

# Dimensões da Psicolinguística na ALFAL

**Marcio Martins Leitão**  
**Marcus Maia**  
organizadores

**UMA OBRA DO PROJETO 7  
PSICOLINGUÍSTICA E INTERFACES**

**COLEÇÃO ALFAL:  
ALÉM DAS FRONTEIRAS**

**COORDENADORES:**  
Dermeval da Hora  
Ângela Helmer



 LÍQUIDO EDITORIAL

copyright © 2022 Marcio Martins Leitão e Marcus Maia

Todos os direitos autorais dos textos publicados neste livro estão reservados aos autores e foram cedidos para uso da Editora Terracota Ltda., exclusivamente para a publicação desta obra. O conteúdo dos textos é de inteira responsabilidade de seus autores.

**Capa e Diagramação:**

Líquido Editorial

**Editor responsável**

Carlos Augusto Baptista de Andrade

**Conselho Editorial**

Adolfo Elizaincín - Universidad de la República – Uruguay

Ángela Helmer - University of South Dakota - USA

Angelita Martínez - Universidad de la Plata - Argentina

Beatriz Arias Álvarez - Universidad Nacional Autónoma de México - UNAM

Benjamim Corte-Real - Universidade Nacional de Timor-Leste

Carlos Augusto Batista de Andrade – Universidade Cruzeiro do Sul

Cleide Antônia Rapucci - UNESP – Assis

Dermeval da Hora - Universidade Federal da Paraíba - Brasil

Elisa Battisti - Universidade Federal do Rio Grande do Sul - Brasil

Gerson Albuquerque - Universidade Federal do Acre - Brasil

Guaraciaba Micheletti - Universidade Cruzeiro do Sul e USP - Brasil

Guillermo Arias Beatón - Cátedra Vigotski - Universidad de Havana - Cuba

Marlise Vaz Bridi – Universidade de São Paulo – UCP - Brasil

Moisés de Lemos Martins - Universidade do Minho - Portugal

Neusa Barbosa Bastos – PUC-SP e Universidade Presbiteriana Mackenzie

Regina Pires de Brito – Universidade Presbiteriana Mackenzie

**Dados Internacionais de Catalogação na Publicação (CIP)  
(Câmara Brasileira do Livro, SP, Brasil)**

Dimensões da psicolinguística na ALFAL [livro eletrônico] / Associação de Linguística e Filologia da América Latina - ALFAL ; Marcio Martins Leitão, Marcus Maia, organizadores. -- São Paulo, SP : Líquido Editorial e Consultoria, 2022. -- (Coleção ALFAL : além das fronteiras / coordenação Dermeval da Hora, Ângela Helmer) PDF

Vários autores.  
Bibliografia.  
ISBN 978-85-92943-10-3

1. Filosofia 2. Línguas e linguagem 3. Linguística 4. Psicolinguística I. Associação de Linguística e Filologia da América Latina - ALFAL. II. Leitão, Marcio Martins. III. Maia, Marcus. IV. Hora, Dermeval da. V. Helmer, Ângela. VI. Série.

22-121700

CDD-401.9

**Índices para catálogo sistemático:**

1. Psicolinguística 401.9

Eliete Marques da Silva - Bibliotecária - CRB-8/9380

Todos os direitos desta edição reservados à Líquido Editorial

[www.liquidoeditorial.com.br](http://www.liquidoeditorial.com.br)

# **ASSOCIAÇÃO DE LINGUÍSTICA E FILOGIA DA AMÉRICA LATINA – ALFAL**

## **DIRETORIA**

Presidente: Dermeval da Hora  
Secretária: Angelita Martínez  
Tesoureira: Ángela Helmer

## **VOGAIS**

Adolfo Elizaincín  
Beatriz Arias Álvarez  
Constanza Moya Pardo  
Elisa Battisti  
Martín Hummel  
Mireya Cisneros Estupiñán

## **UMA OBRA DO PROJETO 7**

### **PSICOLINGUÍSTICA E INTERFACES**

Márcio Martins Leitão  
Marcus Maia

## **COLEÇÃO ALFAL: ALÉM DAS FRONTEIRAS**

### **COORDENADORES:**

Dermeval da Hora  
Ángela Helmer



# **Dimensões da Psicolinguística na ALFAL**

**Marcio Martins Leitão  
Marcus Maia**  
organizadores

**2022**

## Sumário

APRESENTAÇÃO – <i>Marcio Martins Leitão (UFPA) e Marcus Maia (UFRJ)</i> .....	7
A APRENDIZAGEM DE LÍNGUAS BASEADA EM EVIDÊNCIAS EXPERIMENTAIS: POSSÍVEIS INTERFACES ENTRE A PSICOLINGÜÍSTICA E O ENSINO/APRENDIZAGEM DE LÍNGUAS ADICIONAIS - <i>Luiz Amaral (UMass, Amherst), Raquel Fellet Lawall (UFJF), Grazielle Frangiotti (UFSC), Paula Garcia de Freitas (UFPr), Ricardo Augusto de Souza (UFMG)</i> .....	13
PRINCIPAIS TEORIAS DE AQUISIÇÃO DA LINGUAGEM E SUA RELAÇÃO COM O BILINGUISMO - <i>Yeneisy Guilarte Quintela (UFF)</i> .....	36
BILINGÜISMO Y PROSODIA: ADAPTACIÓN TONAL DE PRÉSTAMOS DEL ESPAÑOL EN LENGUAS OTOMANGUES - <i>Mario E. Chávez Peón (CIESAS, México) &amp; Marcela San Giacomo Trinidad (UNAM)</i> .....	54
O PROCESSAMENTO DA VÍRGULA POR SURDOS BILÍNGUES LIBRAS/PORTUGUÊS - <i>Hercules Rodrigues dos Santos (UFPA) &amp; Francisca Maria Carvalho</i> .....	89
A INFLUÊNCIA DOS RECURSOS DE COESÃO E DA COERÊNCIA NO PROCESSAMENTO DA LEITURA – <i>Marcio Martins Leitão (UFPA) &amp; Antonia Simões (UFPA)</i> .....	111
AN EXPERIMENTAL STUDY ON COMPLEX SENTENCES WITH <i>BEFORE</i> AND <i>AFTER</i> – <i>Sara dos Santos Ribeiro (UFRJ)</i> .....	134
TIPOLOGIA ESTRUTURAL DAS ORAÇÕES CAUSAIS INTRODUZIDAS POR <i>PORQUE</i> e <i>JÁ QUE</i> - <i>Sabrina Lopes dos Santos (UFRJ)</i> .....	164
A ACEITABILIDADE DE “O MESMO” COMO ANAFÓRICO CORREFERENCIAL E SUAS IMPLICAÇÕES NA COMPREENSÃO EM LEITURA - <i>Bruna Alexandra Franzen (UFSC) &amp; Ana Cláudia de Souza (UFSC)</i> .....	185
RASTREAMENTO OCULAR NA LEITURA DE SENTENÇAS DE TÓPICO-COMENTÁRIO E SUJEITO-PREDICADO NO PB - <i>Lorrane da Silva Neves Medeiros Ventura (UFRJ)</i> .....	205

