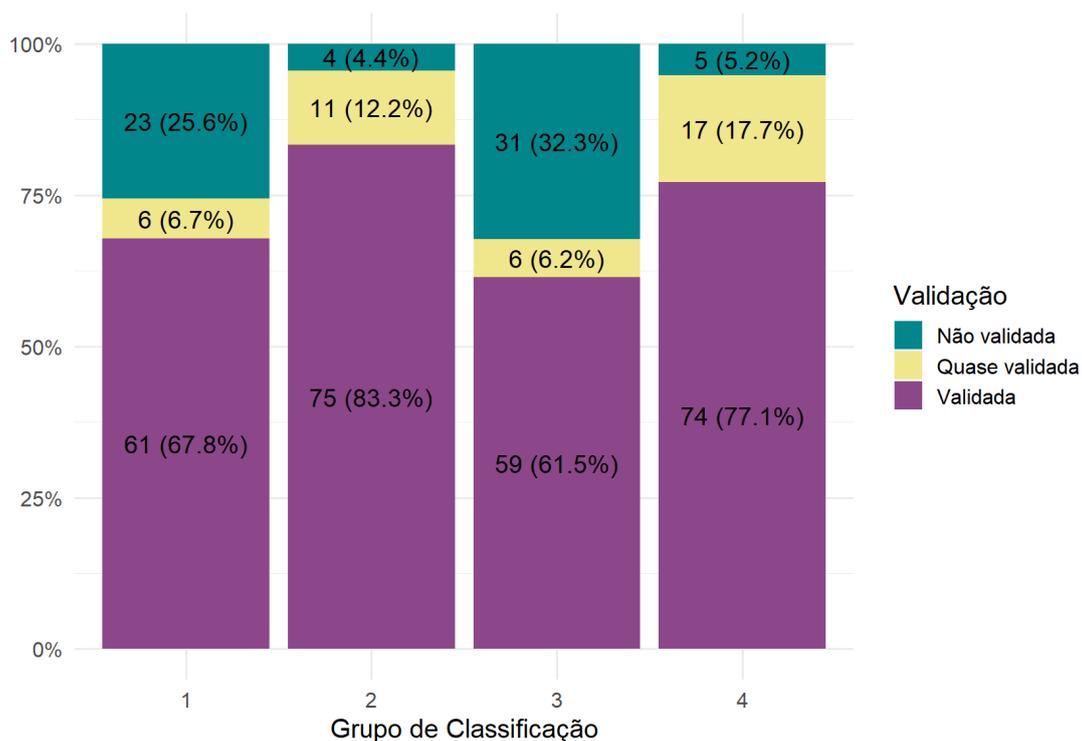


Figura 2: Distribuição da Validação por Grupo de classificação

5.3 Análise de concordância

Como nossa base de dados está no formato de medidas repetidas (falantes acentuando diversas pseudopalavras) e uma mesma palavra-alvo gerou diversas pseudopalavras, o pressuposto de independência entre as observações não é verdadeiro. Logo, optamos por medir a concordância entre a acentuação dos falantes através de uma estatística do tipo Kappa, separando as análises em cada nível das seguintes variáveis: grupo de classificação, estrutura da palavra e segmento modificado.

Busca-se investigar o grau de concordância entre os falantes quanto à classificação das 372 pseudopalavras em oxítone, paroxítone e proparoxítone. Como os níveis da variável resposta (tonicidade da produção) têm uma distribuição desbalanceada, houve a necessidade de utilizar um fator corretor no cálculo da estatística Kappa. Assim, para esse conjunto de dados, foi adotada a

estatística de AC1 de Gwet, medida que, de acordo com a literatura, pode ser utilizada em amostras desbalanceadas (Gwet 2002, Gwet 2008).

AC1 de Gwet: Coeficiente para n experimentos com r avaliadores que utilizam um sistema de classificação em Q categorias

$$AC1 = \frac{p_a - p_{e\gamma}}{1 - p_{e\gamma}}$$

com

$$p_a = \frac{1}{n} \sum_{i=1}^n \left(\sum_{q=1}^Q \frac{r_{iq}(r_{iq} - 1)}{r - 1} \right)$$

$$p_{e\gamma} = \frac{1}{Q - 1} \sum_{q=1}^Q \pi_q(1 - \pi_q)$$

$$\pi_q = \frac{1}{n} \sum_{i=1}^n \frac{r_{iq}}{r}$$

Temos que

$p_{e\gamma}$ = probabilidade de concordância ao acaso

p_a = probabilidade de concordância geral

π_q = probabilidade de classificação na categoria q

r_{iq} = probabilidade de classificação do i -ésimo experimento na categoria q

Quanto mais próximo de 1, menor a probabilidade de a concordância acontecer devido ao acaso, ou seja, melhor a concordância entre os falantes na classificação das pseudopalavras. Altman (1991) define os seguintes pontos de corte para a interpretação da estatística Kappa (e, conseqüentemente, do AC1).

Tabela 3: Pontos de corte para medidas do tipo Kappa

$AC1 < 0,20$	pobre
$0,2 \leq AC1 < 0,4$	razoável
$0,4 \leq AC1 < 0,6$	moderada
$0,6 \leq AC1 < 0,8$	boa
$AC1 \geq 0,8$	muito boa

Além da estimação da medida geral (sem segmentar a base de acordo com alguma característica), separou-se as pseudopalavras em cada uma das categorias de três variáveis escolhidas e calculou-se a medida AC1 de concordância em cada um dos níveis das mesmas.

Observando o resultado da medida sem abertura dos níveis (vide Tabela A.10), os falantes apresentaram uma concordância boa na atribuição do acento tônico pois a estatística de Gwet produziu um coeficiente de, aproximadamente, 0,68.

Já na Tabela A.11 vemos que há semelhança de concordância entre os grupos similares e dissimilares, sendo nos dissimilares os maiores níveis de concordância (0,75 e 0,71). Porém, em todos os casos pode se considerar que a concordância foi boa.

Uma hipótese para o fato de os falantes concordarem mais nos grupos dissimilares seria de que quanto menos referência entre a pseudopalavra e uma palavra-alvo, maior a chance de o falante utilizar outra característica para definir a tonicidade (como por exemplo a estrutura da palavra).

Em relação à estrutura da palavra, há uma maior concordância nas respostas dos falantes cujas pseudopalavras estão no nível CV-CV-CV, como visto na Tabela A.12. Dado que dentro da estrutura CV-CV-CV 85% das pseudopalavras foram classificadas como paroxítonas, parece existir uma concordância alta entre os falantes de seguirem essa tonicidade.

