



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

Joyse Ashley Vitorino de Medeiros

Acesso Lexical: uma dupla rota para o Português Brasileiro

2013

Joyse Ashley Vitorino de Medeiros

Acesso Lexical: uma dupla rota para o Português Brasileiro

Dissertação apresentada à
Universidade Federal do Rio
Grande do Norte para
obtenção do título de Mestre
em Psicobiologia.
Orientador: Sidarta Ribeiro
Coorientadora: Aniela França

Natal - RN

2013

Catálogo da Publicação na Fonte
Universidade Federal do Rio Grande do Norte - UFRN

Medeiros, Joyse Ashley Vitorino De.

Acesso Lexical: uma dupla rota para o Português Brasileiro / Joyse Ashley Vitorino De Medeiros. - Natal, 2013.

56 f: il.

Orientador: Prof. Dr. Sidarta Ribeiro.

Coorientadora: Profa. Dra. Aniela França.

Dissertação (Mestrado) - Universidade Federal do Rio Grande do Norte. Centro de Biociências. Pós-Graduação em Psicobiologia.

1. Morfologia - Dissertação. 2. Psicolinguística - Dissertação. 3. Acesso lexical - Dissertação. I. Ribeiro, Sidarta. II. França, Aniela. III. Universidade Federal do Rio Grande do Norte. IV. Título.

RN/UF/BSE01

CDU 81'366

Título: Acesso Lexical: uma dupla rota para o Português Brasileiro.

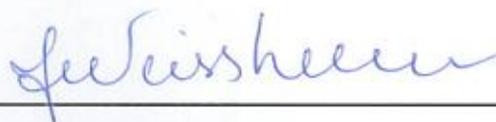
Autor: Joyse Ashley Vitorino de Medeiros

Data da defesa: 27 de março de 2013

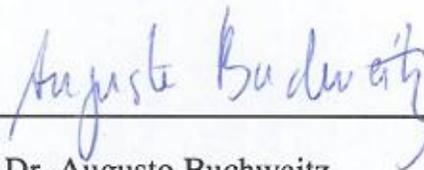
Banca Examinadora:



Prof. Dr. Sidarta Tollendal Gomes Ribeiro
Universidade Federal do Rio Grande do Norte



Prof. Dra. Janaina Weissheimer
Universidade Federal do Rio Grande do Norte



Prof. Dr. Augusto Buchweitz
Pontifícia Universidade do Rio Grande do Sul

AGRADECIMENTOS

Ao meu orientador, Sidarta Ribeiro, pela orientação sempre gentil e solícita.

À Aniela França, nossa guia no admirável mundo novo da Psicolinguística.

Ao Marcio Leitão e José Ferrari Neto, por toda amizade e orientação paralela.

Aos colegas Cleanto, Rafael, Sérgio, Fábio, Hermany e Rute por todas as vezes que ajudaram com *Excel*, com o julgamento de estranhezas das palavras por muitas outras questões e por serem tão abraçáveis e sábios.

Às meninas do ICe (Juliana, Bruna, Melek e Kelly) por tornar as tardes mais aconchegantes.

Ao José Henrique, pela rotina de Matlab perfeita para o projeto.

Ao Daniel Brandão, pela infinita paciência e toda a incomensurável ajuda.

Ao Adriano, pela busca incansável às construções malucas do meu parser yoda.

À mamãe, por ser a melhor mãe do mundo!

RESUMO:

Identificar as unidades que constituem as expressões da linguagem continua a ser uma questão importante na pesquisa psicolinguística contemporânea. Seria o modo como acessamos as palavras influenciado pela estrutura do léxico? Modelos de processamento não lexicais sugerem que, no momento do acesso lexical, as palavras são decompostas em unidades menores, isto é, morfemas. O pressuposto central destes modelos é que as palavras são armazenadas na memória, na forma de unidades significativas mínimas (ex.: amigável: amig/a/ável/). Dentro dessa linha de investigação, palavras compostas são especialmente importantes devido a sua dupla natureza: embora tenham um significado atômico, elas também podem apresentar estruturas segmentáveis. O objetivo desta dissertação foi verificar se há diferença na forma como reconhecemos palavras simples e compostas do Português Brasileiro (PB), utilizando testes de decisão lexical. Esta pesquisa pode lançar luz sobre como a mente representa diferentes categorias de palavras; além disso, pode ajudar a elucidar se diferenças superficiais na estrutura das línguas têm influência no processamento linguístico. Utilizamos testes de decisão lexical para indagar (i) se há diferenças nos tempos médios de reação e acurácia de resposta entre palavras simples e compostas, e (ii) se essas diferenças se correlacionam com a frequência de uso dessas palavras. Os resultados obtidos com o teste de decisão lexical I fornecem evidências de que há um papel para a decomposição no reconhecimento de palavras no PB, pois nas frequências baixa e média as palavras compostas produziram tempos de reação significativamente menores que as simples. No entanto, nas altas frequências dos experimentos de decisão lexical I e II, foram as palavras simples que apresentaram latências de resposta mais curtas. Esse tipo de resultado, por sua vez, se coaduna com as predições de modelos de listagem plena. Dessa forma, para explicar esses resultados se faz necessário sugerir um mecanismo de acesso lexical em dupla rota, em que cada tipo de palavra é acessado mais rapidamente dependendo da sua frequência e propriedades morfológicas.

Palavras-chave: Acesso Lexical, Compostos, Morfologia, Psicolinguística.

ABSTRACT:

Identifying which units constitute the expressions of language remains an important issue in contemporary linguistic research. Is the way by which we have access to words influenced by the lexicon structure? Non lexicalist processing models suggest that, at the time of lexical access, words are decomposed into smaller units, i.e. morphemes. The central assumption of these models is that the words are stored in memory in the form of minimal meaningful units (e.g: friend/ly/). Within this research domain, compound words are especially important due to their dual nature: although they have an atomic meaning, they can also have a segmentable structure. Our objective, in lexical decision tests, was to determine whether there was any difference in the way we recognize Brazilian Portuguese (BP) single or compound words. This research may shed some light on how the mind represents these different word categories and if superficial differences on the structure of the languages influence linguistic processing. We used lexical decision tests to ask whether (i) there are differences in mean reaction times and response accuracy between simple and compound words, and (ii) whether these differences correlate with the frequency of use of these words. The results obtained with the lexical decision test I provide evidence that there is a role for decomposition in word recognition in BP, since low and medium frequency compounds yielded significantly lower reaction times than single words. However, for the words with high frequencies in the lexical decision tests I and II, it was the single words that presented shorter response latencies. This kind of evidence is consistent with predictions of full listing models. In order to explain these results we suggest a dual route mechanism of lexical access, in which each word type will be more quickly accessed depending on its frequency and morphological properties.

Keywords: Lexical Access, Compounds, Morphology, Psycholinguistics.

SUMÁRIO

1. Introdução.....	7
2. Objetivos.....	20
3. Hipóteses.....	21
4. Materiais e Métodos.....	22
Decisão Lexical.....	22
Participantes.....	23
Estímulos.....	23
Procedimento.....	25
Análises dos dados.....	26
5. Resultados.....	27
6. Discussão.....	35
7. Conclusão.....	43
8. Referências.....	44
9. Apêndices	53
Estímulos utilizados nos experimentos.....	54
Análises estatísticas das palavras utilizadas nos experimentos.....	55

INTRODUÇÃO

O que acontece no cérebro quando produzimos, ouvimos ou lemos uma palavra? Como fazemos a correspondência entre a informação sensorial recebida, seja ela acústica, visual ou tátil, e as representações armazenadas no léxico? E ainda, a forma como reconhecemos as palavras é influenciada pelo modo como esse léxico está organizado?

Para tentar explicar o léxico podemos compará-lo a um dicionário. Um dicionário que contenha informações sobre cada um dos itens lexicais, em todos os níveis linguísticos: representação fonológica (ex: modo e ponto de articulação das vogais e consoantes), padrões morfossintáticos (ex: o gênero e número dos nomes e pronomes, a estrutura argumental dos verbos, etc.), informações acerca do significado de cada palavra, além de orientações pragmáticas sobre o emprego adequado em cada contexto de uso.

Contudo, essa analogia não é muito precisa, pois existem diferenças consideráveis entre o modo como as informações linguísticas são registradas em um dicionário ao modo que são em nosso cérebro. Diferentemente dos dicionários, o léxico mental é fluído, pois podemos atualizá-lo a todo momento, criando novas representações ou mesmo modificando representações já existentes. Além disso, o modo de organização do mecanismo de consulta não segue a simplória ordem alfabética, mas pode se basear em princípios de relação em todos os níveis linguísticos (semântico, sintático, morfológico ou fonológico).

Existem três principais abordagens psicolinguísticas acerca do processamento de palavras (figura 1). Modelos de listagem plena (*full listing* ou *non-decompositional*), na sua maioria de natureza conexionista, defendem que todas as palavras são armazenadas integralmente, não sofrendo decomposição ou influência de fatores morfológicos (ex: haveria registros independentes para cantar/canta/cantamos/cantou/cantaria/etc.) (BUTTERWORTH, 1983; RUMELHART & MCCLELLAND, 1986; BYBEE, 1995). Modelos de decomposição plena (*full parsing*), por sua vez, sugerem que as palavras presentes no léxico mental são morfológicamente (de)compostas, ou seja, são formadas a partir de regras de

composição morfológica e tem seu significado construído composicionalmente a partir da soma de suas partes (ex: haveria registro do morfemas e afixos - “cant”+ar/+a/+amos/+ou/+aria/etc.- e as regras de formação para suas flexões) (HALLE & MARANTZ, 1993; TAFT, 1974). Existem ainda os modelos de dupla rota, que propõem que nosso cérebro trabalha com dois sistemas de armazenamento, cada qual com uma base própria de representação e mecanismos de acesso lexical. O modelo de Pinker e Prince (1992), por exemplo, afirma que os verbos irregulares são armazenados e processados sob sua forma plena (ex.: *take, took, taken*)¹, enquanto os regulares seriam representados por morfemas e regras de afixação (ex.: *love + ed*). Uma evidência para essa hipótese é que, na língua inglesa, todos os verbos mais frequentes são irregulares e todos os mais infrequentes são regulares (PINKER, 1999). A versão do modelo dual de Caramazza, Laudanna e Romani (1988) defende que a determinação da via de acesso lexical se baseia na frequência do item lexical: Itens frequentes seriam armazenados e processados sob sua forma plena, enquanto os infrequentes seriam representados por morfemas e regras de afixação. No modelo de dupla rota de Schreuder e Baayen (1999) itens de ortografia irregular² ou semanticamente opacos³ possuem representações plenas e, portanto, não são gerados por regras, enquanto que itens regulares e/ou transparentes podem ter ou não uma representação plena dependendo de sua frequência. Itens frequentes (mesmo que morfológicamente complexos) vão tender a desenvolver representações plenas. Todos esses modelos preconizam que o acesso direto à representação plena será mais rápida. No entanto, o processamento de morfemas permite que haja o reconhecimento de itens muito infrequentes e/ou novos.

¹ Exemplos do português para verbos irregulares haveria representações plenas de casa item lexical (faze/faço/fiz), enquanto para os regulares haveria apenas a representação da raiz e das regras de formação de palavras (verbo comer = “com” + “o” presente; primeira pessoa, + “i” pretérito; primeira pessoa).

² Palavras em que a fonologia não pode ser computada diretamente de sua ortografia tendo que ser memorizada (ex.: exótico (z), trouxe (ss), girafa (j), etc.), são conhecidos também como palavras de ortografia profunda.

³ Transparência semântica é um termo relativo à facilidade em que se é possível inferir o significado de uma palavra composta a partir dos significados isolados de suas raízes, ex. a palavra “mata-moscas” é facilmente reconhecida como algo que designa um instrumento para matar moscas. No entanto, o termo “pé-rapado” tem uma relação mais arbitrária/convencional com seu significado (indivíduo pobre).

Três modelos de representação lexical

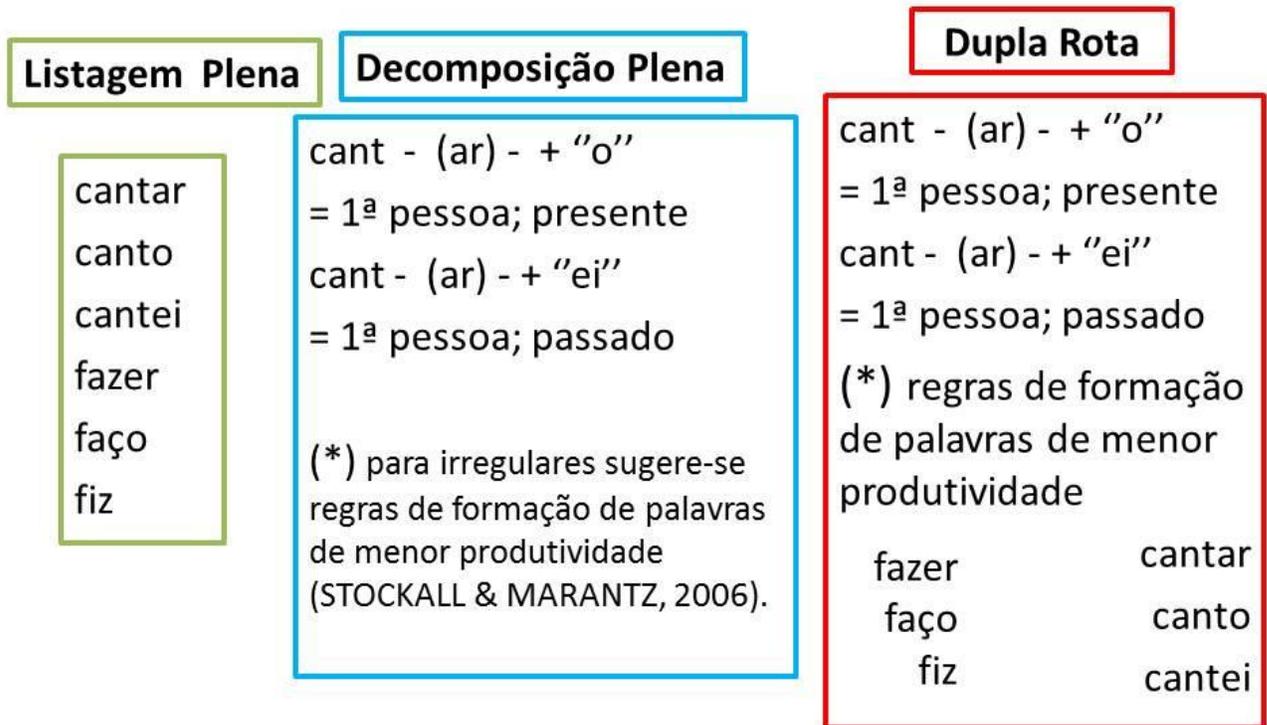


Figura 1. Esquema da representação lexical de cada modelo.