HÁBITOS DE LEITURA NO ENVELHECIMENTO SADIO: SUA CORRELAÇÃO COM AS FUNÇÕES EXECUTIVAS E A MEMÓRIA SEMÂNTICA - <i>Milena Socorro Rocha Gaspar Veja (UFPB)</i> , <i>José Ferrari Neto (UFPB)</i> .....	232
BIGRAMAS E TRIGRAMAS EM PSEUDOPALAVRAS DO PORTUGUÊS BRASILEIRO: VARIÁVEIS ESSENCIAIS NO PROCESSAMENTO ORTOGRÁFICO - <i>Gustavo Lopez Estivalet (UFPB)</i> .....	267
BIODATAS .....	287

# **BIGRAMAS E TRIGRAMAS EM PSEUDOPALAVRAS DO PORTUGUÊS BRASILEIRO: VARIÁVEIS ESSENCIAIS NO PROCESSAMENTO ORTOGRÁFICO**

**Gustavo Lopez Estivalet**

Universidade Federal da Paraíba, Brasil  
gustavoestivalet@hotmail.com

## **INTRODUÇÃO**

O reconhecimento das palavras durante a leitura depende de uma série de variáveis, tais como: tipo de letra, espaçamento entre as letras, número de letras, número de sílabas, sílaba tônica, frequência da palavra, complexidade semântica e contexto. Algumas dessas variáveis estão associadas ao processamento ortográfico (Davis & Bowers, 2006) enquanto outras estão associadas ao processamento linguístico (Monsell et al., 1989). Nesse sentido, observa-se que as pesquisas em psicolinguística relacionadas aos processos de leitura no português brasileiro (PB) têm se preocupado exclusivamente com fenômenos associados ao processamento linguístico, mas ignorado aspectos importantes relacionados ao processamento ortográfico.

Portanto, este trabalho foi realizado com as seguintes finalidades: a) preencher esta lacuna, b) evidenciar a necessidade e a importância de estudos que considerem os aspectos ortográficos durante a leitura e c) pesquisar as variáveis ortográficas específicas relacionadas ao processamento psicolinguístico visual. Diretamente associados ao processamento ortográfico estão fatores como: frequência da palavra (Taft, 1991), vizinhança ortográfica (Andrews, 1997), tamanho da palavra (New et al., 2006) e frequência de n-gramas (Shaoul et al., 2013). A esses, poder-se-iam adicionar fatores associados ao processamento visual como: cor da fonte/fundo, tipo de letra, tamanho das letras, espaçamentos entre letras, palavras e linhas (Vinckier et al., 2011).

Sendo assim, o objetivo principal deste trabalho foi investigar as seguintes variáveis ortográficas no processamento ortográfico de pseudopalavras do PB: frequência de n-gramas, vizinhos ortográficos, número de letras e número de

sílabas. Os objetivos específicos foram: i) testar empiricamente as variáveis ortográficas no processamento de pseudopalavras do PB, ii) validar o motor de geração de pseudopalavras do Léxico do Português Brasileiro (LexPorBR) (Estivalet & Meunier, 2015) e iii) salientar a importância das estruturas grafotáticas no processamento ortográfico do PB.

Como metodologia, aplicou-se um experimento psicolinguístico com tarefa de decisão lexical visual com pseudopalavras como estímulos experimentais. A utilização de pseudopalavras como estímulos experimentais permite descartarem-se efeitos de ativação semântica e uniformiza efeitos de procura e ativação do estímulo, pois pseudopalavras não estão armazenadas no léxico mental e não possuem representações semânticas (McClelland & Rumelhart, 1981). Desse modo, é possível analisarem-se variáveis associadas unicamente ao processamento ortográfico, tais como: frequências dos n-gramas, vizinhança ortográfica e tamanho das pseudopalavras.

A hipótese de base é que a frequências de n-gramas e o tamanho da palavra influenciam a leitura em etapas precoces do reconhecimento lexical, enquanto a vizinhança ortográfica influencia somente etapas tardias (Grainger & Whitney, 2004). A hipótese da presente investigação é que, por um lado, a frequência de n-gramas, o tamanho da palavra, e a vizinhança ortográfica influenciam somente etapas precoces do reconhecimento lexical, por outro lado, somente fatores linguísticos influenciam etapas tardias (Besner & Johnston, 1992). A hipótese alternativa é que há uma interação entre esses diferentes fatores durante o processamento ortográfico e linguístico para o reconhecimento lexical (McClelland & Rumelhart, 1981).

Os resultados apontaram diferenças significativas de frequência de bigramas e trigramas, vizinhança ortográfica e número de letras, suportando uma dissociação entre o processamento ortográfico como uma etapa precoce e o processamento linguístico como uma etapa tardia do acesso lexical e reconhecimento das palavras.

## FUNDAMENTAÇÃO TEÓRICA

Com o objetivo de se discutir como funciona o sistema de leitura da mente humana, pode-se perguntar: “a mente humana lê palavras inteiras?” Essa foi a provocação do artigo *Does the human mind read words as a whole?* (Grainger & Whitney, 2004), onde os autores exploram a ativação de n-gramas para o processamento ortográfico. Os mesmos autores mostraram que macacos babuínos treinados em função de recompensa podem aprender a discriminar palavras de pseudopalavras do inglês sem possuírem referencial semântico. Mesmo quando confrontados com novas palavras, eles aplicaram os conhecimentos probabilísticos adquiridos durante o treinamento para o reconhecimento correto de novas palavras (Grainger et al., 2012). Esse estudo recebeu críticas, mas demonstrou que a aprendizagem e a consistência de bigramas no processamento das estruturas grafotáticas permitem a diferenciação de palavras e pseudopalavras (Bains, 2012).