Por fim, em relação ao segmento modificado, vemos que há semelhança na medida de concordância entre os dois níveis de mudança (0.67 e 0.68). Ou seja, os falantes concordam na tonicidade da produção de forma muito parecida entre os dois níveis como visto na Tabela A.13.

Portanto, a análise de concordância parece indicar que as variáveis grupo de classificação e estrutura da palavra podem ser importantes para a atribuição do acento tônico, visto que seus níveis apresentaram estatísticas de Gwet diferentes entre si.

6. Análise inferencial

Nesta seção apresentamos a análise das variáveis linguísticas, extralinguísticas e experimentais com o intuito de caracterizar o processo de acentuação tônica. Para entender quais variáveis de fato têm influência, construímos modelos multinomiais mistos em que o efeito aleatório considerado foi especificado pelo indicador do falante. Assim, este tipo de modelo permite considerar a correlação das respostas do mesmo falante e o desbalanceamento na quantidade de respostas de cada um. Além disso, essa estrutura possibilita também a inclusão de variáveis relacionadas às características dos falantes em conjunto com as variáveis explicativas correspondentes à palavra alvo.

As seguintes variáveis foram consideradas no modelo completo:

- **Variável resposta:** Tonicidade de produção
- **Efeito aleatório:** Efeito de cada falante
- **Variáveis independentes incluídas como efeitos fixos:** tonicidade alvo, estrutura da palavra, grupo de classificação, segmento modificado, bloco de apresentação, aleatorização, música, línguas, idade, gênero, escolaridade, área de formação

Com isso, foi formulado um modelo de regressão logística multinomial considerando o nível Paroxítone como categoria de referência. Esta escolha baseia-se no fato de ela ser considerada a acentuação tônica mais comum na língua portuguesa. É importante notar que o “resultado” deste modelo são duas componentes, uma para comparar o nível Oxítone em relação a Paroxítone, e a outra para comparar o nível Proproxítone em relação a Paroxítone. A seguir,

apresentamos as formulações matemáticas de cada componente do modelo no contexto do trabalho.

$$\ln\left[\frac{\pi_{ijOxítona}}{\pi_{ijParoxítona}}\right] = X' \beta_{(O-Pa)} + b_{i(O-Pa)}$$

com

C1

$i = 1, \dots, 34$, indicador do falante,

$j = 1, \dots, n_i$: j -ésima pseudopalavra acentuada pelo i -ésimo falante,

X' : matriz associada aos efeitos fixos cujas colunas são as observações de cada variável independente para cada falante,

$\beta_{(O-Pa)}$: vetor de parâmetros desconhecidos associados aos efeitos fixos do modelo para o nível Oxítona/Paroxítona,

$b_{i(O-Pa)}$: efeito aleatório relacionado ao i -ésimo falante correspondentes ao nível Oxítona/Paroxítona.

$$\ln\left[\frac{\pi_{ijProparoxítona}}{\pi_{ijParoxítona}}\right] = X' \beta_{(Pro-Pa)} + b_{i(Pro-Pa)}$$

C2

com

$i = 1, \dots, 34$, indicador do falante,

$j = 1, \dots, n_i$: j -ésima pseudopalavra acentuada pelo i -ésimo falante,

X' : matriz associada aos efeitos fixos cujas colunas são as observações de cada variável independente para cada falante,

$\beta_{(Pro-Pa)}$: vetor de parâmetros desconhecidos associados aos efeitos fixos do modelo para o nível Proparoxítona/Paroxítona,

$b_{i(Pro-Pa)}$: efeito aleatório relacionado ao i -ésimo falante correspondentes ao nível Proparoxítona/Paroxítona.

Devido ao experimento relacionado à validação da similaridade/dissimilaridade entre palavra-alvo e pseudopalavra, no qual algumas dessas relações não foram consideradas validadas (Seção 3), avaliamos se a exclusão de pseudopalavras não validadas afetaria o resultado do modelo ajustado à base total.

Retirando algumas amostras aleatórias diferentes, sempre contendo 25% de pseudopalavras não validadas, observou-se que o resultado do modelo ajustado com cada uma destas amostra é distinto. Logo, entendemos que não seria possível remover apenas uma porcentagem das pseudopalavras cuja similaridade/dissimilaridade não foi validada. Além disso consideramos que a inclusão de pseudopalavras não validadas acabaria descaracterizando ou deslegitimando o processo de validação. Assim, as pseudopalavras não validadas não foram incluídas na modelagem.

6.1 Modelo selecionado

Considerando todas as variáveis independentes descritas na seção anterior foi realizado um processo de seleção de variáveis *backward* de forma manual, considerando o valor-p para cada coeficiente do modelo de forma simultânea.

Adotando 5% como nível de significância, verificou-se que a maior parte das variáveis relacionadas aos falantes não apresentaram significância no modelo ajustado. Depois de retirar uma a uma as variáveis não significantes, avaliamos retirar a variável extralinguística escolaridade, já que essa variável apresentou significância em apenas um nível e possuía apenas uma observação em um dos níveis. Comparamos o critério de informação de Akaike (AIC) dos modelos com e sem escolaridade; o modelo sem escolaridade apresentou menor AIC, logo, concluímos que o modelo sem a variável escolaridade é melhor.

Portanto, o modelo selecionado foi aquele que inclui as variáveis que apresentaram significância em pelo menos um nível -excluindo escolaridade-, que foram: tonicidade da palavra-alvo, estrutura da palavra, grupo de classificação, segmento modificado, bloco de apresentação e aleatorização.

As Tabelas A.14 e A.15, respectivamente, contêm as estimativas, erros padrão e valores-p dos efeitos fixos obtidos pelo modelo final ajustado. A Tabela A.14 apresenta os valores para a componente do modelo relacionada aos níveis

Oxítona/Paroxítona (C1) e a Tabela A.15 apresenta os valores para a componente relacionada aos níveis Proparoxítona/Paroxítona (C2). É importante observar que no modelo multinomial as variáveis são selecionadas de forma conjunta e, portanto, nem sempre os coeficientes das variáveis serão representados em ambos os componentes ou apresentam níveis de significância que levem às mesmas conclusões em ambas as comparações.

6.2 Interpretações dos parâmetros

O modelo multinomial ajustado considerou a denominada parametrização da casela de referência nos seguintes níveis de referência:

Tonicidade alvo = oxítona

Estrutura = CV-CV-CV,

Grupo de classificação = 1,

Segmento modificado = consoante

Bloco de apresentação = 1

Aleatorização = sim

A interpretação da componente do modelo relacionada à comparação entre a probabilidade de uma palavra ser classificada como Oxítona em relação a ser classificada como Paroxítona (formulação C1) pode ser realizada a partir da saída apresentada na Tabela A.14.

