

VARIÁVEIS LEXICAIS E ORTOGRÁFICAS NO ACESSO LEXICAL DAS PALAVRAS DO PORTUGUÊS BRASILEIRO

LEXICAL AND ORTHOGRAPHIC VARIABLES IN THE LEXICAL ACCESS OF WORDS IN BRAZILIAN PORTUGUESE

Gustavo Lopez Estivalet¹

RESUMO

Quais variáveis influenciam o acesso lexical no português brasileiro? O presente trabalho teve como objetivo analisar a influência das principais variáveis lexicais e ortográficas no acesso lexical: a. frequência lexical, b. vizinhança ortográfica, c. categoria gramatical, d. número de letras e e. número de sílabas. Modelos de leitura de via dupla postulam diferentes rotas para o processamento lexical e ortográfico, nos quais essas variáveis influenciam diferentemente as etapas do processamento das palavras. Portanto, para uma melhor compreensão do léxico mental, as influências dessas variáveis devem ser analisadas no português brasileiro. Para tanto, aplicou-se um experimento psicolinguístico com tarefa de decisão lexical na modalidade visual entre palavras e pseudopalavras selecionadas e criadas a partir do Léxico do Português Brasileiro. Os resultados apontaram as seguintes diferenças significativas: palavras foram reconhecidas mais rapidamente do que pseudopalavras; palavras de alta frequência foram reconhecidas mais rapidamente do que palavras de média frequência, que por sua vez, foram reconhecidas mais rapidamente do que palavras de baixa frequência; palavras com alta vizinhança ortográfica foram reconhecidas mais rapidamente do que palavras de baixa vizinhança. Esses resultados sugerem que essas variáveis influenciam o processamento das palavras no acesso lexical. Ainda, os resultados das variáveis número de letras e número de sílabas indicam efeitos graduais no reconhecimento das palavras. Assim, a presente investigação contribui para uma melhor compreensão dos processos de leitura durante o acesso lexical, assim como valida a seleção, criação e controle de estímulos com o uso de corpora na pesquisa em psicolinguística experimental no português brasileiro.

Palavras-chave: Psicolinguística; Acesso Lexical; Reconhecimento de Palavras; Léxico Mental; Decisão Lexical.

ABSTRACT

Which variables influence lexical access in Brazilian Portuguese? The present work aimed to analyze the influence of the main lexical and orthographic variables on lexical access: a. lexical frequency, b. orthographic neighborhood, c. grammatical category, d. number of letters, and e. number of syllables. Dual-route models of reading postulate different routes for lexical and orthographic processing, in which these variables differently influence the stages of word processing. Therefore, for a better understanding of the mental lexicon, the influences of these variables should be analyzed in Brazilian Portuguese. To this end, it was applied a psycholinguistic experiment with a lexical decision task in visual modality between words and pseudowords selected and created from the Brazilian Portuguese Lexicon. The results showed the following significant differences: words were recognized faster than pseudowords; high-frequency words were recognized faster than medium-frequency words, which were recognized faster than low-frequency words; high orthographic neighborhood words were recognized faster than low orthographic neighborhood words. These results suggest that these variables influence word processing in lexical access. Furthermore, the results of the variables number of letters and number of syllables indicate gradual effects on word recognition. Thus, the present investigation contributes to a better understanding of reading processes during lexical access, as well as, validate the selection, creation, and control of stimuli with the use of corpora in experimental psycholinguistic research in Brazilian Portuguese.

Keywords: Psycholinguistics; Lexical Access; Word Recognition; Mental Lexicon; Lexical Decision.

¹ Departamento de Letras Estrangeiras Modernas, Universidade Federal da Paraíba (UFPB) - Laboratório de Processamento Linguístico (LAPROL). Email: gustavoestivalet@hotmail.com.

Introdução

Os avanços das pesquisas em psicolinguística têm demonstrado que uma série de variáveis influencia diferentemente o processamento lexical e ortográfico durante o reconhecimento das palavras e seu acesso lexical. Trabalhos com metodologias *online* em palavras isoladas do inglês, utilizando tarefas simples de decisão lexical, têm apontado principalmente as seguintes variáveis: frequência lexical (MONSELL; DOYLE; HAGGARD, 1989), vizinhança ortográfica (ANDREWS, 1997) e número de letras (NEW *et al.*, 2006).

Contudo, a maior parte desses trabalhos tem sido realizada em inglês, holandês, francês, alemão e espanhol (MARIAN *et al.*, 2012), mas poucas pesquisas são feitas no português, e, menos ainda, no português brasileiro (PB). Possivelmente, isso se deve à escassez de recursos metodológicos disponíveis e à falta de publicações em psicolinguística experimental relacionadas ao léxico mental e acesso lexical no PB (ESTIVALET; MEUNIER, 2017). Sendo assim, faz-se necessário o aprofundamento da investigação empírica dessas variáveis no reconhecimento de palavras no PB.