Assim, tem-se dado especial atenção à pertinência dos bigramas, trigramas e suas frequências para definirem-se as estruturas grafotáticas e unidades de armazenamento para o processamento ortográfico nas diferentes línguas (Shaoul et al., 2013). O alfabeto do PB possui 26 letras (i.e., abcdefghijklmnopqrstuvwxyz), desconsiderando-se a incidência de letras com diacríticos (i.e., àáãäéêíóôõúç), podem-se formar 676 bigramas ( $26 \times 26$ ) e 17.576 trigramas ( $26 \times 26 \times 26$ ), sendo os trigramas são mais vastos e complexos que os bigramas. Assim, na formação de (pseudo)palavras, trigramas apresentam maior consistência ortográfica do que bigramas, pois cada trígama é concatenado a partir da sobreposição de duas letras (e.g., amor, bigramas: #a\_am\_mo\_or\_r#, trigramas: #am\_amo\_mor\_or#, onde o símbolo “#” significa borda da palavra). Logo, pseudopalavras construídas a partir das frequências de bigramas devem ser associadas à alta vizinhança ortográfica e pseudopalavras construídas a partir das frequências de trigramas à baixa vizinhança ortográfica (Davis, 2005).

Observa-se que os bigramas e os trigramas das palavras estão associados a duas variáveis fundamentais para o processamento ortográfico: frequências de n-gramas e vizinhança ortográfica (Grainger & Whitney, 2004). Salienta-se que as frequências de n-gramas não correspondem às frequências das palavras.

Enquanto a frequência da palavra diz respeito à frequência de recorrência da palavra inteira em um corpus específico, a frequência de n-gramas diz respeito à soma das frequências dos n-gramas que compõem uma palavra específica (Taft, 1991). No que diz respeito aos vizinhos ortográficos, os resultados apontam que palavras com alta vizinhança ortográfica possuem reconhecimento lexical com tempos de reação (RTs) mais baixos do que palavras com baixa vizinhança ortográfica, diferentemente, palavras com alta vizinhança ortográfica apresentam mais erros do que palavras com baixa vizinhança ortográfica (Andrews, 1997).

De acordo a Davis (2005), a frequência de n-gramas de uma (pseudo)palavra é calculada a partir da soma das frequências dos n-gramas que constituem a (pseudo)palavra. As frequências *type* de cada n-grama de uma língua determinada são calculadas a partir do número de palavras diferentes que possuem o n-grama em um corpus específico. Sendo assim, as frequências *type* de cada bigrama do PB são calculadas a partir do número de palavras diferentes do PB que possuem cada bigrama, e, as frequências *type* de cada trigrama do PB a partir do número de palavras do PB que possuem cada trigrama, em um corpus específico (Estivalet & Meunier, 2017). Observa-se que a mesma lógica pode ser aplicada a frequência *token* de n-gramas. Enfim, a frequência de bigramas de uma (pseudo)palavra é calculada a partir da soma das frequências dos bigramas que constituem esta (pseudo)palavra, e, a frequência de trigramas da (pseudo)palavra a partir da soma das frequências dos trigramas da (pseudo)palavra (Shaoul et al., 2013).

Por exemplo, retomando-se o exemplo acima de acordo ao LexPorBR, os bigramas da palavra “amor” possuem as seguintes frequências: #a = 23.430, am = 20.537, mo = 11.510, or = 18.521, r# = 9.428, totalizando em 83.462 a sua frequência de bigramas; as frequências de trigramas são: #am = 1.273, amo = 2.195, mor = 1.141, or# = 2.214, totalizando em 6.823 a sua frequência de trigramas (Estivalet & Meunier, 2015). Tendo em vista que existem pouco mais de 600 bigramas no PB, as frequências de bigramas são mais elevadas e menos variadas do que as frequências de trigramas que, por sua vez, são menos elevadas e mais variadas devido ao grande número de trigramas, mais de 17.000. Dessa observação, deriva-se que as estruturas grafotáticas são mais consistentes e estáveis a partir

dos trigramas que se sobrepõem em duas letras, enquanto as frequências e ativações de unidade ortográficas são mais consistentes e estáveis a partir dos bigramas, que por sua vez, se sobrepõem em apenas uma letra (Duyck et al., 2004)

No que diz respeito à direção de análise das palavras, observa-se que palavras do PB possuem o significado semântico associado ao radical na borda esquerda e os traços morfossintáticos associados aos sufixos derivacionais e flexionais na borda direita das palavras. Portanto, pseudopalavras construídas da esquerda para direita tendem a possuir maior consistência ortográfica em relação a prefixos e radicais, enquanto pseudopalavras construídas da direita para esquerda tendem a possuir maior consistência ortográfica em relação aos sufixos derivacionais e flexionais (McClelland & Rumelhart, 1981).

Em relação ao processamento ortográfico em função do número de letras, é consenso que os RTs apresentam um comportamento em forma de curva de “U”. Palavras do inglês apresentam RTs mais lentos e decrescentes para palavras até 4 letras e crescentes acima de 9 letras, palavras entre 5 e 8 letras possuem RTs mais rápidos e equivalentes (New et al., 2006). Esse comportamento é explicado por um lado pelo fato de existir um número elevado de palavras até 4 letras no inglês, resultando em um grande número de palavras semelhantes e provocando maior competitividade lexical para o acesso lexical e reconhecimento. Por outro lado, palavras com 9 ou mais letras são menos frequentes e possuem maior tempo de processamento segmental e visual. Esse mesmo comportamento pode ser predito para o PB, com a ressalva que palavras de línguas latinas são normalmente maiores, logo, deslocando a base da curva em “U” para palavras um pouco maiores (i.e., entre 6 e 10 letras).

De forma semelhante ao número de letras, a predição é que o processamento ortográfico em função do número de sílabas deve também apresentar um comportamento em forma de curva de “U”. Contudo, tendo em vista que o processamento silábico inclui aspectos estruturais e de ritmo temporal, deve haver maiores diferenças entre palavras com mais sílabas de forma gradual (Levelt & Wheeldon, 1994).

Enfim, motores geradores de pseudopalavras empregam diferentes técnicas e algoritmos para a criação de pseudopalavras de uma língua específica. Algumas

técnicas utilizadas atualmente criam pseudopalavras através dos seguintes métodos: anagramas, transposição de letras, ocorrência de sílabas, bigramas ou trigramas, frequência transicional, julgamento de aceitabilidade, entre outros. Por exemplo, o gerador de pseudopalavras multilíngue Wuggy emprega um algoritmo que gera pseudopalavras de uma língua específica (alemão, basco, espanhol, francês, holandês, inglês, servo) a partir dos padrões das estruturas silábicas e frequências transicionais de um conjunto de palavras apresentadas ao programa (Keuleers & Brysbaert, 2010). Diferentemente, o WordGen gera pseudopalavras de uma língua específica (alemão, francês, holandês, inglês) a partir da especificação de sete restrições linguísticas: a) número de letras, b) número de vizinhos ortográficos, c) frequência, d) frequência de bigramas sem posição, e) frequência mínima de bigramas sem posição, f) frequência de bigramas por posição inicial e final, e, g) transparência ortográfica (Duyck et al., 2004).

