

**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO**  
**FACULDADE DE LETRAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

***REFRIGERANTE PARA REFRIGERAR?***  
**UM ESTUDO DE EEG SOBRE O CAMINHO NEUROFISIOLÓGICO DA**  
**DIACRONIA NO ACESSO LEXICAL**

**Julia Cataldo Lopes**

**Rio de Janeiro**  
**Agosto de 2023**  
**Dissertação de mestrado**

***REFRIGERANTE PARA REFRIGERAR?***  
**UM ESTUDO DE EEG SOBRE O CAMINHO NEUROFISIOLÓGICO DA**  
**DIACRONIA NO ACESSO LEXICAL**

Por

**JULIA CATALDO LOPES**

Aluna do curso de mestrado  
do Programa de Pós-Graduação em Linguística

Dissertação de mestrado em Linguística,  
apresentada ao Programa de Pós-  
Graduação em Linguística da Universidade  
Federal do Rio de Janeiro.

Orientadora: Profa. Dra. **Aniela Improta  
França (UFRJ)**

Faculdade de Letras da UFRJ

Agosto de 2023

**CATALDO LOPES, Julia.** *Refrigerante para refrigerar?* Um estudo de EEG sobre o caminho neurofisiológico da diacronia no acesso lexical. **Universidade Federal do Rio de Janeiro, 2023. Dissertação de mestrado em Linguística.**

BANCA EXAMINADORA

---

Presidente Profa. Dra. Aniela Improta França  
Universidade Federal do Rio de Janeiro (UFRJ)

---

Profa. Dra. Marije Soto  
Universidade Federal do Rio de Janeiro (UFRJ)

---

Prof. Dr. Thiago Oliveira da Motta Sampaio  
Universidade Estadual de Campinas (Unicamp)

Suplentes:

Prof. Dr. Gustavo Lopez Estivalet, UFPB

Defendida a dissertação em: 7 de agosto de 2023

## CIP - Catalogação na Publicação

L864r      Lopes, Julia Cataldo  
Refrigerante para refrigerar? Um estudo de EEG  
sobre o caminho neurofisiológico da diacronia no  
acesso lexical / Julia Cataldo Lopes. -- Rio de  
Janeiro, 2023.  
125 f.

Orientadora: Aniela Improta França.  
Dissertação (mestrado) - Universidade Federal do  
Rio de Janeiro, Faculdade de Letras, Programa de Pós  
Graduação em Linguística, 2023.

1. Acesso lexical. 2. Diacronia. 3. Opacidade  
semântica. 4. Neurolinguística. 5. EEG. I. França,  
Aniela Improta, orient. II. Título.

*Aos meus pais, Roberta e Alex,  
pelo amor, apoio e companheirismo  
constantes e incondicionais.*

## Agradecimentos

Aos meus pais, Roberta e Alex, sem os quais nada disso seria possível. Vocês são incríveis, não consigo imaginar pais melhores. Obrigada por absolutamente tudo. Amo vocês.

À minha família: irmãos (Victor, Eduardo, Ana Clara), avós (Sheila, Ivonete e, *in memoriam*s, Roberto e Francisco), tios (Alessandra, Marcelo, Marcella, Laranja, Dido, Vivi, Arlete, Papy), primos (Vinicius, Daniel, Nina, Letícia) e cunhada (Carol). O amor e apoio emocional que recebo de vocês me faz ser quem sou e me ajuda a seguir em frente.

Aos meus amigos da vida: Ingrid, Duda, Laura, Júlia, Ana, Andrei, Renata, Thársus, Krishynan (que inclusive ajudou dando uma consultoria técnica sobre o EEG), Matheus R., Flavinha, Aninha, Breno, Nayara, Dani, Matheus L., Mitt. Compartilhar com vocês os momentos – felizes e tristes – deixa os dias melhores.

Aos meus amigos de laboratório (Diane, Leo, Mayara, Stefanie, Jéssica, Emily, Bernardo, Lara, Renata, Ana Luiza, Wellington, Isadora, Ana Letícia, Maria, Olívia, Mayda), do mestrado (Vital, Moíra) e da Unicamp (Marô, Fernando, Natália – e de quebra Daniel). Obrigada pelo aprendizado, pelas risadas, pela ajuda com os experimentos e, aos mais próximos, pelas confidências e suporte emocional, que me estimulam e me ajudam a passar pelas dificuldades da vida pessoal e profissional.

Aos amigos que, enquanto professores, me ajudaram a conseguir participantes para os experimentos: Luciano e Ana Luiza. Também a todos que participaram dos experimentos e que chamaram seus amigos para participar.

Aos professores da banca, Marije e Thiago. Thiago me recebeu no meu primeiro dia no laboratório e desde então é um amigo e mestre. Marije me conhece também desde que eu era IC e me ensinou e ajudou tanto durante o mestrado que é, mesmo que não oficialmente, minha co-orientadora. Oficialmente, é uma amiga e mentora.

À Aniela, minha orientadora, que me ensinou muitíssimo ao longo de todos esses anos, desde a primeira aula de Linguística I na graduação, e sempre com brilho nos olhos. Ani, você me ajudou a chegar a todos os lugares em que estou. Obrigada pela compreensão, pelo suporte e pelas conversas.

Ao Lucas, meu namorado, que chegou chegando e, inacreditavelmente, deu check em mais boxes do que eu tinha. Obrigada pelo companheirismo, pelas conchinhas e por me alimentar enquanto eu estava focada em escrever essa dissertação. Te encontrar foi um dos maiores presentes que a vida já me deu. Aos meus sogros, Danielle e Flávio, que me acolheram na casa deles durante as semanas em que rodei os experimentos, e à Cacau e à Luna, que deixaram os dias mais alegres.

Obrigada a todos, vocês fazem a vida valer a pena.

## RESUMO

CATALDO LOPES, Julia. *Refrigerante para refrigerar? Um estudo de EEG sobre o caminho neurofisiológico da diacronia no acesso lexical*. Universidade Federal do Rio de Janeiro, 2023. Dissertação de mestrado em Linguística.

Quando bebemos um refrigerante, não percebemos que estamos nos refrigerando. Na verdade, a palavra *refrigerante* vem de *refrigerar*, mas provavelmente não percebemos mais isso. Será, porém, que a nossa faculdade de linguagem guarda alguma memória daquela relação linguística que era transparente em séculos passados, mas que hoje se tornou opaca? Caso seja esse o caso, se apresentarmos a palavra *refrigerar* a um participante de um experimento, isso poderia ajudar no acesso à *refrigerante*, que seria apresentado a seguir.

Assim, no presente trabalho, avaliamos o acesso lexical de palavras que são, hoje, semanticamente opacas em relação àquelas que as originaram. Nosso objetivo é verificar se essas palavras possuem identidade morfológica entre si ou se a distância diacrônica entre as duas nos faz ajustar suas raízes e, assim, abrir diferentes entradas em nosso léxico mental (STOCKALL; MARANTZ, 2006; FRANÇA, 2008).

Para verificar isso, realizamos experimentos comportamentais e neurofisiológicos com EEG, para avaliar o custo do processamento que sucede os milissegundos da compreensão do *onset* de uma palavra. Tendo a decisão lexical como tarefa experimental, lançamos mão de um paradigma de *priming* aberto multimodal com SOA longo, por serem essas condições que aparentemente inibem o acesso decomposicional a *targets* semanticamente opacos (HEYER; KRONISHOVA, 2018). Tivemos como variáveis dependentes o tempo de resposta comportamental, a taxa de acurácia da resposta comportamental e as amplitudes de onda do componente neurofisiológico N400 (300 a 500ms após *onset* do *target*). Como variáveis independentes, tivemos i) o tipo de relacionamento entre *prime* e *target*: morfológico intacto (*refrigerar-refrigeração*), morfológico mas semanticamente obscuro (*refrigerar-refrigerante*) ou nenhum relacionamento (*refrigerar-cadeira*); e ii) a quantidade de camadas morfológicas dos *targets*: 1 camada (*refrigerar-refrigeração*) ou 2 camadas (*refrigerar-refrigerantezinho*).

Ambos os nossos resultados comportamentais e neurofisiológicos apontaram um maior custo de acesso lexical para *refrigerante*, quando comparado a *refrigerar*, indicando que essas palavras não possuem identidade morfológica entre si. Interpretamos esse resultado como denotativo de uma espécie de Garden Path lexical que pode ser levado no início a levantar a hipótese de identidade entre *prime* e alvo, mas logo tem que ajustar a aposta para um outro contrato semântico. Lançamos mão da teoria da Morfologia Distribuída (HALLE; MARANTZ, 1993) para explicar melhor esses resultados: a herança do *prime* falharia e, quando nosso parser percebe o Garden Path, ele retorna à Enciclopédia para renegociar o sentido da nova raiz diante da qual percebeu estar. Por fim, obtivemos resultados contrastantes entre o teste comportamental e o neurofisiológico no que diz respeito ao custo de processamento de palavras com diferentes quantidades de camadas morfológicas. O primeiro, que avalia um *wrap-up* do processamento da palavra, se mostrou sujeito a um efeito de tamanho de palavra: quanto mais camadas, mais custoso é o acesso lexical, conforme prevê a hipótese do *Affix Stripping* (TAFT; FOSTER, 1975). Já o EEG, que capta o processamento de camadas internas à palavra, não demonstrou diferença no processamento de palavras com uma ou duas camadas morfológicas, quando essas palavras têm a história derivacional, ou seja, compartilham o merge entre raiz e o primeiro morfema categorizador. Este merge caracteriza o que conhecemos como o ponto da arbitrariedade saussuriana e o compartilhamento deste ponto caracteriza o status de identidade lexical.

Palavras-chave: acesso lexical; opacidade semântica; diacronia; EEG; Morfologia Distribuída.

## ABSTRACT

CATALDO LOPES, Julia. ***Restaurant to restore? An EEG study on the neurophysiological path of diachrony in lexical access.*** Universidade Federal do Rio de Janeiro, 2023. Master Thesis in Linguistics.

When we enter a restaurant, we don't realize that we are going to restore ourselves. In fact, the word restaurant comes from restore, but we probably don't realize that anymore. Could it be, however, that our faculty of language keeps some memory of that linguistic relationship that was transparent in past centuries, but that today has become opaque? If that's the case, if we present the word restore to a participant in an experiment, it could help in accessing the restaurant that would be presented next.

Thus, in the present work, we evaluated the lexical access of words that are, today, semantically opaque in relation to those that originated them. Our objective is to verify if these words have morphological identity with each other or if the diachronic distance between the two makes us adjust their roots and, thus, open different entries in our mental lexicon (STOCKALL; MARANTZ, 2006; FRANÇA, 2008).