Destaca-se que o efeito de frequência é um dos principais fenômenos encontrados, replicados e discutidos nos estudos empíricos do léxico mental. Igualmente, os efeitos de tamanho de palavras (número de letras e número de sílabas), assim como vizinhança ortográfica, são efeitos recorrentes em psicolinguística, que por sua vez, têm sido explicados através de diferentes modelos empíricos (TAFT, 1991). Enquanto modelos seriais predizem esses efeitos independentes em diferentes etapas do processamento lexical (BESNER; JOHNSTON, 1992), modelos interativos predizem a saliência desses efeitos a partir das interações dos diferentes nós de processamento (RUMELHART; MCCLELLAND, 1982).

Monsell *et al.*, (1989) apresentaram um estudo com experimentos e simulações, evidenciando que os efeitos de frequência podem acontecer em diferentes etapas da tarefa de decisão lexical: i. decodificação ortográfica, ii. identificação lexical e iii. processo de decisão. Taft (1991) revisou uma série de experimentos em inglês com manipulação de frequências de lexema e lema. O autor propôs um modelo de decomposição obrigatória em que o acesso lexical é guiado por ambas frequências, sendo a frequência de lexema interpretada como a frequência que os radicais se combinam com os sufixos. O modelo de decomposição obrigatória é composto por três etapas: i. decomposição, ii. ativação e iii. recombinação. Na primeira etapa, palavras morfologicamente complexas são decompostas a partir de evidências estruturais e de saliência; em seguida, os morfemas são procurados no léxico mental e seus traços lexicais, semânticos e morfossintáticos são ativados; enfim, os morfemas são

recombinados para verificação da existência da palavra. Portanto, efeitos de frequência de lema acontecem na primeira e na segunda etapa durante a decomposição, procura e ativação dos morfemas no léxico mental, enquanto os efeitos de frequência de lexema acontecem somente na terceira etapa durante a verificação da recombinação dos morfemas e existência da palavra.

New *et al.* (2006) mostraram que os efeitos de tamanho de palavras também são pertinentes para o acesso lexical, os autores demonstraram que os tempos de reconhecimento lexical de palavras do inglês obedecem uma curva em U: palavras de 3 e 4 letras apresentam tempos maiores, palavras de 5 a 8 letras apresentam os menores tempos (sendo 7 letras o menor tempo) e palavras de 9 a 13 apresentam tempos maiores. Andrews (1997) demonstrou através de uma série de experimentos a estreita relação entre vizinhança ortográfica e frequência lexical. Palavras com muitos vizinhos apresentam tempos de reconhecimento mais rápidos do que palavras com poucos vizinhos, pois apesar da concorrência lexical, a ativação de outras palavras semelhantes é um forte indício da existência da palavra alvo.

Com o objetivo de contemplar estes diferentes achados empíricos sobre o processamento de palavras e acesso lexical, Besner e Johnson (1992) propuseram um modelo complexo de reconhecimento de palavras de via dupla. Na via lexical, as palavras são reconhecidas diretamente através da familiaridade visual; já na via de análise visual, as palavras são processadas através de diferentes componentes: i. análise ortográfica, ii. silabação, iii. similaridade ortográfica/fonológica, iv. frequência lexical e v. ativação de traços morfossintáticos.

Neste sentido, o presente trabalho teve como objetivo principal investigar a influência dessas três variáveis (frequência lexical, vizinhança ortográfica e categoria gramatical) no processamento de palavras do PB. Como objetivos secundários, exploraram-se outras duas variáveis pertinentes: número de sílabas (MEHLER *et al.*, 1981) e categoria gramatical (CARAMAZZA; HILLIS, 1991). Portanto, esta pesquisa justifica-se a partir de dois aspectos: i. investigação das variáveis lexicais e ortográficas no processamento das palavras do PB, e, ii. verificação das normas de frequência e vizinhança ortográfica do *corpus* Léxico do Português Brasileiro (LexPorBR)² (ESTIVALET; MEUNIER, 2015).

2 O Léxico do Português Brasileiro (LexPorBR) é um léxico baseado em palavras construído com o objetivo de auxiliar e guiar a seleção, o controle e a criação de palavras e pseudopalavras do português brasileiro para o desenvolvimento de experimentos psicolinguísticos. Ele foi elaborado a partir do corpus do NILC de textos escritos e contém mais de 42 milhões de palavras *token*, 215 mil palavras *type* e 21 categorias de informações lexicais e gramaticais (ESTIVALET; MEUNIER, 2015; 2017). O LexPorBR pode ser acessado, consultado e baixado gratuitamente no site: <http://www.lexicodoportugues.com>.