Atualmente, nenhum desses geradores automáticos de pseudopalavras satisfaz as necessidades de criação e manipulação de estímulos controlados para a utilização em experimentos psicolinguísticos do PB. Com o objetivo de preencher esta lacuna, promover e facilitar a criação e seleção de pseudopalavras do PB, Estivalet e Meunier (2015) criaram um motor de geração de pseudopalavras do PB acesso e utilização fácil, aberto e gratuito a partir dos dados gerados e coletados no corpus LexPorBR<sup>59</sup>. Esse instrumento gera pseudopalavras do PB a partir da especificação de quatro critérios: a) número de letras, b) tipo (bigramas ou trigramas), c) categorial gramatical (adjetivo, advérbio, gramatical, numeral, substantivo, verbo ou todas) e d) número de pseudopalavras.

Como resultado, o motor gera o número de pseudopalavras desejado com o número de letras especificado a partir da concatenação de bigramas ou trigramas da categoria gramatical selecionada. O motor gera as pseudopalavras a partir da superposição de bigramas (uma letra) ou trigramas (duas letras) em duas direções, esquerda-direita e direita-esquerda. Ao final, são disponibilizadas as pseudopalavras criadas acompanhadas das seguintes informações: i) categoria gramatical, ii) frequência de bigrama ou trigrama da pseudopalavra, iii) log10 da

---

<sup>59</sup> <http://www.lexicodoportugues.com/pseudowords.php>.

frequência calculada da pseudopalavra e iv) número de letras da pseudopalavra (Estivalet & Meunier, 2017). Sendo assim, um dos objetivos específicos do presente trabalho foi validar esse motor de geração de pseudopalavras do PB e a pertinência das informações computadas e fornecidas pelo mesmo no processamento de pseudopalavras do PB.

## **METODOLOGIA**

### **Participantes**

No total, 34 participantes entre 18 e 46 anos (idade média 28,4 anos, 17 mulheres) realizaram o experimento. Todos participantes eram falantes do PB como língua materna, estudantes universitários, destros, não possuíam problemas de audição, de visão (corrigidos por óculos ou lentes de contato), de linguagem e cognitivos. Eles não conheciam os objetivos da pesquisa e assinaram um termo de consentimento para participação de forma voluntária. O experimento seguiu a Declaração de Helsinki, sendo aprovado pelo Comitê de Ética de Pesquisas com Seres Humanos IRB: 00009118.

### **Materiais**

Foram criadas 720 pseudopalavras a partir das estruturas grafotáticas do PB através do motor de criação de pseudopalavras do LexPorBR (Estivalet & Meunier, 2017). Destas, 360 pseudopalavras foram criadas a partir da concatenação de bigramas (BIG) e 360 pseudopalavras a partir da concatenação de trigramas (TRI). Em cada um desses subgrupos, 180 pseudopalavras foram criadas com a concatenação de bigramas ou trigramas da esquerda para direita (ED) e 180 pseudopalavras com a concatenação de bigramas ou trigramas da direita para esquerda (DE). Em cada um destes subgrupos, foram criadas 60 pseudopalavras de alta frequência (AF), 60 pseudopalavras de média frequência (MF) e 60 pseudopalavras de baixa frequência (BF), de acordo com a soma das frequências dos bigramas ou trigramas concatenados em cada pseudopalavra (Davis, 2005). Enfim, em cada um desses subgrupos, 30 pseudopalavras foram criadas a partir

das frequências de bigramas ou trigramas de substantivos (SUB) e 30 pseudopalavras foram criadas a partir das frequências de bigramas ou trigramas de verbos (VER). Ainda, foram selecionadas 720 palavras existentes conforme Estivalet (2020) para contrabalancear as respostas positivas na tarefa de decisão lexical.

As diferentes condições experimentais das pseudopalavras foram controladas e pareadas em relação à frequência, aos vizinhos ortográficos, ao número de letras e ao número de sílabas. A TABELA 1 apresenta as médias controladas e pareadas nas diferentes condições experimentais.

TABELA 1 – Médias das frequências, vizinhos ortográficos, número de letras e número de sílabas dos estímulos controlados e pareados nas diferentes condições experimentais.

Cat. Gram.	Frequência	Estrutura	Freq. Zipf	Vizinhos	Letras	Sílabas
Substantivos	Alta	Bigrama	6,55	1,77	7,82	3,24
		Trigrama	6,77	0,13	7,20	3,56
	Média	Bigrama	4,37	1,33	7,08	3,05
		Trigrama	4,52	0,05	7,12	3,65
	Baixa	Bigrama	3,14	1,83	7,17	3,47
		Trigrama	3,28	0,08	7,26	3,62
Verbos	Alta	Bigrama	6,67	1,34	7,06	3,48
		Trigrama	6,03	0,71	7,89	3,49
	Média	Bigrama	4,48	1,11	7,93	3,80
		Trigrama	4,73	0,15	7,92	3,73
	Baixa	Bigrama	3,24	1,53	7,20	3,46
		Trigrama	3,45	0,34	7,25	3,30

## Procedimentos

O experimento foi criado, aplicado e teve os dados coletados através do programa *E-Prime v2.0 Professional* (Psychology Software Tools Inc., Sharpsburg, PA, EUA). Os participantes foram testados individualmente em uma sala silenciosa. Primeiramente, um ponto de fixação aparecia no centro da tela por 500 ms; em seguida, a (pseudo)palavra alvo era apresentada no centro da tela por 4.000 ms ou até a realização da resposta do participante; após, uma tela em branco era apresentada por 500 ms entre os estímulos; enfim, a mesma sequência recomeçava com a apresentação de outro estímulo.

Os estímulos foram apresentados no centro da tela de um computador portátil com tela LCD de 15" em letras minúsculas, brancas contra um fundo preto, fonte Courier New e tamanho 18 pt. Os participantes foram instruídos a realizarem suas respostas o mais rapidamente e corretamente possível em uma tarefa de decisão lexical entre palavras e pseudopalavras na modalidade visual. Eles utilizaram uma tecla verde do computador com uma das mãos para palavras existentes e uma tecla vermelha do computador com a outra mão para pseudopalavras inexistentes. Antes de iniciar a coleta dos dados experimentais, os participantes passaram por uma sessão de treino com 10 estímulos, houve um intervalo de um minuto no meio do experimento. No total, o experimento durou em torno de 24 minutos.