To verify this, we carried out a behavioral and a neurophysiological experiment with an EEG, to evaluate the processing cost that follows the milliseconds of understanding the onset of a word. Having the lexical decision set as the experimental task, we used a multimodal *priming* paradigm with a long SOA, as these conditions apparently inhibit a decompositional access to semantically opaque *targets* (HEYER; KRONISHOVA, 2018). We had as dependent variables the behavioral response time, the accuracy rate of the behavioral response and the wave amplitudes of the neurophysiological component N400 (300 to 500ms after onset of the *target*). As independent variables, we had i) the type of relationship between *prime* and *target*: intact morphological relationship (to restore-restoration), morphological but semantically unclear relationship (to restore-restaurant) or no relationship (restore-sailor); and ii) the number of morphological layers of the *targets*: 1 layer (to restore-restaurant) or 2 layers (to restore-little restaurant).

Both our behavioral and neurophysiological results pointed to a higher cost of lexical access for restaurant, when compared to restore, indicating that these words do not have morphological identity with each other. We interpret this result as denoting a kind of lexical Garden Path that can be taken at first to raise the hypothesis of identity between *prime* and *target*, but soon having to adjust the bet to another semantic contract. We make use of the theory of Distributed Morphology (HALLE; MARANTZ, 1993) to better explain these results: the inheritance of the *prime* would fail and, when our parser perceives the Garden Path, it returns to the Encyclopedia to renegotiate the meaning of the new root before which it perceived be. Finally, we obtained contrasting results between the behavioral and neurophysiological tests regarding the cost of processing words with different amounts of morphological layers. The first, which evaluates a word processing *wrap-up*, proved to be subject to a word size effect: the more layers, the more expensive the lexical access, as predicted by the Affix Stripping hypothesis (TAFT; FOSTER, 1975). The EEG, which captures the processing of layers internal to the word, showed no difference in the processing of words with one or two morphological layers, when these words have a derivational history, that is, they share the merge between the root and the first categorizing morpheme. This merge characterizes what we know as the point of Saussurian arbitrariness and the sharing of this point characterizes the status of lexical identity.

Keywords: lexical access; semantic opacity; diachrony; EEG; Distributed Morphology.



### **Financiamento:**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES, Brasil), pelo Programa de Excelência Acadêmica (PROEX), e com apoio da Fundação Carlos Chaga Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ, RJ), pelo Programa Mestrado Nota 10.



### **Ética em pesquisa:**

O projeto de pesquisa apresentado e discutido nesta dissertação, junto aos dois experimentos conduzidos, foi aprovado pelo Comitê de Ética em Pesquisa do Instituto de Estudos e Saúde Coletiva da Universidade Federal do Rio de Janeiro (IESC – UFRJ) em 5 de outubro de 2022.

Título da pesquisa: O processamento de palavras no cérebro Número do parecer: 5.685.040 CAAE: 60759022.0.0000.5286
--

O parecer e os termos de consentimento livre e esclarecido (TCLE) estão disponíveis no link OSF abaixo.

### **Ciência aberta e reproduzível:**

O projeto de experimentação foi pré-registrado na plataforma Open Science Framework (OSF) através do DOI 10.17605/OSF.IO/8564X e está disponível no seguinte link: <https://osf.io/8564x>. O pré-registro é uma maneira de incitar a ética em pesquisa e a confiabilidade no processo de coleta e análise dos resultados.

As versões experimentais, os dados coletados dos participantes e as análises estatísticas estão disponíveis na mesma plataforma, no seguinte link: <https://osf.io/m6fwx/>. Essa prática permite a verificação da metodologia e dos resultados dos estudos, incitando a reprodutibilidade científica.

## SUMÁRIO

LISTA DE FIGURAS .....	12
LISTA DE TABELAS .....	13
LISTA DE GRÁFICOS .....	14
1. Introdução .....	15
2. Processamento e representação de palavras .....	29
2.1 Lexicalismo e Morfologia Distribuída.....	29
2.2 Teorias de Processamento e estudos experimentais.....	33
2.3 Transparência e opacidade semântica .....	38
2.4 Stimulus Onset Asynchrony (SOA) .....	43
3. Neurofisiologia das palavras .....	45
3.1 Processamento visual de palavras .....	47
3.2 Hipótese da ativação contínua .....	53
3.3 Fisiologia cerebral e medidas do EEG.....	55
4. Experimento comportamental.....	59
4.1 Justificativa, objetivo e variáveis.....	59
4.2 Materiais e métodos .....	60
4.3 Desenho experimental e procedimentos .....	64
4.4 Hipóteses e previsões.....	66
4.5 Participantes.....	67
4.6 Resultados.....	67
4.7 Discussão .....	73
5. Experimento neurofisiológico (EEG).....	76
5.1 Justificativa, objetivo e variáveis.....	76
5.2 Materiais e métodos .....	76
5.3 Desenho experimental e procedimentos .....	76
5.3.1 Coleta dos dados neurofisiológicos .....	78
5.4 Hipóteses e previsões.....	79
5.5 Participantes.....	80
5.6 Resultados.....	81
5.6.1 Resultados comportamentais do EEG (RT de aperto de botão e ACC).....	81
5.6.2 Pré-processamento dos dados neurofisiológicos .....	87
5.6.3 Resultados neurofisiológicos do EEG (amplitudes de ondas).....	88
5.7 Discussão .....	99
5.7.1 Discussão dos resultados comportamentais do EEG.....	99

5.7.2	Discussão dos resultados neurofisiológicos do EEG.....	101
6.	Considerações finais .....	106
	Bibliografia.....	111
	ANEXOS .....	122
A.	Materiais experimentais .....	122

## LISTA DE FIGURAS

- Figura 1. Palavras em Guarani Mbyá, uma língua aglutinante. Fonte: Maia et al., 2019:60. ...21
- Figura 2. A arquitetura da gramática na Morfologia Distribuída, exemplificada com a palavra refrigeração. Fonte: Adaptado de Scher, Bassani, Minussi, 2013:19. .... 32
- Figura 3. a) Diagrama exemplificando SOA. Fonte:  
<https://commons.wikimedia.org/wiki/File:SOA-diagram.svg>. Nenhuma modificação feita, acessado em: 31/07/2023. b) Diagrama exemplificando a diferença entre SOA e ISI. Fonte:  
<http://www.eegget-it.nl/erp.html#>. Nenhuma modificação feita, acessado em: 31/07/2023.... 43
- Figura 4. A Casa de Lichtheim de 1874. Fonte: França, 2011:II. .... 46
- Figura 5. Hipótese de Competição (acima) e de Inibição (abaixo) para eliminação de representações concorrentes durante o acesso lexical de teach, do inglês ensinar. Fonte: Adaptado de França, 2005:62. .... 54
- Figura 6. Esquematização do limiar para que um potencial de ação seja gerado e do período refratário sofrido por um neurônio após tal potencial de ação. Fonte: Freitas, 2017:15. .... 56
- Figura 7. Geração de um ERP através de média de todos os potenciais de ação gerados por um dado estímulo, ao longo de vários *trials* (apresentações ao participante). Em cada par de ondas da imagem, a superior representa o ERP analisado e, a inferior, a onda diferencial com o ruído filtrado. Fonte: <http://www.eegget-it.nl/erp.html#>. Nenhuma modificação feita, acessado em: 31/07/2023. .... 58
- Figura 8. Protocolo de apresentação do experimento na tela do computador, em linha do tempo. Fonte: autora. .... 64
- Figura 9. No canto superior esquerdo, a pesquisadora está inserindo gel condutor de eletricidade entre o eletrodo e o escalpo do participante. No canto superior direito, vê-se o participante com a touca e o conjunto de eletrodos conectados, de um lado, no escalpo do participante e, de outro, no amplificador. Na tela do computador, vê-se ondas elétricas captadas do escapo do participante e amplificadas. No canto inferior, vê-se uma participante lendo a tela de instruções do experimento, pronta para inicia-lo. Suas mãos e dedos já estão posicionados nos botões do dispositivo Chronos que usará para registrar suas respostas. .... 77
- Figura 10. (2a) Sistema internacional 10-20 de posicionamento de eletrodos com 32 canais na touca actiCap (BrainProducts). Fonte: Bastos, Adamatti e Billa, 2016:3. (2b) Lógica e métrica utilizada para divisão dos eletrodos do sistema 10-20 pelo escalpo. Fonte: <https://info.tmsi.com/blog/the-10-20-system-for-eeg>, visitado em 13/07/2023. .... 78
- Figura 11. Mapa de delimitação das regiões cerebrais de interesse (ROIs), com seus respectivos canais (eletrodos): ROI Frontal com eletrodos F3, Fz, F4; ROI Fronto-temporal esquerda com eletrodos F7, FT9, T7; ROI Fronto-temporal direita com eletrodos F8, FT10, T8; ROI Fronto-central com eletrodos FC5, FC1, FC2, FC6; ROI Central com eletrodos C3, Cz, C4; ROI Centro-parietal com eletrodos CP5, CP1, CP6, CP2; ROI Parietal com eletrodos P3, P7, P4, P8; e ROI Occipital com eletrodos Oz, O1, O2. Fonte: Silva-Nasser, 2022:76. ... 88

## LISTA DE TABELAS

Tabela 1. Níveis de descrição de Marr para estudo de sistemas cognitivos. Fonte: Brennan, 2022, p.3, tradução própria.....	19
Tabela 2. Três células experimentais, com exemplos dos materiais experimentais. Primes se mantiveram constantes em todas as condições e eram palavras não-derivadas, enquanto targets eram os estímulos de interesse, palavras derivadas, que mudavam de acordo com a condição experimental.....	63
Tabela 3. Acurácia da decisão lexical por condição experimental, no teste comportamental. .	68
Tabela 4. RT da decisão lexical por condição experimental, no teste comportamental. ....	69
Tabela 5. RT da decisão lexical por tipo de relacionamento, no teste comportamental. ....	71
Tabela 6. RT da decisão lexical por quantidade de camadas morfológicas, no teste comportamental. ....	72
Tabela 7. Acurácia da decisão lexical por condição experimental, na análise comportamental do EEG. ....	82
Tabela 8. Tempos de resposta da decisão lexical por condição experimental, na análise comportamental do EEG. ....	83
Tabela 9. Tempos de resposta da decisão lexical por tipo de relacionamento, na análise comportamental do EEG. ....	85
Tabela 10. Tempos de resposta da decisão lexical por quantidade de camadas morfológicas, na análise comportamental do EEG. ....	86
Tabela 11. Comparações dos modelos 1, 2 e 3 com modelos nulos, em cada ROI. Células coloridas de acordo com a intensidade da significância estatística. ....	91
Tabela 12. Todos os materiais experimentais utilizados nos testes comportamental e neurofisiológico reportados. ....	122