A hipótese de base é que somente as duas primeiras variáveis citadas acima (i.e., frequência lexical e vizinhança ortográfica) influenciam o processamento e reconhecimento de palavras escritas do PB (TAFT, 1991). Diferentemente, a hipótese da presente investigação é que todas as cinco diferentes variáveis lexicais e ortográficas influenciam diferentemente o processamento e reconhecimento das palavras (BAAYEN, 2008). A hipótese alternativa é que interações entre as informações de frequência, ortográficas e semânticas definem o processamento e reconhecimento das palavras (RUMELHART; MCCLELLAND, 1982).

Para tanto, foi aplicado um experimento psicolinguístico *online* com uma tarefa de decisão lexical com 720 palavras selecionadas a partir do *corpus* LexPorBR (ESTIVALET; MEUNIER, 2015). Os resultados apresentaram diferenças significativas entre palavras de alta, média e baixa frequência (MONSELL; DOYLE; HAGGARD, 1989) e palavras com alta e baixa vizinhança ortográfica (ANDREWS, 1997). Ainda, foram encontradas diferenças significativas em relação ao número de letras (NEW et al., 2006) e número de sílabas (MEHLER *et al.*, 1981). Esses resultados são discutidos à luz de modelos psicolinguísticos de reconhecimento das palavras, sugerindo evidências a favor de modelos que consideram a complexidade e influência dessas diferentes variáveis no sistema de processamento e reconhecimento das palavras no acesso lexical (BESNER; JOHNSTON, 1992).

2. Metodologia de pesquisa

2.1 Participantes

Foram testados 24 participantes (12 mulheres) falantes do PB como língua materna entre 18 e 42 anos (idade média 27,1 anos). Todos participantes eram estudantes de cursos universitários, destros, possuíam audição normal, visão normal ou corrigida por óculos/lentes de contato e não possuíam histórico de problemas cognitivos ou de linguagem. Os participantes não sabiam os objetivos do estudo e assinaram um termo de consentimento para participação no experimento como voluntários. O experimento foi conduzido de acordo com a Declaração de Helsinki (WORLD MEDICAL ASSOCIATION, 2013), tendo o protocolo aprovado pelo Comitê de Ética de Pesquisas com Seres Humanos IRB: 00009118.

2.2 Materiais

Os participantes realizaram um experimento com uma tarefa de decisão lexical entre palavras e pseudopalavras na modalidade visual. Foram investigadas três variáveis fatoriais: a. frequência lexical

(alta, média e baixa), b. vizinhança ortográfica (alta e baixa), c. categoria gramatical (substantivos e verbos); e, duas variáveis numéricas: d. número de letras (3-15) e e. número de sílabas (1-6). A Tabela 1 apresenta exemplos de estímulos experimentais.

Tabela 1: Exemplos de palavras nas condições experimentais..

Cat. Gram.	Viz./Freq.	Alta frequência	Média frequência	Baixa frequência
Substantivos	Alta	campanha	barcos	abrigo
	Baixa	juros	beijos	fico
Verbos	Alta	acabei	somaram	deram
	Baixa	ocupa	exercida	pede

Fonte: elaboração própria.

Foram selecionadas 720 palavras do PB como estímulos experimentais, 360 substantivos e 360 verbos. Dentro de cada categoria gramatical, 180 palavras foram de alta vizinhança ortográfica (>4) e 180 palavras de baixa vizinhança ortográfica (<4), de acordo com o N de Coltheart (COLTHEART *et al.*, 1977; YARKONI; BALOTA; YAP, 2008). Dentro de cada conjunto de vizinhança ortográfica, 60 palavras foram de alta frequência (>4), 60 palavras de média frequência (>3 e <4) e 60 palavras de baixa frequência (<3), de acordo com a escala *zipf* (VAN HEUVEN *et al.*, 2014). As variáveis investigadas (frequência lexical, vizinhança ortográfica, categoria gramatical, número de letras e número de sílabas) foram retiradas do léxico LexPorBR (ESTIVALET; MEUNIER, 2015). Destaca-se que essas métricas foram contabilizadas a partir do Corpus do NILC³ (Núcleo Interinstitucional de Linguística Computacional) contendo mais de 42 milhões de palavras *token* e 215 mil palavras *type* (PINHEIRO; ALUÍSIO, 2003). A escala *zipf* é uma linearização da distribuição das frequências lexicais que pode ser comparada entre diferentes *corpora* (VAN HEUVEN *et al.*, 2014); o número de letras foi calculado a partir do número de caracteres das palavras; o número de sílabas foi calculado a partir da aplicação do silabador *Petrus* do NILC (MARQUIAFAVEL; BOKAN; ZAVAGLIA, 2014); e as normas de vizinhança ortográfica foram calculadas a partir do pacote ‘vwr’ (KEULEERS, 2013). Todos estímulos experimentais foram selecionados e controlados em i. frequência lexical, ii. número de letras, iii. número de sílabas, iv. vizinhança ortográfica e v. OLD20, conforme a Tabela 2. As 720 pseudopalavras foram criadas a partir do motor de geração de pseudopalavras do *corpus* LexPorBR (ESTIVALET; MEUNIER, 2017).

