

## **EFEITO DA FREQUÊNCIA DE OCORRÊNCIA NA IDENTIFICAÇÃO DE PADRÕES ACENTUAIS**

TOKEN FREQUENCY EFFECTS ON STRESS IDENTIFICATION

**Maria Mendes Cantoni<sup>1</sup>**

mmcantoni@gmail.com

**RESUMO:** Neste estudo, foram avaliados efeitos da frequência de uso na percepção de padrões acentuais do português brasileiro. Em um experimento de identificação de palavras com ambiguidade na posição do acento, observou-se uma relação direta da frequência de ocorrência com a probabilidade de identificação da palavra (palavras mais frequentes eram mais percebidas que as menos frequentes). Por outro lado, não foi identificada influência da frequência de tipo, nem mesmo em palavras com valores de frequência de ocorrência equivalentes. Esses resultados são considerados favoráveis à hipótese de imprevisibilidade do acento no português brasileiro.

**PALAVRAS-CHAVE:** Frequência de Uso; Acento; Percepção da Fala.

**ABSTRACT:** In this study, effects of frequency of usage on the perception of stress patterns in Brazilian Portuguese were evaluated. In an experiment of word identification with ambiguity in the stress location, a direct relation between token frequency and the probability of word identification was observed (more frequent words were perceived more often than less frequent ones). On the other hand, no influence of type frequency was identified, not even in words with equivalent token frequency values. These results are considered favorable to the unpredictability hypothesis of stress in Brazilian Portuguese.

**KEYWORDS:** Frequency of Usage; Stress; Speech Perception.

### **INTRODUÇÃO**

A influência do uso na construção da gramática e na atuação de fenômenos linguísticos tem, especialmente desde a década de 80, atraído a atenção dos linguistas. Com o desenvolvimento dos Modelos Baseados no Uso (e.g. Bybee 1985, 2001, 2010; Langacker 1987, 2000; cf. Barlow & Kemmer 2000), que assumem que as representações linguísticas são múltiplas e que a gramática é determinada

---

<sup>1</sup> Doutora em Linguística, Universidade Federal de Minas Gerais – UFMG.

dinamicamente a partir das experiências comunicativas, a frequência lexical ou de uso passou a ser incorporada na construção de teorias sobre a linguagem. A frequência de uso como termo geral engloba dois tipos de frequência lexical, geralmente avaliados por meio de **corpora** representativos de uma variedade linguística. A frequência de tipo (**type frequency**) corresponde ao número de itens de um **corpus** que apresentam um determinado padrão (ou tipo), sendo que padrões podem ser palavras, morfemas, sílabas, etc. Já a frequência de ocorrência (**token frequency**) corresponde a quantas vezes um padrão efetivamente ocorre em um *corpus*. Para os Modelos de Uso, a frequência de tipo e de ocorrência exercem impacto, respectivamente, sobre a produtividade de padrões e sobre a força lexical (Bybee 1985, 2001).

São bem documentados na literatura os efeitos exercidos pela frequência de uso em diversos aspectos da linguagem (cf. Bybee & Hopper 2001). Sabe-se também que, dada a natureza complexa da própria linguagem, tais efeitos muitas vezes interagem com efeitos relacionados à própria gramática. Como exemplo, mencionaremos alguns estudos. Rosenzweig e Postman (1958) demonstraram que a compreensibilidade das palavras aumenta diretamente em função de seu tamanho e de sua frequência de uso. Pagel et al. (2007), em análise comparativa, defendem a frequência de uso das palavras seria crucial para os processos de mudança, uma vez que as línguas começariam a divergir inicialmente em partes menos usadas de seu vocabulário e que palavras de alta frequência tendem a preservar formas morfológicas antigas. Algumas palavras mudam o suficientemente devagar de forma a permitir a manutenção de formas análogas por centenas de anos. Nielsen (2011) investigou a influência da frequência lexical e da densidade da vizinhança lexical na identificação de palavras a partir de estímulos auditivos e visuais. Os resultados obtidos mostram que, no paradigma de reconhecimento exclusivamente auditivo, palavras mais frequentes foram reconhecidas melhor, enquanto que no paradigma auditivo-visual, palavras menos frequentes foram reconhecidas melhor. Além disso, palavras pertencentes a vizinhanças lexicais densas foram reconhecidas mais acuradamente que palavras de vizinhanças pouco densas.