## RESULTADOS

Apenas as pseudopalavras foram analisadas. Dois participantes apresentaram taxas de erro maiores que 15% e foram retirados (5,39%). Quatro pseudopalavras (\*pompam, \*eviada, \*aberros, \*sedi) apresentaram taxas de erro superiores a 50% e foram eliminadas (0,20%). RT menores que 300 ms e maiores que 2.000 ms foram considerados fora da tarefa e retirados (1,28%). Enfim, 5,07% das respostas foram erradas e eliminadas. No total, 11,52% dos dados foram retirados da análise de RT.

Pseudopalavras 876(369) foram significativamente respondidas mais lentamente do que palavras 708(262) ( $t(31) = 8,69$ ,  $p < 0,001$ ). Através de uma análise de variância (ANOVA), foi analisado o RT como variável dependente em função de a) frequência (AF, MF e BF), b) estrutura (BIG e TRI), c) direção (ED e DE) e d) categoria gramatical (SUB e VER) como variáveis independentes fatoriais, com interações entre duas variáveis.

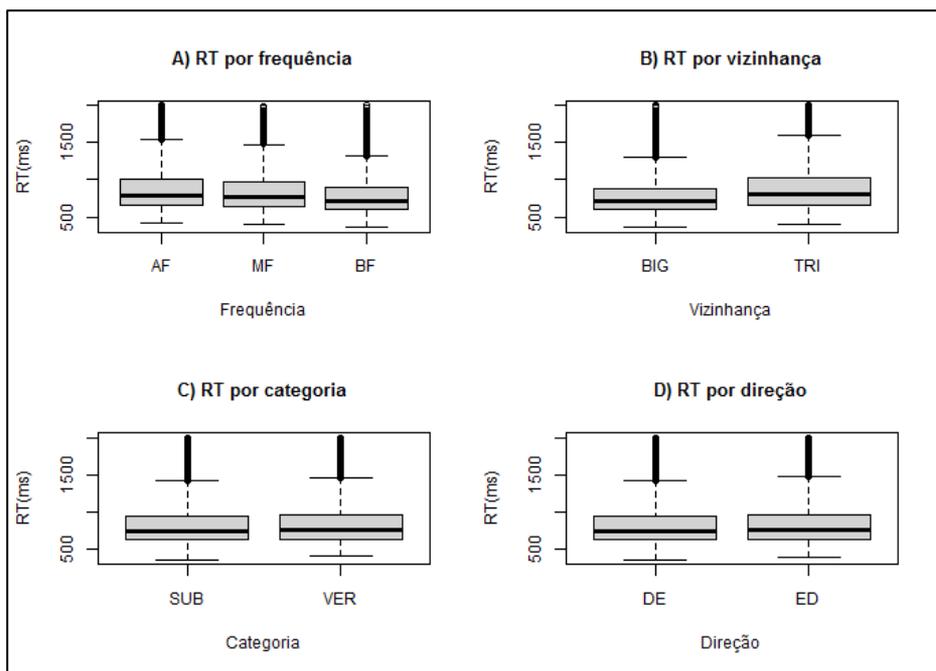
Os resultados apontaram efeitos principais significativos de frequência ( $F(2,62) = 107,21$ ,  $p < 0,001$ ), estrutura ( $F(1,31) = 144,10$ ,  $p < 0,001$ ) e categoria ( $F(1,31) = 4,51$ ,  $p < 0,05$ ), assim como um efeito significativo de interação entre frequência e categoria ( $F(2,62) = 9,11$ ,  $p < 0,001$ ). A análise *post-hoc* desta interação indicou

uma diferença significativa entre as categorias gramaticais somente em pseudopalavras de baixa frequência: AF ( $t(31) = 0,97$ ,  $p = 0,34$ ), MF ( $t(31) = 0,004$ ,  $p = 0,99$ ), BF ( $t(31) = 4,64$ ,  $p < 0,001$ ). Sendo assim, a variável direção não apresentou nenhuma pertinência em relação aos RTs e foi desconsiderada para apresentação das médias dos RTs, desvios-padrão e taxas de erro na TABELA 2 e na FIGURA 1.

TABELA 2 – Médias dos RTs, desvios-padrão e taxas de erro nas diferentes condições experimentais.

Cat. Gram.	Frequência Estrutura	Baixa		Média		Alta	
		RT(ms)	Erro(%)	RT(ms)	Erro(%)	RT(ms)	Erro(%)
Substantivos	Bigramas	798(251)	0,31	773(243)	0,19	729(213)	0,36
	Trigramas	931(328)	0,13	907(308)	0,21	805(273)	0,12
Verbos	Bigramas	829(269)	0,50	787(247)	0,30	752(234)	0,45
	Trigramas	930(312)	0,73	891(291)	0,39	861(280)	0,44

FIGURA 1 - RTs em função de A) frequência (AF: alta frequência, MF: média frequência, BF: baixa frequência), B) vizinhança ortográfica (BIGgramas: alta vizinhança, TRigramas: baixa vizinhança), C) categorias gramaticais (SUBstantivos e VERbos) e D) direção de formação (ED: esquerda para direita, DE: direita para esquerda).