Com um teste de decisão lexical⁴, Taft e Foster (1975) produziram um trabalho seminal em que se verificava a existência de uma rota decomposicional no reconhecimento de palavras e, conseqüentemente, se unidades menores que as palavras possuem representações independentes. No experimento, os autores encontraram latências de resposta mais tardias para bases presas de palavras prefixadas (“*juvenate*” de “*rejuvenate*”) quando comparadas a pseudobases presas, ou seja, partes de palavras que parecem ser afixadas (“*pertoire*” de “*repetoire*”). Da mesma

⁴ A tarefa de decisão lexical é uma avaliação psicológica do quão rapidamente uma seqüência de letras é classificada como uma palavra real (que pertence ao corpo da língua) ou uma não palavra ou, ainda, uma pseudopalavra (palavra construída sob as mesmas regras de fonotáticas de uma língua).

forma, pseudopalavras formadas por uma base presa⁵ verdadeira e um afixo real (“*dejuvenate*”) levam mais tempo para serem rejeitadas que pseudopalavras formadas por uma base presa falsa e um afixo real (“*depertoire*”). Martin (1982, *apud* CARAMAZZA, 1988) relata um efeito de sobreposição em que quanto maior for a coincidência de uma não palavra com uma palavra real, maior será o tempo que ela levará para ser rejeitada. No entanto, no trabalho de Taft e Foster (1975) a sobreposição que os itens possuíam com palavras reais tinha a mesma extensão para ambos os grupos. Dentro deste paradigma experimental modelos de listagem plena não preveem diferenças nos tempos de reação de pseudopalavras, pois nenhuma delas possui uma representação mental e, portanto, deveriam ser rejeitadas similarmente. Portanto, para explicar esse resultado é necessário assumir que características morfológicas (tais como raízes e afixos) são processadas independentemente durante o acesso lexical, pois possuem representações independentes. Uma vez que, nos experimentos relatados, a única diferença entre as não palavras testadas era se tratavam de bases presas legítimas ou não. A partir desses resultados Taft e Foster (1975) propuseram um mecanismo de reconhecimento de palavras - “Affix Stripping” – no qual durante o reconhecimento cada palavra fosse reduzida a suas menores unidades de análise que ainda carregassem significado, e que cada morfema fosse listado (representado) individualmente no léxico mental. Caramazza *et al.* (1988) também sugeriram uma série de experimentos para verificar a existência da rota decomposicional e por a prova o mecanismo “Affix Stripping”. Neste trabalho foi utilizada a metodologia de decisão lexical e foi verificado também especificamente o efeito da presença de estruturas morfológicas em pseudopalavras. Os estímulos⁶ foram parametrizados em quatro categorias: pseudopalavras com raízes verdadeiras + sufixos verdadeiros (“*cantevi*”), pseudopalavras com raízes verdadeiras + sufixos falsos (“*cantovi*”), pseudopalavras com raízes falsas + sufixos verdadeiros (“*canzevi*”), pseudopalavras com raízes falsas + sufixos falsos (“*canzovi*”). Para o modelo de Taft e Foster (1975) os estímulos apresentariam a seguinte gradação nas latências de

⁵ Segmento que carrega significado, mas não constitui uma palavra plena em si, nem apresenta-se independentemente dentro da língua (“*rejuvenescer*”; *re* = prefixo “*tornar/fazer novamente*; *juvenescer* = relativo a jovem).

⁶ Estímulos elaborados a partir de palavras italianas.

resposta do teste de decisão lexical: *cantevi* > *canzevi* > *cantovi* = *canzovi*, pois os itens que possuíssem sufixos falsos seriam rapidamente descartados sem que ao menos fosse verificada a legitimidade das raízes, aqueles com sufixos verdadeiros passariam por uma segunda etapa de verificação da raiz e caso tanto raiz como sufixo se pareassem com uma representação verdadeira do léxico esse item ainda passaria por uma última verificação de legitimidade de combinação desses elementos. O modelo proposto por Caramazza *et al.* (1988) - *Augmented Addressed Morphology* (AAM) - apresenta uma pequena diferença sugerindo que a gradação nas latências de resposta seria: *cantevi* > *cantovi* = *canzevi* > *canzovi*. Por que neste paradigma um item que não possuísse nenhum morfema real seria o primeiro a ser rejeitado, sendo seguido pelos itens que possuíssem ao menos um morfema verdadeiro, e, por último aqueles itens que possuíssem dois morfemas verdadeiros só seriam rejeitados quando sua combinação fosse verificada como ilegal, porque aqueles dois itens possuem representações individuais, mas não existem combinados. Os resultados encontrados seguem o padrão previsto pelo modelo AAM de Caramazza apontando que unidades morfológicas como raízes e afixos também permitem o acesso lexical, e que novas palavras, diferentemente do que propõe o Modelo de *Affix Stripping* não passam inicialmente apenas por um mecanismo de remoção/validação de afixos, pois “*cantovi*” e “*canzevi*” levam o mesmo tempo para serem rejeitados (“*cantovi*” deveria ser prontamente rejeitado), mas aqui apenas possuir uma raiz real fez com que o item levasse mais tempo para ser rejeitado. De acordo com Caramazza *et al.* (1988) itens morfológicamente ilegais seriam rejeitados em um nível “pré-lexical” do processamento enquanto aqueles morfológicamente legais passariam por um nível “lexical” seguinte em que se é determinado se a ativação conjunta daquelas unidades é válida, ou seja, constitui uma palavra na língua.

Trabalhos como o de Penke *et al.* (1997) também corroboram os modelos de dupla rota mostrando resultados específicos a cada tipo de palavra. No referido artigo, com o uso da técnica de extração de potenciais relacionados a eventos (do inglês: *event related potentials* - ERPs⁷), foi encontrada uma resposta eletrofisiológica

⁷ Nesta técnica é apresentado um estímulo enquanto observa-se a atividade com ele sincronizada no eletroencefalograma, o potencial relacionado ao processamento do evento (o ERP). O ERP tem uma amplitude menor que o registro encefalográfico de fundo e é, portanto, geralmente extraído por médias

semelhante, tanto para verbos irregulares do alemão conjugados no passado de forma errônea, quanto para plurais irregulares malconstruídos. Essa resposta eletrofisiológica era uma negatividade que se distribuía na porção anterior do lobo temporal esquerdo (do inglês: *lateral anterior negativity* - LAN⁸) 200ms após estimulação. Em um experimento de *priming*⁹ envolvendo também o processamento de verbos da língua alemã, Sonnenstuhl, Eisenbeiss e Clahsen (1999) encontraram uma quantidade similar de facilitação quando o prime era um verbo regular conjugado no passado e o alvo continha sua raiz. Por outro lado, verbos irregulares conjugados no passado produzem efeito de priming muito menor em relação a sua própria raiz. Por exemplo, “geschlafen” (verbo regular conjugado no passado, “dormir”) facilita a ativação de sua raiz “schlaf” tanto quanto ela própria (“schlaf”). No entanto, “geöffnet” (verbo irregular conjugado no passado, “abrir”) facilita sua raiz “öffn” muito menos que ela mesma “öffn”. Münte *et al.* (1999) demonstraram que, na língua inglesa, verbos regulares não flexionados (“walk”) primados por suas formas flexionadas (“walked”), apresentavam uma resposta eletrofisiológica característica, denominada N400¹⁰, de menor amplitude quando comparados a outros verbos regulares não primados ou quando comparados a verbos irregulares (“teach”) primados por suas formas conjugadas no passado (“taught”). E, ainda, também em relação a outras condições de controle, como palavras que possuíam alguma sobreposição fonológica ou ortográfica como, por exemplo, “card” primando “car”. Essa redução na amplitude do N400 nas formas regulares sugere que houve uma pré-ativação das palavras regulares, fornecendo indícios que elas são decompostas morfológicamente, dessa forma permitindo que as raízes fiquem ativadas, produzindo assim uma facilitação. Pinker e Ullman (2002) trazem casos de lesões neurológicas onde é notável a dissociação no processamento de verbos regulares e

geradas em análises computacionais. As flutuações de voltagem geradas pelos neurônios não envolvidos no processamento do estímulo de interesse (o registro de fundo, que nesse caso chamamos de ruído) serão randômicas em relação ao momento de início do estímulo e, assim, se cancelarão na amostragem média, deixando apenas o registro da atividade relacionada ao evento. Assim, para se obter um *ERP* há de se realizar o teste muitas vezes.

⁸ Uma *LAN* é um *ERP* sensível a processamento sintático e é predominantemente localizada sobre o córtex anterior esquerdo (FRIEDERICI *et al.* 1996).

⁹ A técnica de reativação, ou *priming*, consiste em apresentar um elemento (o prime) e depois, com a apresentação de um segundo elemento (o alvo), verificar se a apresentação do prime facilitou o reconhecimento do alvo.

¹⁰ Negatividade presente no *ERP* que tem sua amplitude máxima cerca de 400ms após a apresentação do estímulo (KUTAS & HILLYARD, 1980).

irregulares. Em um paciente com lesões em regiões perisilvianas do córtex, na porção anterior do seu hemisfério esquerdo, apresentava sintomas de agramatismo, enquanto outro paciente com uma lesão na região temporoparietal esquerda apresentava sintomas de anomia. Resultados dos testes de flexão verbal mostraram que o paciente agramático apresentava maior dificuldade com a flexão de verbos regulares que irregulares, enquanto o paciente anômico apresentava maior dificuldade com a flexão de verbos irregulares e aplicava de forma errônea a regra regular de flexão nos verbos irregulares. Esses autores argumentam que abordagens baseadas apenas em sistemas de associação de padrão (modelos conexionistas como os de listagem plena) não explicam tais dissociações neurais, pois em casos como esses, lesões nas redes deveriam sempre prejudicar de forma mais intensa o processamento dos itens irregulares, devido ao fato deles existirem em menor número.

Todos esses resultados demonstram que existem diferenças entre o processamento de palavras regulares e irregulares, sugerindo assim um possível processamento em paralelo dessas categorias, ou seja, uma dupla rota.

No Português Brasileiro também foi encontrada evidência para a rota decomposicional nos trabalhos de Pessoa (2008) e Garcia *et al.* (2012). Com uma metodologia de decisão lexical, Pessoa verificou que palavras prefixadas (reaplicar) são processadas mais lentamente do que as não prefixadas (resumir). Ou seja, palavras em que o “re” aparece com conteúdo morfêmico são acessadas mais lentamente do que as em que o “re” não aparece como morfema. Garcia, Maia e França (2012), utilizando decisão lexical com *priming*, demonstraram que palavras que partilham da mesma identidade morfológica (i.e. palavras que possuem as mesmas raízes) facilitam o acesso umas das outras. Maia, Lemle e França (2007), com a técnica de rastreamento ocular, relatam evidências tanto para uma rota decomposicional como também para uma outra via de acesso mais direta (uma rota lexical) na leitura de palavras do português brasileiro.

Bertram e Hyona (2003) investigaram se a maneira com a qual palavras do Finlandês são processadas depende de sua extensão. Com uma metodologia de rastreamento ocular e uma tarefa de leitura automonitorada, eles constataram que palavras compostas longas, com primeiros constituintes de alta frequência, produzem

fixações mais curtas se comparadas a palavras compostas longas com primeiros constituintes de baixa frequência. Em palavras compostas curtas uma redução no tempo de fixação só era encontrada se a frequência da própria palavra composta (e não de seus constituintes) fosse alta. Com esse resultado os autores sugerem que o processamento de palavras compostas longas e curtas ocorre de forma diferenciada, sendo, para as primeiras, dependente da frequência do constituinte inicial, pois essa é a informação que é obtida logo com a fixação inicial. Por outro lado, para palavras compostas curtas é a frequência da palavra toda que influi, pois, nesse caso, essa informação pode ser obtida com somente uma fixação.

Em um estudo sobre o grego e o polonês, Kehayia *et al.* (1999) constataram que há decomposição morfológica no reconhecimento de palavras compostas em ambas as línguas. Esse resultado foi obtido pela facilitação no reconhecimento das palavras compostas pelo *priming* de palavras que continham sua primeira ou segunda raiz (ex.: em grego, *domata* “tomate” facilitou o reconhecimento de *domatosalata* “salada de tomate”). Foi verificado, ainda nesse trabalho, que ambas as raízes de uma palavra composta facilitam seu reconhecimento. Contudo, a primeira raiz tem um efeito facilitador mais forte que a segunda, mesmo que essa última tenha mais conteúdo compartilhado com o significado da palavra composta, ou seja, a posição da raiz que possui um maior conteúdo semântico relacionado ao significado da palavra como um todo não influi no seu tempo de reconhecimento (não foi encontrado influência de *Headedness*).

Palavras compostas são a forma mais fácil e eficaz de se criar e transferir novas representações. Pela construção de novos itens lexicais que compartilham o significado daqueles já existentes, novas palavras compostas podem, a princípio, ser compreendidas a primeira vista (LIBBEN, 2006, p. 2.)¹¹.

¹¹ Tradução da autora de: “... *compounding offers the easiest and most effective way to create and transfer new meanings. By building new lexical items upon the meanings of existing items, novel compounds can, in principle, be understood upon first presentation*” (LIBBEN, 2006, p. 2).

Libben (2006) propõe três modelos básicos para a representação de palavras compostas (figura 2): (a) maximização da eficiência computacional; (b) maximização da eficiência de armazenamento e (c) maximização de oportunidade de armazenamento e computação.

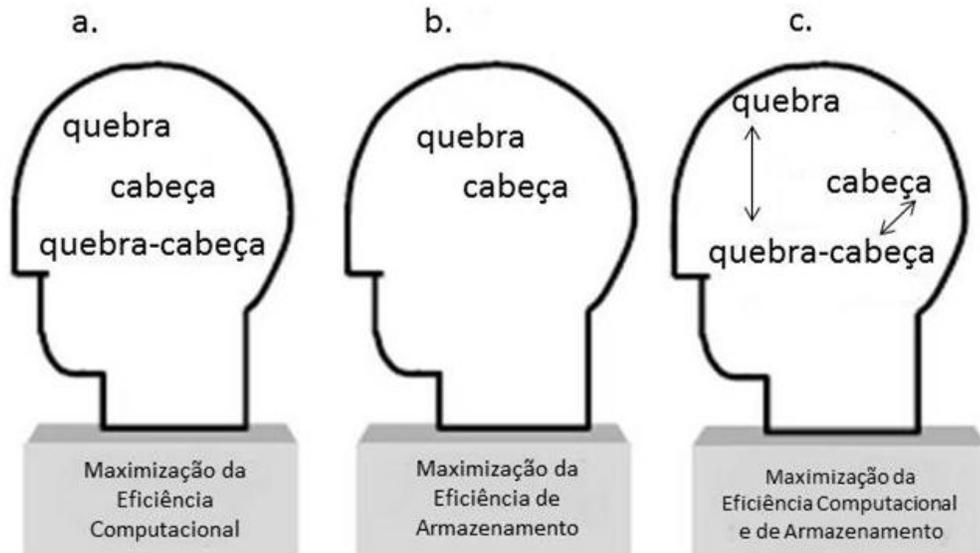


Figura 2. Modelos básicos para a representação de palavras compostas (extraído e adaptado de Libben, 2006). (a) há representação da palavra composta plena independente dos seus constituintes.