## LISTA DE GRÁFICOS

Gráfico 1. RT da decisão lexical por condição experimental, no teste comportamental.....	69
Gráfico 2. RT da decisão lexical por tipo de relacionamento, no teste comportamental. ....	71
Gráfico 3. RT da decisão lexical por quantidade de camadas morfológicas, no teste comportamental. ....	72
Gráfico 4. Tempos de resposta por condição experimental, na análise comportamental do EEG. ....	83
Gráfico 5. Tempos de resposta da decisão lexical por tipo de relacionamento, na análise comportamental do EEG. ....	85
Gráfico 6. Tempos de resposta da decisão lexical por quantidade de camadas morfológicas, na análise comportamental do EEG. ....	86
Gráfico 7. Amplitudes médias (em $\mu\text{V}$ ) por condição experimental, em cada ROI.....	92
Gráfico 8. Amplitudes médias (em $\mu\text{V}$ ) por tipo de relacionamento, em cada ROI. ....	93
Gráfico 9. Amplitudes médias (em $\mu\text{V}$ ) por quantidade de camadas morfológicas, em cada ROI. ....	94
Gráfico 10. Amplitudes (em $\mu\text{V}$ ), em função do tempo (em ms), plotadas para todas as condições experimentais, em cada e todos os ROI.....	95
Gráfico 11. Amplitudes (em $\mu\text{V}$ ) das condições MI1, MO1 e SR1, em função do tempo (em ms), nas ROIs FC, Central e CP. ....	96
Gráfico 12. Amplitudes (em $\mu\text{V}$ ) das condições MI2, MO2 e SR2, em função do tempo (em ms), nas ROIs FC, Central e CP. ....	97
Gráfico 13. Amplitudes (em $\mu\text{V}$ ) das condições MI1, MI2, MO1 e MO2, em função do tempo (em ms), nas ROIs FC, Central e CP. ....	98
Gráfico 14. Amplitudes (em $\mu\text{V}$ ) dos tipos de relacionamento MI, MO, SR e PSEUD, em função do tempo (em ms), nas ROIs FC, Central e CP. ....	98

## 1. Introdução

---

Hatchlings arrive in the world with an elaborate set of innate behaviors. First is the complex behavior that allows the young bird to emerge from the egg. Once a bird has hatched, a variety of additional behaviors indicate how much of its early life is dependent on inherited information. In some species, hatchlings automatically crouch down in the nest when a hawk passes overhead but are oblivious to the overflight of an innocuous bird. (...) Despite the relatively scant attention paid to this aspect of memory, evolved and inherited neural associations are the most important component of the stored information in the brain that determines whether or not an individual survives long enough to reproduce (PURVES *et al.*, 2017:682).

Todo falante nativo tem um vasto conhecimento tácito sobre o funcionamento de sua língua, conseguindo brincar com ela de diversas maneiras: entendemos as ambiguidades e nos decidimos sobre os sentidos que devemos acatar; inferimos as palavras que virão em seguida; percebemos leves desvios de pronúncia; fazemos concordância à distância, mesmo depois de encaixes complexos – isso só para mencionar algumas das nossas muitas aptidões linguísticas. Porém, na verdade, antes de sermos adultos e falantes experientes – ainda quando somos bebês de até uns 10 meses – já somos foneticistas universais, pois conseguimos distinguir a diferença entre quaisquer fonemas, por exemplo as oclusivas dentais do inglês ou do português contrastando com as oclusivas retroflexas, do hindu ou do asturiano (WERKER; TEES, 1984; WERKER; LOGAN, 1985). Depois, afunilamos nosso conhecimento só para contrastes relevantes e internos à nossa língua mãe, e então nos tornamos exímios nos espaços alofônicos.

No domínio das palavras, desde pequenos percebemos os mecanismos de construção de sentido e os reproduzimos quando queremos criar ou compreender palavras novas. A criatividade lexical é um fenômeno generalizado na infância: a frequência e facilidade com que as crianças inventam palavras indicam que elas já perceberam a existência de dispositivos de formação de palavras e aplicam com desenvoltura os princípios que regem esse sistema de formação de palavras. Quando somos ainda jovens usamos esses neologismos principalmente para preencher uma lacuna no vocabulário convencional, ou seja, para expressar um significado que ainda não tenha feito parte da experiência linguística dos mais jovens. Muitas gírias têm esta história de nascimento (MOTSIOU, 2012; POIX, 2019).

Por exemplo, se viajante é alguém que fica viajando muitas vezes e palestrante é alguém que fica dando palestras, quem é um ficante? É alguém que fica com você sem você ter que chamar de namorado, o qual, por sua vez, teria uma característica mais fixa. Um ficante é nomeado pela ação de ficar com alguém, fisicamente durante um tempo curto, dando uns beijinhos que não significam um compromisso. Se caminhonete é um caminhão pequeno e colchonete é um colchão mais leve, o que significa um peguete? A noção de brevidade que vimos em *ficante* também está contida no verbo *pegar* forçado a entrar na semântica de estabelecer relacionamento romântico com alguém, mais um sufixo -ete (um frequentativo): um peguete é alguém que se *pega* de leve muitas vezes, mas cuja importância não é assim tão grande. Mesmo uma pessoa mais velha, que nunca teve um ficante ou um peguete, consegue facilmente entender como funcionam os relacionamentos amorosos de hoje em dia, decompondo estas palavras mesmo que inconscientemente. As raízes vão junto com um sentimento proporcionado pelos sufixos -ante e -ete, que revelam a impermanência, repetição e ligeireza destas relações.

Mais importante para esta dissertação, o que poderíamos dizer de *refrigerante*? Algo que refrigera? Da mesma maneira, o que seria *restaurante*? Algo que restaura? Dizemos que essas palavras foram idiomatizadas em nossa língua, isto é, que ganharam um sentido diferente do que teriam tido por derivação composicional do conceito original de sua raiz: seria o caso de *refrigeração*, que evoluiria de *refrigerar* de forma bem comportada, composicional. Como se dá esse processo e como ele é percebido por falantes nativos é justamente o que queremos descobrir nesta dissertação!

Não há como negar que, sincronicamente, desconhecemos a história de formação – o caminho diacrônico – de quase todas as palavras. Na escola não somos formalmente ensinados que *refrigerante* passou por um processo derivacional e compartilha a mesma raiz *refriger-*, que vem do verbo *refrigerar* e que passou por um processo derivacional que juntou um prefixo frequentativo *re-* à raiz *fri-* e à outra raiz *ger-*. Como a linguagem é predominantemente reflexa, a grande maioria das pessoas, por mais letrada que seja, nem se dá conta disso até que alguém chame a atenção para esse fato pela *primeira* vez. Também não pensamos que bebemos um *refrigerante* para nos *refrigerar*. Estamos simplesmente matando a sede ou a vontade de algo doce, seja no verão ou no inverno, quando não precisamos de refrigeração.



Contudo, uma vez que a faculdade da linguagem nos permite um valiosíssimo conhecimento gramatical de formação de palavras, será que o nosso parser<sup>1</sup> enxerga uma relação entre essas palavras e as que as derivaram, mesmo sem nos darmos conta disso? Quando vemos um Guaraná ser servido em nosso copo e vamos processar a palavra *refrigerante*, nós a decomparamos até a raiz *refriger-*, de *refrigerar*, ou criamos uma nova raiz *refrigerant-*? Quais são os morfemas que constituem essa palavra? Em outras palavras, poderíamos nos perguntar se temos uma *memória* não consciente de que *refrigerante* vem de *refrigerar* e de que *restaurante* vem de *restaurar*. Será que teríamos, através do nosso conhecimento implícito da língua mãe, uma forma de reconstruir a genética das palavras e relacionar pares como *refrigerar-refrigerante*, de forma semelhante ao que fazemos com os fonemas até os 10, 11 meses de vida? Seríamos nós também morfólogos universais?

Mas que mecanismo de *memória* estaríamos usando, se estamos justamente dizendo que seria uma habilidade não consciente? Seria esta uma memória filogenética, que nos permitiria garantir um conhecimento de mundo que pode ser construído por desenvolvimento e conectividade neural, ao longo dos milhares de anos da evolução de uma espécie? Se respondermos que sim, estaríamos pensando em uma morfologia preservada por *imprinting*, como nos referimos na epígrafe desta dissertação. É por conta do *imprinting* que filhotes de pássaros quebram instintivamente seus ovos quando estão prontos para nascer e, mesmo antes de qualquer interação com outros pássaros de sua espécie, já sabem diferenciar aves predadoras ao olhar sua silhueta (PURVES et al, 2018).

Podemos, então, dividir as coisas que sabemos entre: (i) as que nascemos já conhecendo; (ii) as para as quais nascemos já tendo uma predisposição genética, mesmo que precisemos de um pouco de interação com o meio para desenvolvê-las, como por exemplo andar e falar – são geralmente atividades que podemos fazer de maneira inconsciente (LENNEBERG, 1967); e (iii) as que aprendemos explicitamente através da interação com o meio em que vivemos (habilidades essas que, obviamente, serão deslanchadas pelo código genético da espécie, mas cujo conteúdo vem do mundo (ANDRADE; FRANÇA; SAMPAIO, 2018).

Ainda dentro dessa discussão, seria a nossa capacidade de brincar com as partes integrantes de uma palavra e criar novos verbetes algo instintivo, proveniente dos mecanismos linguísticos inscritos em nossa memória filogenética, ou seria algo que aprendemos

---

<sup>1</sup> Parser: sistema de computação sintática da faculdade da linguagem (cf. FRAZIER, 1979; GORRELL, 1995).

estritamente a partir das experiências com os dados linguísticos com os quais temos contato? Aplicando ao nosso exemplo, o conhecimento da estrutura morfológica diacrônica de palavras como *refrigerante* (i.e., o fato de que ela certamente vem de *refrigerar*) seria algo inscrito biologicamente em nós ou seria algo que podemos acessar apenas se aprendermos formalmente ou ainda se (conscientemente) deduzirmos?

É valioso remarcar que, se no passado a alusão a um conhecimento genético-filológico sempre foi um pouco jocosa, hoje existe uma hipótese que não nos permite abandonar esta explicação tão rapidamente assim. Um campo instigante que está começando a ser investigado versa sobre o armazenamento de memórias em geral (incluindo a memória linguística) e tenta esclarecer sobre os mecanismos celulares de armazenamento de conteúdo no corpo humano. Trata-se de uma investigação sobre os mecanismos neurais e áreas cerebrais que nos permitem lembrar de situações que vivemos no passado e do significado das palavras. No entanto, temos presente em cada uma das nossas células um mecanismo de registro de informações muito eficaz evolutivamente, que guarda elementos como nosso DNA. Seria o caso de guardarmos nossas memórias da mesma maneira? Elas não seriam, então, apenas produto da ativação e conectividade de áreas cerebrais específicas, como também estariam codificadas e armazenadas em células específicas (AKHLAGHPOUR, 2022; GALLISTEL, 2021; GERSHMAN *et al.*, 2021). Ainda estamos longe de confirmar ou refutar hipóteses como essa, que saem da ficção científica e passam a apontar para explicações que certamente estarão fora do escopo desta pesquisa. Contudo, estarão sempre dentro da lista desta pesquisadora de como seguir imaginando e investigando.

Pousando de novo na questão central dessa dissertação, poderíamos nos questionar se o conhecimento instintivo de que *refrigerante* vem de *refrigerar* está guardado biologicamente em nossas células e/ou nas conexões neurais tácitas de nosso cérebro, ou se só compreendemos que essas palavras são relacionadas se, pela experiência, conscientemente somos chamados atenção para esse fato. Esclarecimentos sobre essa questão trarão luz para uma outra pergunta, ainda em aberto na psico e na neurolinguística: como nós armazenamos as palavras na nossa mente, isto é, como funciona nosso dicionário mental. Não seria nem um pouco eficaz o classificarmos alfabeticamente, como fazem o Aurélio e o Houaiss. Outras maneiras seriam por seu sentido ou por sua forma.