Percebe-se que, mantidas as demais variáveis em seus respectivos níveis de referência, a probabilidade da pseudopalavra ser classificada como paroxítona é maior que a probabilidade da mesma ser classificada como oxítona se: a tonicidade da palavra-alvo for paroxítona (valor-p <0,001) ou proparoxítona (valor-p <0,001) em vez de oxítona.

Ao contrário, a probabilidade da pseudopalavra ser classificada como oxítona é maior que a probabilidade de ela ser classificada como paroxítona se: o

segmento modificado for uma vogal em vez de consoante (valor-p = 0,004), a estrutura da palavra for do tipo CV-CV-CVC ao invés de CV-CV-CV (valor-p < 0,001), a palavra pertencer ao grupo 3 (valor-p <0,001) ou ao grupo 4 (valor-p <0,001) em vez do grupo 1.

A interpretação da componente do modelo relacionada à comparação entre a probabilidade de uma palavra ser classificada como Proparoxítônica em relação a ser classificada como Paroxítônica (formulação C2) pode ser realizada a partir da saída apresentada na Tabela A.15.

Assim, mantidas as demais variáveis em seus respectivos níveis de referência, a probabilidade da pseudopalavra ser classificada como paroxítônica é maior que a probabilidade da mesma ser classificada como proparoxítônica se: a tonicidade da palavra-alvo for paroxítônica em vez de oxítônica (valor-p = 0,003); a palavra pertencer ao grupo 2 (valor p <0,001) ou ao grupo 4 (valor p <0,001) em vez do grupo 1; se pertencer ao bloco de apresentação 2 em vez do bloco de apresentação 1 (valor p <0,001).

Ao contrário, a probabilidade da pseudopalavra ser classificada como proparoxítônica é maior que a probabilidade de ela ser classificada como paroxítônica se: a palavra-alvo for proparoxítônica em vez de oxítônica (valor p <0,001), a estrutura for do tipo CV-CV-CVC ao invés de CV-CV-CV (valor-p < 0,001), o segmento modificado for uma vogal ao invés de uma consoante (valor-p = 0,012).

Podemos resumir que, na componente oxítônica/paroxítônica, a estrutura da pseudopalavra é o que mais influencia na produção oxítônica (a probabilidade de produção oxítônica -e não paroxítônica-, quando a estrutura da palavra é CV-CV-CVC, aumenta em aproximadamente 73 vezes se comparada à probabilidade de produção oxítônica -e não paroxítônica- com estrutura CV-CV-CV), além da tonicidade da palavra-alvo (quando a mesma difere de oxítônica, a probabilidade de produção oxítônica diminui entre 75% e 90%, se comparada à da tonicidade alvo oxítônica) e o segmento modificado (se a mudança ocorre em uma vogal, a probabilidade de produção oxítônica -e não paroxítônica- aumenta em 20% se comparada a probabilidade de produção oxítônica -e não paroxítônica- com

mudança em consoante). Outro ponto de destaque vai para a influência dos grupos de classificação: para os grupos 3 (similar de baixa frequência) e 4 (dissimilar de baixa frequência), a probabilidade de produção oxítona -e não paroxítona- aumenta em 74% e 52%, respectivamente, quando comparada ao grupo 1 (similar de alta frequência).

Na componente proparoxítona/paroxítona, podemos destacar a tonicidade alvo proparoxítona, que aumenta a probabilidade de produção proparoxítona -e não paroxítona- em mais de 400% (em relação a probabilidade de produção proparoxítona com tonicidade alvo oxítona), a estrutura da palavra CV-CV-CVC, que aumenta a probabilidade de produção proparoxítona -e não paroxítona- em 500% (em relação à probabilidade de produção proparoxítona com estrutura CV-CV-CV) e os grupos 2 (dissimilar de alta frequência) e 4 (dissimilar de baixa frequência), que diminui em aproximadamente 70% e 75% a probabilidade de atribuição proparoxítona (quando comparada à probabilidade de produção proparoxítona de pseudopalavras classificadas no grupo 1).

É possível analisar os resultados em níveis diferentes dos definidos na casela de referência. Para a componente 1, o cálculo utilizado nessa comparação entre os níveis da variável grupo de classificação é explicitado a seguir.

$$\frac{\frac{P(Y_{ij}=Oxítona|Grupo=m)}{P(Y_{ij}=Paroxítona|Grupo=m)}}{\frac{P(Y_{ij}=Oxítona|Grupo=n)}{P(Y_{ij}=Paroxítona|Grupo=n)}} = \frac{e^{\beta_{Grupo m}}}{e^{\beta_{Grupo n}}} = e^{\beta_{Grupo m} - \beta_{Grupo n}}$$

Os índices m e n representam os níveis da variável grupo, e devem ser diferentes de 1, pois as comparações com o grupo 1 já foram apresentadas.

A Tabela 4 apresenta os resultados numéricos dessas comparações para a componente 1 do modelo. Por meio dessa tabela, pode-se observar que a chance da pseudopalavra ser classificada como oxítona (e não paroxítona) é maior se ela for do grupo 3 do que se ela for do grupo 2. Além disso, a chance da pseudopalavra ser classificada como oxítona (e não paroxítona) é menor se ela for do grupo 4 do que se ela for do grupo 2. Por fim, a chance da

pseudopalavra ser classificada como oxítona (e não paroxítona) é a mesma se ela for do grupo 4 ou do grupo 3.

Tabela 4: Comparativos entre os níveis da variável Grupo – Componente 1

Comparação	K = exp(Cálculo)	Valor -p
oxítona~grupo 3 <i>versus</i> grupo 2	1,57	< 0,001
oxítona~grupo 4 <i>versus</i> grupo 2	0,87	< 0,001
oxítona~grupo 4 <i>versus</i> grupo 3	0,79	0,131

Para a componente 2, o cálculo utilizado nessa comparação entre os níveis da variável grupo de classificação é explicitado a seguir.

$$\frac{\frac{P(Y_{ij}=Proparoxítona|Grupo=m)}{P(Y_{ij}=Paroxítona|Grupo=m)}}{\frac{P(Y_{ij}=Proparoxítona|Grupo=n)}{P(Y_{ij}=Paroxítona|Grupo=n)}} = \frac{e^{\beta_{Grupo m}}}{e^{\beta_{Grupo n}}} = e^{\beta_{Grupo m} - \beta_{Grupo n}}$$

Os índices m e n representam os níveis da variável grupo, e devem ser diferentes de 1, pois as comparações com o grupo 1 já foram apresentadas anteriormente.