Neste estudo, procuramos avaliar se a frequência de uso exerce um impacto também no processamento prosódico da fala, mais especificamente na identificação do padrão acentual das palavras.

## 2. O ACENTO

O acento<sup>2</sup> é um fenômeno prosódico associado à proeminência acústica e perceptual de unidades sonoras (geralmente as sílabas) em um determinado domínio (palavras ou sentenças, por exemplo). No português brasileiro, o acento tem como principal correlato acústico a duração e, adicionalmente, pode ser relacionado a padrões de mudança na frequência fundamental (fo) e na intensidade<sup>3</sup>. Opera no domínio das palavras e apresenta três possibilidades de alternância entre sílabas fortes (s, **strong**) e fracas (f, **weak**), dependendo da localização da sílaba proeminente com relação ao fim da palavra: as palavras no português brasileiro podem ser proparoxítonas (s-w-w), paroxítonas (w-s-w) ou oxítonas (w-w-s).

Um ponto controverso do sistema sonoro do português brasileiro se refere à natureza previsível ou imprevisível do acento: se pode ser localizado por meio de um conjunto de fatores linguísticos condicionantes, ou algoritmo de atribuição acentual, no primeiro caso; ou se sua ocorrência estaria associada às próprias formas lexicais, no segundo caso.

Tradicionalmente, as descrições acentuais do português brasileiro argumentam pela previsibilidade com base na predominância do padrão paroxítono sobre os outros dois padrões acentuais possíveis (sendo ao proparoxítono frequentemente atribuído o caráter de excepcionalidade) e buscam estabelecer as condições métricas e fonotáticas determinantes para a localização do acento, na forma de um conjunto de regras (ou algoritmo de atribuição) (Bisol 1992; Lee 1995; cf. Araújo 2007).

Por outro lado, o grande número de exceções às regras de atribuição acentual reforçam a hipótese de imprevisibilidade do acento: quanto menor o grau de regularidade ou previsibilidade de um padrão (como o acento), mais necessário se torna especificá-lo no léxico. Espera-se, portanto, que esteja sujeito a efeitos de frequência de uso (Cantoni 2013).

Neste estudo, procurou-se considerar a hipótese de imprevisibilidade do acento avaliando a possível influência da frequência de uso na percepção de palavras em situação de ambiguidade.

---

<sup>2</sup> Consideraremos neste estudo somente o acento primário.

<sup>3</sup> Segundo Cantoni (2013), as sílabas acentuadas são caracterizadas por maior duração, abaixamento na curva de fo e maior intensidade em comparação com as sílabas não acentuadas.

### **3. PERCEPÇÃO E IDENTIFICAÇÃO DE PALAVRAS**

Uma das tarefas fundamentais envolvidas na percepção linguística é a identificação de palavras, um processo complexo, que transcorre em diversas etapas e é influenciado por diversos fatores. Requer o processamento acústico e a segmentação do contínuo da fala, muitas vezes na ausência de pistas claras que indiquem as fronteiras das palavras. Está sujeito ao conhecimento fonotático probabilístico específico da língua, sendo também guiado por informações fonéticas, prosódicas e lexicais (Jusczyk & Luce 2002). A proeminência acentual guia a segmentação do contínuo da fala e inicia o processo de identificação de palavras (Cutler & Norris 1988). Sabe-se que o processo de identificação de uma palavra depende tanto de sua correspondência com o estímulo sonoro quanto de sua relação com as palavras lexicalmente próximas, que irão competir no processo de ativação (Dahan & Magnuson 2006). Os contextos semântico e sintático também apresentam papel importante no processo de identificação (Kim et al. 2012).

Neste estudo, iremos avaliar o processo de identificação de palavras em situação de ambiguidade sonora, em que as informações fonéticas não servem para decidir entre os itens em competição. Além disso, as palavras serão testadas em isolamento, de forma a serem desprovidas de contexto semântico e sintático. Espera-se, desta forma, que o fator predominante na identificação das palavras seja a frequência de uso, com observaram Connine e colegas (1993) em estudo sobre identificação de palavras envolvendo ambiguidade no vozeamento de consoantes, no inglês. Mais especificamente, espera-se que a frequência de ocorrência exerça maior influência na identificação de palavras com ambiguidade na posição do acento do que a frequência de tipo. Ou seja, espera-se que palavras mais frequentes sejam mais reconhecidas que as palavras menos frequentes, independente de seu padrão acentual. Além disso, no caso de palavras igualmente frequentes, a hipótese da imprevisibilidade do acento seria reforçada por um efeito não significativo ou fraco da frequência de tipo, ao passo que o acento previsível dependeria do favorecimento da identificação das paroxítonas nessa situação.