Pela semântica, agruparíamos a palavra globalização, por exemplo, com palavras como nação e internacional. Pela forma, a agruparíamos com globo, global e globalizar, com as quais

ela divide raiz + o *primeiro* morfema categorizador, neste caso, o nominalizador Ø (glob+Ø<sub>N</sub>). A *primeira* maneira depende mais de memória de armazenamento que de operações linguísticas: cada uma dessas palavras teria entradas lexicais diferentes em nosso léxico mental e não precisaríamos estar sempre construindo suas derivações a partir de seu morfema base (BUTTERWORTH, 1983; HAY; BAAYEN, 2005; JACKENDOFF, 1983). A segunda, por outro lado, se apoia na hipótese diametralmente oposta: guardaríamos apenas o essencial na mente, ou seja, o significado da raiz glob- e de cada um dos sufixos existentes no português; a cada vez que processássemos global, globalizar e globalização, incluiríamos ao significado da raiz os sufixos, com os significados que carregam, por merge (TAFT, FORSTER, 1975; 1976; FORSTER, 1979; TAFT, 1979; PYLKÄNNEN *et al.*, 2000; STOCKALL, MARANTZ, 2006; EMBICK *et al.*, 2001; PYLKÄNNEN *et al.*, 2002; PYLKKÄNEN, MARANTZ, 2003; PYLKÄNNEN *et al.*, 2003).

É possível perceber que essas apostas se referem tanto a modelos de *representação* linguística quanto a modelos de *processamento* linguístico<sup>2</sup>. Comparando-os com os níveis de descrição propostos por David Marr (1982), modelos de representação (e derivação) se debruçariam no nível *computacional*, enquanto modelos de processamento estariam preocupados com o nível *algorítmico* de investigação científica (Tabela 1).

<b>Computacional</b>	O problema que um sistema cognitivo está resolvendo, incluindo seus inputs e outputs.
<b>Algorítmico</b>	Os passos através dos quais um sistema cognitivo resolve o problema, gerando os outputs corretos para os inputs disponibilizados.
<b>Implementacional</b>	Como um sistema físico (por exemplo um circuito neural) executa um certo algoritmo.

Tabela 1. Níveis de descrição de Marr para estudo de sistemas cognitivos. Fonte: Brennan, 2022, p.3, tradução própria.

<sup>2</sup> Historicamente, *primeiro* surgiram os modelos representacionais e, com a Revolução Cognitiva de 1950 e a interação entre a Linguística e a Psicologia, eles deram origem a modelos de processamento. Em 19xx, no entanto, houve a Crise da DTC (Teoria da Complexidade Derivacional, SAMPAIO, 2015), que nos fez compreender que complexidade derivacional não é necessariamente sinônimo de complexidade de processamento. Desde então, os estudos em representação e processamento andam tanto juntos quanto separados, de maneira que essas áreas hoje se informam mutuamente.

No nível computacional de representação, a hipótese Lexicalista de armazenamento e derivação é um exemplo de teoria que propõe a organização do léxico por palavras completas (CHOMSKY, 1995). A Morfologia Distribuída e a Morfologia Exoesqueletal, por outro lado, são exemplos de teorias que propõem a existência de uma sintaxe dentro das palavras, de maneira que o armazenamento e a derivação se dariam pelos morfemas constituintes de cada palavra (HALLE; MARANTZ, 1993; BORER, 2005a; BORER, 2005b).

Já no nível algorítmico de Marr, tais modelos teóricos corresponderiam, respectivamente, às hipóteses lexicalista e decomposicional de processamento lexical. Ambas têm como objetivo explicar como realizamos o acesso lexical, isto é, como juntamos forma com significado para ativar um conceito. Pelo modelo lexicalista, acessaríamos um significado a cada vez que nos deparássemos com uma palavra inteira. Já pelo modelo decomposicional, o primeiro passo realizado quando ouvimos/lemos/vemos uma palavra seria decompô-la em seus morfemas constituintes, para depois realizar o acesso lexical por sua raiz e, em seguida, adicionar composicionalmente por merge o sentido de cada um de seus afixos. Há ainda modelos chamados de dupla-rota, que apostam na coexistência de ambos os mecanismos apresentados (processamento por inteiro e serial). A escolha da utilização de um e não outro aconteceria de acordo com sua eficácia dadas as características específicas da palavra acessada.

Um conceito importante nessa discussão é o de palavra. Estamos aqui falando em processamento de palavras, mas ainda hoje a conceituação desse termo não é evidente. Para falantes de português, talvez a definição mais aceita seria “elemento de uma sentença que é separado por espaços”. No entanto, de acordo com a mesma, como classificaríamos palavras como guarda-chuva, pé de moleque e salário-mínimo<sup>3</sup>? Ao analisar línguas aglutinantes, fica ainda mais difícil sustentar essa definição. Na figura 1, por exemplo, vemos dois exemplos de recursividade verbal em Guarani Mbyá. É possível perceber que a palavra *xeayvukuaapota* comporta o que, em português, seriam 1 pronome + 3 verbos diferentes. Já *ndaexakuaapotaxei* comporta 1 pronome + 4 verbos + 2 marcadores de negação. Assim, se fossemos aplicar a definição mencionada de palavra ao Guarani Mbyá, 1 palavra dessa língua corresponderia a, respectivamente, 4 e 7 palavras do português.

---

<sup>3</sup> Definição de ‘salário-mínimo’, segundo o Dicionário Online de Português: “indivíduo que recebe o salário mínimo, o valor mínimo, regulamentado por lei, como remuneração pelo trabalho realizado”. Disponível em: <<https://www.dicio.com.br/salario-minimo/>>, acessado em 10 de julho de 2023.

<p> <i>xe-[[[ayvu]-kuaa]-pota] pende ayvu</i>            1SG-falar-saber-querer 2PL língua            ‘Eu quero[saber [falar a sua língua]]’         </p> <p> <i>nd-a-[[[[exa]-kuaa]-pota]-xe]-i jagua</i>            NEG-1SG-ver-saber-querer-desejar-NEG cachorro            ‘Eu não desejo[querer[ saber [ver (reconhecer) um ca-            chorro]]]’ (Dooley: 64)         </p>
--

Figura 1. Palavras em Guaraní Mbyá, uma língua aglutinante. Fonte: Maia *et al.*, 2019:60.

Aplicada ao nosso tema, poderíamos nos perguntar se *refrigerante* e *refrigerar* são acessados por inteiro – e nesse caso com certeza possuiriam entradas lexicais diferentes em nosso léxico mental – ou se são processados serialmente, por suas partes – e, nesse caso, ainda estaria em questão a delimitação das partes constituintes de cada uma dessas palavras.

Diante de questões e apostas teóricas tão diversificadas, como as mencionadas, colocá-las à prova parece uma ótima ideia. Considerando que todos esses debates são sobre características de uma habilidade biológica (a linguagem) comum a todo ser humano, um método de investigação acertado e disponível atualmente é o que envolve o nível implementacional de Marr (Tabela 1). Daí surge a relevância da experimentação psico- e neurolinguística para a compreensão dos mecanismos biológicos da faculdade da linguagem.

De fato, já foram várias as descobertas relacionadas tanto ao comportamento quanto ao cérebro do ser humano durante o processamento de linguagem (produção e compreensão). Para tanto, foi fundamental o relacionamento da linguística com a psicologia, a neurologia, a filosofia e a computação. Temos, hoje, inúmeros instrumentos que nos ajudam a entender os mecanismos biológicos envolvidos na linguagem. Por questões éticas, jamais abriríamos o crânio de uma pessoa e mexeríamos nele apenas para ver o que está acontecendo enquanto ela fala/escreve/sinaliza ou ouve/lê/vê sinais. No entanto, podemos apresentar um estímulo linguístico a uma pessoa e pedi-la para realizar uma tarefa de julgamento que será medida por uma resposta mecânica, como o aperto de um botão. Assim estaremos analisando o *comportamento* desse ser durante seu uso de linguagem.

Mais sofisticadamente, já temos máquinas como o eletroencefalograma (EEG), a ressonância magnética funcional (fMRI) e o magnetoencefalograma (MEG), que nos permitem delimitar as áreas cerebrais envolvidas nesse processamento (*onde* a linguagem está ancorada neuralmente) e a linha do tempo dessas computações (*quando* isso acontece depois que somos expostos a um estímulo linguístico, ou seja, quanto tempo demora para realizarmos cada etapa

desse processamento). Cada um dos métodos hoje disponíveis tem seus pontos fortes e fracos, de maneira que a comunidade científica se beneficia muito de investigações de um mesmo fenômeno por diferentes técnicas. O EEG, por exemplo, tem altíssima precisão temporal, mas pouca precisão espacial, enquanto fMRI é o extremo oposto. Já o MEG apresenta ambas essas precisões ao mesmo tempo. Não é incomum, no entanto, um efeito ser encontrado mais fortemente apenas por um(uns) desse(s) método(s) (cf. LAU *et al.*, 2013 para efeitos diferentes em diferentes técnicas; e cf. LAU *et al.*, 2016 para um estudo usando essas 3 técnicas).

Percebe-se, portanto, que as maneiras que temos atualmente de entender *como* a linguagem é processada passam na realidade pelo *onde* e o *quando* esse evento acontece (BRENNAN, 2022). Um método mais recente que tenta analisar diretamente o *como* veio com o avanço da computação, que permitiu a geração de modelos computacionais que mostram as maneiras mais eficazes de se realizar uma tarefa linguística. Sendo o corpo humano um organismo que, por questões evolutivas, busca a maior eficiência pelo menor custo, faz sentido testar modelos que busquem simular o processamento mais eficaz de linguagem. Esse é, no entanto, um caminho contrário ao dos estudos neurofisiológicos. Enquanto EEG, fMRI e MEG observam o que acontece no cérebro após um estímulo linguístico e inferem que as atividades encontradas se relacionam à linguagem, os modelos computacionais são alimentados com um resultado e buscam as maneiras pelas quais é possível se chegar a ele, isto é, os modelos. A neurolinguística e mesmo a linguística teórica podem se beneficiar imensamente de uma combinação dessas técnicas, mas ainda é cedo para afirmar que são realmente complementares. Afinal, a própria premissa de utilização dos modelos computacionais para estudar linguagem é falseável.