³ O corpus do NILC (Núcleo Interinstitucional de Linguística Computacional) é formado por textos escritos, contendo diversos gêneros textuais, conforme a seguinte distribuição do número de palavras *token*: texto didático 182.679, enciclopédia 286.317, ensaio 2.191.189, texto jornalístico 30.179.462, texto legal 1.067.575, literário 923.133, revista 153.78 (PINHEIRO; ALUÍSIO, 2003). O corpus do NILC pode ser acessado, consultado e baixado nos sites: <https://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>.

Tabela 2: Médias das características lexicais das palavras nas condições experimentais.

Cat.	Frequência	Vizinhança	Frequência	Letras	Sílabas	Vizinhos	OLD20
Substantivos	Alta	Alta	4,61	6,97	2,53	7,30	1,49
		Baixa	4,28	7,53	2,97	1,33	2,15
	Média	Alta	3,52	6,07	2,43	6,50	1,57
		Baixa	3,45	8,60	3,53	0,73	2,56
	Baixa	Alta	2,76	6,70	2,40	7,23	1,52
		Baixa	2,90	8,30	3,43	0,67	2,62
Verbos	Alta	Alta	4,39	6,03	2,50	6,27	1,51
		Baixa	4,23	7,53	3,10	1,63	1,86
	Média	Alta	3,43	6,17	2,57	6,63	1,54
		Baixa	3,45	8,37	3,70	1,53	1,90
	Baixa	Alta	2,83	6,43	2,70	5,53	1,61
		Baixa	2,88	8,60	3,70	1,47	2,02

Fonte: elaboração própria.

Com o objetivo de se testar um grande número de (pseudo)palavras do PB (720 palavras e 720 pseudopalavras), os 1.440 estímulos foram divididos em oito blocos de 180 palavras. A partir desses oito blocos, construíram-se 12 listas experimentais contendo quatro blocos cada uma (i.e., 720 estímulos), seguindo-se o contrabalanceamento de quadrado-latino (BALOTA *et al.*, 2007). As listas possuíam os seguintes critérios: i. não houve duas palavras seguidas começando com a mesma letra e ii. no máximo três palavras ou pseudopalavras com as mesmas condições experimentais em sequência. Esses critérios foram estabelecidos com o objetivo de minimizarem-se possíveis efeitos de *priming* ortográfico/fonológico, assim como das características das variáveis pesquisadas. O experimento final contou com 720 estímulos de teste e 10 estímulos de prática, houve um intervalo de um minuto no meio do experimento e o tempo médio de realização do experimento foi de 18 minutos.

2.3 Procedimentos

O experimento foi construído e aplicado através do programa E-Prime® 2.0 Professional (Psychology Software Tools, Inc., Sharpsburg, PA, USA) (SCHNEIDER; ESCHMAN; ZUCCOLOTTO, 2012). Os participantes foram testados individualmente em local silencioso. Cada estímulo foi apresentado na seguinte sequência: primeiro, um ponto de fixação era apresentado por 500 ms; segundo, a palavra ou pseudopalavra era apresentada por 2.000 ms ou até a resposta do participante; terceiro, uma tela preta era apresentada por 500 ms; enfim, a mesma sequência recomeçava com a apresentação de um novo estímulo. Os estímulos foram apresentados no centro da tela LCD 15" de um computador portátil em letras minúsculas e brancas sobre um fundo preto,

em fonte Courier New e tamanho 18 pt. A mensuração dos tempos de reação (RT) começava com a apresentação da palavra ou pseudopalavra e terminava quando o participante realizava sua resposta. Os participantes utilizavam uma tecla verde do teclado do computador para palavras existentes e uma tecla vermelha do teclado do computador para pseudopalavras inexistentes. Eles foram orientados a realizarem suas respostas o mais corretamente e rapidamente possível.