A Tabela 5 apresenta os resultados numéricos dessas comparações para a componente 2 do modelo. É possível observar que a chance da pseudopalavra ser classificada como proparoxítona (e não paroxítona) é maior se ela for do grupo 3 do que se ela for do grupo 2. Além disso, a chance da pseudopalavra ser classificada como proparoxítona (e não paroxítona) é a mesma se ela for do grupo 4 ou do grupo 2. Por fim, a chance da pseudopalavra ser classificada como oxítona (e não paroxítona) é maior se ela for do grupo 4 do que se ela for do grupo 3.

Tabela 5: Comparativos entre os níveis da variável Grupo – Componente 2

Comparação	K = exp(Cálculo)	Valor -p
proparoxítona~grupo 3 <i>versus</i> grupo 2	4,49	< 0,001
proparoxítona~grupo 4 <i>versus</i> grupo 2	0,27	0,318
proparoxítona~grupo 4 <i>versus</i> grupo 3	4,19	< 0,001

6.3 Avaliação do modelo

Não há, na literatura pesquisada, explicações de métodos para analisar resíduos e encontrar pontos aberrantes em modelos multinomiais, apenas para modelos logísticos (Tapia *et al*, 2018). Como alternativa para avaliar o desempenho do modelo, foram calculadas medidas de desempenho de sua capacidade preditiva.

Portanto, para avaliar o ajuste do modelo, calculamos o Coeficiente de Predição Positiva -CPP- (probabilidade de classificação no nível q dado que a tonicidade de produção foi avaliada no nível q) e o Coeficiente de Predição Negativa -CPN- (probabilidade de não ser classificada no nível q dado que a tonicidade de produção não foi avaliada no nível q) para cada nível da variável resposta. As medidas são feitas agrupando dois níveis da variável dependente e os comparando com o terceiro. Os valores estimados e seus correspondentes intervalos de confiança estão apresentados nas Tabelas A.16, A.17 e A.18, para os níveis oxítona, paroxítona e proparoxítona, respectivamente.

Vemos que para as pseudopalavras oxítona e paroxítona, os valores de CPP e CPN são superiores a 0,83 - considerados altos -, o que indica um bom ajuste do modelo aos dados. Observa-se na Tabela A.18 - para as pseudopalavras proparoxítonas -, que o valor de CPP é baixo (0,11), o que pode ser explicado pela baixa frequência dessa classificação na base.

Outras métricas também analisadas foram: Valores Preditivos Positivos (probabilidade de tonicidade de produção no nível q dado que a classificação foi no nível q) e Valores Preditivos Negativos (probabilidade de tonicidade de produção não foi avaliada no nível q dado que a classificação não foi no nível q). Os resultados foram, de modo geral, altos, e concluímos então que o ajuste do modelo está satisfatório.

6.4 Análise dos componentes aleatórios

Em modelos mistos, os efeitos aleatórios seguem o pressuposto de distribuição normal. Para analisar essa suposição, foi feito o gráfico dos quantis teóricos da distribuição normal contra os valores preditos dos efeitos aleatórios do modelo (Figura B.7). Embora a figura aponte uma distribuição com caudas um pouco pesadas, a literatura indica que essa é uma premissa robusta a possíveis afastamentos do pressuposto, logo, podemos considerar que a distribuição dos efeitos aleatórios é satisfatória.

7. Conclusão

A partir da análise inferencial concluímos que há indícios de que existem outros fatores, além da estrutura da palavra, que interferem na acentuação tônica. Por meio dos modelos foi evidenciado o papel dos grupos de classificação e da tonicidade da palavra-alvo na determinação da tonicidade de produção. Além disso, os dados analisados indicam que características dos falantes não são importantes no processo de atribuição tônica.

O processo de validação das pseudopalavras se mostrou relevante, uma vez que pseudopalavras não validadas interferem na seleção do modelo. Levando isso em conta, concluímos que o processo de validação proposto pela pesquisadora é, de fato, coerente, e apenas as pseudopalavras validadas e quase validadas são propícias para modelagem.

Com relação ao experimento, a interferência na aleatorização durante o processo de coleta dos dados mostrou-se significativa para o modelo final, portanto acreditamos que seria importante uma replicação do estudo, no qual se garantisse a aleatorização de todos os blocos para todos os falantes.

APÊNDICE A

Tabelas

Tabela A.1: Perfil dos participantes de acordo com a Naturalidade.

Naturalidade	n
São Paulo, SP	16 (47%)
Outros municípios de SP	12 (35%)
Outras UF	6 (18%)
Total	34 (100%)

Tabela A.2: Frequência da variável dicotômica Área de formação

Área de formação	Total
Outro	19 (56%)
Letras	15 (44%)
Total	34 (100%)

Tabela A.3: Escolaridade dos participantes pela Área de formação.

Escolaridade	Área de formação		
	Outro	Letras	Total
Fundamental Completo	1 (3%)	0 (0%)	1 (3%)
Superior Incompleto	6 (18%)	11 (32%)	17 (50%)
Superior Completo	5 (15%)	1 (3%)	6 (18%)
Pós-Graduação (Completo ou Incompleto)	7 (20%)	3 (9%)	10 (29%)
Total	19 (56%)	15 (44%)	34 (100%)

Tabela A.4: Frequência da variável dicotômica Línguas.

Conhecimento em línguas	n
Sim	26 (76%)
Não	8 (24%)
Total	34 (100%)

Tabela A.5: Frequência da variável dicotômica Música

Conhecimento em música	n
Sim	21 (62%)
Não	13 (38%)
Total	34 (100%)

Tabela A.6: Frequência da Tonicidade de produção por nível de Tonicidade da palavra-alvo

Tonicidade produção	Tonicidade da palavra-alvo			Total
	oxítona	paroxítona	proparoxítona	
Oxítona	2.642 (54%)	1.939 (39%)	323 (7%)	4.904 (100%)
Paroxítona	2.154 (30%)	2.824 (39%)	2.185 (31%)	7.163 (100%)

Tonicidade da palavra-alvo

Tonicidade produção	oxítona	paroxítona	proparoxítona	Total
Proparoxítona	49 (11%)	72 (16%)	323 (73%)	444 (100%)

Tabela A.7: Frequência da Tonicidade da palavra-alvo por nível de Tonicidade de produção