#### 4. METODOLOGIA EXPERIMENTAL

O experimento realizado consistiu em na identificação de palavras em situação de ambiguidade. Foram testados 12 sujeitos, ambos os sexos, nível universitário, 20 a 30 anos, falantes nativos do português, sem problemas de audição ou fala. Os estímulos utilizados foram diferentes sequências de sílabas fracas e fortes com segmentação ambígua, que pudessem ser interpretadas como duas palavras distintas, uma paroxítona e outra oxítona, a depender do **parsing** realizado no momento da percepção (e.g. ...pɔsipɔsipɔsipɔsipɔ... → ...pɔ[sipɔ][sipɔ][sipɔ][sipɔ]s... “cipó” ou ...[pɔsi][pɔsi][pɔsi][pɔsi]pɔs... “posse”). Foram selecionados 19 pares de palavras oxítona-paroxítona, com frequência próxima (ambas pouco frequentes) e com frequência distante (uma frequente e outra pouco frequente). As palavras utilizadas no experimento e sua respectiva frequência de ocorrência, obtida no banco de dados do Projeto Aspa<sup>4</sup>, são mostradas no Quadro 1, a seguir. Como se vê na coluna “Estímulo”, foi adotado um código de identificação para cada estímulo correspondendo às três primeiras letras da palavra oxítona do par.

Estímulo	Oxítona		Paroxítona	
	Palavra	Frequência de Ocorrência	Palavra	Frequência de Ocorrência
man	manoel	7.958	noelma	16
dav	davi	1.373	vida	112.365
caj	caju	873	juca	1.922
tum	tumor	2.212	morto	20.475
dal	dali	2.161	lida	1.775
tup	tupi	1.420	pito	206
bab	babá	1.283	baba	432
mar	marrom	1.111	roma	9.689
mac	macau	878	calma	3.501
jab	jabuti	422	botija	43
fub	fubá	357	bafo	205
tut	tutor	312	torto	783
cip	cipó	271	posse	15.246
gui	guichê	227	chegue	2.470
cas	cassis	125	cisca	10

4 [www.projetoaspa.org.br](http://www.projetoaspa.org.br)

sum	sumô	127	moço	1.639
caq	caqui	107	quica	22
mas	maçom	20	soma	7.913
cja	cajá	59	jaca	127

**Quadro 1:** Pares de estímulos e seus valores de ocorrência

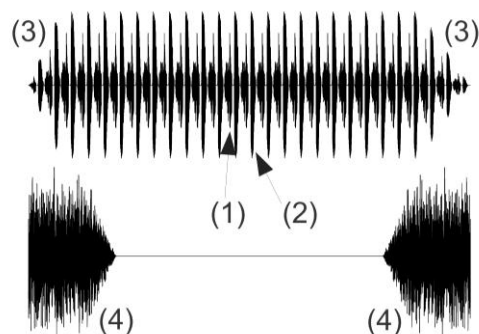
Os estímulos foram construídos a partir de sílabas isoladas<sup>5</sup>, produzidas pela mesma locutora (jovem, falante nativa da variedade de Belo Horizonte). As sílabas isoladas foram alteradas uniformemente em duração e intensidade para corresponderem em proporção aos valores das sílabas fortes e fracas. Os valores de referência foram obtidos de cinco palavras paroxítonas e oxítonas gravadas previamente, para as quais se fez uma média dos valores das tônicas entre si e das átonas entre si. Por meio do **software** Audacity (Audacity Team 2008), a sílaba forte da sequência foi amplificada em 3dB e a sílaba fraca teve, então, sua amplitude escalonada em 0,5 com relação à forte. Além disso, a duração das sílabas foi alterada (sem alterar o tom; Audacity: função “Alterar tempo”) de forma que as sílabas fortes passaram a ter 0,4 s e as fracas, 0,3 s. A frequência fundamental, ainda que relevante para distinção de sílabas tônicas e átonas, não foi alterada nos estímulos, por ser o principal parâmetro acústico responsável pela entonação, cujos efeitos não foram considerados neste experimento. As sílabas forte e fraca foram então concatenadas e repetidas 25 vezes, cada uma. Por fim, duas estratégias foram tomadas para mascarar o início e o fim das sequências, de forma a evitar que os sujeitos fossem influenciados pela ordem inicial ou final para identificação do estímulo: o início e o fim das sequências foram atenuados regressiva (**fade-in**) e progressivamente (**fade-out**) e sobrepostos a 3 s de ruído browniano. A FIG. 1, a seguir, apresenta o sinal construído para um dos estímulos utilizados no experimento.