(b) apenas os constituintes são armazenados.

(c) constituintes e palavras compostas são representadas com conexões entre si.

Libben (2006) explica que a maximização de oportunidade (figura 2. exemplo c) é a ideia mais parcimoniosa uma vez que trabalhos recentes sugerem isso (JAREMA, 2006) e, também, por haver evidências do vínculo entre palavras compostas e seus constituintes (SANDRA, 1990; MARSLEN-WILSON *et al.*, 1994; ZWITSERLOOD, 1994; LIBBEN, 1998; KEHAYA, 1999; DUÑABEITIA *et al.*, 2009) e, ainda, porque não é possível a todas as palavras compostas ter seu significado depreendido diretamente da soma dos significados de seus constituintes (ex.: pé-rapado, composto opaco). O autor explica que a vantagem de um sistema redundante é que no acesso lexical

“não seria necessário que nada fosse decidido a priori. Ao invés disso, todas as representações que poderiam ser ativadas, seriam ativadas. Dessa

forma, novas palavras só são processadas em termos de seus morfemas constituintes, porque não há representação da palavra plena para ativar, e palavras reais podem apresentar gradações de ativação da palavra plena e ativação dos constituintes. Para palavras muito frequentes, é esperado que a ativação da palavra plena seja mais forte e mais rápida. Para palavras menos frequentes, a rota morfológica pode, de fato, "chegar primeiro." (LIBBEN, 2006, p.9)¹².

Libben (2006) aponta também que uma estratégia lexical padrão (que inicialmente reconheceria as palavras como plenas e só depois, caso necessário, verificaria seus constituintes) não seria eficaz no reconhecimento de novas palavras compostas, principalmente em línguas cuja probabilidade de se encontrar um novo composto é alta. Por outro lado, uma estratégia padrão unicamente de decomposição seria ineficaz por gerar muitos erros relacionados ao quadro de leitura (ex.: "sirigaita" poderia ser lida como "siri" e "gaita").

Fiorentino e Poeppel (2007) investigando também a estrutura das representações lexicais e o(s) seu(s) mecanismo(s) de acesso realizaram um teste de decisão lexical com palavras da língua inglesa. Utilizando palavras compostas de mesma frequência, silabidade e número de letras que palavras simples foi possível investigar a legitimidade da hipotética estrutura morfológica interna das palavras compostas. Pois, estas últimas, caso fossem decompostas (acessadas pelos seus constituintes) apresentariam maior frequência e menor silabidade e número de letras. Algo que poderia lhes conferir facilitação no momento do acesso lexical. Esses autores queriam testar, em especial modelos de dupla rota e modelos de decomposição plena. Eles sugeriram as seguintes hipóteses: para os modelos de dupla rota haveriam dois subtipos: os de decomposição inicial e os de decomposição tardia. Modelos de

¹² Tradução de minha autoria de: *"it does not require anything to be decided. Rather, all representations that can be activated, will be activated. Novel words will only be processed in terms of their constituent morphemes because there is no whole-word representation to activate, and real words may show graded trade-offs between whole-word and constituent activation. For very frequent words, whole-word activation would be expected to be both stronger and faster. For less frequent words, the morphological route might, in fact, 'get there first'.*(LIBBEN, 2006, p.9)."

decomposição inicial (figura 3, modelo I) assumem que ambas as rotas (de decomposição e lexical) são ativadas simultaneamente. Neste paradigma vários fatores (transparência, lexicalidade, produtividade e frequência) influenciariam qual rota será mais veloz de acordo a estrutura da palavra (simples, composta, derivada) e categoria (substantivo, verbo, preposição) (SCHREUDER & BAAYEN, 1995; CARAMAZZA, LAUDANNA, & ROMANI, 1988). Outra versão do modelo de dupla rota seria o modelo de decomposição tardia (figura 3, modelo II), o qual prevê que os constituintes são ativados apenas após o acesso a palavra plena. E, isso, somente em algumas circunstâncias, como, por exemplo, quando há relação semanticamente transparente entre a palavra plena e seus componentes (figura 3, modelo III) (por exemplo, GIRAUDO & GRAINGER, 2000, ver DIEPENDAELE, SANDRA, & GRAINGER, 2005 para maior discussão). Por outro lado, o modelo de decomposição plena prevê que todas as palavras são automaticamente decompostas em seus componentes, e que todo o processamento é feito por via decomposicional (TAFT, 2004; STOCKALL & MARANTZ, 2006).

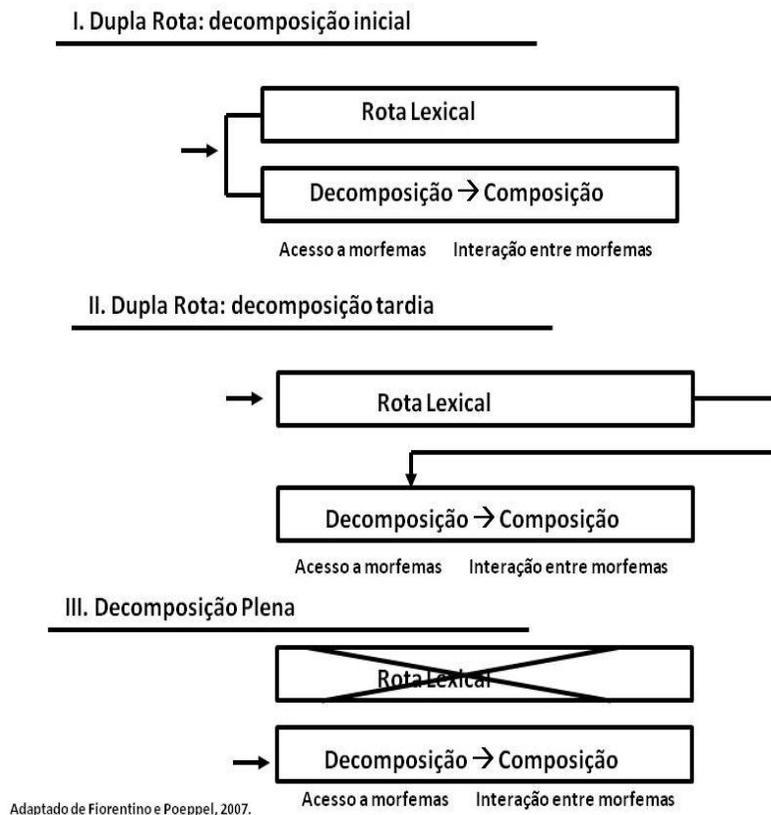


Figura 3. Três modelos básicos sobre o processamento de itens morfológicamente complexos, tais como os compostos.

Fiorentino e Poeppel (2007) encontraram que palavras compostas são avaliadas corretamente, como palavras reais, em taxas mais altas que palavras simples (ou seja, que palavras simples tiveram mais erros) e que também apresentaram tempos de reação menores, independentemente do nível de frequência testado (baixa, média e alta). Ainda nesse mesmo estudo, o componente M350¹³, obtido por magnetoencefalografia (MEG), supostamente envolvido com o acesso lexical, foi o único sensível à manipulação dos estímulos e apresentou um pico inicial anterior para as palavras compostas em relação àquele encontrado nas palavras simples. Dados esses resultados, esses autores supõem que haja diferença qualitativa no processamento dos itens cuja estruturação morfológica é mais complexa, uma vez que

¹³ Equivalente do N400 (que é obtido com eletroencefalografia). Ele é mais precoce, tendo sua amplitude máxima cerca de 350ms após a apresentação do estímulo.

as compostas sempre apresentaram latências de resposta menores, favorecendo modelos que incorporem a decomposição inicial. O fato de haver evidência para decomposição morfológica mesmo em palavras bastante frequentes (que poderiam ter uma representação lexicalizada e, portanto, serem acessadas por uma rota lexical) corroboram os modelos de decomposição plena. Os autores reiteram, ainda, que seus resultados também são compatíveis com alguns modelos de dupla rota que postulem a existência de representações lexicais com estrutura morfológica interna, as quais podem ser acessadas, simultânea e independentemente, através da ativação de constituintes morfêmicos.

Tendo em vista a discussão mencionada nos parágrafos anteriores, o objetivo do presente trabalho é verificar se o processamento de palavras simples difere em relação ao de palavras compostas no Português Brasileiro, assim como ocorre em Fiorentino e Poeppel (2007), com a língua inglesa. Tendo em vista que (i) a identificação das unidades constitutivas das expressões lexicais continua a ser uma questão importante na pesquisa linguística contemporânea (CHOMSKY, 2000), bem como (ii) a verificação da universalidade dos mecanismos de processamento linguístico e/ou sua proporção e também (iii) a mensuração da influência das variações das línguas na arquitetura da mente. Tudo isso em função de contribuir para que sejam elaboradas abordagens e técnicas de ensino mais adequadas para cada tarefa, momento do desenvolvimento e particularidades da língua em questão. Para tanto foram realizados dois testes de decisão lexical. Durante a execução dessas tarefas monitoramos o tempo de reação dos participantes. Responder tais perguntas pode (i) revelar indícios de como são representadas na mente diferentes categorias de palavras, (ii) lançar luz sobre como ocorrem os seus processamentos, e (iii) apontar se diferenças superficiais nas estruturas das línguas influem no processamento linguístico.

OBJETIVOS

Objetivo Geral

Elucidar características dos possíveis mecanismos envolvidos no reconhecimento de palavras simples e compostas do Português Brasileiro.

Objetivos Específicos

- i. Verificar se há diferenças nos tempos médios de reação em testes de decisão lexical para palavras simples e compostas do Português Brasileiro;
- ii. Verificar se há diferenças na acurácia de resposta em testes de decisão lexical para palavras simples e compostas do Português Brasileiro;
- iii. Verificar se essas diferenças se correlacionam com a frequência de uso dessas palavras;
- iv. Verificar qual modelo de representação lexical melhor explica os resultados obtidos.

HIPÓTESES

O presente estudo tem como principal objetivo verificar se a forma como o léxico está estruturado influencia os mecanismos de acesso lexical. Para tanto é necessário buscar os efeitos comportamentais das características dos possíveis mecanismos envolvidos no reconhecimento de palavras simples e compostas do Português Brasileiro.

Apontaremos as seguintes hipóteses:

H0: A representação lexical não leva em conta unidades morfológicas tais como raízes e afixos tendo, portanto, apenas representações plenas. Uma vez dentro da mesma faixa de frequência, os tempos de reação e acurácia são iguais para palavras simples e compostas (a exemplo dos modelos de listagem plena).

H1: A representação lexical conta apenas com unidades morfológicas (tais como raízes e afixos), a computação do sentido da palavra é obtida apenas no momento do acesso lexical. Os tempos de reação serão menores para palavras compostas, pois a análise morfológica dos compostos em dois constituintes de maior frequência isolada e menor número de letras e sílabas lhe concederá facilitação no acesso lexical e sua acurácia de resposta também será maior (a exemplo dos modelos de decomposição plena).

H2: A representação lexical possui formas plenas e unidades morfológicas (tais como raízes e afixos), os tempos de reação dependem da estrutura morfológica, frequência de uso e da densidade da vizinhança ortográfica dos itens. Poderá haver facilitação advinda da análise morfológica dos itens, mas também por influência de uma grande vizinhança ortográfica (com baixas frequências). A acurácia de resposta será maior e o tempo de resposta menor dependendo da estrutura morfológica da palavra, sua faixa de frequência e vizinhança ortográfica (a exemplo dos modelos de dupla rota).

MATERIAIS E MÉTODOS

A fim de elucidar algumas das questões que levantamos nesse trabalho realizamos dois testes de decisão lexical. Em todos os experimentos foram registrados os tempos de reação dos voluntários. Em ambos os testes pretendemos verificar se existe alguma diferença perceptível (pelos métodos utilizados) no modo como palavras simples e compostas são acessadas.

Decisão Lexical

A tarefa de decisão lexical é uma avaliação do quão rapidamente uma sequência de letras é classificada como uma palavra real (que pertence ao corpo da língua) ou uma não palavra¹⁴. Uma sequência de letras é apresentada (em um monitor) ao voluntário que, em seguida, deve decidir, o mais rápido e acuradamente possível, se o item apresentado era uma palavra verdadeira ou não. As latências para responder ao teste são tomadas como seu tempo de reação. Essa medida pode nos dizer algo sobre a complexidade do processamento da tarefa em questão (quanto mais tempo levar, mais complexo diz-se do processamento analisado). Além disso, outra importante medida obtida dessa tarefa é a acurácia de resposta, que é a proporção em que os voluntários responderam corretamente, aceitando as palavras reais e rejeitando as não palavras (que no nosso caso serão pseudopalavras). Foram realizados dois testes de decisão lexical. Ambos com os mesmos parâmetros somente diferindo no conjunto de estímulos. No primeiro teste de decisão lexical havia três níveis de frequência (baixa, média, alta), no segundo apenas um: (alta II).

¹⁴ Nesse estudo todas as não palavras serão pseudopalavras. As pseudopalavras contrastam com as nonsense que fazem uma sequência cuja fonotática é anômala na língua em questão, por exemplo, *buerat*. Já as pseudopalavras por seguirem as regras fonotáticas da língua poderiam ser palavras, por exemplo, *butera*. Na decisão lexical as nonsense são rejeitadas rapidamente, pois logo são reconhecidas como não palavras. Contudo, as pseudopalavras levam mais tempo para serem avaliadas, pois elas têm, ao menos foneticamente, a possibilidade de ser uma palavra real, pela interseção que têm em comum com outros itens lexicais.

Participantes

O primeiro teste de decisão lexical foi realizado por 41 falantes nativos do português brasileiro (24 mulheres), destros, com visão normal ou corrigida, sem problemas mentais declarados, que se voluntariaram a participar destes experimentos. Todos os participantes eram adultos graduados ou graduandos, com idade entre 18 a 30 anos, média de 23,3 anos. O segundo teste de decisão lexical foi realizado por 39 novos participantes (25 mulheres, média de idade de 22,3) com as mesmas características do grupo anterior. Os experimentos foram conduzidos em salas disponibilizadas pelo departamento de Biofísica e Farmacologia da UFRN.