3. Resultados e discussão

Encontrou-se uma diferença significativa entre os RT de palavras 679(234) ms e pseudopalavras 844(338) ms, $t(46) = 4,72$, $p < 0,001$ (teste-t de Student, distribuição bicaudal, amostras pareadas), assim como uma diferença significativa entre as taxas de erro de palavras 1,48% e pseudopalavras 2,51%, $t(46) = 2,33$, $p < 0,05$ (teste-t de Student, distribuição bicaudal, amostras pareadas). Esse efeito é recorrente em tarefas de decisão lexical, pseudopalavras são descartadas mais lentamente que palavras são reconhecidas devido ao tempo de procura do estímulo no léxico mental. Ainda, pseudopalavras apresentaram maior taxa de erro porque eventualmente são classificadas como palavras potenciais da língua (DUYCK *et al.*, 2004). Portanto, destaca-se que as pseudopalavras geradas pelo LexPorBR efetivamente respeitaram as estruturas ortotáticas do PB, resultando em confusão com palavras potenciais (ESTIVALET; MEUNIER, 2017). A seguir, somente as palavras existentes foram analisadas.

RT abaixo de 300 ms e acima de 1.800 ms foram considerados fora da tarefa e descartados (0,46%), uma palavra experimental (i.e., emana) foi removida porque apresentou taxa de erro superior a 50% (0,06%) e respostas incorretas foram removidas (2,96%), no total 3,47% dos dados foram descartados para análise dos RT (BAAYEN, 2008). Essa limpeza assegura uma melhor distribuição normal dos dados para análise dos mesmos. As médias dos RT, desvios padrão e taxas de erro são apresentados na Tabela 3.

Tabela 3: Médias dos RT, desvios padrão e de erro nas condições experimentais fatoriais.

Cat. Gram.	Freq. Viz.	Alta		Média		Baixa	
		RT(ms)	Erro(%)	RT(ms)	Erro(%)	RT(ms)	Erro(%)
Substantivos	Alta	586(152)	0,06	649(185)	0,23	696(211)	0,60
	Baixa	629(173)	0,02	729(232)	0,13	761(243)	0,20
Verbos	Alta	614(157)	0,13	649(191)	0,35	704(220)	0,80
	Baixa	615(164)	0,09	665(197)	0,12	724(233)	0,24

Fonte: elaboração própria.

As variáveis fatoriais foram avaliadas através de uma análise de variância (ANOVA, F1:

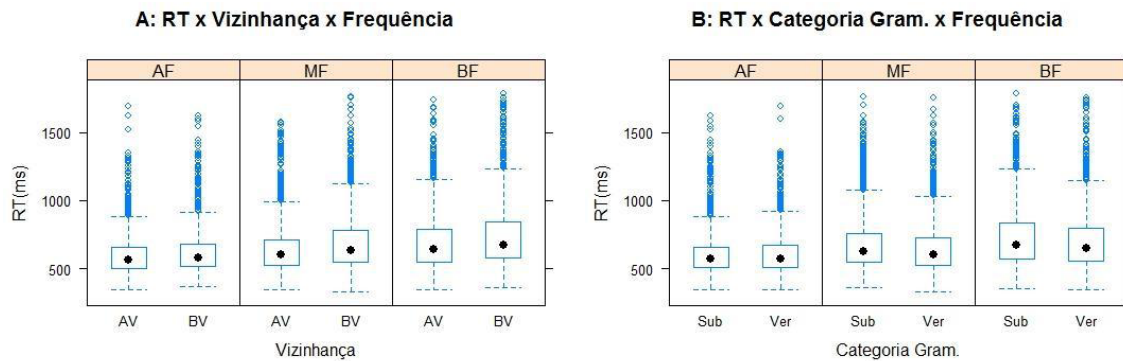
participantes) contendo RT como variável dependente; a. frequência lexical (alta, média e baixa), b. vizinhança ortográfica (alta e baixa) e c. categoria gramatical (substantivos e verbos) como variáveis independentes fixas; e participantes como variável aleatória (BAAYEN, 2008). Foram encontrados os seguintes efeitos principais significativos: frequência lexical $F(2,23) = 298,07$, $p < 0,001$, vizinhança ortográfica $F(1,23) = 110,02$, $p < 0,001$ e categoria gramatical $F(1,23) = 6,73$, $p < 0,05$, e, os seguintes efeitos de interação significativos: frequência lexical e categoria gramatical $F(2,23) = 10,76$, $p < 0,001$ e vizinhança e categoria gramatical $F(1,23) = 49,28$, $p < 0,001$, conforme a Figura 1.