Alguns desses métodos já foram utilizados para estudar o processamento de palavras em português, conforme será discutido no capítulo 2. É imprescindível, no entanto, já mencionarmos trabalhos que influenciaram diretamente o estudo apresentado nessa dissertação. França *et al.* (2008) investigaram a hipótese de haver uma diferença entre o acesso semântico à parte arbitrária de uma palavra (i.e., à concatenação de sua raiz com seu *primeiro* morfema categorizador) e o acesso à parte composicional da palavra. Por exemplo, na palavra *globalização*, a parte que carrega a arbitrariedade saussureana é [glob+a-], enquanto a parte composicional é [-li+za+ção]. Para tanto, eles realizaram um teste de *priming* multimodal, dividindo seus estímulos em dois grupos, um em que *prime* e *target* tinham relacionamento fonológico entre si (GF) e um em que *prime* e *target* tinham relacionamento morfológico entre

si (GM). Cada um desses grupos continha estímulos com 3 tamanhos: GF apresentava palavras com 3 diferentes quantidades de sílabas, enquanto GM apresentava palavras com 3 diferentes quantidades de camadas morfológicas. Os autores acreditavam que essa diferenciação de tamanho os ajudaria a entender melhor a derivação lexical, uma vez que fazia o ponto da arbitrariedade saussureana variar nos estímulos do GF, mas não nos do GM.

Analizando os tempos de resposta comportamental e as ondas elétricas produzidas pelos participantes, os autores descobriram que, enquanto no GF o acesso a *targets* menores era mais rápido que o acesso a *targets* com mais sílabas, no GM o tamanho dos *targets* (i.e., quantidade de camadas morfológicas) não gerou nenhuma relevância estatisticamente significativa nos tempos de resposta e nas amplitudes. Além disso, os tempos de resposta foram globalmente menores para os estímulos do GM. Os autores sugeriram, então, a morfologia como um fator acelerador do reconhecimento de palavras em português e analisaram seus resultados à luz da Morfologia Distribuída (doravante MD). A concatenação da raiz com o *primeiro* morfema categorizador aconteceria durante a terceira fase descrita pela MD, a Enciclopédia, quando o significado de uma palavra seria pareado com sua forma, de maneira que esse seria o momento do acesso lexical por excelência, o momento da arbitrariedade saussureana, que representa um esforço cognitivo. Após essa *primeira* computação, as demais camadas morfológicas da palavra seriam processadas composicionalmente de forma menos custosa.

No entanto, como todos os *targets* do GM eram relacionados também semanticamente com seus respectivos *primes*, França *et al.* (2008) se perguntaram na conclusão de seu estudo o que aconteceria com pares de palavras que gerassem dúvidas nos falantes quanto ao seu relacionamento semântico sincrônico. Eles exemplificam a questão com as palavras *restaurante*, *refrigerante* e *liquidação*, que parecem não ter mais reconhecidas dentro delas as palavras *restaurar*, *refrigerar* e *líquido*, respectivamente. Será que, nesses casos quando há uma relação semântica diacrônica mas não sincrônica entre palavras da mesma família morfológica, a morfologia também é um fator acelerador do reconhecimento lexical?

Trata-se, enfim, da pergunta central dessa dissertação – se realmente enxergamos *refrigerar* dentro de *refrigerante*, *restaurar* dentro de *restaurante* e *líquido* dentro de *liquidação*! Contudo, já havíamos começado essa investigação durante a graduação da presente autora, trabalho que foi reportado tanto em sua monografia quanto em um artigo (cf. CATALDO; FRANÇA, 2021; CATALDO LOPES, 2020). Naquele projeto, realizamos um teste de *priming* para investigar possíveis diferenças entre as rotas semântica e morfológica de

processamento durante o acesso lexical. Tínhamos três condições experimentais: uma em que *prime* e *target* eram apenas semanticamente relacionados (*refrigerar-gelado*), uma em que *prime* e *target* era morfológica e semanticamente relacionados (*refrigerar-refrigeração*) e uma em que *prime* e *target* eram morfológica mas não semanticamente relacionados, hoje em dia (*refrigerar-refrigerante*). Essa última condição abarcava justamente a questão levantada por França *et al.* (2008). Dizemos que ela possui semântica opaca, justamente por não ser imediatamente transparente a existência do conceito de *refrigerar* dentro de *refrigerante*.

Escolhemos como tarefa experimental a decisão lexical e expomos os participantes a 24 pares de palavras de interesse e 24 pares de pseudopalavras, como distratores. Uma análise do tempo de resposta dos participantes indicou: (i) um processamento de memória associativa para pares de palavras com relação apenas semântica entre si (e.g., *refrigerar-geladeira*); (ii) um processamento linguístico computacional para pares de palavras com relação morfológica entre si, seja ela sincronicamente transparente (e.g., *refrigerar-refrigeração*) ou sincronicamente opaca (e.g., *refrigerar-refrigerante*); e (iii) a existência de novas entradas lexicais para palavras sincronicamente opacas em relação às que as geraram (e.g. *refrigerante*, em relação a *refrigerar*). Isso sugere que nós não tenhamos um conhecimento inconsciente de que *refrigerante* vem de *refrigerar* e que criemos uma nova raiz *refrigerant-* em nosso léxico mental e memória lexical. No entanto, uma vez que essa peça do léxico (a raiz) fosse acessada, o restante do processamento se daria de maneira composicional, conforme previam modelos decompositivos de processamento.

Dessa forma, encontramos em dois modelos teóricos um suporte teórico para explicar nossos resultados e aprofundar nossas perguntas: o modelo da Morfologia Distribuída, assim como França *et al.* (2008), o modelo de processamento lexical do *Affix Stripping* (TAFT; FOSTER, 1975). A tese de Taft e Foster é de que cada morfema seja armazenado individualmente no léxico mental, de forma que as palavras complexas sejam decompostas e tenham seus morfemas analisados antes do acesso lexical. Os resultados de Cataldo e França (2021) apontaram justamente para uma diferença de custo temporal entre o momento do acesso lexical (mais custoso) e a adição composicional de morfemas ao conjunto [raiz+morfema categorizador] (estatisticamente muito menos custosa).

Como acontece com a maioria dos estudos, a conclusão desse trabalho e sua apresentação em diversos congressos nos gerou ainda mais perguntas e decidimos continuar sua investigação realizando modificações no design experimental e adicionando mais um teste



comportamental e um teste neurofisiológico com EEG ao projeto. Discutiremos essas adições nessa dissertação. Acreditamos que a observação eletroencefalográfica será valiosíssima uma vez que nos permite analisar, com enorme precisão, o curso temporal e o custo do processamento das nossas palavras de interesse, excluindo fatores derivados do comportamento dos participantes. Isso ocorre porque esse equipamento amplifica e grava as ondas cerebrais imediatamente posteriores à apresentação de estímulos linguísticos. Dessa maneira, caso o mesmo padrão de onda seja observado após a apresentação de diversos estímulos de um mesmo tipo, propomos que tal padrão seja um Potencial (elétrico) Relacionado a Evento (linguístico, no caso) ou, em inglês *Event-Related Potential* (ERP)<sup>4</sup>.

Estudos anteriores na área apontam o N400 como um potencial relevante para essa investigação, por ele geralmente ser atrelado ao momento do acesso lexical (FRANÇA *et al.*, 2008; KUTAS; HILLYARD, 1980). Assim como acontece com as outras assinaturas neurofisiológicas, seu nome é dado de acordo com suas características: ele apresenta amplitude negativa aproximadamente aos 400ms após a apresentação do estímulo de interesse.

Dessa maneira, no presente trabalho temos como objetivo analisar o processamento de palavras do tipo *refrigerar-refrigerante*, que não possuem mais relação semântica sincrônica, mas que ainda mantêm peças morfológicas superficialmente relacionadas. Faremos isso através de dois testes experimentais: um comportamental e um neurofisiológico com EEG. Ambos terão o mesmo design experimental: serão um teste de *priming* multimodal (*prime* auditivo e alvo visual), com decisão lexical como tarefa experimental (dizer se uma palavra existe ou não). A escolha do *priming* multimodal se justifica na tentativa de diminuir e talvez mesmo eliminar fatores ortográficos da Gráfico. Isso se faz necessário porque, no português, dificilmente uma palavra morfológicamente relacionada a outra não compartilhará com esta semelhança ortográfica e fonológica. Já a decisão lexical foi escolhida como tarefa justamente por elicitar o processamento lexical e focar a atenção do participante na classificação do “diferente”, isto é, das pseudopalavras. Apostamos que, dessa maneira, o processamento das palavras que existem (i.e., que fazem parte de nossas condições de interesse) seja o mais próximo possível do processamento natural de palavras isoladas.

---

<sup>4</sup> Os ERPs também são chamados de *componentes neurofisiológicos* e *assinaturas neurofisiológicas ou neurais*, justamente por, em geral, demarcarem o efeito do processamento de um tipo específico de estímulo.

No estudo comportamental, analisaremos como variáveis dependentes o tempo de resposta e a acurácia e, no teste de EEG, além desses dois fatores, analisaremos a latência e amplitude das ondas, no momento que vai de 300 a 500ms após a apresentação do estímulo-alvo (MORRIS; STOCKALL, 2012).

Como variáveis dependentes, teremos: i) o tipo de relacionamento entre *prime* e *target*: Morfológico Intacto (MI – como em *refrigerar-refrigeração*); morfológico obscuro (MO – como em *refrigerar-refrigerante*); e uma condição sem (nenhum tipo de) relacionamento, nem semântico, nem morfológico (SR – como em *refrigerar-caneta*). Essa última condição é uma adição em relação ao estudo anterior, que tem como objetivo funcionar como baseline e como controle; ii) a quantidade de camadas morfológicas da palavra de interesse: 1 camada ou 2 camadas (*refrigerar-refrigerante* e *refrigerar-refrigerantezinho*, respectivamente). Esse fator visa a avaliação do custo de uma adição composicional a uma raiz já acessada. Uma análise neurofisiológica do mesmo nos permitirá entender se o custo dessa adição composicional será refletido em um N400 com amplitude maior para estímulos com 2 camadas morfológicas (conforme previsto pela hipótese do *Affix Stripping*, com a exceção de que ela foi baseada em testes comportamentais) ou se, após o ponto da arbitrariedade saussureana, o custo da adição composicional não será nem mesmo refletido neurofisiologicamente (como foi reportado em FRANÇA *et al.*, 2008).

Teremos, portanto um design 3x2. Outros elementos importantes de nosso design se justificam na literatura experimental disponível. Parece ser o caso de escolhas experimentais influenciarem fortemente os resultados encontrados para o processamento de palavras semanticamente opacas: quanto mais encoberta é a apresentação do *prime* e quanto menor é o tempo entre a apresentação do *prime* e do *target* (SOA)<sup>5</sup>, mais decomposicional é o processamento das opacas (HEYER; KORNISHOVA, 2018; FELDMAN; BARAC-CIKOJA; KOSTIC, 2002). Como obtivemos resultados positivos de decomposição para palavras semanticamente opacas na *primeira* etapa (já concluída) desse projeto (cf. CATALDO E FRANÇA, 2021), resolvemos nessa segunda etapa codificar nossa experimentação com as condições mais adversas ao processamento decomposicional para ver se, mesmo face a elas, ele se mantém.

---

<sup>5</sup> Essa variável é conhecida como SOA – Stimulus Onset Asynchrony.