Estímulos

Foi empregado um conjunto de 120 itens no primeiro teste de decisão lexical (30 palavras compostas verdadeiras; 30 pseudopalavras compostas; 30 palavras simples verdadeiras; 30 pseudopalavras simples). As palavras verdadeiras foram retiradas do Corpus Brasileiro, versão 1.0 PUCSP/SP FAPESP disponível no *Sketch Engine*¹⁵ com 1.1 bilhão de palavras (tokens). No segundo experimento de Decisão Lexical novas sete palavras simples verdadeiras e 7 compostas verdadeiras foram selecionadas para a tarefa, para as pseudopalavras escolhemos 7 compostas e 7 simples do primeiro teste de decisão lexical. As 64¹⁶ pseudopalavras dos testes de Decisão Lexical (entre simples e compostas) foram criadas em consonância às regras fonotáticas do PB. As pseudopalavras compostas foram criadas seguindo o modelo utilizado por Fiorentino & Poeppel, sendo o primeiro constituinte em uma palavra real (dissilábica) e o segundo de duas sílabas inventadas (ex.: fila-bera). Uma questão de suma importância no design experimental foi moldar os estímulos de forma que os resultados pudessem ser atribuídos somente à morfologia das palavras (e, mais distalmente, à língua avaliada), para tanto foram parametrizados as seguintes variáveis: todos os itens variavam entre 8

¹⁵ www.sketchengine.co.uk <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>

¹⁶ 60 pseudopalavras para os testes de decisão lexical e mais 4 (duas pseudopalavras simples e duas pseudopalavras compostas) para o treino.

e 10 letras. A maior parte das palavras tinha 4 sílabas e todas as pseudopalavras possuíam 4 sílabas. O número de fonemas das palavras compostas também foi controlado e variavam entre: 4:4; 4:5; 5:4; 5:5.

No teste de Decisão Lexical I havia três níveis de frequência¹⁷: Baixa, Média e Alta para as palavras verdadeiras (simples e compostas). No teste de Decisão Lexical II havia apenas um nível de frequência: Alta II (as palavras desse grupo tinham em média o dobro da frequência do grupo de Alta frequência do teste I). Todas as palavras compostas utilizadas são paroxítonas, ou seja, a sílaba tônica era a penúltima sílaba da segunda palavra da composta (exemplo: vira-**lata**). Também evitou-se, tanto quanto possível, o uso de palavras semanticamente relacionadas, polissêmicas ou homófonas. França (2008) verificou com um experimento de decisão lexical que a extensão da raiz, independentemente da extensão da palavra, influi na latência do seu reconhecimento. Por isso tentamos controlar também essa variável.

Diferentemente das compostas utilizadas no trabalho de Fiorentino & Poeppel (2007) as palavras compostas utilizadas no presente trabalho foram formadas por um processo de composição por justaposição, ou seja, possuem um hífen. Foi feita essa opção devido ao fato de que as palavras do PB formadas por aglutinação tendem a perder letras ou fonemas e/ou são formadas por bases não recorrentes na língua e com isso muitas acabam perdendo a identidade de palavra composta (ex.: aguardente, arquipélago). Seguindo esse raciocínio as compostas por justaposição do PB se assemelham muito mais as compostas do inglês que as formadas por aglutinação, pois nelas ambos os constituintes mantêm todas as suas letras e são bases livres¹⁸.

¹⁷ Avaliada pelo número de ocorrências no Corpus Brasileiro. É importante ressaltar ainda que essas categorias (baixa, média e alta) não se aplicam a totalidade do corpus, mas sim que dentro do conjunto de palavras selecionadas há uma gradação de frequência. Ou seja, essa frequência é relativa. Uma palavra do nosso grupo de alta frequência não será categorizada como tal se comparada a todo o resto do corpus, mas sim quando contrastada ao conjunto de palavras selecionadas para os experimentos.

¹⁸ Sequências fônicas que constituem palavras nas línguas.

Procedimento

Após ter sido orientado que teria que julgar a legitimidade de uma sequência de letras, o voluntário se sentava de frente a tela do computador e permanecia só durante todo o experimento (cerca de 10 minutos), em uma sala silenciosa de iluminação e temperatura confortáveis. O experimento era executado por uma rotina¹⁹ de MATLAB²⁰ que gerava uma tela com fundo negro onde eram apresentadas as palavras e as pseudopalavras e era registrada a resposta e o tempo de reação do participante. Em ambos experimentos uma cruz (+) de fixação aparecia no centro da tela por 500ms, logo em seguida uma sequência de letras (fonte em caixa baixa, tamanho 30) surgia e permanecia por 1 segundo, mas mesmo após sua saída o voluntário ainda tinha mais 2 segundos para dar sua resposta antes que a próxima palavra chegasse. A distância do participante em relação ao monitor era de aproximadamente 45 centímetros. O voluntário deveria responder apertando a tecla ← para as pseudopalavras e a tecla → para as verdadeiras. Antes de iniciar o experimento de decisão lexical havia um treino²¹ que poderia ser repetido caso o voluntário assim desejasse. A ordem em que os elementos eram apresentados era aleatória.

Todos os procedimentos acima descritos foram previamente autorizados pelo Comitê de Ética em Pesquisa da UFRN, através do parecer favorável 165/2011.

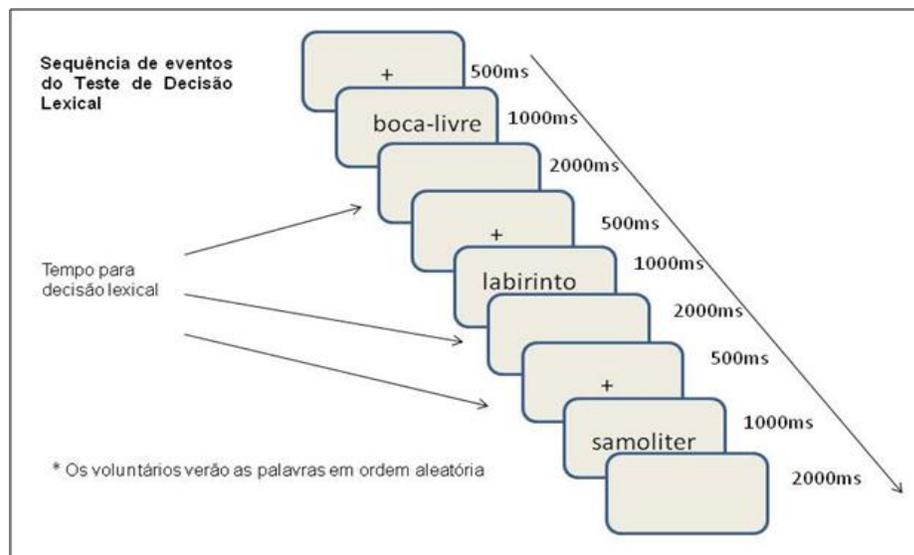


Figura 4. Sequência de eventos e janela temporal do experimento de decisão lexical.

¹⁹ desenvolvida no Instituto do Cérebro por José Henrique Targino Dias Góis.

²⁰ MATLAB (R2010a) Software (MATLAB@) - Mathworks.

²¹ No treino havia duas palavras de cada condição que não foram utilizadas nos experimentos.

Análises dos dados

Foram removidas as respostas que excediam o limite de dois desvios padrão da média de cada condição experimental (compostas de baixa; média; alta frequência; simples de baixa; média; alta frequência; pseudopalavras compostas; pseudopalavras simples) por sujeito e por item, separadamente em relação ao tempo de reação e acurácia de resposta. No teste de decisão lexical I foram retirados dados de quatro participantes ($n = 41$, - 9,7% dos dados) no tempo de reação e acurácia de resposta. No teste de decisão lexical II foram retirados também dados de quatro participantes ($n = 39$, -10% dos dados) no tempo de reação e apenas de dois na acurácia de resposta ($n = 39$, -5% dos dados). No teste de Decisão Lexical I foram removidos também dados de tempo de reação de quatro palavras (uma em cada seguinte condição: composta de baixa frequência; composta de média frequência; simples de baixa frequência; simples de média frequência). Enquanto para análise da acurácia de resposta foram removidas seis palavras (uma composta de baixa frequência; duas compostas de média frequência; duas simples de baixa frequência; uma simples de média frequência). Somente foram analisados dados dos participantes que obtiveram mais de 75% de acerto na tarefa. As análises estatísticas foram realizadas no SPSS²². Para a análise dos dados quantitativos foram realizadas ANOVAS e Testes T com correção de Bonferroni. Para os dados categóricos Testes de Wilcoxon com correção de Bonferroni.

²² The IBM® **SPSS**® Statistics 19.

RESULTADOS

Teste de Decisão Lexical I

Análises dos tempos de reação por classe de palavra

Em uma análise inicial dos dados com uma ANOVA de medidas repetidas de dois fatores (Status X Classe de Palavra), com dois níveis para cada fator (Verdadeira X Pseudopalavra; Composta x Simples), não foi encontrada diferença entre os tempos de reação da classe de palavras compostas (M 0,963 ms) em comparação com as simples (M 0,974 ms) ($F(1,36)=2,826$, $p=0,101$), mas sim no status da palavra: comparando as palavras verdadeiras (M 0,869 ms) com as pseudopalavras (M 1,067 ms) ($F(1,36)=86,442$, $p<0,001$), um efeito de lexicalidade²³. Foi encontrada também uma interação significativa entre os dois fatores ($F(1,36)=18,525$, $p<0,001$).

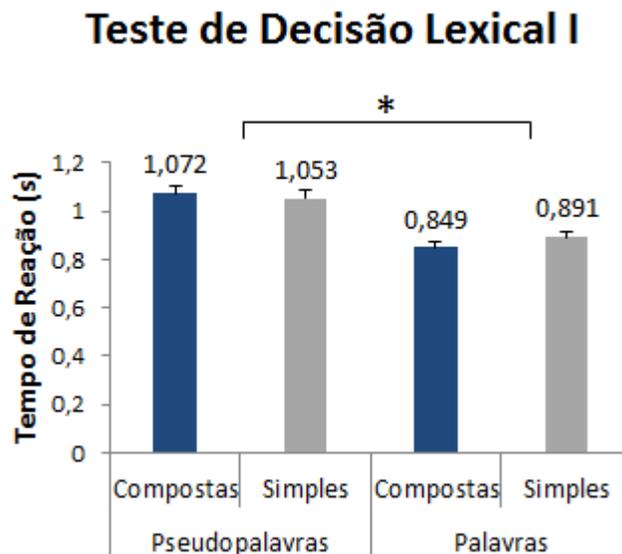


Gráfico 1. Tempo de reação de palavras compostas e simples, verdadeiras e pseudopalavras no experimento de decisão lexical I.

²³ Palavras reais, por possuírem uma representação lexical, são reconhecidas mais rapidamente que não palavras.

Análises dos tempos de reação nos três níveis de frequência

Foi realizada uma ANOVA de medidas repetidas de dois fatores (Classe de Palavra X Frequência) com cada fator tendo respectivamente dois e três níveis (Palavra Composta x Palavra Simples; Baixa X Média X Alta Frequência). Esse teste estatístico revelou diferença significativa para Classe de Palavra ($F(1,36)=37,386$, $p<0,001$), para Frequência ($F(2,72)=73,988$, $p<0,001$) e em uma interação desses dois fatores ($F(2,72)=2,297$, $p<0,001$) como podemos ver no seguinte gráfico:

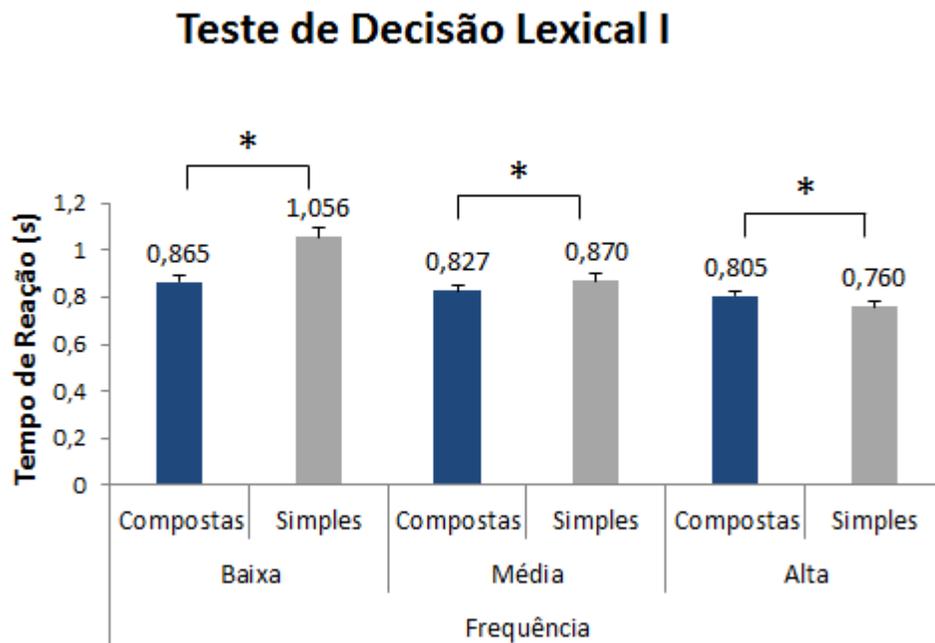
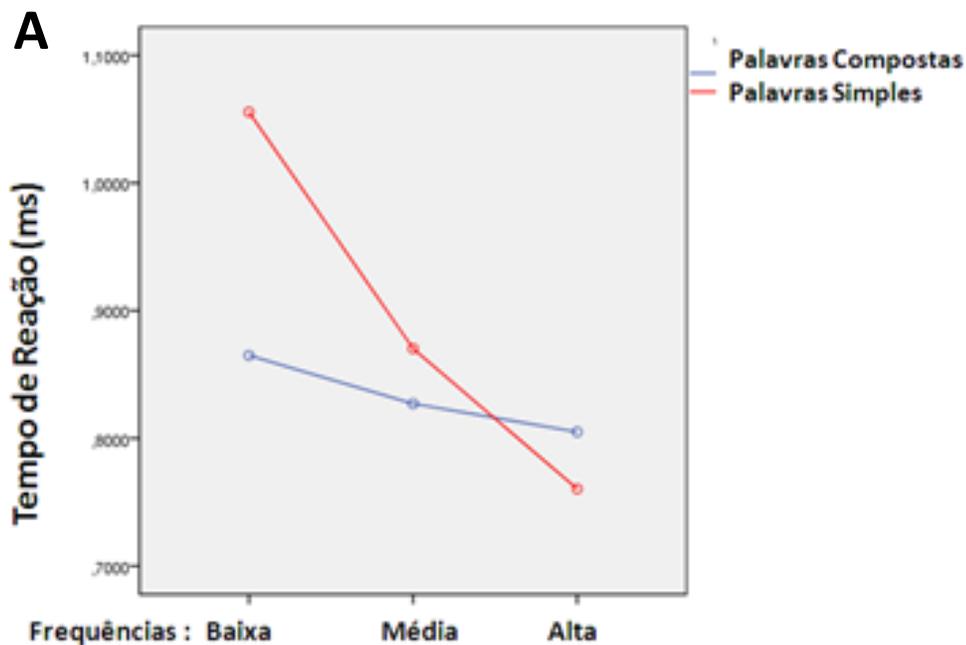


Gráfico 2. Tempo de reação de palavras verdadeiras compostas e simples nos três níveis de frequência do experimento de decisão lexical I.