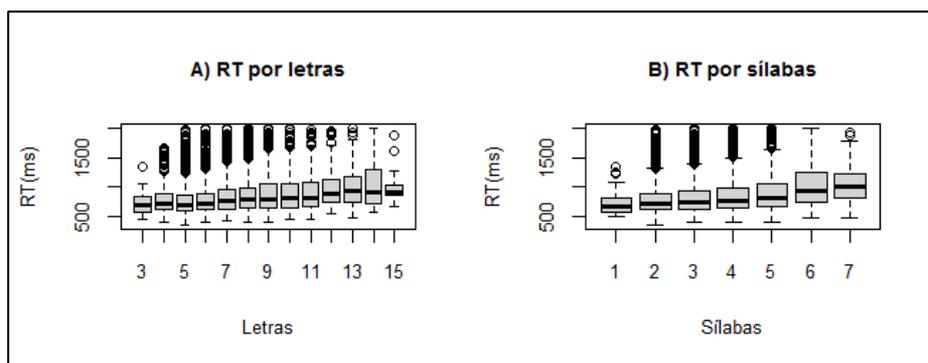


Em relação às taxas de erro, pseudopalavras 2,06% apresentaram significativamente mais erros que palavras 1,79% ( $z(31) = 3,23$ ,  $p < 0,01$ ). Através de uma análise de desvio de qui-quadrado binomial ( $\chi^2$ ), foi analisado o número de erros como variável dependente em função das mesmas variáveis independentes acima. Os resultados apontaram efeitos principais significativos de frequência ( $\chi^2(2) = 13,92$ ,  $p < 0,001$ ), categoria gramatical ( $\chi^2(1) = 62,06$ ,  $p < 0,001$ ) e direção ( $\chi^2(1) = 9,83$ ,  $p < 0,01$ ), assim como efeitos significativos de interação entre frequência e estrutura ( $\chi^2(2) = 6,45$ ,  $p < 0,05$ ) e entre estrutura e categoria gramatical ( $\chi^2(2) = 17,69$ ,  $p < 0,001$ ). A análise *post-hoc* da interação entre frequência e estrutura indicou uma diferença significativa entre as estruturas somente em BF: AF ( $z(31) = 0,55$ ,  $p = 0,59$ ), MF ( $z(31) = 1,09$ ,  $p = 0,28$ ) e BF ( $z(31) = -3,35$ ,  $p < 0,01$ ); e, na interação entre estrutura e categoria gramatical em SUB ( $z(31) = 3,59$ ,  $p < 0,05$ ) e VER ( $z(31) = -2,24$ ,  $p < 0,05$ ).

Em seguida, através de uma análise estatística de regressão linear multifatorial, foi analisado o RT como variável dependente em função do e) número de letras, f) número de sílabas, g) frequência e h) vizinhos ortográficos como variáveis

independentes contínuas. Os resultados apontaram que as quatro variáveis explicam 24% da variação dos dados ( $R^2 = 0,24$ ,  $F(4,710) = 57,02$ ,  $p < 0,001$ ); o número de letras é o fator mais preditivo ( $\beta = 32,63$ ,  $t(710) = 8,63$ ,  $p < 0,001$ ), seguido da vizinhança ortográfica ( $\beta = 10,73$ ,  $t(710) = 5,86$ ,  $p < 0,001$ ), frequência ( $\beta = 4,06$ ,  $t(710) = 2,38$ ,  $p < 0,05$ ) e número de sílabas ( $\beta = -3,81$ ,  $t(710) = -0,51$ ,  $p = 0,61$ ). As distribuições e os comportamentos dos RTs em função do número de letras e do número de sílabas são apresentados na FIGURA 2.

FIGURA 2 – RTs em função do A) número de letras e do B) número de sílabas.



## DISCUSSÃO

A presente pesquisa permitiu observar que as frequências dos n-gramas das (pseudo)palavras do PB influenciam diretamente o comportamento do reconhecimento de (pseudo)palavras em relação ao processamento ortográfico. Enquanto bigramas apresentam a co-ocorrência de apenas uma letra, sugerindo uma menor variabilidade de unidades e menor consistência nas estruturas grafotáticas, trigramas apresentam a co-ocorrência de duas letras e uma maior variabilidade das unidades e maior consistência das estruturas grafotáticas recorrentes do PB. Logo, fica clara a pertinência das frequências de bigramas e de trigramas para uma melhor compreensão das estruturas grafotáticas do PB, assim como para a seleção, o controle e a manipulação dessas variáveis em pesquisas

psicolinguísticas sobre o processamento ortográfico no PB (Estivalet & Meunier, 2015).

## **VARIÁVEIS FATORIAIS**

Pseudopalavras criadas a partir de bigramas possuem alta vizinhança ortográfica e pseudopalavras criadas a partir de trigramas possuem baixa vizinhança ortográfica. Os resultados apresentaram um efeito principal de estrutura com pseudopalavras com alta vizinhança (BIG) sendo respondidas em média 109 ms mais rapidamente do que pseudopalavras com baixa vizinhança (TRI) (Davis & Bowers, 2006). Além disso, foi encontrada uma diferença significativa de estrutura nas taxas de erro somente em baixa frequência, indicando uma maior interferência da competição lexical em pseudopalavras do PB que apresentam uma estrutura grafotática pouco frequente. Esses resultados vão de encontro aos resultados de palavras existentes que propõem que uma alta vizinhança ortográfica (BIG) facilita o acesso lexical (Andrews, 1997), e no caso do presente experimento, facilita a rejeição da pseudopalavra.

Em seguida, pseudopalavras de AF foram respondidas em média 32 ms mais rapidamente do que pseudopalavras de MF, e, pseudopalavras de MF foram respondidas em média 52 ms mais rápido do que pseudopalavras de BF (Monsell et al., 1989). Portanto, esses resultados corroboram a hipótese de que a frequência de bigramas ou trigramas das (pseudo)palavras calculada a partir da soma dos bigramas ou trigramas que compõem as (pseudo)palavras é um fator preditivo para o processamento ortográfico (Davis, 2005).

De forma interessante, houve uma interação significativa entre categoria gramatical e frequência, apontando uma diferença de 38 ms entre pseudopalavras formadas a partir de n-gramas de SUB mais rápidas do que VER, somente em BF. A predição do presente trabalho era justamente que SUB seriam respondidos mais rapidamente que VER devido ao maior processamento dos traços morfossintáticos de VER. Isso é, independente do reconhecimento e acesso semântico da raiz, os bigramas ou trigramas finais dos VER são consistentes com

os sufixos flexionais verbais, induzindo a um maior tempo de processamento desses traços morfossintáticos para a rejeição dessas pseudopalavras (Taft, 1991).

No que diz respeito à direção de formação das pseudopalavras (ED ou DE), esta variável não apresentou resultados significativos, sendo pseudopalavras formadas da ED respondidas 12 ms mais rapidamente do que pseudopalavras formadas da DE. Assim, aparentemente, tanto palavras formadas da ED como da DE acabam por apresentar a mesma consistência grafotática.

## **VARIÁVEIS CONTÍNUAS**

De acordo aos resultados de New et al. (2006) para o inglês, o processamento das pseudopalavras do PB em função do número de letras também apresentou médias de RTs mais rápidos em pseudopalavras de 5 (755 ms) letras, com RTs mais elevados para palavras de 4 letras (775 ms) e 6 letras (789 ms) e assim sucessivamente para palavras maiores. Diferentemente de nossa predição, esses resultados não foram maiores em função das características das palavras do PB, mas parecem estar associados diretamente às características do processamento ortográfico (Grainger & Whitney, 2004). Ainda, os resultados não apresentaram exatamente uma curva em “U”, mas sim em “V” com RTs decaindo linearmente em pseudopalavras entre 3-5 e aumentando linearmente em pseudopalavras de 5-13 letras (Estivalet, 2020).