---

<sup>5</sup> Optou-se por gravar as sílabas isoladamente, e não as próprias palavras, a fim de evitar que possíveis detalhes fonéticos finos característicos de cada palavra e decorrentes da ordenação das sílabas (e.g. transições de formantes) pudessem influenciar na tarefa de identificação dos estímulos.

Oscilograma da sequência *cipó/posse*

- (1) Sílabas fracas escaladas em 0,5 de amplitude
- (2) Sílabas fortes amplificadas em 3dB
- (3) *Fade-in e fade-out* no início e fim da sequência
- (4) Sobreposição de 3s de ruído browniano, com *fade-out e fade-in*



**Figura 1:** Exemplo de estímulo utilizado no experimento

A tarefa proposta consistiu em um teste de identificação, em que os sujeitos escutavam estímulos auditivos e deveriam relatar livremente a palavra ouvida, sem qualquer tipo de direcionamento por parte da pesquisadora. As respostas que não corresponderam a nenhuma das duas palavras associadas ao estímulo (34 casos) não foram consideradas na análise dos dados. Para avaliar o desempenho da tarefa, foi calculada uma taxa de identificação por sujeito e por palavra. A taxa de identificação foi determinada como a porcentagem de respostas do sujeito ou da palavra que correspondessem a uma das duas palavras esperadas (p. ex., 100% se o sujeito respondesse apenas as palavras esperadas ou se um dos estímulos fosse identificado somente com as palavras esperadas). Com relação aos estímulos, destacaram-se três estímulos com baixas taxas de identificação, “dali-lida” (dal), “guichê-chegue” (gui) e “cassis-cisca” (cas). Durante a análise dos dados, verificamos que a exclusão destes estímulos não alterou significativamente os resultados, pelo que optamos por mantê-los na análise. Além disso, dois sujeitos apresentaram taxa de identificação próxima de 50% – o que significa que quase a metade de suas respostas não foram aproveitadas. Juntos, responderam por 17 do total de 34 palavras sem correspondência. Possivelmente seu desempenho foi comprometido por algum fator, por exemplo, por não terem compreendido bem a tarefa. Contudo, durante a análise dos dados, verificamos que os resultados também não eram alterados significativamente com a retirada desses sujeitos, portanto, optamos por mantê-los na análise dos dados.