Tonicidade da palavra-alvo

Tonicidade produção	oxítona	paroxítona	proparoxítona	Total
oxítona	2.642 (55%)	1.939 (40%)	323 (11%)	4.904 (39%)
paroxítona	2.154 (44%)	2.824 (58%)	2.185 (77%)	7.163 (57%)
proparoxítona	49 (1%)	72 (2%)	323 (11%)	444 (4%)
Total	4.845 (100%)	4.835 (100%)	2.831 (100%)	12.511 (100%)

Tabela A.8: Frequência do Grupo de classificação por Tonicidade de produção

Grupo de classificação

Tonicidade produção	1.Similar de alta freq.	2.Dissimilar de alta freq.	3.Similar de baixa freq.	4.Dissimilar de baixa freq.	Total
oxítona	1.077 (22%)	1.152 (23%)	1.308 (27%)	1.367 (28%)	4.904 (100%)
paroxítona	1.805 (25%)	1.818 (25%)	1.760 (25%)	1.780 (25%)	7.163 (100%)
proparoxítona	153 (34%)	48 (11%)	167 (38%)	76 (17%)	444 (100%)

Tabela A.9: Frequência da Tonicidade de produção por Estrutura da palavra.

Estrutura da palavra

Tonicidade produção	CV-CV-CV	CV-CV-CVC	Total
oxítona	748 (10%)	4.156 (79%)	4.904 (39%)
paroxítona	6.200 (85%)	963 (19%)	7.163 (57%)
proparoxítona	323 (5%)	121 (2%)	444 (4%)
Total	7.271 (100%)	5240 (100%)	12.511 (100%)

Tabela A.10: Estatística de Gwet para a base total (sem abertura por nível de uma variável).

Estatística	Valor da Estatística
Gwet's AC1	0,68 (0,653;0,702)

Tabela A.11: Estatística de Gwet para cada nível da variável Grupo de classificação.

Grupo de classificação	Valor da Estatística
Similar de Alta Frequência	0,66 (0,613;0,715)
Similar de Baixa Frequência	0,59 (0,529;0,646)
Dissimilar de Alta Frequência	0,75 (0,717;0,793)
Dissimilar de Baixa Frequência	0,71 (0,662;0,750)

Tabela A.12: Estatística de Gwet para cada nível da variável Estrutura da palavra.

Estrutura da palavra	Valor da Estatística
CV-CV-CVC	0,66 (0,618;0,698)
CV-CV-CV	0,76 (0,734;0,792)

Tabela A.13: Estatística de Gwet para cada nível da variável Segmento modificado.

Segmento modificado	Valor da Estatística
Vogal	0,67
Consoante	0,68

Tabela A.14: Valores estimados, erros padrão e níveis de significância correspondentes aos parâmetros da componente C1

Coefficientes	Estimativa	Erro padrão	Estatística	Valor-p	K = exp(Estimativa)
oxítone~(Intercepto)	-1,950	0,271	-7,199	< 0,001	0,142
oxítone~tonicidade_alvo paroxítone	-1,392	0,075	-18,455	< 0,001	0,248
oxítone~tonicidade_alvo proparoxítone	-2,014	0,106	-19,057	< 0,001	0,133
oxítone~estrutura_palavra CV-CV-CVC	4,293	0,080	53,827	< 0,001	73,186
oxítone~grupo 2	0,105	0,099	1,056	0,291	1,111
oxítone~grupo 3	0,556	0,103	5,386	< 0,001	1,743
oxítone~grupo 4	0,419	0,098	4,256	< 0,001	1,520
oxítone~segmento_modificado vogal	0,187	0,064	2,909	0,004	1,206
oxítone~bloco_apresentacao 2	0,427	0,092	4,621	< 0,001	1,532
oxítone~bloco_apresentacao 3	0,141	0,094	1,506	0,132	1,151
oxítone~bloco_apresentacao 4	0,400	0,096	4,146	< 0,001	1,492

oxítona~aleatorizacao nao	-0.638	0,166	-3,835	< 0,001	0,528
---------------------------	--------	-------	--------	---------	-------

Tabela A.15: Valores estimados, erros padrão e níveis de significância correspondentes aos parâmetros da componente C2

Coeficientes	Estimativa	Erro padrão	Estatística	Valor-p	K = exp(Estimativa)
proparoxítona~(Intercepto)	-3,685	0,351	-10,501	< 0,001	0,025
proparoxítona~tonicidade_alvo paroxítona	-0,621	0,211	-2,940	0,003	0,537
proparoxítona~tonicidade_alvo paroxítona	1,669	0,186	8,947	< 0,001	5,307
proparoxítona~estrutura_palavra CV-CV-CVC	1,814	0,148	12,266	< 0,001	6,135
proparoxítona~grupo 2	-1,338	0,184	-7,252	< 0,001	0,262
proparoxítona~grupo 3	0,163	0,152	1,075	0,282	1,177
proparoxítona~grupo 4	-1,139	0,162	-7,036	< 0,001	0,320
proparoxítona~segmento_modificado vogal	0,294	0,118	2,498	0,012	1,342
proparoxítona~bloco_apresentacao 2	-0,794	0,196	-4,057	< 0,001	0,452
proparoxítona~bloco_apresentacao 3	-0,216	0,158	-1,365	0,172	0,806
proparoxítona~bloco_apresentacao 4	-0,179	0,159	-1,125	0,260	0,836
proparoxítona~aleatorizacao nao	0,217	0,273	0,794	0,427	1,242

Tabela A.16: Medidas de desempenho do modelo para as pseudopalavras Oxítonas

	Estimativa	Intervalo de Confiança (95%)
Coefficiente de Predição Positiva	0,87	(0,86; 0,88)
Coefficiente de Predição Negativa	0,88	(0,87; 0,89)
Valores Preditivos Positivos	0,85	(0,84; 0,86)
Valores Preditivos Negativos	0,89	(0,89; 0,90)

Tabela A.17: Medidas de avaliação do modelo para as pseudopalavras Paroxítonas

	Estimativa	Intervalo de Confiança (95%)
Coefficiente de Predição Positiva	0,89	(0,88; 0,90)
Coefficiente de Predição Negativa	0,83	(0,82; 0,84)
Valores Preditivos Positivos	0,85	(0,84; 0,86)
Valores Preditivos Negativos	0,87	(0,86; 0,88)

Tabela A.18: Métricas de avaliação do modelo para as pseudopalavras

Proparoxítonas

	Estimativa	Intervalo de Confiança (95%)
Coefficiente de Predição Positiva	0,11	(0,08; 0,14)