Testes *post hoc* de comparação dos grupos de palavras em cada nível de frequência mostraram haver uma diferença significativa das médias dos tempos de reação em todos os níveis: na baixa frequência palavras compostas (M 923 ms) tiveram um tempo de reação médio menor se comparadas as palavras simples (M 1,113 ms) ($t=-8,888$, gl (36), $p < 0,001$). O mesmo se aplica para a frequência média: palavras compostas (M 884 ms), palavras simples (M 925 ms) ($t= - 2,667$, gl (36), $p = 0,011$). Contudo, na alta frequência as palavras simples (M 821 ms) é que apresentaram um tempo médio menor em comparação com as compostas (M 867 ms) ($t= 3,490$, gl (36), $p = 0,001$). Com a correção de Bonferroni com o $p < 0,016$ todos os três testes foram significativos. O gráfico 3a mostra a discrepante influência da frequência entre palavras simples e compostas.



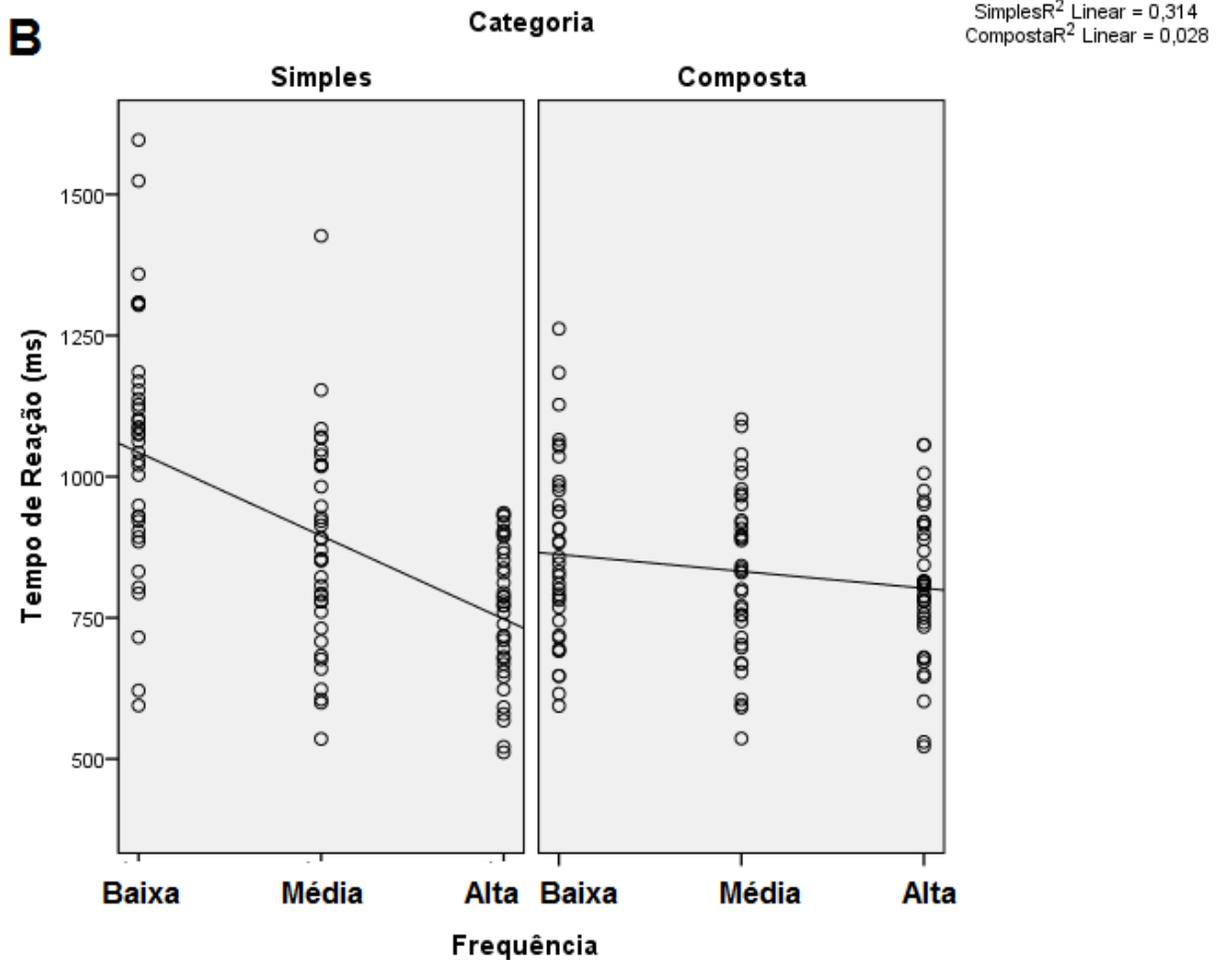


Gráfico 3. **A.** Tempo de reação de palavras verdadeiras (compostas e simples) no experimento de decisão lexical I. **B.** Correlação do tempo de reação com a frequência.

Uma avaliação mais profunda entre os fatores categoria de palavra e frequência revela que para palavras simples existe uma correlação moderada que é responsável por 31% da variação dos dados ($r = -0,560$, $n=111$, $p < 0,001$, bicaudal). Essa correlação é negativa, a medida que a frequência aumenta os tempos de reação diminuem. Enquanto para palavras compostas não há correlação ($r = -0,166$, $n=111$, $p = -0,081$, bicaudal).

Análises da acurácia de resposta por classe de palavra

Para avaliar a taxa de acerto das respostas, que são dados de distribuição binomial, decidimos utilizar o teste Wilcoxon, que é o equivalente não paramétrico do teste t , indicado para designs de pares de medidas repetidas como esta. Após a correção de Bonferroni de $p < 0,025$, confirmamos que houve diferença significativa entre ambos os pares: pseudopalavras compostas (92,9% acerto) X pseudopalavras simples (95,8% acerto) ($z=3,317$, $N - \text{empates} = 11$, $p=0,001$, bicaudal) e para as palavras simples (96% acerto) X palavras compostas (97,8 % acerto) ($z=2,449$, $N - \text{empates} = 6$, $p=0,014$, bicaudal).

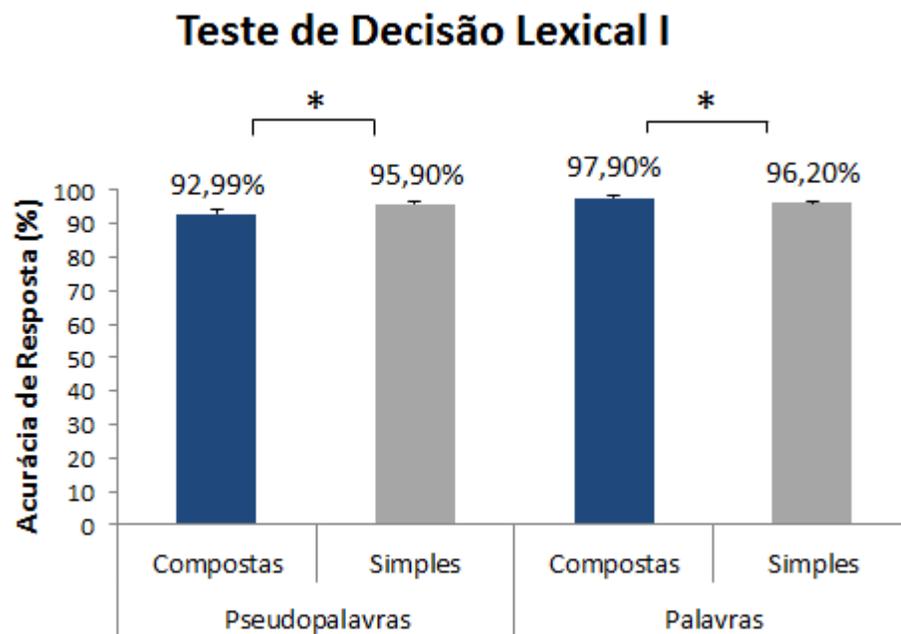


Gráfico 4. Acurácia de resposta de palavras compostas e simples, verdadeiras e pseudopalavras no experimento de decisão lexical I.

Análises da acurácia de resposta nos três níveis de frequência

Neste nível de análise pelo teste de Wilcoxon de medidas pareadas o resultado foi significativo apenas na baixa frequência, com maior taxa de acerto para as palavras compostas ($z=5,099$, N - empates = 36, $p<0,001^{24}$, bicaudal). Mas não houve diferença significativa na média ($z=1,414$, N - empates = 2, $p<0,157$, bicaudal) e na alta frequência ($z=1,732$, N - empates = 3, $p<0,083$, bicaudal).

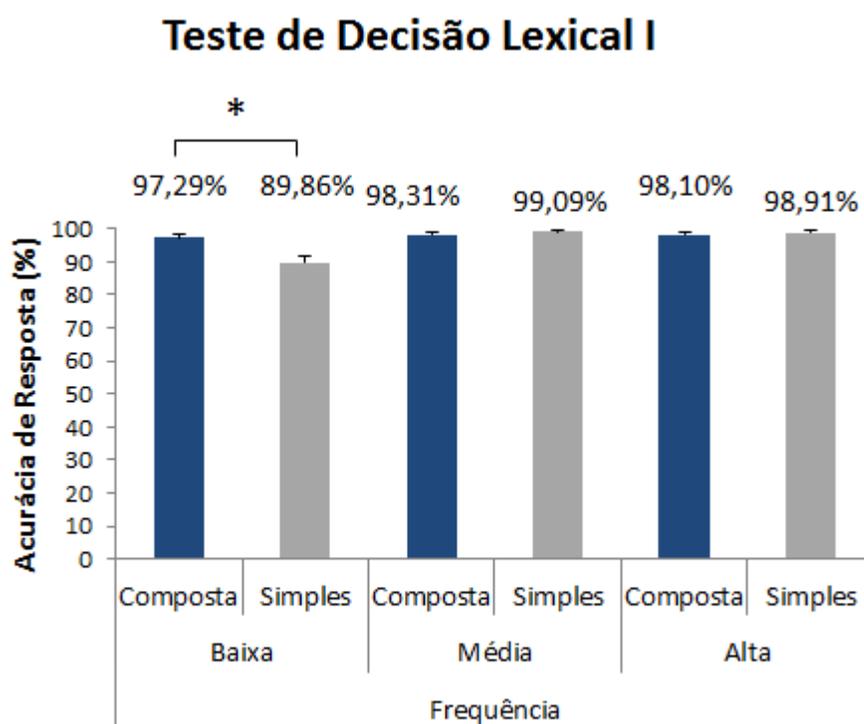


Gráfico 5. Acurácia de resposta de palavras verdadeiras compostas e simples nos três níveis de frequência no experimento de decisão lexical I.

²⁴ Significativo mesmo com a correção de Bonferroni com $p=0,016$.

Experimento de Decisão Lexical II

Análises dos tempos de reação por classe de palavra

Nesse experimento as palavras simples e compostas empregadas tinham o dobro da frequência do grupo de alta frequência do experimento de decisão lexical I. Uma ANOVA de medidas repetidas de dois fatores (Status X Categoria), com dois níveis para cada fator, respectivamente, (Verdadeira X Falsa; Composta X Simples) demonstrou efeitos significativos para Status ($F(1,34)=35,883$, $p<0,001$) e para Categoria ($F(1,34)=13,286$, $p=0,001$), mas não para uma interação entre esses dois fatores ($F(1,34)=2,175$, $p=0,149$). Contudo, testes *post hoc* demonstraram haver diferença significativa entre apenas os tempos de reação médios das palavras verdadeiras: com as palavras simples apresentando um tempo de reação médio menor (M 0,867 ms) que as compostas (M 0,950 ms) ($t=3,349$, $gl= 34$, $p=0,002$). Correção de Bonferroni: $p=0,025$.

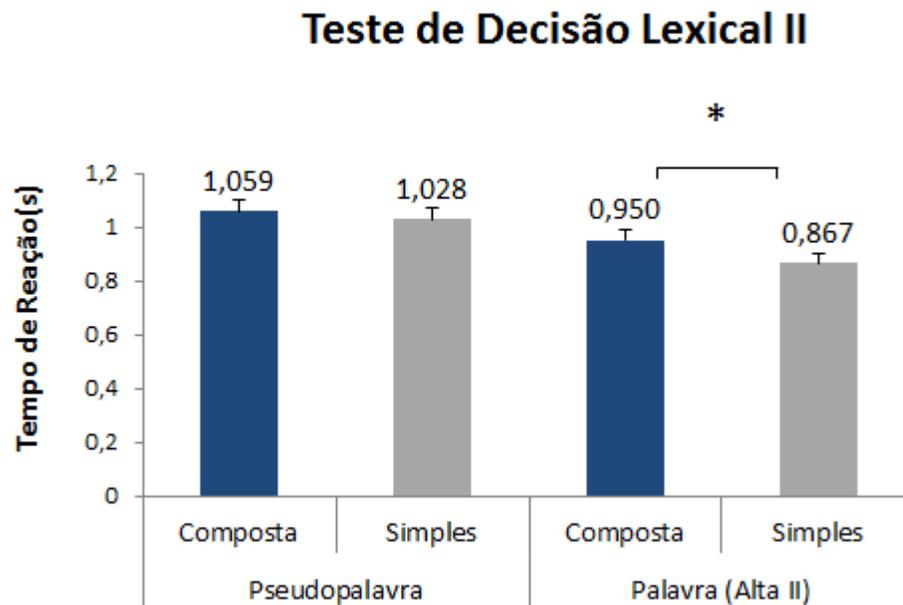


Gráfico 6. Tempo de reação de palavras compostas e simples, verdadeiras e pseudopalavras no experimento de decisão lexical II.