#### 4. RESULTADOS

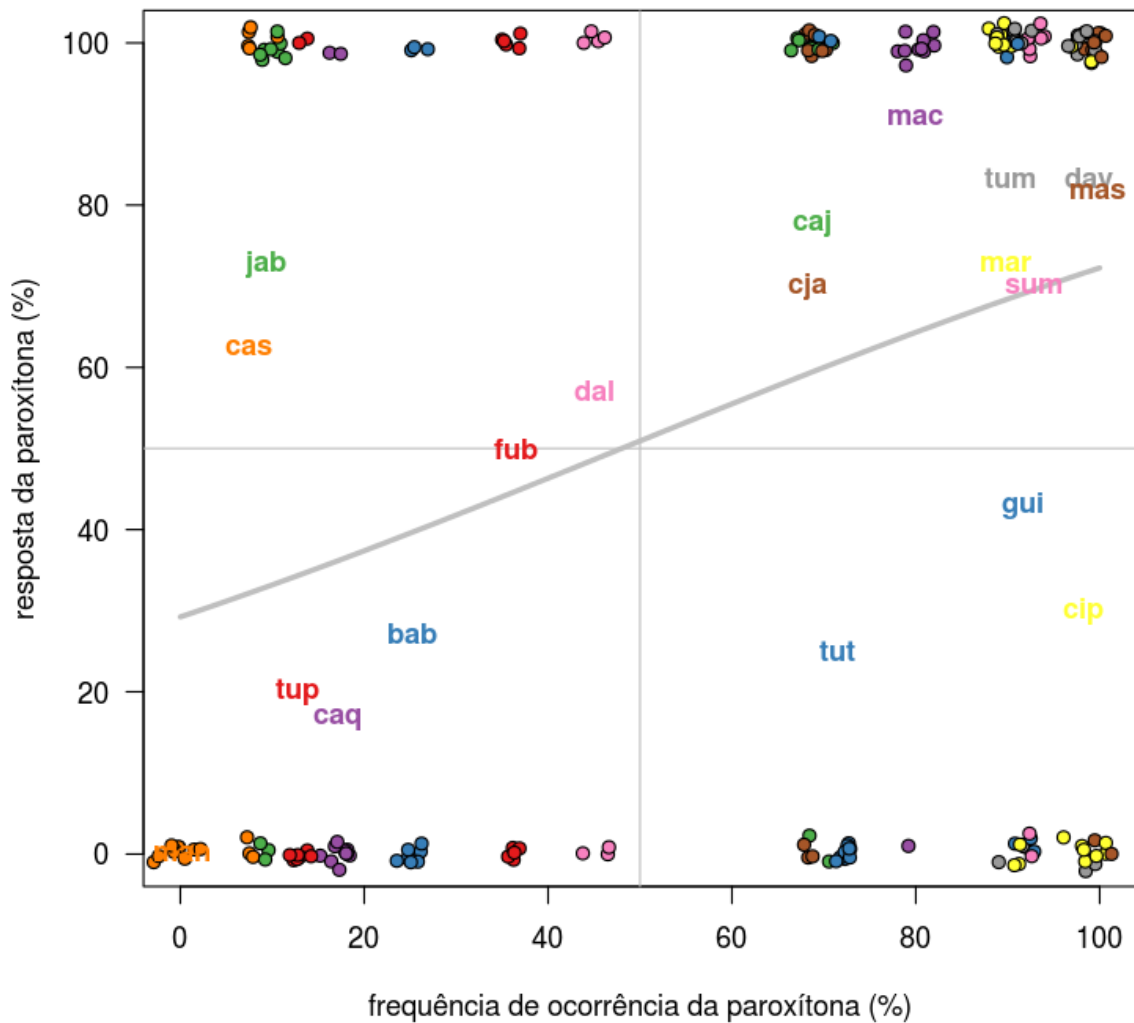
A variável dependente neste experimento é a palavra identificada para cada estímulo (oxítone ou paroxítone) e foi considerada em termos do percentual da “resposta paroxítone” com relação à oxítone para cada estímulo ambíguo. A variável independente é a frequência de ocorrência das palavras associadas ao estímulos, que foi avaliada como a “frequência relativa da palavra paroxítone”, ou a proporção da frequência da paroxítone com relação à frequência da oxítone, calculada pela fórmula:

$$\frac{\textit{frequênciaabsolutadaproxítone}}{\textit{frequênciaabsolutadaproxítone} + \textit{frequênciaabsolutadaoxítone}}$$

A análise estatística dos dados foi realizada por meio do R (R Development Core Team 2011). Os dados obtidos foram ajustados com base em um modelo generalizado linear com efeitos mistos (função “lmer”, pacote “lme4”; Bates 2010), visto que a variável dependente “resposta da paroxítone” é percentual, com uma distribuição de erros binomial. A variável independente “frequência relativa da paroxítone” de cada estímulo foi considerada como um efeito fixo e contínuo no modelo. Os sujeitos foram considerados como um efeito aleatório no modelo, o que equivale a assumir que cada sujeito apresenta uma tendência particular quanto à proporção de escolha da forma paroxítone.

A seguir, a Fig. 2 mostra a distribuição dos dados obtidos e os dados ajustados ao modelo linear.