Como hipóteses e previsões, temos: i) a existência de entradas lexicais diferentes para palavras sincronicamente transparentes e opacas, como *refrigerar-refrigeração* e *refrigerar-refrigerante*. Para tanto, prevemos tempos de resposta e amplitudes diferentes para essas duas condições; ii) um curso decomposicional de processamento para todas as palavras, tanto antes quanto após o acesso lexical (respectivamente, decomposição para retirada das partes constituintes da palavra e posterior composicionalidade do sentido de cada morfema adicionado à [raiz+morfema categorizador]. Para tanto, prevemos tempos de resposta e amplitudes maiores para as condições com 2 camadas morfológicas, em comparação com as de 1 camada morfológica, como manifestação da pequena adição computacional sofrida. Por exemplo, o par *refrigerar-refrigerantezinho* teria um custo adicional em relação a *refrigerar-refrigerante*, por conta da adição de uma camada morfológica [zinho], que também se manifestaria na comparação das amplitudes e tempos de respostas gerados pelos estímulos dessas condições. Aqui nos baseamos na proposta do *Affix Stripping* (TAFT; FOSTER, 1975) que, ao propor uma decomposição da palavra em seus morfemas constituintes, prevê justamente um custo de processamento (mesmo que pequeno) para cada adição morfológica de afixos. Escolhemos apostar no *Affix Stripping* porque os resultados comportamentais reportados em Catado e França (2021) mostravam clara e estatisticamente uma diferença de processamento entre pares de palavras morfolologicamente relacionadas (independentemente de seu nível de transparência semântica) com 1 e 2 camadas morfológicas; iii) ausência de facilitação (*priming*) para as condições em que *prime* e alvo não têm nenhum tipo de relacionamento entre si, como *refrigerar-caderno*. Assim, prevemos tempos de resposta e amplitudes maiores para os estímulos dessas condições, quando comparados com estímulos com a mesma quantidade de camadas morfológicas, mas com algum tipo de relacionamento entre *prime* e alvo (morfológico intacto ou morfológico semanticamente opaco).

Uma vez introduzida a área em que o presente trabalho se enquadra, junto a seus objetivos, a metodologia escolhida e nossas hipóteses e previsões, resta-nos mergulhar mais fundo nas discussões. No próximo capítulo (capítulo 2), discutiremos os modelos teóricos do Lexicalismo e o da Morfologia Distribuída – sendo que o último foi escolhido como o que melhor explica as operações sintáticas e semânticas que buscamos apreender com os experimentos – (seção 2.1), as teorias de processamento com seus estudos experimentais, demonstrando a variedade de designs experimentais encontrados na literatura, junto a sua influência nos resultados obtidos (seção 2.2). Em seguida, focaremos em duas variações

experimentais de suma importância para esse trabalho: estudos que analisaram a transparência e opacidade semântica (seção 2.3) e, por fim, estudos que variam o SOA (*stimulus onset asynchrony*), tempo de apresentação entre o *prime* e o alvo em estudos de *priming* (seção 2.3).

Em seguida, aprofundaremos o âmbito neurofisiológico da pesquisa com palavras (capítulo 3), apresentando o que já foi descoberto sobre o processamento visual de palavras na seção 3.1, sobre a fisiologia do cérebro humano e do funcionamento do EEG na seção 3.2 e, finalmente, sobre o ERP de interesse para esse estudo na seção 3.3: o N400. Apresentaremos pesquisas relevantes para descoberta e caracterização do N400, assim como pesquisas que analisaram o N400 manipulando estímulos semanticamente transparentes e opacos.

Passaremos, então, para a apresentação dos experimentos: o comportamental no capítulo 4 e o neurofisiológico no capítulo 5. Ambos os capítulos contarão com uma justificativa e o objetivo do teste (seção x.1), a apresentação dos materiais experimentais (seção x.2), do desenho experimental e dos procedimentos de coleta (seção x.3), a discussão das hipóteses e previsões (seção x.4), a caracterização dos participantes (seção x.5), a explanação dos resultados (seção x.6) e sua discussão (seção x.7).

Por fim, concluiremos o texto no capítulo no capítulo 6, mencionando algumas perguntas em aberto e possíveis caminhos futuros de investigação. No capítulo 7, estão disponíveis as referências bibliográficas e, em anexo, todos os estímulos experimentais utilizados nos testes e gráficos de amplitudes em todas as regiões cerebrais de interesse.

## 2. Processamento e representação de palavras

---

What are the building blocks out of which everything that exists is made? What forces or interactions or “fit” binds them into the familiar objects we see around us every day? (MARSHAL; ZOHAR, 1997:3).

### 2.1 *Lexicalismo e Morfologia Distribuída*

Conforme a linguística foi sendo estabelecida como uma ciência cognitiva e, portanto, natural, perguntas como a de Marshal e Zohar, na epígrafe do capítulo, passaram a fazer muito sentido quando dirigidas ao objeto desse estudo e dessa ciência: a faculdade da linguagem: “Quais são os primitivos a partir dos quais tudo o que existe no mundo é feito?”

Se todo sistema precisa de peças que funcionem segundo certas regras, quais seriam as peças mínimas da linguagem? Até onde podemos dividi-la? Quais são os tijolos que, combinados uns aos outros, constroem os sentidos que veiculamos dentro da nossa consciência e para outras pessoas?

Uma resposta rápida para essa pergunta poderia ser as palavras. Afinal, à *primeira* vista parecem ser unidades pequenas, carregadoras de sentido e que, combinadas com outras de natureza similar, geram sentidos maiores e mais complexos. De fato, durante séculos, os filósofos, gramáticos e poetas acreditaram que a palavra era o tijolinho da linguagem: a unidade mínima carregadora de sentido e que, combinada com outras de natureza similar, gerava sentidos maiores e mais complexos. Essas são teorias *lexicalistas*, que propõem uma arquitetura da gramática com duas computações separadas: uma que monta traços que formam palavras constituindo o léxico do indivíduo, e outra que tira do léxico palavras que são combinadas umas com outras na computação sintática.

Nestes modelos lexicalistas, a montagem de palavras precede as operações da sintaxe. A título de exemplo, no Programa Minimalista (CHOMSKY, 1995), palavras já completamente constituídas entram em um espaço de trabalho transitório chamado Numeração, que contém o conjunto de palavras a serem concatenadas (*merge*) na fase em curso. A operação *Selecionar* escolhe uma palavra da Numeração e a introduz na computação para ser concatenada a uma outra palavra ou sintagma, que possivelmente será sujeita a operações sintáticas como o

deslocamento. Portanto, nos modelos lexicalistas, a sintaxe lida com palavras pré-formadas, aplicando a elas as operações de *concatenar* e *mover* (FRANÇA; LEMLE, 2006).

Entretanto, basta uma breve incursão tipológica para que esse modelo seja questionado. Estudar o que há de comum e de diferente nas diversas línguas existentes traz a consciência de que conceitos que parecem muito intuitivos para falantes de uma língua não o são para os de outras. A definição de palavra parece ser um exemplo, conforme exemplificamos na introdução dessa dissertação com o exemplo dos verbos recursivos do Guarani Mybá, uma língua que aglutina morfemas. Apoiando-se no lexicalismo, não há como dar conta da adequação explicativa ao se considerar línguas como essa.

Surgiu, então uma teoria que toma o morfema como unidade mínima sintática e semântica da linguagem, a Morfologia Distribuída (doravante MD; HALLE; MARANTZ, 1993; MARANTZ, 1997; HARLEY; NOYER, 1999). Trata-se de uma das teorias decomposicionais mais discutidas, que completa 30 anos neste ano de 2023. Ela traz uma solução arquitetônica interessante que assume que as unidades básicas que entram na computação sintática não são palavras, mas sim apenas os seus traços abstratos, sem conteúdo fonológico. Nas palavras de Pylkkänen (2002, p.12): “[...] on the basis of lexical integrity alone, it seems unwarranted to draw the strong conclusion that entirely different modules of grammar must be responsible for the construction of complex entities such as *joyfulness* as opposed to complex entities such as *the girl ran*”.

A MD deu conta dessa ideia ao propor que as palavras, com seus traços sintáticos, traços morfofonológicos e propriedades semânticas, não seriam então pré-montadas, mas sim formadas dinamicamente a partir da distribuição de tarefas entre três diferentes listas que participam de três diferentes estágios na derivação da estrutura sintática e em sua interpretação.

O primeiro estágio é a *Lista 1, que contém traços abstratos sem substância fônica*. O que costuma ser atribuído a esta lista é a contraparte abstrata dos itens funcionais, tais como: verbalizador, determinante, relacionador, pessoa, número, tempo, complementizador, agentivizador, aplicativo, além de posições vagas (place-holders) reservadas para a inserção tardia das raízes, junto à posição que elas assumirão na estrutura sintática (isso se dará de acordo com a tipologia da língua). Ou seja, fica definido e salvaguardado já na lista 1 o local de inserção da raiz (FRANÇA; LEMLE, 2006; HARLEY, NOYER, 1999).

O segundo estágio é a *Lista 2, constituída de Peças de Vocabulário de dois tipos: funcionais ou raízes*. Estas peças são inseridas no output da sintaxe, contanto que sejam

compatíveis com a estrutura de traços gerada na Lista 1. Assim, é do lado da morfofonologia, no componente morfologia, que acontece a inserção de Peças do Vocabulário, com substância, formando uma representação fonológica subjacente, que é o input da fonologia, com todas as suas regras. Isso constitui uma característica importante e singular desse modelo: a inserção fonológica tardia, característica que cede ao *morfema*, na MD, um conceito diferente daquele que ele recebe na maior parte das demais teorias. De acordo com Silva e Medeiros (2016, p. 111): “[não] devemos confundir o que é um morfema para a Morfologia Distribuída – um feixe de traços morfosintáticos abstratos, sem contraparte fonológica – com o que é um morfema no sentido tradicional – um amálgama de propriedades semânticas, gramaticais e fonológicas”.

Na MD, a inserção lexical tardia obedece ao princípio do subconjunto: uma peça só pode ser inserida em um nó se sua especificação for igual às especificações dos traços provenientes da computação sintática, ou se ao menos contar com um subconjunto dessas especificações. Por esta razão, pode haver Peças de Vocabulário sub-especificadas em relação aos traços fornecidos pela sintaxe. Por exemplo, a Lista 1 pode selecionar um feixe de traços {raiz, nominalizador, sing, masc}. Uma estrutura vai ser formada e as especificações sintáticas podem ser vistas nos nós terminais da estrutura. Agora, na hora de inserir os itens de vocabulário na Lista 2 a peça de vocabulário vogal temática disponível para a raiz [cliente\_] é o [e], comum de dois gêneros. Assim a Lista 1 especificou mais do que a Lista 2, mas ainda assim, as peças [client+N+e] ganharam como as que melhor atenderiam às especificações que vieram da Lista 1.