Análise da acurácia de resposta

O teste Wilcoxon somente encontrou uma diferença significativa na taxa de acerto de resposta entre o par de palavras verdadeiras: compostas (95% acerto) X simples (99% acerto) ($Z=2,828$, $N - \text{empates} = 8$, $p=0,005$). Correção de Bonferroni: $p=0,025$.

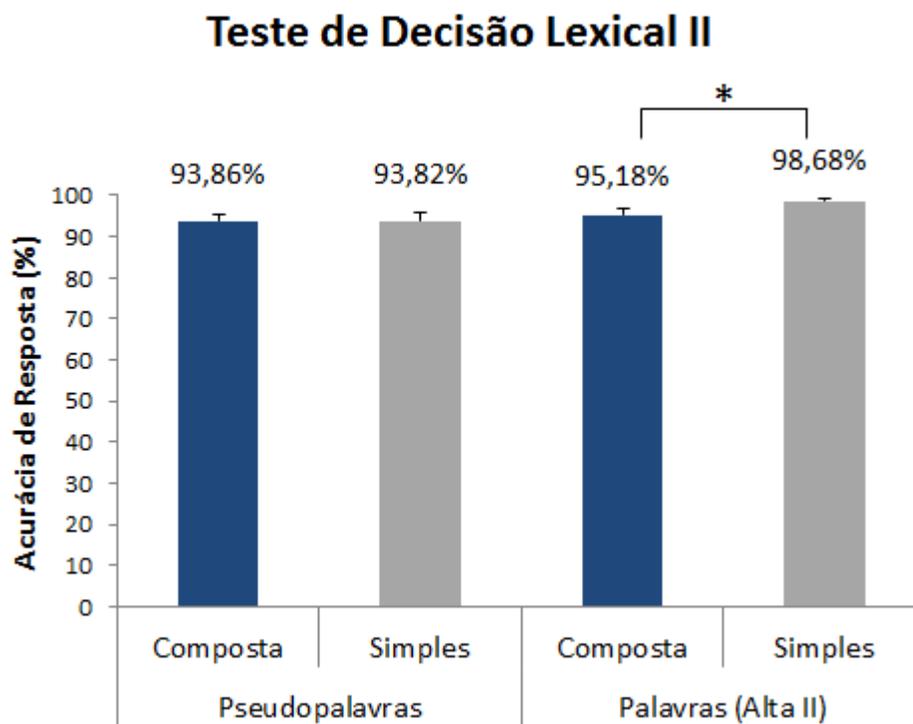


Gráfico 7. Acurácia de resposta de palavras compostas e simples, verdadeiras e pseudopalavras do experimento de decisão lexical II.

DISCUSSÃO

Os resultados obtidos com os testes de decisão lexical I e II fornecem evidências de que há um papel para a decomposição no reconhecimento de palavras também no Português Brasileiro, pois na baixa e média frequência as palavras compostas tiveram tempos de reação significativamente menores que as simples. No entanto, os tempos de reação das palavras simples nas altas frequências dos experimentos de decisão lexical I e II se coadunam com as previsões de modelos de listagem plena (BYBEE, 1995), que sugerem que com o aumento da frequência de uma palavra, suas representações lexicais se fortaleçam. Portanto, a melhor forma de explicar esses resultados é admitir a existência de uma dupla rota, em que cada tipo de palavra será reconhecido mais rapidamente dependendo da sua estrutura morfológica e sua frequência (figura 5).

Em relação a acurácia de resposta, apesar de somente haver diferença significativa nos extremos de frequência, os resultados também sustentam a hipótese da dupla rota, pois na baixa frequência do experimento de decisão lexical I houve uma maior taxa de acerto para as compostas enquanto na alta frequência do experimento de decisão lexical II houve uma maior taxa de acerto para as simples. Os resultados de Frost (1994), com decisão lexical, também ilustram que quando um item depende mais da rota lexical para ser reconhecido ele tem seu tempo de resposta aumentado na baixa frequência. Esse autor encontrou latências maiores para itens fonologicamente irregulares²⁵ quando comparados a outros regulares. Os resultados de Lehtonen *et al.* (2003, 2006) (onde examinou-se, respectivamente, o finlandês e o sueco) estão em perfeita consonância ao presente trabalho, pois eles também encontraram evidência para decomposição morfológica para substantivos flexionados de baixa frequência, enquanto para aqueles de alta frequência os autores sugerem que foi empregada uma rota lexical.

²⁵ Irregular aqui se refere a ortografia e não a conjugação.

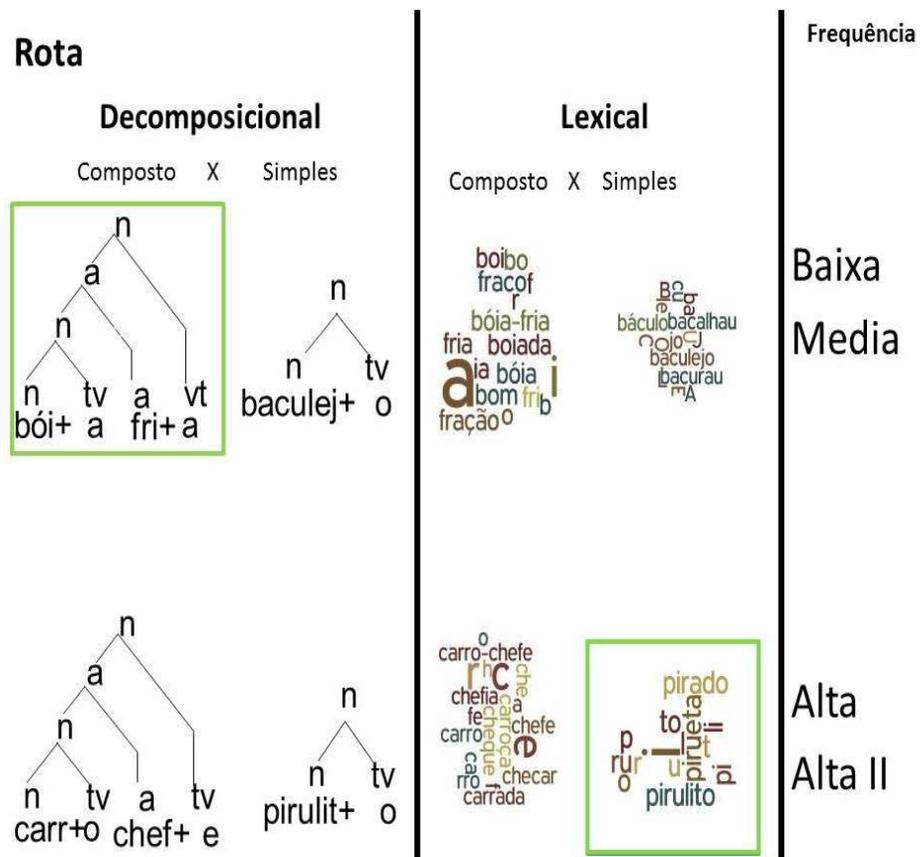


Figura 5. Esquema das rotas Decomposicional e Lexical nos níveis de frequência. Os resultados deste trabalho sugerem que empregamos dois mecanismos no reconhecimento de palavras do PB: uma rota decomposicional que procura por unidades morfológicas tais como raízes e afixos e uma rota lexical que é mais sensível a frequência e a vizinhança ortográfica das palavras²⁶. Contudo, é provável que tanto palavras simples quanto complexas são analisadas simultaneamente por uma rota lexical e outra decomposicional. A rota que será mais veloz/eficaz para cada tipo de palavra (composta ou simples) parece ser resultado da computação da extensão e frequência dos morfemas das palavras, mas também do peso de cada vizinhança ortográfica somada a frequência desses vizinhos.

Atkinson e Joula (1973, *apud* BALOTA & CHUMBLEY, 1984) propuseram que, na tarefa decisão lexical, uma sequência de letras poderia apresentar uma medida de familiaridade e significação (do inglês “familiarity and meaningfulness dimension”) que

²⁶ Originalmente um vizinho ortográfico é definido como uma palavra de mesma extensão que outra que difira de outra em apenas uma letra. Neste trabalho estende-se essa definição a palavras com sobreposição.

se baseia na sua sobreposição fonológica e ortográfica a outros itens lexicais. O contínuo de distribuição desses valores apresenta uma margem de sobreposição onde algumas pseudopalavras podem apresentar um valor que a leve a ser interpretada como uma palavra real, enquanto uma palavra real pode ser tomada por uma falsa. Os resultados dos grupos de baixas frequências desse presente trabalho se coadunam com as predições do modelo de reconhecimento de memória de Atkinson e Joula. Visto que palavras simples de baixa frequência apresentaram latências de resposta maiores em relação a palavras compostas. Contudo, uma vez que nas altas frequências a medida de familiaridade e significação pareça ser sobrepujada pelos efeitos da frequência, mantemos a versão da dupla rota. A figura 6 é um esquema da relação dos itens compostos e simples com suas respectivas vizinhanças ortográficas onde sugere-se que itens simples de baixa frequência se assemelhem muito a não palavras por ter uma vizinhança menor. Enquanto itens compostos teriam uma familiaridade e significação²⁷ maior tendo em vista que são duas outras palavras juntas. É importante ressaltar, no entanto, que apesar dos itens compostos das altas frequências também serem elementos formados por duas unidades sua medida de familiaridade e significação provavelmente seja menor que a das palavras simples, uma vez que seus tempos foram maiores nesses níveis de frequência indicando que essa medida é influenciada não somente por estrutura morfológica, mas também outras variáveis aqui não contempladas.

²⁷ como proposto por Atkinson e Joula (1973).

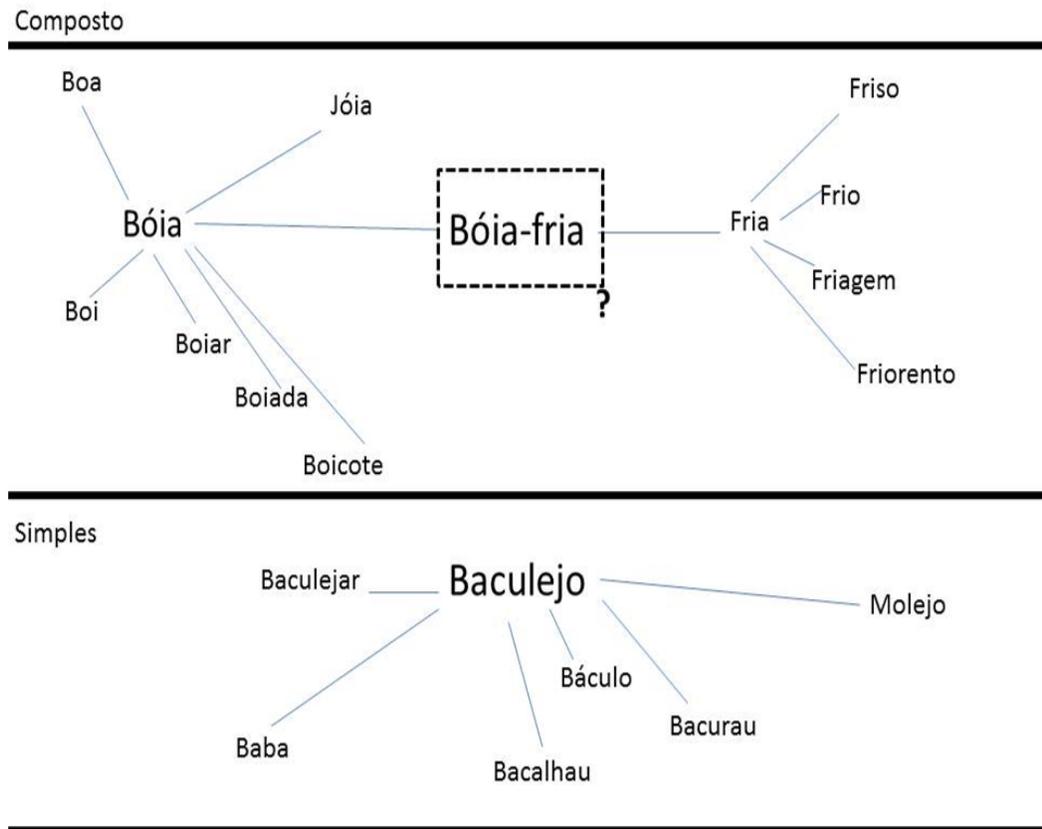


Figura 6. Esquema da vizinhança ortográfica de dois itens de baixa frequência.

Figura 6. Esquema da vizinhança ortográfica de dois itens de baixa frequência: um composto e um simples. Nas baixas frequências a palavra composta aparentemente está inserida em uma vizinhança ortográfica maior, tendo, portanto, uma familiaridade maior que a palavra simples.

A diferença entre os resultados do presente trabalho em relação aos obtidos com a língua inglesa (FIORENTINO & POEPEL, 2007), sob os mesmos paradigmas experimentais, talvez seja reflexo da menor produtividade que a composição tem no Português Brasileiro. Feldman *et al.* (2007) argumenta que, ao menos no que se refere à decisão lexical visual, as semelhanças entre as línguas são maiores que as divergências e que estas últimas podem derivar das diferenças na densidade de vizinhos ortográficos, extensão média da palavra, produtividade dos processos de derivação, flexão, composição e tamanho de família morfológica²⁸. Ela cita que, por exemplo, em uma tarefa de decisão lexical, palavras do Espanhol de mesma frequência

²⁸ Número de palavras formadas com um mesmo morfema (ex.: beijar, beijou, beijando, beijo, beijoca, beijoqueiro, beija-flor, etc.).

com grandes vizinhanças ortográficas foram reconhecidas mais rapidamente que aquelas com menor densidade de vizinhos (PEREA & ROSA, 2000). Baayen (2007), citando Baayen e Moscoso del Prado Martin (2005), argumenta que verbos irregulares²⁹ tendem a ter mais vizinhos semânticos que verbos regulares e que em Tabak *et al.* (2005a) as diferenças de processamento encontradas também foram atribuídas a esse fator. Libben (2006) sugere um modelo de representação lexical em que ocorra a “maximização de oportunidade”. E que, para o caso específico dos compostos, haja representações da forma plena e de seus constituintes independentemente, mas que podem ser relacionados devido a transparência semântica. Baayen (2007) menciona uma série de trabalhos onde palavras regulares morfologicamente complexas demonstraram possuir representações plenas (TAFT, 1979; SERENO & JONGMAN, 1997; BAAYEN *et al.*, 1997, 2002; BIEN *et al.*, 2005). Bem como a evidência da coexistência de formas regulares de presente e passado em Stemberger e Middleton (2003). Em consonância, os resultados de Alegre e Gordon (1999) indicam que somente palavras regulares infrequentes sejam processadas por uma rota decomposicional.