**Figura 2:** Percentual de respostas paroxítonas em função da frequência de ocorrência relativa das paroxítonas

Legenda:	bab (babá-baba)	gui (guichê-chegue)
man (manoei-noelma)	mar (marrom-roma)	cas (cassis-cisca)
dav (davi-vida)	mac (macau-calma)	sum (sumô-moço)
caj (caju-juca)	jab (jabuti-botija)	caq (caqui-quica)
tum (tumor-morto)	fub (fubá-bafo)	mas (maçom-soma)
dal (dali-lida)	tut (tutor-torto)	cja (cajá-jaca)
tup (tupi-pito)	cip (cipó-posse)	

Na Fig. 2, o eixo horizontal indica a frequência de ocorrência relativa da palavra paroxítona associada aos estímulos. O eixo vertical indica a porcentagem de respostas paroxítonas. As respostas dos sujeitos para cada estímulo correspondem aos pontos desenhados em 0% (resposta = palavra oxítone) ou em 100% (resposta = palavra paroxítone). Note-se que as cores dos estímulos se repetem. Os estímulos foram representados em cores diferentes e os pontos foram ligeiramente espaçados, para melhor visibilidade. Os códigos de três letras (as três primeiras letras

da palavra oxítone do estímulo) representam a resposta média de cada estímulo. A reta em cinza é a curva de regressão obtida no ajuste ao modelo. As duas retas cinza no centro do gráfico são as referências para 50% de frequência de ocorrência da paroxítone (reta vertical) e 50% de probabilidade de escolha da paroxítone (reta horizontal).

Como se pode ver na Fig. 2, a curva de regressão apresenta inclinação positiva e é significativamente diferente de zero ( $z = 4.26$ ,  $p < 0.0001$ ), o que comprova a existência de um efeito da frequência de ocorrência diretamente relacionado à probabilidade de escolha da palavra. Além disso, observa-se que ponto em que a curva de regressão cruza a reta de referência vertical (ponto teórico em que a frequência de ocorrência da paroxítone seria igual à da oxítone) encontra-se muito próximo à reta de referência horizontal (ponto teórico em que a probabilidade de identificação da paroxítone seria igual à da oxítone). Mais precisamente, neste ponto, o percentual de resposta da paroxítone é 51%, que não é significativamente diferente de 50% ( $z = 0.24$ ,  $p > 0.81$ ). Se houvesse uma tendência a favor das paroxítonas, seria esperada uma resposta à paroxítone significativamente maior que 50% quando a frequência relativa da paroxítone fosse 50% – ou seja, seria esperado que a reta de correlação fosse deslocada do centro para a parte superior do gráfico, o que não ocorreu.

Pode-se observar uma grande dispersão dos resultados ao redor da curva de regressão, com destaque para os estímulos “jabuti-botija” (jab), “cassis-cisca” (cas), “tutor-torto” (tut), “guichê-chegue” (gui) e “cipó-posse” (cip), que se encontram muito deslocados da curva. Por fim, destaca-se que, na avaliação do efeito aleatório do modelo (R: função “ranef”, pacote “lme4”), “tendência por informante”, foram obtidos valores nulos, o que indica que tal efeito não é significativo.

## 5. CONCLUSÕES

Em conclusão, procurou-se avaliar neste estudo se a frequência de uso exerce um papel na percepção de padrões acentuais no português brasileiro, no contexto teórico da hipótese de imprevisibilidade do acento. No experimento realizado, avaliamos a identificação de palavras paroxítonas e oxítonas em contexto de ambiguidade. Os resultados obtidos indicam que, no contexto experimental criado, a identificação do estímulo ambíguo como paroxítone ou oxítone depende direta e unicamente da frequência de ocorrência das palavras paroxítone e oxítone associadas

ao estímulo. Este resultado reforça as evidências de que o uso linguístico influencia o acesso lexical e de que as informações sobre a frequência de ocorrência das palavras são armazenadas no léxico.

Além disso, os resultados não evidenciaram influência da frequência de tipo na identificação das palavras: o modelo estatístico elaborado a partir dos dados prevê que, quando a frequência de ocorrência dos itens paroxítono e oxítono fosse semelhante, haveria igual probabilidade de escolha entre eles. Este resultado, junto com a influência da frequência de ocorrência, podem ser interpretados como indício de que as informações acentuais no português brasileiro estão atreladas às palavras individualmente, favorecendo a hipótese de imprevisibilidade do acento.