	Estimativa	Intervalo de Confiança (95%)
Coeficiente de Predição Negativa	1,00	(1,00; 1,00)
Valores Preditivos Positivos	0,66	(0,53; 0,77)
Valores Preditivos Negativos	0,97	(0,96; 0,97)

APÊNDICE B

Figuras

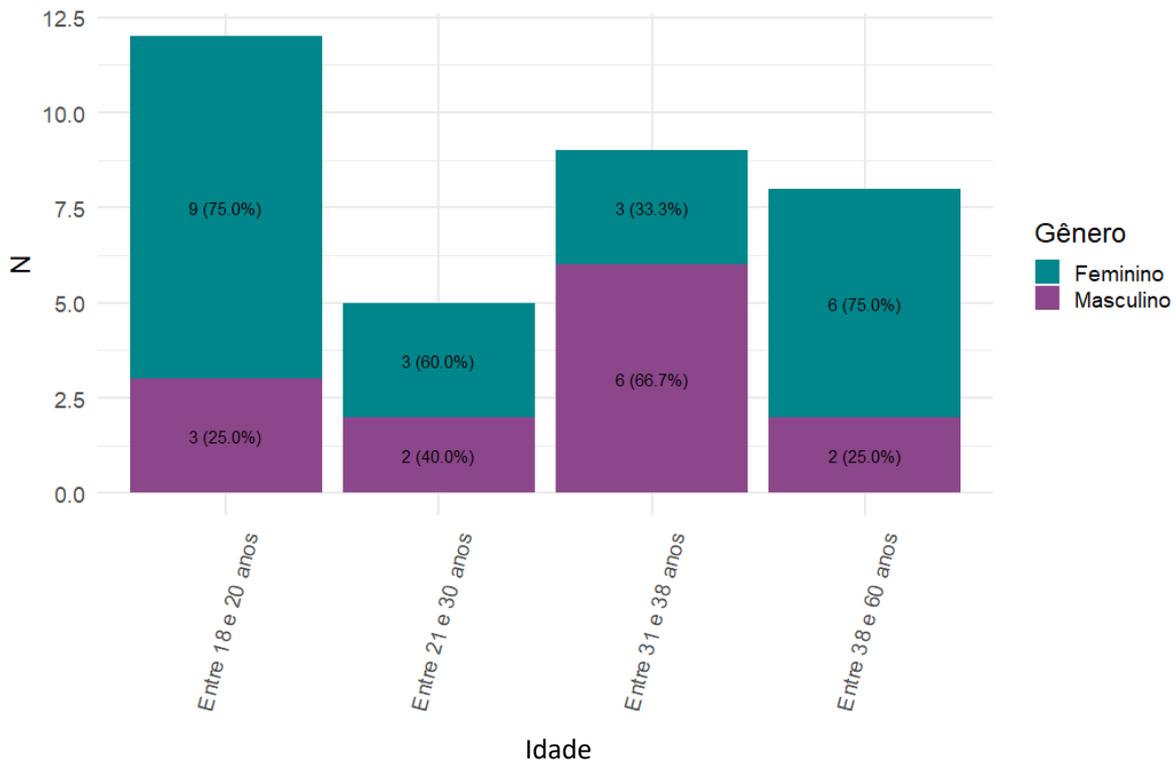


Figura B.1: Distribuição de Gênero por Idade

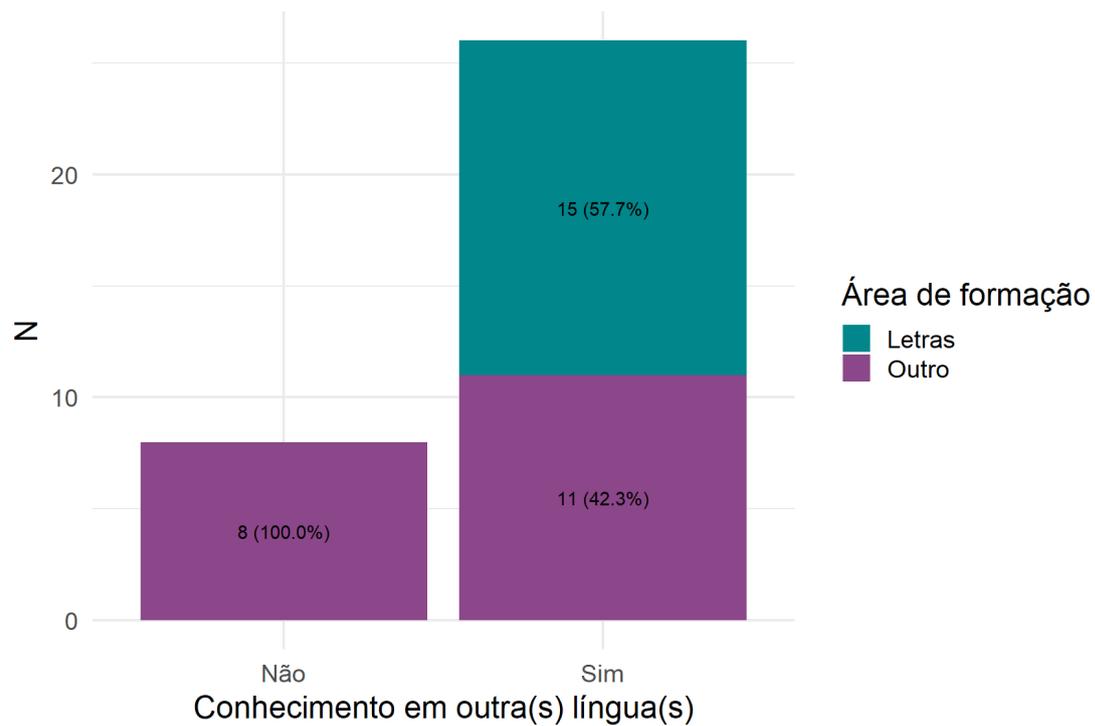