O terceiro e último estágio é a *Lista 3, denominada Enciclopédia, o componente interpretativo que fornece a parte idiossincrática da leitura semântica*. Ou seja, este é o local onde se dá a negociação da arbitrariedade saussureana (arbitrariedade na relação entre forma e significado no sentido próprio de Saussure), quando, logo na *primeira* fase, uma raiz é concatenada a um morfema categorizador. Se houver outras camadas sintáticas, elas serão concatenadas fase a fase e, na Forma Lógica, receberão as leituras composicionais dos traços abstratos, leituras estas que serão integradas com a leitura idiossincrática dada na Enciclopédia (FRANÇA, LEMLE 2006; HARLEY, NOYER 1999).

A partir da primeira fase, o composto pode ainda receber outras camadas de concatenação, voltando o ciclo derivacional para a Lista 1. Porém, esse ciclo não passará mais pela Lista 3, pois não haverá mais negociação semântica de base arbitrária, entre o conteúdo da raiz e o primeiro morfema categorizador. A partir deste ponto, há apenas concatenações

composicionais. A título de exemplo, aplicando as previsões da MD, na primeira fase da derivação de *refrigeração*, *refrigerar* é lido pela Enciclopédia e a raiz [refrig-] é computada no contexto de um verbalizador, formando *refrigerar*. Depois, *refrigerar* volta para a derivação e pega na Lista 1 o feixe de traços {nominalizador, sing, fem} entra na Lista 2 e insere o nominalizador [ção] como o melhor candidato para as especificações de traços que estão nos terminais sintáticos formados a partir da Lista 1. Neste ponto, o composto não volta à enciclopédia porque a base semântica já havia sido acordada na primeira fase da derivação que formou *refrigerar*. A distância semântica entre *refrigerar* e *refrigeração* depende apenas de um cálculo composicional, i.e., a partir do significado de *refrigerar*, que é verboso, é possível fazer-se um ajuste no nominalizador [ção], chegando-se composicionalmente à *refrigeração* (figura 2).

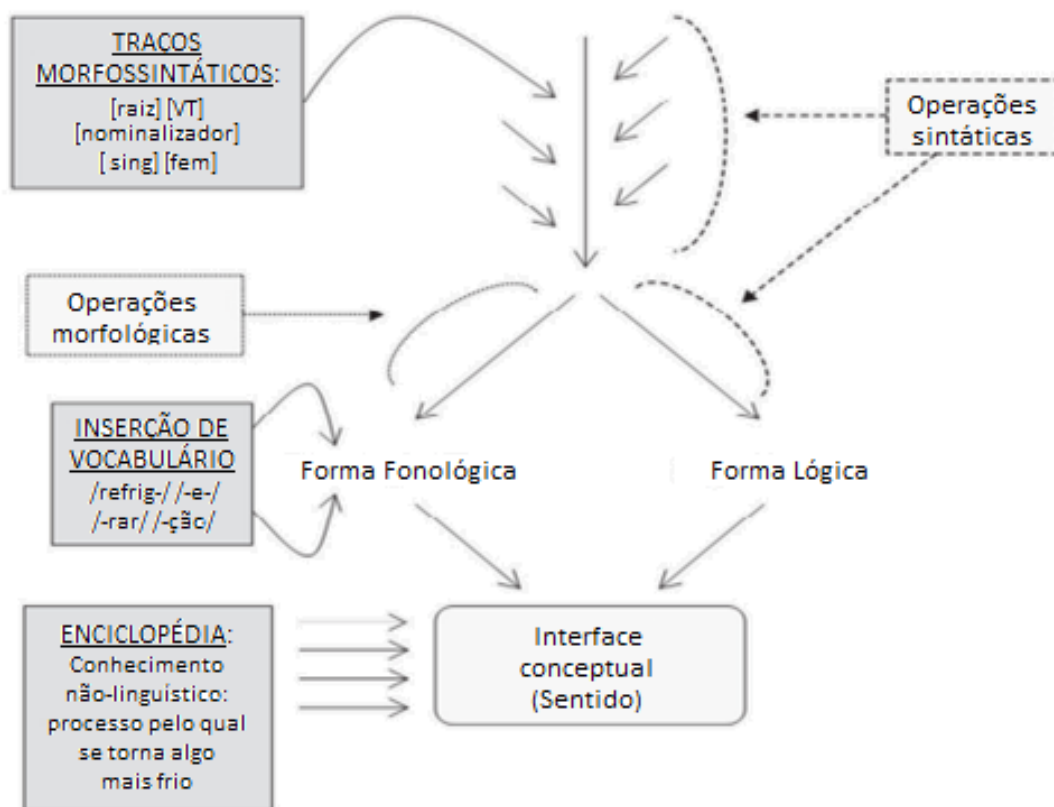


Figura 2. A arquitetura da gramática na Morfologia Distribuída, exemplificada com a palavra *refrigeração*. Fonte: Adaptado de Scher, Bassani, Minussi, 2013:19.

A MD, com sua organização e capacidade de previsibilidade, trouxe para a área de estudos linguísticos uma grande possibilidade de ser aferida empiricamente. Portanto, ela atinge



consonância entre Teoria e resultados experimentais, permitindo uma adequação além da explicativa (neurofisiológica) a dados descritivos, além, também, de acabar incitando novas perguntas experimentais. Assim, uma associação entre pesquisa teórica e empírica, geralmente difícil de ser feita, pode se tornar mais corrente. Todavia, é necessário aqui fazer uma ressalva: apesar de a MD ser cada vez mais corroborada pela testagem empírica, ainda é uma teoria jovem e, portanto, é cedo para entender exatamente como ela se implementa na prática. As hipóteses e conclusões levantadas pelos pesquisadores, por mais sofisticadas que sejam, ainda precisam ser ainda muito mais testadas e desenvolvidas. Como parte desse esforço, elencaremos, no próximo capítulo, alguns dos resultados experimentais mais relevantes já reportados.

## *2.2 Teorias de Processamento e estudos experimentais*

Enquanto ciência que tem como objeto a capacidade para a linguagem, comum a todo ser humano (isto é, a pessoas que falam as mais diferentes línguas), a psico- e a neurolinguística trouxeram uma nova frente para a discussão teórica sobre a definição de palavra e sobre o átomo da linguagem. Se permanece um fato que a hipótese lexicalista não dá conta de línguas aglutinantes como o Guaraní Mbyá, é também verdade que a análise linguística teórica e a comparação tipológica não são suficientes para descreditar a possibilidade de estarmos tratando de um parâmetro e não de um princípio da linguagem (CHOMSKY, 1981).

Os estudos da cognição da linguagem vieram, portanto, nos ajudar a entender como a representação e a derivação linguística funcionam durante o processamento de linguagem, no módulo cognitivo da linguagem que todo ser humano possui. Afinal, a descoberta de como uma frase é processada e de como se atribui um significado a uma sequência de sons ou de riscos em um papel passa pela identificação das etapas e dos elementos constituintes desse processo. Isso é possível porque o avanço da tecnologia nos permitiu um acompanhamento online e preciso do processamento, isto é, a aferição de medidas “durante o processamento cognitivo que uma pessoa realiza inconscientemente enquanto recebe um estímulo linguístico oral ou

escrito” (KENEDY, 2015, p. 148)<sup>6</sup>. Isso se dá através dos experimentos cognitivos (comportamentais e neurofisiológicos). Tais testes geraram e foram gerados a partir de três teorias de processamento.

A primeira delas, advinda de teorias linguísticas como o lexicalismo, propõe um modelo de listagem plena, em que o acesso lexical seria feito direto às palavras já montadas, sem a necessidade de estoque e remontagem de suas partes (BUTTERWORTH, 1983; JACKENDOFF, 1983; BYBEE, 1995; HAY; BAAYEN, 2005; BAAYEN *et al.*, 2011). Nesse modelo, a memória é essencial e o cérebro teria uma capacidade de armazenamento virtualmente ilimitada. Isso porque uma palavra como globo seria memorizada de maneira independente de todas as suas possíveis formas derivadas, como global, globalizar e globalização. Precisam também ser memorizadas todas as situações de uso de cada uma dessas palavras com outras palavras. De acordo com Cataldo e França (2021, p.5):

Seria, assim, formada uma rede de itens – menos ou mais reforçados pela frequência de cada um e pela frequência de suas famílias – que poderiam ser acessados diretamente em qualquer acepção. Portanto, o caminho tomado nessa abordagem começa pela semântica e vai na direção de cima para baixo (*top-down*), supostamente garantindo a eficiência necessária ao uso online imediato e plenamente contextualizado da linguagem (CATALDO; FRANÇA, 2021:5).

Já a segunda hipótese está em consonância com o proposto pela MD: o cérebro humano conseguiria decompor as palavras que processa em morfemas, além de juntar tais morfemas para produzir novos compostos, em um exemplo de criatividade linguística. Essa capacidade seria determinada pela genética da espécie, podendo inclusive estar presente em outros sistemas cognitivos. Dessa maneira, a operação básica de concatenar (*merge*, em inglês) não seria computacionalmente custosa e poderia ser eficientemente utilizada a todo tempo. Ela seria responsável pela sintaxe utilizada para juntar palavras em uma frase, assim como pela sintaxe utilizada para juntar morfemas em palavras complexas. Haveria, portanto, o que chamamos de computação morfossintática plena (TAFT; FOSTER, 1976; TAFT, 1979; FOSTER, 1979; PYLKKANEN *et al.*, 2000; 2002; 2003; EMBICK *et al.*, 2001; PYLKKANEN; MARANTZ,

---

<sup>6</sup> Já o processamento *offline* conta com uma reflexão consciente do participante sobre o estímulo ao qual foi apresentado.

2003; RASTLE; DAVIS; NEW, 2004; STOCKALL; MARANTZ, 2006; KAZANINA *et al.*, 2008; McCORMICK; BRYLSBAERT; RASTLE, 2009; BEYERSMANN; CASTLES; COLTHEART, 2013; PINTO; VILLALVA; COSTA, 2015; SAGUIE; MAIA; SANTOS, 2018; CATALDO; FRANÇA, 2021; CABRAL; SOTO, 2023). Nas palavras de Cataldo e França (2021, p. 6), haveria uma “[...] decomposição completa das partes na direção de baixo para cima (*bottom-up*), indo da sintaxe à semântica através de regras computacionais explícitas”.

Como consequência de um modelo como esse, não precisaríamos de tanta memória lexical (como apostam os modelos de listagem plena). Pelo contrário, o mais eficaz seria guardarmos o menor número possível de léxico na mente e submetermos os morfemas e os pareamentos de forma e significado já estocados à atuação de regras de merge, tão eficazes e pouco custosas quanto descrito acima. O estudo seminal e pioneiro desse modelo foi o de Taft e Foster (1975), que propôs a existência do chamado *affix stripping*: a tese de que todos os afixos de uma palavra seriam decompostos automaticamente, antes mesmo que o sentido de sua raiz fosse acessado.

Mas como um estudo psico e/ou neurolinguístico poderia chegar a conclusões desse tipo? Geralmente, são feitos estudos que propiciam a ativação lexical, isto é, o processamento de uma palavra, com ativação de seu significado. Uma das maneiras de se realizar isso é apresentando palavras a um participante e dando a ele uma tarefa como a de dizer se essas palavras existem ou não em sua língua, a chamada tarefa de decisão lexical. Essa tarefa elicitava o processamento lexical e foca a atenção do participante na classificação do “diferente”, isto é, das pseudopalavras ou não-palavras. É uma maneira de aproximar o processamento das palavras que existem (que constituem as condições de interesse) do processamento natural de palavras isoladas.