Corroborando o efeito de frequência, a presente tarefa de decisão lexical em PB encontrou diferenças significativas entre o reconhecimento de palavras de alta, média e baixa frequência lexical (MONSELL; DOYLE; HAGGARD, 1989). Palavras de alta frequência foram reconhecidas 62 ms mais rapidamente que palavras de média frequência $t(46) = 2,75$, $p < 0,05$ e palavras de média frequência foram reconhecidas 49 ms mais rapidamente que palavras de baixa frequência $t(46) = 2,32$, $p < 0,05$ (testes-t de Student pareados com correção de Bonferroni). Este resultado vai de encontro a teorias que propõem a organização do léxico mental em função das frequências lexicais.

Em seguida, encontrou-se o efeito de vizinhança ortográfica, em que palavras com alta vizinhança foram reconhecidas 39 ms mais rápido que palavras com baixa vizinhança. Este resultado facilitador em palavras com mais vizinhos ortográficos está de acordo com uma série de estudos apresentados por Andrews (1997), palavras com muitos vizinhos ortográficos ativam simultaneamente uma série de outras palavras e são rapidamente reconhecidas. Não houve efeito de interação entre frequência lexical e vizinhança ortográfica, sugerindo que esses efeitos são independentes (TAFT, 1991), conforme pode ser verificado na Tabela 3 e na Figura 1A. Observa-se que o efeito de vizinhança ortográfica ocorre principalmente em palavras pequenas, com até cinco letras, pois palavras grandes desassociam-se facilmente em função de sua complexidade ortográfica (MARIAN *et al.*, 2012).

Logo após, verificou-se que verbos são reconhecidos apenas 13 ms mais rapidamente que substantivos, esta diferença se deve às características do processamento morfológico entre substantivos derivados e verbos flexionados (CARAMAZZA; HILLIS, 1991). Apesar de os verbos apresentarem mais complexidade nos traços morfossintáticos, substantivos derivados apresentam maior complexidade estrutural e semântica durante seu processamento e reconhecimento. Mais importante que o efeito principal, houve uma interação entre frequência lexical e categoria gramatical, sugerindo que essas categorias gramaticais possuem um comportamento em palavras de alta frequência e outro comportamento em palavras de média e baixa frequência, conforme pode ser verificado na Figura 1B.

Figura 1: Variáveis fatoriais: a. frequência lexical, b. vizinhança ortográfica e c. categoria gramatical. AF: alta frequência, MF: média frequência, BF: baixa frequência, AV: alta vizinhança, BV: baixa vizinhança, Sub: substantivos, Ver: verbos.



Fonte: elaboração própria.

Esses resultados sugerem que os verbos são reconhecidos mais rapidamente que os substantivos somente em média frequência $t(46) = 2,24, p < 0,05$ e baixa frequência $t(46) = 2,83, p < 0,05$, mas não possuem diferenças em alta frequência, sugerindo um sistema de processamento lexical de duas vias (BESNER; JOHNSTON, 1992). Enfim, também houve um efeito de interação entre vizinhança ortográfica e categoria gramatical, justificado através do fato que substantivos possuem naturalmente mais vizinhos ortográficos do que verbos no PB (ESTIVALET; MEUNIER, 2015).

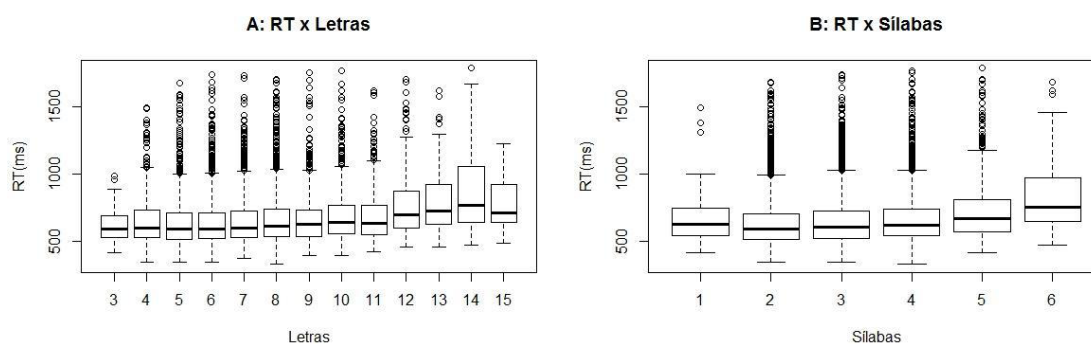
Na sequência, uma regressão linear múltipla foi calculada contendo os RT como variável dependente, e, d. número de letras (3-15) e e. número de sílabas (1-6) como variáveis independentes numéricas. Foi encontrado um efeito principal significativo $F(2,716) = 50,77, p < 0,001$ (número de letras $t(716) = 2,36, p < 0,05$; número de sílabas $t(716) = 2,51, p < 0,05$), com $R^2 = 0,12$, de acordo com a Figura 2.