A variável número de sílabas não apresentou o mesmo comportamento que a variável número de letras, refutando a nossa hipótese em relação ao processamento temporal silábico (Levelt & Wheeldon, 1994). Apesar de haver um aumento linear dos RTs em relação ao aumento do número de sílabas (1-6 sílabas), as médias desses aumentos não foram significativas devido a grande variância dos resultados. Uma possível interpretação para esses resultados não significantes em relação ao processamento do número de sílabas é que o processamento silábico é referente ao processamento fonológico temporal (McClelland & Rumelhart, 1981). Logo, tendo em vista que o presente experimento explorou o processamento ortográfico das pseudopalavras na modalidade visual, independentemente da representação lexical e conceitual, as

unidades fonológicas temporais não apresentam contribuição significativa em relação ao processamento ortográfico precoce, mas apenas em etapas mais tardias do processamento linguístico (Besner & Johnston, 1992).

Enfim, as variáveis contínuas frequência e vizinhança ortográfica também apresentaram significância preditiva em relação ao processamento das pseudopalavras do PB. Nesse sentido, a contribuição da vizinhança ortográfica vai de encontro aos resultados de Andrews (1997). Por um lado, (pseudo)palavras com mais vizinhos ortográficos (BIG) e estrutura grafotática mais frequente são rejeitadas mais rapidamente, por outro lado, (pseudo)palavras com menos vizinhos ortográficos (TRI) e estrutura grafotática mais consistente são rejeitadas mais lentamente em função da competição lexical, pois se parecem mais com palavras existentes.

### **(PSEUDO)PALAVRAS DO PB**

As pseudopalavras foram rejeitadas 168 ms mais lentamente do que as palavras foram reconhecidas, corroborando um dos principais efeitos encontrados na psicolinguística experimental. A hipótese de base que sustenta esse efeito é que as palavras são organizadas no léxico mental em uma espécie de lista de acordo com a sua frequência, palavras mais frequentes estão no topo da lista e são reconhecidas mais rapidamente. Assim, para a rejeição de pseudopalavras, é necessário percorrer-se toda a lista de palavras do léxico mental, resultando em RTs mais lentos (Taft, 1991). Ainda, pseudopalavras apresentaram mais erros do que as palavras devido à premissa de que palavras são reconhecidas pela consistência de por sua forma ortográfica/fonológica, mas principalmente a partir do acesso lexical e representação semântica, diferentemente, as pseudopalavras possuam estrutura grafotática semelhante às palavras existentes, sendo confundidas durante o rápido e automático processamento ortográfico com possíveis palavras (Grainger & Whitney, 2004).

Os resultados estatisticamente significativos encontrados no presente estudo em função das variáveis pesquisadas – frequência de bigrama ou trigrama, vizinhança ortográfica e tamanho em letras das pseudopalavras – permitem confirmar o bom

funcionamento e validar o motor de geração de pseudopalavras do PB do LexPorBR (Estivalet & Meunier, 2015). Os estímulos experimentais geradas por essa ferramenta apresentaram resultados interessantes, pertinentes e significativos em relação à criação de pseudopalavras a partir das frequências de bigramas e trigramas e da consistência grafotática das (pseudo)palavras do PB (Duyck et al., 2004). Nesse sentido, destaca-se que a construção de pseudopalavras a partir de bigramas ou de trigramas está diretamente associada à manipulação de vizinhança ortográfica, pois quanto menor o n-grama (bigramas), menor a consistência estrutural e, portanto, maior a frequência de suas unidades, assemelhando-se a palavras de alta vizinhança; quanto maior o n-grama (trigramas), maior a consistência grafotática e menor a frequência de suas unidades, assemelhando-se a palavras de baixa vizinhança (Shaoul et al., 2013).

Ou seja, por um lado, a partir de concatenação de bigramas (i.e., sobreposição de uma letra), obtêm-se pseudopalavras que possuem altas frequências de constituição grafotática, por outro lado, a partir da concatenação de trigramas (i.e., sobreposição de duas letras) obtêm-se pseudopalavras que possuem alta consistência grafotática. Portanto, enquanto a frequência de bigramas apresenta uma boa predição em relação às letras subsequentes, a frequência de trigrama apresenta uma boa predição em relação à consistência grafotática. Outras investigações com o foco específico nestas variáveis, assim como em unigramas e quadrigrams, devem ser realizados para uma melhor compreensão desta variável.

Portanto, as variáveis analisadas no presente estudo (i.e., número de letras, frequência de bigramas/trigramas e vizinhança ortográfica) estão diretamente associadas ao processamento precoce visual e ortográfico das (pseudo)palavras do PB, tendo em vista que apresentaram diferenças significativas nos resultados em função de suas manipulações e independente da existência de uma representação lexical e semântica no léxico mental. Logo, a sugestão é que essas variáveis sejam observadas, controladas e manipuladas em futuros estudos psicolinguísticos sobre o processamento visual e ortográfico, acesso lexical e reconhecimento de palavras.

## CONSIDERAÇÕES FINAIS

Através da presente pesquisa, ficou claro que o reconhecimento das palavras durante a leitura é composto por pelo menos dois processos distintos: i) processamento precoce visual ortográfico das letras que compõem as (pseudo)palavras (Grainger et al., 2012) e ii) processamento linguístico tardio do acesso lexical e semântico-conceptual para o reconhecimento/rejeição das (pseudo)palavras (Monsell et al., 1989). Nesse sentido, os presentes resultados nos revelaram uma série de resultados comportamentais interessantes e pertinentes em relação a unidades maiores que letras, os bigramas e os trigramas, na primeira etapa do reconhecimento/rejeição de (pseudo)palavras, ou seja, no processamento ortográfico. Mais do que isso, os resultados do presente estudo permitiram uma melhor compreensão do processamento ortográfico como uma etapa posterior ao processamento visual (Vinckier et al., 2011) e anterior ao processamento linguístico para reconhecimento, sugerindo um modelo em que a) o tamanho das palavras em letras (i.e., ortográfico), b) a similaridade formal (i.e., vizinhança ortográfica) e c) a frequência/consistência grafotática (i.e., bigramas e trigramas), são variáveis pertinentes unicamente a etapa do processamento ortográfico (Besner & Johnston, 1992).