Os resultados apresentados conseguem apreender um efeito estatístico que é global para a população – a correlação entre a identificação da palavra e a frequência de ocorrência. Observa-se, contudo, uma grande dispersão nos resultados, de forma que alguns estímulos se afastam mais que os outros da curva de correlação (cf. Fig. 2). Este fato pode ser devido a especificidades do **corpus** de referência utilizado: o banco dados do Projeto Aspa foi formado a partir do Banco de Português do Projeto DIRECT-PUC-SP<sup>6</sup>, compilado a partir de textos escritos, com predominância da variedade culta e do registro formal. De modo geral, é representativo do português brasileiro contemporâneo, contudo, por sua composição, muitas palavras podem apresentar valores de frequência que não correspondem à experiência linguística dos falantes da língua. Estudos futuros utilizando outros **corpora** de referência poderiam reavaliar os resultados deste experimento com dados de frequência mais próximos à frequência dos sujeitos.

## AGRADECIMENTOS

Agradeço a Thaïs Cristófaru por seus comentários e discussão dos resultados e a Rafael Laboissière pelo apoio na análise dos dados.

---

6 [www.lael.pucsp.br/direct](http://www.lael.pucsp.br/direct)

## REFERÊNCIAS

- ARAÚJO, G. *O acento em português: abordagens fonológicas*. São Paulo: Parábola, 2007.
- AUDACITY Team. *Audacity 1.2.6*. 2008. Disponível em: <audacity.sourceforge.net/>. Acesso: 01/05/2010.
- BARLOW, M.; KEMMER, S. (Ed.). *Usage-based models of language*. Stanford: CSLI, 2000.
- BATES, D. *lme4: Mixed-effects modeling with R*. Springer, 2010. Disponível em: lme4.r-forge.r-project.org/lmmwR/lrgprt.pdf. Acesso em 03/11/2012.
- BISOL, L. O acento e o pé métrico binário. *Cadernos de Estudos Lingüísticos*, v. 22, p. 69-80, 1992.
- BYBEE, J. *Language, Usage and Cognition*. Cambridge: Cambridge University Press, 2010.
- BYBEE, J. *Morphology: a study of the relation between meaning and form*. Amsterdam: John Benjamins, 1985.
- BYBEE, J. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.
- BYBEE, J.; HOPPER, P. (Ed.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001.
- CONNINE, C.; TITONE, D.; WANG, J. Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 19, n. 1, p. 81-94, Jan. 1993.
- CUTLER, A.; NORRIS, D. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, v. 14, n. 1, p. 113-121, 1988.
- DAHAN, D.; MAGNUSON, J. Spoken Word Recognition. In: TRAXLER, M.; GERNSBACHER, M. A. (Ed.). *Handbook of Psycholinguistics*, 2nd ed. 2006.
- KIM, D.; STEPHENS, J. D. W.; PITT, M. A. How does context play a part in splitting words apart? Production and perception of word boundaries in casual speech. *Journal of Memory and Language*, v. 66, n. 4, p. 509-529, May 2012.
- LANGACKER, R. A Dynamic Usage-Based Model. In: BARLOW, M.; KEMMER, S. (Ed.). *Usage-based models of language*. Stanford: CSLI, 2000. p. 1-63.
- LANGACKER, R. *Foundations of cognitive grammar*. Stanford: Stanford University, 1987. 2v.
- LEE, S. H. *Morfologia e fonologia lexical do português do Brasil*. Tese (Doutorado) - Campinas, Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, 1995.
- NIELSEN, K. Effect of lexical frequency and neighborhood density on audiovisual spoken recognition. *ICPhS XVII*, 2011.
- PAGEL, M.; ATKINSON, Q.; MEADE, A. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, v. 449, p. 717-720, 2007.

JUSCZYK, P. W.; LUCE, P. Speech Perception and Spoken Word Recognition: Past and Present. *Ear and Hearing*, v. 23, p. 2-40, 2002.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. v. 3.4.0, Vienna - Áustria, 2017. Disponível em: <<http://www.r-project.org/>>. Acesso: 23/05/2017.

ROSENZWEIG, M.; POSTMAN, L. Frequency of usage and the perception of words. *Science*, v. 127, p. 263-266, 1958.