Conforme encontrado em New *et al.* (2006), os resultados do presente estudo também apresentaram uma curva em U dos RT em relação ao número de letras, palavras de 3-6 letras possuem RT decrescentes e palavras de 7-15 letras possuem RT crescentes em função do número de letras, conforme a Figura 2A. Este efeito deve ser explicado em função da alta vizinhança ortográfica, assim como da maior homonímia e polissemia que palavras pequenas apresentam (MARIAN *et al.*, 2012). Importante, verifica-se que até nove letras, o reconhecimento das palavras em função do número de letras possui um comportamento semelhante, apresentando maiores diferenças e variâncias de 9-15 letras. Este efeito pode ser justificado devido à necessidade de mais de uma sacada ocular para a leitura de palavras maiores de nove letras (BESNER; JOHNSTON, 1992) e determinação da sílaba tônica.

Finalmente, da mesma forma que o comportamento do número de letras, os RT em relação ao número de sílabas também apresentou uma curva em U, com RT decrescentes para palavras de 1-2 sílabas e RT crescentes para palavras de 3-6 sílabas (MEHLER *et al.*, 1981). A mesma explicação dada acima em relação ao número de letras pode ser utilizada para o comportamento dos RT em função do número de sílabas. É necessário salientar que o processamento silábico inclui também um componente fonológico associado ao pé prosódico e sílaba tônica, logo, apresentando diferenças de RT ligados ao processamento temporal de leitura das palavras (BALOTA *et al.*, 2007).

Em geral, os resultados apresentados acima corroboram os efeitos psicolinguísticos de frequência lexical, vizinhança ortográfica e número de letras (TAFT, 1991), assim como a presente hipótese em relação aos efeitos de categoria gramatical e número de sílabas.

Figura 2: Variáveis numéricas d. número de letras e e. número de sílabas.



Fonte: elaboração própria.

Considerações finais

O principal objetivo do presente trabalho foi verificar a influência das três principais variáveis lexicais e ortográficas, frequência lexical, vizinhança ortográfica e número de letras (TAFT, 1991), no processamento de palavras isoladas do PB. Além dessas variáveis, também foram verificadas outras duas variáveis de interesse, categoria gramatical e número de sílabas, através de um experimento psicolinguístico com uma tarefa de decisão lexical em um grande número de palavras do PB (BALOTA *et al.*, 2007). Os resultados apresentaram efeitos principais significativos assim como algumas interações entre essas variáveis.

O efeito de frequência lexical tem sido discutido em diversos modelos de reconhecimento de palavras (MONSELL; DOYLE; HAGGARD, 1989). Modelos associativos de via única têm argumentado que a frequência lexical é a principal variável preditiva do resultado de interações

entre a forma e o significado das palavras (RUMELHART; MCCLELLAND, 1982). Contudo, estes modelos têm enfrentado dificuldades para explicar os diferentes efeitos de vizinhança ortográfica que emergem a partir da manipulação desta variável, assim como as diferenças entre o processamento de diferentes categorias gramaticais. O presente estudo sugere que os efeitos de vizinhança ortográfica estão presentes de forma independente da frequência lexical (MARIAN *et al.*, 2012). Ainda, a presente pesquisa encontrou diferenças no processamento e reconhecimento de palavras de diferentes categorias gramaticais, que por sua vez podem ser explicadas através do processamento dos traços morfossintáticos presentes nas diferentes categorias gramaticais, tais como substantivos derivados e verbos flexionados (CARAMAZZA; HILLIS, 1991).

Nesse sentido, o presente estudo contribui com evidências a favor de modelos de processamento de duas vias para o reconhecimento de palavras (BESNER; JOHNSTON, 1992). Enquanto algumas palavras são reconhecidas diretamente pela sua forma inteira no léxico mental, outras palavras são processadas através de seus constituintes linguísticos discretos. De uma forma geral, palavras de alta frequência são processadas pela via direta lexical enquanto palavras de média e baixa frequência são reconhecidas através do processamento ortográfico e fonológico. Ainda, substantivos derivados são acessados através da via direta no léxico mental e verbos flexionados são processados pelo componente morfossintático decomposicional.

Sendo assim, conforme proposto no modelo serial de Besner e Johnston (1992), os efeitos de frequência, vizinhança ortográfica, tamanho das palavras e categoria gramatical das palavras ocorrem em diferentes etapas do processamento lexical. Primeiramente, ocorrem os efeitos de tamanho de palavra em relação ao número de letras (unidades simbólicas a serem decodificados) e ao número de sílabas (unidades fonológicas a serem processadas); em seguida, ocorrem os efeitos de vizinhança ortográfica/fonológica em que outras palavras são ativadas, gerando competição lexical; após, ocorrem os efeitos de frequência de lexema e lema; enfim, ocorre o efeito de categoria gramatical com a ativação dos traços morfossintáticos.