Em seu trabalho Fiorentino e Poeppel (2007) verificaram que sempre (em todos os níveis de frequência) os itens morfologicamente mais complexos (os compostos) eram reconhecidos mais rapidamente quando comparados aos simples, sugerindo que a análise morfológica dos compostos (em dois constituintes de maior frequência isolada e menor número de letras e sílabas) lhe concedia alguma facilitação no acesso lexical. No entanto esse tipo design experimental pode ocultar a presença de uma rota alternativa (lexical) que seja revelada não apenas em função da frequência, mas também de fatores tais como proporção de vizinhos ortográficos, suas respectivas frequências, número médio de letras por palavra, etc. Todos esses fatores idiossincráticos das palavras isoladamente e de cada língua. Consoante a essa hipótese são os resultados de Bertram, Laine e Karvinen (1999) e Traficante e Burani (2003), respectivamente com o Italiano e o Finlandês, em que o processamento de itens morfologicamente complexos foram mais lentos quando comparados com aqueles mais estruturalmente simples. Assim como os compostos dos grupos de altas

²⁹ Dependentes principalmente da rota lexical.

frequências deste presente trabalho. Da mesma forma, Wurm e Baayen (2005) encontraram efeitos de frequência em itens regulares pertencentes a vários níveis de frequência, assim, contestando a alegação de que o sistema de processamento empregado depende da frequência do item em questão. Feldman *et al.* (2007), defende que as rotas lexical e decomposicional se contrabalanceiam considerando a influência da frequência de um morfema constituinte e a frequência da palavra plena. Essa autora também afirma que em relação ao processamento morfológico é inadequado se limitar a observar apenas a presença ou ausência de efeitos ao invés de examinar sua gradação.

Modelos de listagem plena, apesar de esclarecer os efeitos de frequência, não explicam como os falantes reconhecem novas palavras formadas a partir de morfemas já conhecidos (CARAMAZZA *et al.*, 1988; TAFT & FOSTER, 1975), nem dão conta do processamento de itens novos que não se assemelhem a outros já existentes, mas uma regra pode ser aplicada a qualquer item (PINKER & ULLMAN, 2002). Para modelos decomposicionais a identificação de padrões observados recorrentemente em diferentes itens lexicais faz que essa redundância seja caracterizada pelo sistema como uma regra. Pinker e Ullman (2002) reiteram que o caminho inverso poderia ser feito, itens lexicais irregulares surgiriam, diacronicamente, pela memorização de um item complexo formado com várias camadas como um todo indivisível³⁰. Uma evidência para isso são os resultados de Alegre e Gordon (1999) onde foram encontrados efeitos de frequência para substantivos flexionados somente acima de um limite de frequência de cerca de seis ocorrências por milhão, enquanto para itens morfológicamente simples esses efeitos foram encontrados em todas as faixas de frequência.

Portanto, mesmo modelos de dupla rota podem ainda ser bastante radicais por não pressupor plasticidade e gradação ao processo de representação mental das palavras. Há muitos itens lexicais que não estão completamente inseridos em grupo fechado, principalmente no que se refere a compostos (FIGUEIREDO SILVA & SELL, 2011). Estes itens podem ser percebidos diferentemente dependendo das exigências de cada tarefa (decisão lexical, nomeação, etc.), das experiências distintas de cada

³⁰ Por isso mesmo itens ditos irregulares do inglês apresentariam padrões de formação (Shake - Shook - Shaken / Take - Took - Taken / Goose - Geese - Tooth - Teeth).

pessoa³¹ ou mesmo tendências naturais particulares a um processamento mais analítico ou holístico.

Muitos trabalhos apontam a influência da consciência fonológica³² para o aprendizado da leitura e da escrita (GOUGH, JUEL, GRIFFITH, 1986; STANOVICH, CUNNINGHAM, CRAMMER, 1984; CAPOVILLA, 1999, 2000; GUIMARÃES, 2003). Críticos da educação (ADAMS *et al.*, 2003, SCLIAR-CABRAL, 2003) defendem que o emprego massivo do método construtivista de alfabetização em detrimento ao tradicional método fônico, a partir da década de 1980, é um dos fatores responsáveis pela baixa taxa de alfabetização plena em nosso país³³. Tendo em vista nossos resultados favoráveis à decomposição no reconhecimento de palavras do Português Brasileiro, ou seja, uma vez que aspectos morfológicos têm implicações no acesso lexical, sugerimos que atividades específicas à compreensão da estrutura morfológica das palavras, assim como as fonológicas, sejam cruciais para o desenvolvimento de habilidades de leitura e escrita. Baltazar (2012) traz um exemplo de como o aprendizado de segundas línguas, âmbito em que se é notória a negligência à aspectos morfológicos, também pode se beneficiar da manipulação desse nível de estruturação linguística. Em seus resultados falantes aprendizes de língua inglesa processaram verbos regulares de alta frequência de forma não decomposicional³⁴, mas verbos irregulares sim. Sugiro que esse efeito paradoxal talvez seja devido a maior ênfase que se dá nessa situação educacional aos padrões morfológicos que são encontrados nos verbos irregulares (vide nota de rodapé 27) de forma a auxiliar no aprendizado.

Diante do exposto algumas das possíveis perspectivas futuras deste trabalho é verificar a influência da transparência semântica ou ordem dos constituintes no grau de decomposibilidade das palavras compostas do PB através de tarefas de *priming* ou nomeação. Nessa mesma direção, a reprodução dos experimentos aqui descritos, somada à realização do registro eletroencefalográfico se faz necessário para a

³¹ Exemplo: método de alfabetização.

³² Consciência fonológica é a habilidade de segmentar a língua falada em unidades (i.e. palavras, sílabas e fonemas) e perceber que elas são recorrentes (BYRNE, FIELDING-BARSLEY, 1989 *apud* STAMPA).

³³ De acordo com estimativas do Indicador de Analfabetismo Funcional (INAF, 2011) apenas ¼ da população brasileira alcançou a alfabetização plena (capacidade de compreender e interpretar textos; analisar e relacionar suas partes; distinguir fato de opinião, realizar inferências e sínteses). O estudo também relata que esse valor manteve-se estagnado nos últimos 10 anos.

³⁴ A autora sugere que as regras morfológicas envolvidas na produção do pretérito alomorfe “ed” (que possui três pronúncias diferentes (/id/,/t/ e /d/) ainda não haviam sido internalizadas por esses falantes.

complementação das medidas comportamentais e a confirmação de que estes resultados refletem os mesmos correlatos neurais tal como os previamente descritos na literatura. Além disso, seria oportuna a avaliação de janelas temporais mais precisas, sem a influência de vieses de tarefas extralinguísticas como o aperto de botão.

CONCLUSÃO

Nossos resultados sugerem que o acesso lexical no Português Brasileiro emprega dois mecanismos de processamento: uma via decomposicional e outra lexical, dependente de frequência. Atribuo a divergência encontrada neste trabalho em relação ao de Fiorentino e Poppel (2007), realizado com a língua Inglesa, a fatores particulares de cada item lexical, sua frequência de uso, número de vizinhos ortográficos e produtividade do processo de formação de palavras por composição na nossa língua. Todos esses elementos contribuem para modular qual rota será mais veloz/eficiente. Tais resultados também sugerem que, ao menos no que se refere às palavras compostas, pode-se inferir que há alguma representação em nível morfológico.

REFERÊNCIAS

- ADAMS, M. J.; BEARD, R.; CAPOVILLA, F.; CARDOSO-MARTINS, C.; GOMBERG, J. E.; MORAS, J.; ARAÚJO E OLIVEIRA, J. B. (2003). Relatório Final de Grupo de Trabalho Alfabetização Infantil: os novos caminhos. Brasília, DF: Congresso Nacional, 2003. Disponível em: <http://www2.camara.gov.br/comissões/cec/relatórios/Relat_Final.pdf> Acesso em: 5 jan. 2013.
- ALEGRE, M.; GORDON, P. (1999). Rule based versus associative processes in derivational morphology. *Brain and Language*, 68, 347-354.
- ANDERSON, M. C. (2009). Retrieval. In: BADDELEY, A.; EYSENCK, M. W.; ANDERSON, M.C. *Memory*. New York: Psychology Press.
- ATKINSON, R.C.; JUOLA, J.F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and Performance* (Vol.4). New York: Academic Press.
- BAAYEN, H. & SCHREUDER, R. (1999) War and peace: Morphemes and full forms in a Noninteractive Activation Parallel Dual-Route Model. *Brain and Language*, 68, 27-32
- BAAYEN, R. H. (2007). Storage and computation in the mental lexicon. In: JAREMA, G.; LIBBEN, G. *The Mental Lexicon: core perspectives*. Amsterdam: Elsevier.
- BAAYEN, R. H.; DIJKSTRA, T.; SCHREUDER, R. (1997). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 37, 94-117.
- BAAYEN, R. H.; MOSCOSO DEL PRADO MARTÍN, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, 81, 666–698.
- BAAYEN, R. H.; SCHREUDER, R.; DE JONG, N. H.; KROTT, A. (2002). Dutch inflection: the rules that prove the exception. In: *Storage and Computation in the*

Language Faculty (S. Nootboom, F. Weerman and F. Wijnen, eds.), 61–92. Dordrecht: Kluwer Academic Publishers.

BALOTA, D.A.; CHUMBLEY, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 340-357.

BALTAZAR, L. M. (2012). Frequency Effects and the Processing of Verbal Morphology by L1 and L2 Speakers of English. Dissertação. Universidade Federal de Santa Catarina. Florianópolis, Setembro.

BERTRAM, R.; HYÖNÄ, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*. 48, 615 – 634.

BERTRAM, R.; LAINE, M.; KARVINEN, K. (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: evidence from a morphologically rich language. *Journal of Psycholinguistic Research*, 28, 213-226.

BIEN, H.; LEVELT, W.; BAAYEN, R. H.(2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the USA*, 102, 17876–17881.

BUTTERWORTH, B. (1983). Lexical Representation. In B. Butterworth (Ed.), *Language production*, Vol. 2, London: Academic Press. 257–294.

BYBEE, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10: 425-455.

BYBEE, J. (2010). *Language, Usage and Cognition*. New York: Cambridge.

BYBEE, J.; MCCLELLAND, J.L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22, 381–410.

BYRNE, B.; FIELDING BARNSELY, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Education and Psychology*, 8, p. 313-321.

CAPOVILLA, A. G. S. (1999). Leitura, escrita e consciência fonológica: desenvolvimento e inter-correlações e interações. Tese. São Paulo: Universidade de São Paulo.

CAPOVILLA, A. G. S.; CAPOVILLA, F. C. (2000). *Problemas de leitura e escrita*. São Paulo: Memnon.

CARAMAZZA, A.; LAUDANNA, A.; ROMANI, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297–332.

CHOMSKY, N. (2000). Linguistics and the Brain Science. Em: MARANTZ, A.; MIASHITA, Y.; O'NEIL, W.; (Eds). *Image, Language and Brain*. Cambridge / Massachusetts: MIT.

COLTHEART, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing*. London: Academic Press. 151-216.

CUNHA, A. S. C. (2005). As vantagens dos modelos de ativação mistos no acesso lexical: uma comparação entre os modelos AAM e MRM. *Caderno Seminal Digital*, Ano 12, Nº 4, Vol.4 (Jul/Dez 2005) – ISSN 1806-9142

DIEPENDAELE, K., SANDRA, D., & GRAINGER, J. (2005). Masked cross-modal morphological priming: unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, 20, 75_114.

DUÑABEITIA, J.A., LAKA, I., PEREA, M.; CARREIRAS, M. (2009). Is Milkman a superhero like Batman? Constituent morphological priming in compound words. *European Journal of Cognitive Psychology*, 21(4), 615-640.

FELDMAN, L. B.; BASNIGHT-BROWN, D.M. (2007). Origins of Cross-Languages Differences in Word Recognition. In: JAREMA, G.; LIBBEN, G. *The Mental Lexicon: core perspectives*. Amsterdam: Elsevier.

FIGUEIREDO SILVA, M. C.; SELL, F. F.. (2011). Algumas notas sobre compostos em português brasileiro e em Libras. In: Roberta Pires de Oliveira; Carlos Miotto. (Org.). *Percursos em Teoria da Gramática*. 1ª ed. Florianópolis: Editora da UFSC, 2011, 17-41.

FIORENTINO, R.; POEPEL, D. (2007). Compound Words and Structure in the Lexicon. *Language and Cognitive Processes*. 22 (7), 953 - 1000.

FRANÇA, A. I. (2005). O Léxico Mental em Ação. *Linguística*, Rio de Janeiro, V. 1, n.1, 45 - 80, junho.

FRANÇA, A. I.; LEMLE, M.; GESUALDI, A. R.; CAGY, M.; INFANTOSI, A. F. C. (2008). A neurofisiologia do acesso lexical: palavras em português. *VEREDAS ON-LINE – Psicolinguística– 2/2008*, 34-49 – PPG LINGÜÍSTICA/UFJF – Juiz de Fora, MG- ISSN 1982-2243.

FRIEDERICI, A.D.; HAHNE, A.; MECKLINGER, A. (1996). The temporal structure of syntactic parsing: early and late event-related brain potential effects elicited by syntactic anomalies. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 22:1219–1248.

FROST, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23 (4), 1277-1288.

GAGNÉ, C. L. (2009). Psycholinguistic perspectives, in: *The Oxford Handbook of Compounding*. LIEBER, R.; STEKAUER, P. Orgs. New York: Oxford University Press.

GARCIA, D.C.; MAIA, M.; FRANÇA, A. I. (2012). The time course of word recognition: evidence from brazilian portuguese. *ReVEL*, vol. 10. n. 18.

GIRAUDO, H., & GRAINGER, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, 15, 421 _444.

GOUGH, P.; JUEL, C.; GRIFFITH, P. (1986). Reading, spelling and orthographic cipher. In: GOUGH, P.; EHRI, L.; TREIMAN, R. *Reading acquisition*. London: Lawrence Erlbaum Associates Publishers, 35-48.

GUIMARÃES, S. R. K. (2003). Dificuldade no desenvolvimento da lectoescrita: o papel das habilidades metalinguísticas. *Psicologia: teoria e pesquisa*. Brasília, v.19, n. 1.

INSTITUTO PAULO MONTENEGRO; ONG AÇÃO EDUCATIVA. (2011). INAF Brasil 2011: Indicador de Analfabetismo Funcional: principais resultados. Disponível em: <http://www.ipm.org.br/download/informe_resultados_inaf2011_versao%20final_12072012b.pdf> Acesso dia 20 de setembro de 2012.

JAREMA, G. (2006). Compound representation and processing: a cross-language perspective. In: (2006). *The representation and processing of compound words*. Ed. Gary Libben & Gonia Jarema. New York: Oxford.