Figura B.2: Distribuição da variável Área de formação por Línguas

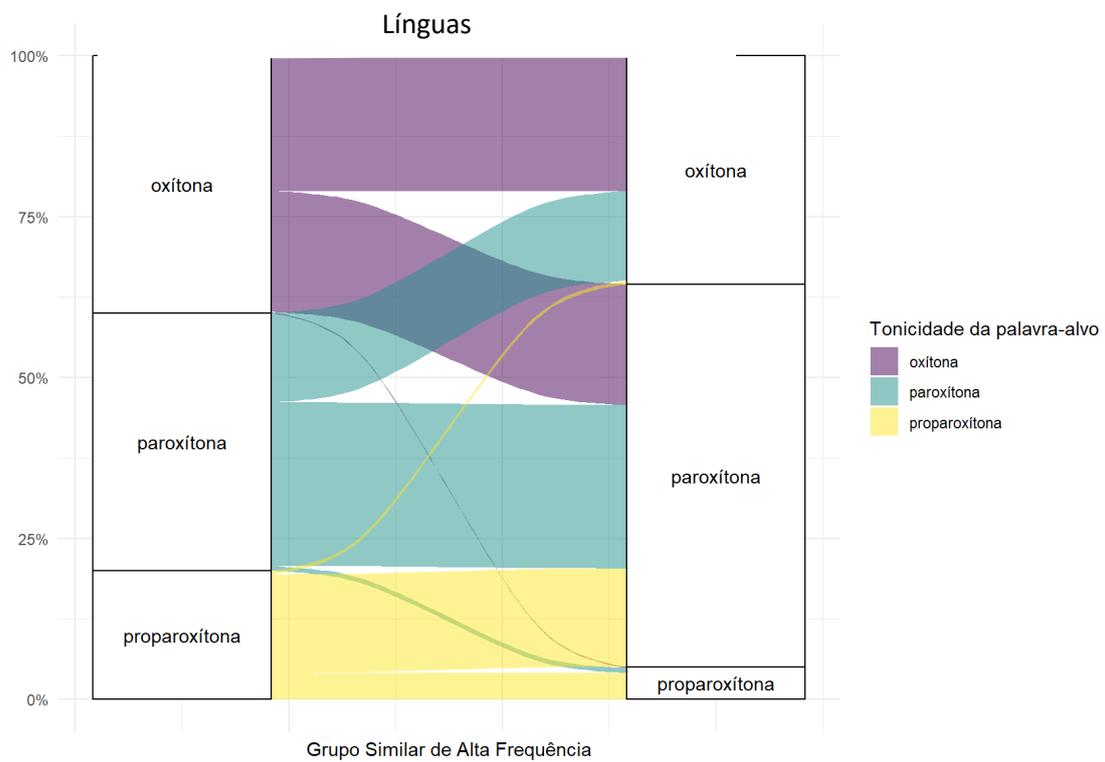


Figura B.3: Fluxo de tonicidade no Grupo Similar de Alta Frequência

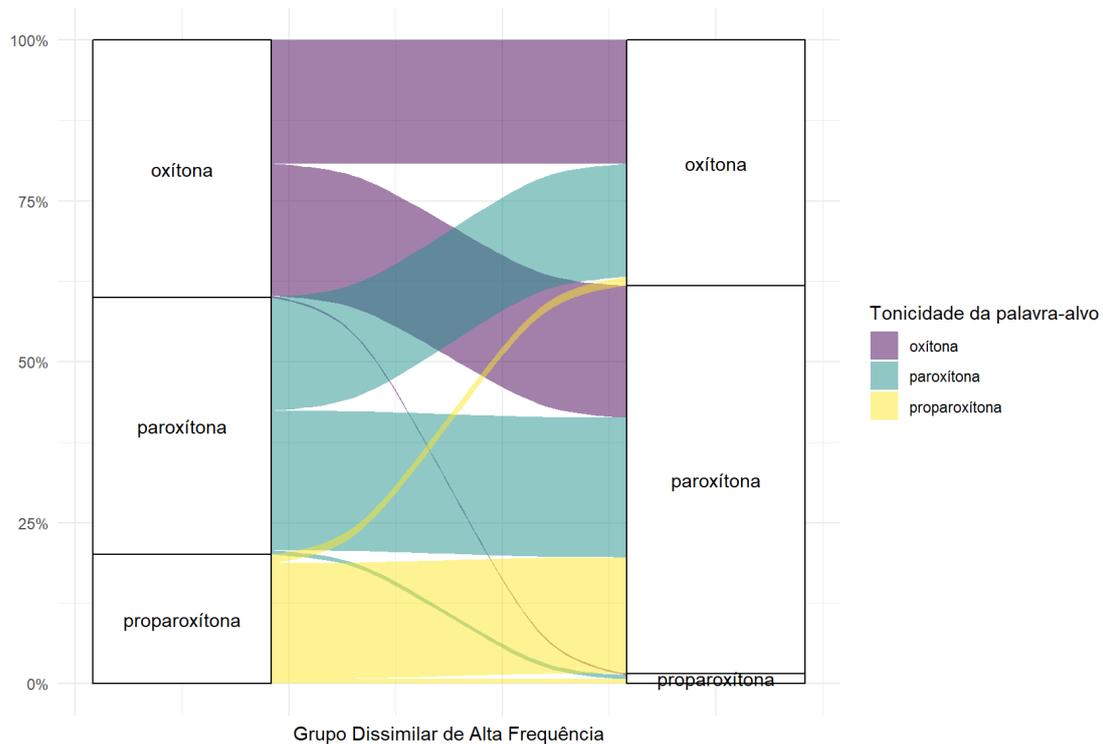


Figura B.4: Fluxo de tonicidade no Grupo Dissimilar de Alta Frequência

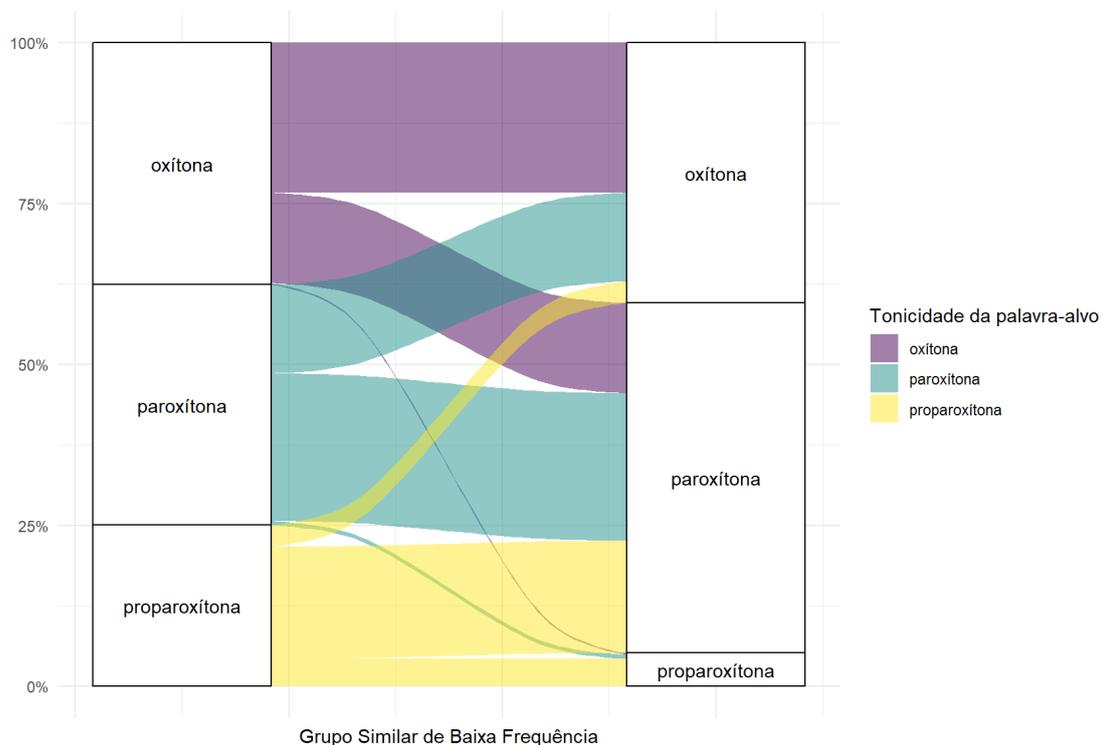