Finalmente, o presente estudo também confirmou que as normas de frequência, vizinhança ortográfica, categoria gramatical, número de letras e número de sílabas estabelecidas pelo LexPorBR são pertinentes e devem ser observadas para a seleção, manipulação e controle de estímulos psicolinguísticos do PB (ESTIVALET; MEUNIER, 2017). Enfim, espera-se que novas pesquisas do processamento e reconhecimento de palavras venham enriquecer a compreensão deste e outros fenômenos, assim como colaborem para o desenvolvimento da psicolinguística experimental no PB.

REFERÊNCIAS

- ANDREWS, Sally. The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, v. 4, n. 4, p. 439–461, dez. 1997.
- BAAYEN, Rolf Harald. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: Cambridge University Press, 2008.
- BALOTA, David *et al.* The English Lexicon Project. *Behavior Research Methods*, v. 39, n. 3, p. 445–459, ago. 2007.
- BESNER, Derek; JOHNSTON, James. Reading and the mental lexicon: on the uptake of visual information. In: MARSLEN-WILSON, William (ed.). *Lexical Representation and Process*. Cambridge/MA: The MIT Press, p. 291–316, 1992.
- CARAMAZZA, Alfonso; HILLIS, Argye. Lexical organization of nouns and verbs in the brain. *Nature*, v. 349, n. 6312, p. 788–790, fev. 1991.
- COLTHEART, Max *et al.* Access to the internal lexicon. In: DORNIC, Stan (ed.). *Attention and Performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates, p. 535–555, 1977.
- DUYCK, Wouter *et al.* WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers*, v. 36, n. 3, p. 488–499, ago. 2004.
- ESTIVALET, Gustavo Lopez; MEUNIER, Fanny. The Brazilian Portuguese Lexicon: An Instrument for Psycholinguistic Research. *PLOS ONE*, v. 10, n. 12, p. e0144016, dez. 2015.
- ESTIVALET, Gustavo Lopez; MEUNIER, Fanny. Corpus psicolinguístico Léxico do Português Brasileiro. *Revista SOLETRAS*, v. 33, n. 1, p. 212–229, jul. 2017.
- KEULEERS, Emmanuel. vwr: Useful functions for visual word recognition research. *R Package*, p. 19, 2013.

MARIAN, Viorica *et al.* CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE*, v. 7, n. 8, p. e43230, ago. 2012.

MARQUIAFÁVEL, Vanessa; BOKAN, Alessandro; ZAVAGLIA, Cláudia. PETRUS: A rule-based grapheme-to-phoneme converter for Brazilian Portuguese. In: BAPTISTA, Jorge *et al.* (ed.). *PROPOR 2014, LNAI 8776*. Heidelberg, Germany: Springer, 2014.

MEHLER, Jacques *et al.* The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, v. 20, n. 3, p. 298–305, jun. 1981.

MONSELL, Stephen; DOYLE, Mark; HAGGARD, Patrick. Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, v. 118, n. 1, p. 43–71, 1989.

NEW, Boris *et al.* Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, v. 13, n. 1, p. 45–52, fev. 2006.

PINHEIRO, Gisele; ALUÍSIO, Sandra. *Corpus NILC: descrição e análise crítica com vistas ao projeto Lacio - WebSérie de Relatórios do Núcleo Interinstitucional de Linguística Computacional NILC - ICMC - USP*. São Carlos, SP: Universidade Federal de São Carlos - UFSCar, 2003.

RUMELHART, David; MCCLELLAND, Jay. An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, v. 89, n. 1, p. 60–94, 1982.

SCHNEIDER, Walter; ESCHMAN, Amy; ZUCCOLOTTO, Anthony. *E-Prime 2.0*. Learning Research and Development Center: University of Pittsburgh, 2012.

TAFT, Marcus. *Reading and the Mental Lexicon*. East Sussex/UK: Lawrence Erlbaum Associates Ltd., 1991.

VAN HEUVEN, Walter. *et al.* SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, v. 67, n. 6, p. 1176–1190, jun.

2014.

WORLD MEDICAL ASSOCIATION. World Medical Association Declaration of Helsinki. *JAMA*, v. 310, n. 20, p. 2191, nov. 2013.

YARKONI, Tal; BALOTA, David; YAP, Melvin. Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, v. 15, n. 5, p. 971–979, out. 2008.