Conforme Davis e Bowers (2006), existem diversas hipóteses em relação ao processamento ortográfico de acordo às posições das letras e o entendimento desses mecanismos é de fundamental importância para uma melhor compreensão da organização e das representações do léxico mental, assim como do processo de acesso lexical e reconhecimento das palavras. Diferentemente da proposta de Grainger e Whitney (2004), os resultados desta investigação não apontam que as palavras são lidas de forma inteira, mas que existe um processamento serial dos n-gramas que compõem as pseudopalavras de acordo com suas frequências e consistência grafotáticas (Taft, 1991). Logo, sugere-se que é justamente a consistência dessas estruturas grafotáticas, traduzidas por suas frequências de unidades de n-gramas, que ofereçam a forma ortográfica/fonética para etapas posteriores do processamento linguístico e acesso lexical. Enfim, é somente nessas etapas tardias que variáveis como frequência da palavras,

número de sílabas, competição lexical e complexidades morfossintática e semântica atuam no acesso lexical para o reconhecimento das palavras.

Enfim, este estudo permitiu validar de forma experimental a pertinência das variáveis e o funcionamento do motor de geração de pseudopalavras do PB do LexPorBR (Estivalet & Meunier, 2017). Sendo assim, sugere-se que futuros estudos em psicolinguística do PB considerem e incorporem a seleção, o controle e a manipulação das variáveis discutidas aqui, assim como tenham como objetivo o aprofundamento da compreensão do processamento ortográfico e linguístico para o reconhecimento das palavras.

## AGRADECIMENTOS

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de pesquisa para realização deste estudo. Agradeço aos pesquisadores F.M. e M.H. pelas valiosas discussões acerca do processamento visual e ortográfico no reconhecimento de (pseudo)palavras. Agradeço a J.F.P pela leitura atenta de uma versão preliminar deste trabalho. Enfim, agradeço à comissão organizadora do XIX Congresso da Associação de Linguística e Filologia da América Latina pela organização do evento.

## REFERÊNCIAS

- ANDREWS, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4(4), 439–461. Disponível em: <https://doi.org/10.3758/BF03214334>. Acesso em: 21 jun. 2021.
- BAINS, W. (2012). Comment on Orthographic Processing in Baboons (Papio papio). *Science*, 337(6099), 1173–1173. Disponível em: <https://doi.org/10.1126/science.1224508>. Acesso em: 21 jun. 2021.

BESNER, D.; JOHNSTON, J. C. (1992). Reading and the mental lexicon: on the uptake of visual information. In W. D. Marslen-Wilson (ed.). **Lexical Representation and Process**, p. 291–316. The MIT Press.

DAVIS, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. **Behavior Research Methods**, 37(1), 65–70. Disponível em: <https://doi.org/10.3758/BF03206399>. Acesso em: 21 jun. 2021.

DAVIS, C. J.; BOWERS, J. S. (2006). Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. **Journal of Experimental Psychology: Human Perception and Performance**, 32(3), 535–557. Disponível em: <https://doi.org/10.1037/0096-1523.32.3.535>. Acesso em: 21 jun. 2021.

DUYCK, W.; DESMET, T.; VERBEKE, L. P. C.; BRYBAERT, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. **Behavior Research Methods, Instruments, & Computers**, 36(3), 488–499. Disponível em: <https://doi.org/10.3758/BF03195595>. Acesso em: 21 jun. 2021.

ESTIVALET, G. L. (2020). Variáveis lexicais e ortográficas no acesso lexical das palavras do português brasileiro. **Revista Linguística**, 16(1), 264–277. Disponível em: <https://doi.org/10.31513/linguistica.2020.v16n1a31539>. Acesso em: 21 jun. 2021.

ESTIVALET, G. L.; MEUNIER, F. (2015). The Brazilian Portuguese Lexicon: An Instrument for Psycholinguistic Research. **PLOS ONE**, 10(12), e0144016. Disponível em: <https://doi.org/10.1371/journal.pone.0144016>. Acesso em: 21 jun. 2021.

ESTIVALET, G. L.; MEUNIER, F. (2017). Corpus psicolinguístico Léxico do Português Brasileiro. **REVISTA SOLETRAS**, 33(1), 212–229. Disponível em: <https://doi.org/10.12957/soletras.2017.29702>. Acesso em: 21 jun. 2021.

GRAINGER, J.; DUFAU, S.; MONTANT, M.; ZIEGLER, J. C.; FAGOT, J. (2012). Orthographic Processing in Baboons (*Papio papio*). **Science**, 336(6078), 245–248. Disponível em: <https://doi.org/10.1126/science.1218152>. Acesso em: 21 jun. 2021.

GRAINGER, J.; WHITNEY, C. (2004). Does the human mind read words as a whole? **Trends in Cognitive Sciences**, 8(2), 58–59. Disponível em: <https://doi.org/10.1016/j.tics.2003.11.006>. Acesso em: 21 jun. 2021.

KEULEERS, E.; BRYBAERT, M. (2010). Wuggy: A multilingual pseudoword generator. **Behavior Research Methods**, 42(3), 627–633. Disponível em: <https://doi.org/10.3758/BRM.42.3.627>. Acesso em: 21 jun. 2021.

LEVELT, W. J. M.; WHELDON, L. (1994). Do speakers have access to a mental syllabary? **Cognition**, 50(1–3), 239–269. Disponível em: [https://doi.org/10.1016/0010-0277\(94\)90030-2](https://doi.org/10.1016/0010-0277(94)90030-2). Acesso em: 21 jun. 2021.

MCCLELLAND, J. L.; RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. **Psychological Review**, 88(5), 375–407. Disponível em: <https://doi.org/10.1037/0033-295X.88.5.375>. Acesso em: 21 jun. 2021.

Monson, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? **Journal of Experimental Psychology: General**, 118(1), 43–71. Disponível em: <https://doi.org/10.1037/0096-3445.118.1.43>. Acesso em: 21 jun. 2021.

NEW, B.; FERRAND, L.; PALLIER, C.; BRYBAERT, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. **Psychonomic Bulletin & Review**, 13(1), 45–52. Disponível em: <https://doi.org/10.3758/BF03193811>. Acesso em: 21 jun. 2021.

SHAUL, C.; WESTBURY, C. F.; BAAYEN, R. H. (2013). The subjective frequency of word n-grams. **Psihologija**, 46(4), 497–537. Disponível em: <https://doi.org/10.2298/PSI1304497S>. Acesso em: 21 jun. 2021.

TAFT, M. (1991). **Reading and the Mental Lexicon**. Lawrence Erlbaum Associates Ltd.

VINCKIER, F.; QIAO, E.; PALLIER, C.; DEHAENE, S.; COHEN, L. (2011). The impact of letter spacing on reading: A test of the bigram coding hypothesis. **Journal of Vision**, 11(6), 8–8. Disponível em: <https://doi.org/10.1167/11.6.8>. Acesso em: 21 jun. 2021.