Figura B.5: Fluxo de tonicidade no Grupo Similar de Baixa Frequência

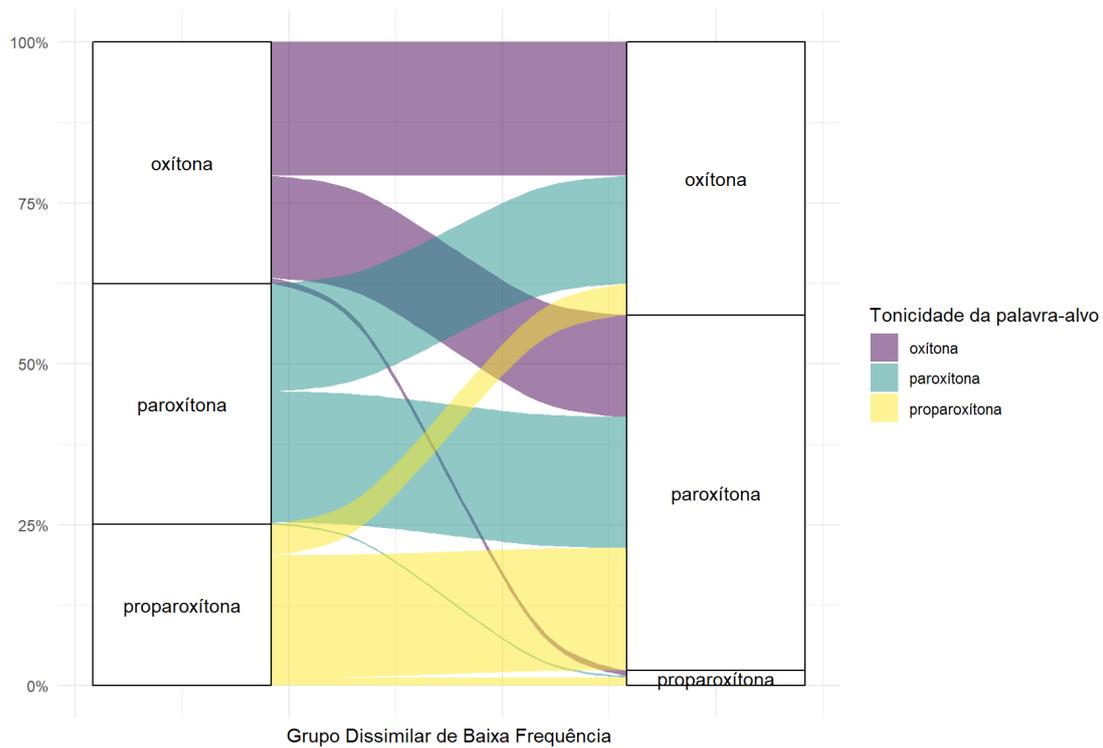


Figura B.6: Fluxo de tonicidade no Grupo Dissimilar de Baixa Frequência

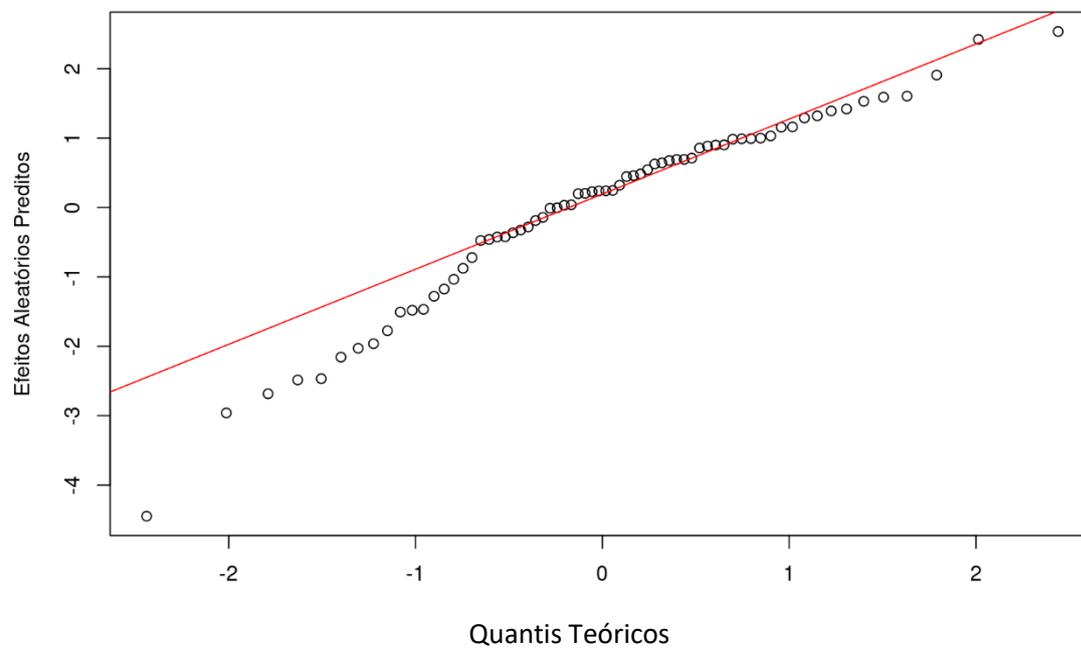


Figura B.7: Distribuição dos efeitos aleatórios do modelo versus os quantis da distribuição normal padrão