JAREMA, G.; BUSSON, C.; NIKOLOVA, R.; LIBBEN, G. (1999). Processing compounds: A Crosslinguistic Study. *Brain and Language*. 68 (1), 363–370.

KEHAYIA, E.; JAREMA, G.; TSAPKINI, K.; PERLAK, D.; RALLI, A.; KADZIELAWA, D. (1999). The Role of Morphological Structure in the Processing of Compounds: The Interface between Linguistics and Psycholinguistics. *Brain and Language*. 68, 370–377.

KUTAS, M.; HILLYARD, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*. 11: 99-116.

LEHTONEN, M.; LAINE, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6, 213–225.

LEHTONEN, M.; NISKA, H.; WANDE, E.; NIEMI, J.; LAINE, M. (2006). Recognition of Inflected Words in a Morphologically Limited Language: Frequency Effects in Monolinguals and Bilinguals. *Journal of Psycholinguistic Research*, Vol. 35, No. 2, March.

- LIBBEN, G. (2006). Why Study Compound Processing? An overview of the issues. Chapter one. In: LIBBEN, G. & JAREMA, G. (Eds.), *The Representation and Processing of Compound Words*. New York : Oxford university Press.
- LIBBEN, G. (1998). 'Semantic transparency in the processing of compounds: Consequences for representation, processing and impairment', *Brain and Language*, 61: 30–44.
- LIBBEN, G., GIBSON, M., YOON, Y., & SANDRA, D. (1997). Semantic transparency and compound fracture. *CLASNET Working Papers*, 9, 1–13.
- MAIA, M; LEMLE, M.; FRANCA, A. I. (2007). Efeito Stroop e rastreamento ocular no processamento de palavras, *Ciências & Cognição*, vol. 4,.2-17.
- MARSLEN-WILSON, W., TYLER, L., WAKSLER, R., & OLDER, L. (1994). Morphology and meaning in the mental lexicon. *Psychological Review*, 101, 333.
- MARTIN, R. (1982). The pseudohomophone effect: the role of visual similarity in non-word decisions. *Quarterly Journal of Experimental Psychology*. 34(A), 395-409.
- MEYER, D. E.; SCHVANEVELDT, R. W. (1971). "Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations" *Journal of Experimental Psychology*, 90(2), 227-234.
- MÜNTE, T. F.; SAY, T.; CLAHSSEN, H.; SCHILTZ, K.; KUTAS, M. (1999). Decomposition of morphologically complex words in English: evidence from event-related brain potentials. *Cognitive Brain Research*. 7, 241–253.
- PENKE, M.; WEYERTS, H.; GROSS, M.; ZANDER, E.; MUNTE, T. F.; CLAHSSEN, F. (1997). How the brain processes complex words: an event-related potential study of German verb inflections. *Cognitive Brain Research*. 6, 37–52.
- PEREA, M.; ROSA, E. (2000). The effects of orthographic neighborhood in reading and laboratory word identification tasks: A review. *Psicologica International Journal of Methodology and Experimental Psychology*, 21, 327-340.

PESSOA, L. S. F. (2008) O Processamento de formas verbais contendo a sílaba “re” inicial com e sem natureza morfológica no Português Brasileiro. Dissertação - João Pessoa - UFPB.

PINKER, S. & ULLMAN, M. (2002) The past and future of the past tense. *Trends in Cognitive Science*, 6, 456-463.

PINKER, S. (1999). *Words and Rules: The Ingredients of Language*. New York: Basic Books.

PINKER, S. PRINCE, A. (1992). Regular and irregular morphology and the psychological status of rules of grammar, in: L.A. Sutton, C. Johnson, R. Shields Eds., *Proc. 17th Annual Meeting of the Berkeley Linguistics Society*, Berkeley Linguistics Society, Berkeley, CA.

RUMELHART, D. E.; MCCLELLAND, J. L. (1986). On learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In: *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 2, D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (eds.), 216–271. Cambridge, MA: MIT Press.

SANDRA, D. (1990) On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology*, 42: 529–67.

SCHREUDER, R., & BAAYEN, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing*, 131 _154. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

SCLIAR-CABRAL, L. (2003). *Guia prático de alfabetização*. São Paulo: Contexto.

SEIDENBERG, M. S., & MCCLELLAND, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.

SERENO, J.; JONGMAN, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25, 117-129.
SONNENSTUHL, I.; EISENBEISS, S.; CLAHSSEN, H. (1999).

Morphological priming in the mental lexicon: evidence from German. *Cognition*. 72, 203-236.

STAMPA, M. 2009. *Aquisição da Leitura e da Escrita: uma abordagem teórica e prática a partir da consciência fonológica*. Rio de Janeiro: Wak.

STANOVICH, K.; CUNNINGHAM, A.; CRAMMER, B. (1984). Assessing phonological awareness in kindergarten children: issues of task comparability. *Journal of Experimental Child Psychology*, 38, p.175-190.

STEMBERGER, J. P.; MIDDLETON, C. (2003). Vowel dominance and morphological processing. *Language and Cognitive Processes*, 18, 369–404.

STOCKALL, L., & MARANTZ, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon*, 1, 85_123.

TABAK, W.; SCHREUDER, R.; BAAYEN, R. H. (2005a). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. In: *Linguistic Evidence — Empirical, Theoretical, and Computational Perspectives* (S. Kepser and M. Reis, eds.), 529– 555. Mouton de Gruyter, Berlin.

TAFT, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7, 263–272.

TAFT, M. (1994) Interactive activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 9, 271-294.

TAFT, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, 57A, 745 _765.

TAFT, M.; FORSTER, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behaviour*, 14, 638–647.

TRAFICANTE, D.; C. BURANI (2003). Visual processing of Italian verbs and adjectives: The role of the inflectional family size. In: *Morphological Structure in Language*

Processing (BAAYEN, R. H. and SCHREUDER, R. eds.), 45-64. Mouton de Gruyter, Berlin.

VERGARA-MARTÍNEZ, M.; DUÑABEITIA, J. A.; LAKA, I.; CARREIRAS, M. (2009). ERP correlates of inhibitory and facilitative effects of constituent frequency in compound word reading. *Brain Research*. 1257, 53 - 64.

WURM, L. H.; BAAYEN, R. H. (2005, June). Surface frequency effects, even for suffixed inflections, below the magic number six. Paper presented at *Perspectives on Morphological Processing*, Cambridge, UK.

ZWITSERLOOD, P. (1994) 'The role of semantic transparency in the processing and representation of Dutch compounds', *Language and Cognitive Processes*, 9: 341–68.

APÊNDICES

Apêndice I

1. Pseudopalavras e palavras utilizadas no Teste de Decisão Lexical I

Pseudopalavras Simples

babifame
boselete
bucodeda
buvachalas
camonepe
caromiga
cebintilha
creladita
fenacodo
ferdabupa
girgoforfo
lanucagem
lecanusa
loguamiza
megulena
merpifabo
micaluxa
nebedove
nedrofeque
nipolado
pacofife
plarocaca
ponaleta
pumovica
ramarbole
ranocidomi
terdalono
tracudona
trameila
vomardere

Pseudopalavras Compostas

água-meva
barco-naro
blusa-mida
cara-pruca
cola-jano
dente-garro
faro-tinzo
fila-bera
forro-beno
frango-loge
lama-peves
lente-salu
linha-pafa
livro-celo
lupa-fuzes
mela-tuca
moça-teda
pega-sano
perna-tina
rasga-vola
resma-culo
tarde-urdo
tipo-naco
toma-bado
torta-paco
touca-pirilo
traça-tena
trena-deno
troca-rofa
veste-inge

Palavras Simples, Baixa Frequência

argueiro
baculejo
berimbelo
catupiri
cocoruto

encafifado
mixirica
mocorongo
peripaque
siriguela

Palavras Simples, Média Frequência

cacarejar
cavanhaque
charlatão
comemorar
insolência
manjedoura
micareta
quitinete
sacolejar
tagarela

Palavras Simples, Alta Frequência

alfinete
beterraba
canivete
crocodilo
espinafre
guardanapo
mastigar
picareta
pirulito
relâmpago

Palavras Compostas, Baixa Frequência

bicho-grilo
bóia-fria
borra-botas
cobra-cega

lava-jato
limpa-fossa
papa-anjo
para-choque
quebra-mola
saca-rolha

Palavras Compostas, Média Frequência

carta-bomba
cata-vento
chupa-cabra
dedo-duro
ferro-velho
gelo-seco
manda-chuva
puxa-saco
sofá-cama
tira-gosto

Palavras Compostas, Alta Frequência

arco-íris
banho-maria
bate-boca
caixa-preta
carro-chefe
conta-gota
luta-livre
salto-alto
surdo-mudo
vira-lata

2. Pseudopalavras e palavras utilizadas no Teste de Decisão Lexical II

Pseudopalavras simples

fenacodo
lecanusa
megulena
pacofife
pirgenemas
saritenho
soturpejar

Pseudopalavras Compostas

cola-jano
dente-garro
fila-bera
frango-loge
mela-tuca
tipo-naco
veste-inge

Palavras Simples, Alta Frequência II

cangaceiro
caramujo
gabarito
gelatina
mandioca
parabólica
relâmpago

Palavras Compostas, Alta Frequência II

arco-íris
bate-papo
carro-chefe
cessar-fogo
fome-zero
porta-mala
roda-viva

*itens removidos das análises por apresentar desvio padrão duas vezes maior que a média.

Decisão Lexical I **Para o tempo de reação:**

borra-botas
mixirica
carta-bomba
sacolejar

para acurácia de resposta:

borra-botas
berimbelo
cocoruto
carta-bomba
cata-vento
sacolejar

Apêndice II

Análises estatísticas das propriedades dos estímulos dos testes de decisão lexical I e II.

Análise Geral do número de letras

Palavras Compostas X palavras Simples ($t=1.720$, $gl=29$, $p=0.096$, bicaudal)

Tabela 1. Número de letras médio das palavras do teste de decisão lexical I

Frequência	Baixa	Média	Alta
Compostas	9	8,888889	9
Simples	8,809524	9,125	8,5

Análise do número das letras por frequência

Baixa frequência

Compostas X Simples ($t=1.342$, $gl=9$, $p=0.213$, bicaudal)

Média frequência

Compostas X Simples ($t=0.000$, $gl=9$, $p=1.000$, bicaudal)

Alta frequência

Compostas X Simples ($t=1.627$, $gl=9$, $p=0.138$, bicaudal)

Alta frequência II

O número de letras das palavras simples e compostas utilizadas no teste de decisão lexical I e II são iguais. ** O número de letras das pseudopalavras foi igual ao número de letras das palavras verdadeiras.

Análise Geral da frequência

Compostas X Simples ($t=-0.045$, $gl=36$, $p=0.964$, bicaudal)

Tabela 2. Frequência média das palavras do teste de decisão lexical I

Frequência	Baixa	Média	Alta
Compostas	15,88889	149,6667	840,4
Simples	10,88889	175	869,1

Tabela 3. Número de letras médio e frequência média das palavras do teste de decisão lexical II

	Número de Letras	Frequência
Compostas	8,714285714	1975,714286
Simples	8,714285714	1935,428571

Análise da frequência por subnível

Baixa frequência

Compostas X Simples ($t=0.723$, $gl=9$, $p=0.488$, bicaudal)

Média frequência

Compostas X Simples ($t=-0.474$, $gl=9$, $p=0.647$, bicaudal)

Alta frequência

Compostas X Simples ($t=-0,235$, $gl=9$, $p=0.819$, bicaudal)

Alta frequência II

Compostas X Simples ($t=0,127$, $gl=6$, $p=0.903$, bicaudal)

A frequência das palavras simples e compostas utilizadas no teste de decisão lexical I e II são iguais.

Análise Geral do número de Sílabas das palavras

Compostas X Simples ($t=0.374$, $gl=36$, $p=0.711$, bicaudal)

Análise do número de sílabas das palavras por nível de frequência

Baixa frequência

Compostas X Simples ($t=-1.500$, $gl=9$, $p=0.168$, bicaudal)

Média frequência

Compostas X Simples ($t=1.000$, $gl=9$, $p=0.343$, bicaudal)

Alta frequência

Compostas X Simples ($t=1.964$, $gl=9$, $p=0.081$, bicaudal)

Alta frequência II

Compostas X Simples ($t=-1.000$, $gl=6$, $p=0.356$, bicaudal)

O número de sílabas das palavras simples e compostas utilizadas no teste de decisão lexical I e II são iguais. * Todas as pseudopalavras tem 4 sílabas.

Todas as palavras compostas e simples utilizadas nos experimentos de decisão lexical I e II foram pareadas por número de letras e sílabas e frequência (todos os $p>0.0800$). Ou seja, para esses parâmetros não havia diferença entre palavras simples e compostas, tanto a nível global como entre os níveis de frequência.

Análise da geral da frequência dos constituintes versus palavras

Média da frequência dos 1^{os} constituintes 37269,89

Média da frequência dos 2^{os} constituintes 45742,92;

Média da frequência das palavras simples 647,78

Média da frequência das palavras compostas 644,84

Primeiros constituintes x palavras simples ($t=4,559$, $gl=36$, $p<0,001$, bicaudal)

Segundos constituintes x palavras simples ($t=3,732$, $gl=36$, $p=0,001$, bicaudal)

Primeiros constituintes x palavras compostas ($t=4,560$, $gl=36$, $p<0,001$, bicaudal)

Segundos constituintes x palavras compostas ($t=3,733$, $gl=36$, $p=0,001$, bicaudal)

Correção de Bonferroni $p<0,0125$, todas as comparações foram significativas. Sempre a frequência média dos constituintes (1 e 2) era maior que das palavras plenas.

Análise de silabidade dos constituintes versus palavras

Média do número de sílabas dos 1^{os} constituintes 2,03

Média do número de sílabas dos 2^{os} constituintes 2,03

Média do número de sílabas das palavras simples 4,03

Média do número de sílabas das palavras compostas 4,05

Primeiros constituintes x palavras simples ($t=-29,799$, $gl=36$, $p<0,001$, bicaudal)

Segundos constituintes x palavras simples ($t=-75,000$, $gl=36$, $p<0,001$, bicaudal)

Primeiros constituintes x palavras compostas ($t=-29,799$, $gl=36$, $p<0,001$, bicaudal)

Segundos constituintes x palavras compostas ($t=-75,000$, $gl=36$, $p<0,001$, bicaudal)

Correção de Bonferroni $p<0,0125$, todas as comparações foram significativas. Sempre o número médio de sílabas dos constituintes (1 e 2) era maior que o das palavras plenas. Em média os constituintes eram dissilábicos e as palavras plenas quadrissilábicas.