Pode-se, ainda, deixar o design experimental mais complexo, utilizando um paradigma experimental como o de *priming*. O objetivo de um paradigma como esse é avaliar se o acesso a uma palavra, por exemplo, é facilitado pela apresentação anterior de um item (palavra, imagem, som etc.) que mantenha algum tipo de relacionamento com ela. Trata-se de uma técnica psicológica antiga, usada inclusive em campanhas políticas e de marketing. Um consumidor será influenciado a comprar um produto se a ele forem associadas experiências e sensações positivas. Assim, o consumidor terá uma sensação de satisfação facilitada/primada por tal produto.

Nos estudos da linguagem, o objetivo é verificar se o acesso lexical a uma palavra é facilitado pelo fato de uma outra ter sido apresentada anteriormente. A *primeira* palavra apresentada geralmente mantém algum tipo de relacionamento (morfológico, semântico, ortográfico, fonológico) com a segunda, a que é analisada. Assim, pode-se verificar se uma via de processamento morfológica, semântica, ortográfica ou fonológica facilita o processamento lexical, o que nos permitiria propor que a ativação lexical passaria por um processo movido pela identificação de características morfológicas, semânticas, ortográficas, fonológicas etc. da palavra.

Por exemplo, no português brasileiro (doravante PB), Garcia (2009) realizou um estudo psicolinguístico de *priming* em que comparou o processamento de pares de palavra com relacionamento morfológico (FILA-fileira), semântico (ORDEM-fileira) ou fonológico/ortográfico (FILÉ-fileira) entre si. Ela utilizou, ainda, a técnica do *priming* encoberto (*masked priming*, em inglês), que mostra o *prime* subliminarmente, de forma que o participante às vezes nem se dá conta, conscientemente, de o ter visto. O processamento visual da pessoa, no entanto, registra invariavelmente tal *prime*.

A pesquisadora avaliou a acurácia e os tempos de resposta da tarefa comportamental que deu a seus participantes, registrando tempos de resposta significativamente menores para pares da condição morfológica, quando comparados aos de relacionamento semântico e fonológico/ortográfico. Para esses últimos pares de condição fonológica, os tempos de resposta foram até maiores que para os da condição controle de relacionamento nulo. Apesar de essa última diferença não ter sido estatisticamente relevante, a autora sugere a interpretação de que a semelhança fonológica entre *prime* e alvo possa inibir o processamento lexical. Essa é uma proposta que vai ao encontro da inserção fonológica tardia proposta pela Morfologia Distribuída. Junto a isso, os baixos tempos de resposta para os pares da condição morfológica levaram Garcia a propor um modelo decomposicional de processamento para o português brasileiro, de acordo com o qual o léxico mental seria estruturado morfológicamente, morfema a morfema. Assim, fatores semânticos tardios não seriam relevantes para os momentos iniciais do processamento.

Conforme discutido na introdução, os resultados França *et al.* (2008) também podem ser interpretados como ratificando a existência de um processamento decomposicional e com inserção fonológica tardia. A razão dessa conclusão é a condição de relacionamento fonológico ter apresentado tempos maiores de processamento conforme mais sílabas eram incluídas na

palavra alvo, mas o mesmo não ter acontecido com a condição de relacionamento morfológico que, por sua vez, registrou acesso lexical sempre no momento em que a raiz da palavra era concatenada com o *primeiro* morfema categorizador, independentemente de quantas camadas composicionais eram adicionadas à palavra depois. Essas conclusões vieram da observação de medidas neurofisiológicas de EEG, cuja interpretação será melhor discutida no próximo capítulo.

Por fim, há, ainda, um terceiro modelo de processamento lexical, chamado de dupla rota. O nome vem da proposta de que o cérebro esteja pronto tanto ao acesso *bottom-up* quanto *top-down*, uma vez que o cérebro é equipado com córtices para o processamento e para o armazenamento, de forma que a utilização de um não exclua a do outro. A escolha da rota utilizada se daria primando pela eficiência e dependeria das características da palavra a ser acessada. Pinker (1991), por exemplo, propõe que verbos conjugados no passado simples do inglês seriam acessados por decomposição quando regulares, isto é, quando sua desinência modo-temporal é marcada pelo sufixo -ed (to walk-walked, to bake-baked, to surf-surfed), já que seria muito fácil e comum identificar suas partes. Já quando o passado fosse irregular, marcado por um alomorfe (to eat-ate, to do-did, to write-wrote), o processamento seria feito de maneira plena, através do acesso a redes de associação semântica (cf. também CARAMAZZA *et al.*, 1988; PINKER; PRINCE, 1992; MARSLER-WILSON *et al.*, 1994; SCHREUDER; BAAYEN, 1995).

Cabral e Soto (2023) é um estudo em português brasileiro que avaliou o processamento de participios passados flexionados vs. derivados, apostando em uma hipótese de dupla rota. Consideraram participios flexionados os verbais e derivados os nominais. Os autores controlaram frequência lexical e acreditaram que participios nominais seriam processados de forma inteira quando possuísem alta frequência, enquanto participio nominais de baixa frequência e verbais passassem por segmentação morfossintática. Encontraram, no entanto, efeitos gerais de alta frequência e *priming* para todas as condições, acabando por confirmar a existência de uma decomposição plena desses estímulos. Ressalvam, no entanto, que seria importante realizar um estudo com participios passados em contexto de sentença, justamente para aferir o nível da influência deste fator na existência ou não de um processamento decomposicional.

Uma vez compreendia a relação dessas hipóteses de processamento com resultados experimentais encontrados na literatura, ainda resta, contudo, uma questão: como essas teorias

de processamento se comportam quando estamos tratando de palavras semanticamente opacas, como as que são o objeto de estudo do presente trabalho? Teorias de listagem plena dirão que palavras opacas serão acessadas inteiras, como qualquer outra palavra; teorias decomposicionais propõem que, mesmo sem um reconhecimento aparente da morfologia regular dessa palavra, nosso módulo linguístico é capaz de perceber esse tipo de padrão; por fim, teorias de dupla rota dirão que esse é justamente o tipo de característica que diferencia o acesso decomposicional do pleno. Essa questão será melhor discutida e exemplificada no próximo módulo.

### 2.3 *Transparência e opacidade semântica*

A transparência semântica é uma característica de palavras cujo sentido é facilmente depreendido pelo sentido de cada uma das partes que compõem tal palavra. A palavra *globalização*, por exemplo, passa por adições composicionais de sufixos a partir da [raiz + morfema categorizador] *glob-o*: *glob-al* junta o significado da raiz *glob-* com o sufixo adjetivizador, garantindo ao nome que modifica ter característica de *glob-*; a adição do verbalizador *-izar* em *glob-al-izar* denota o processo de tornar global (i.e., tornar algo como tendo característica de *glob-*); e, finalmente, a palavra volta a ser um nome com o sufixo *-ção* que, ao invés de denotar o objeto *glob+o*, denotará o processo de tornar um objeto como tendo característica de *glob-*.

Essa depreensão de sentido não é tão evidente, contudo, em uma palavra como *soube*, pretérito perfeito do indicativo do verbo *saber* na primeira pessoa do singular. Isso porque trata-se de um alomorfe do morfema da raiz do verbo, ao contrário do que acontece em uma conjugação regular como a de *comer-eu comi*. Com o avanço dos estudos sobre acesso lexical, a rota de processamento de palavras semanticamente opacas virou uma pauta instigante de pesquisa

Rastle, Davis e New (2004), em um estudo comportamental seminal, analisaram tempos de resposta para o acesso a palavras alvo apenas aparentemente relacionadas a seus *primes*, mas sem relacionamento semântico entre eles, como *corner / corn* do inglês. O sufixo agentivo *-er*, é muito produtivo no inglês (*fertiliz-er* é um algo que fertiliza, *bak-er* é aquele que faz doces), mas não é de fato um morfema/sufixo nessa palavra, pois *corn* significa *milho* e *corner* significa *esquina*. Se decomposta automaticamente *corner* se torna, portanto, uma palavra cujo sentido

é opaco, não imediatamente depreensível pela análise consciente de suas partes. Os pesquisadores observaram, no entanto, um efeito positivo de *priming* para esse tipo de palavra quando o *prime* foi apresentado de forma encoberta. Isso porque, no *priming* encoberto, a informação do *prime* é subliminar e, através do *affix stripping*, *corner* seria imediatamente segmentado em *corn-*, *-er*. O *corn-* residual poderia, então, pré-ativar o alvo *corn*, apresentado de forma aberta logo em seguida, de maneira que os participantes encarariam as palavras falsamente aparentadas como possuindo entradas iguais. Portanto, os autores reportam efeito positivo de *priming* até para palavras que tem relacionamento pseudo-morfológico.

Um resultado contrário foi reportado no português brasileiro, com Maia, Lemle e França (2007). Eles utilizaram um eye-tracker para rastrear as sacadas de olho de participantes que eram expostos ao teste *stroop*<sup>7</sup> aplicado ora à fronteira morfológica e ora à fronteira pseudomorfológica (**mocinho** ou **mocinho**). Depois da apresentação dos estímulos os participantes eram instruídos a responder a cor das letras. Por exemplo, “Qual a cor do i?” [**i** ou **i**] para assim verificarem quanto conhecimento intuitivo a respeito dos morfemas dentro das palavras os falantes de PB possuem. A escolha do efeito *stroop* aplicado à fronteira morfológica foi mais forte e incitou o reconhecimento lexical da palavra escrita de forma mais acurada do que a da condição pseudo-morfológica. Os resultados deste estudo assim parecem corroborar modelos *bottom-up*, apontando para a existência de um processamento morfológico mais consciente para morfemas que foram computados composicionalmente.

Maia e Ribeiro (2015) trabalharam com o acesso lexical a pseudopalavras mono e polimorfêmicas (como *liboramima* e *norbalense*, respectivamente), comparando-o com o de palavras também mono e polimorfêmicas (como *jabuticaba* e *jornaleiro*, respectivamente). O uso das pseudopalavras com sufixos comuns na língua estudada permite que se saiba qual curso segue o processamento linguístico de um participante, quando a ele é apresentada o que a princípio é uma nova palavra do português. Concluíram que o acesso a palavras e pseudopalavras polimorfêmicas segue um curso *bottom-up*, pela possibilidade de decomposição do lexema. Propuseram ser esse o mesmo motivo pelo qual as pseudopalavras e

---

<sup>7</sup> No paradigma do efeito *stroop*, escreve-se o nome de uma cor, como vermelho e muda-se a cor da fonte para uma outra que não vermelho: amarelo, por exemplo. Assim, o participante vê a palavra vermelho escrita em amarelo. Sua tarefa é nomear a cor em que a palavra vermelho está escrita, portanto, ele deve dizer amarelo. Um grupo controle em que a palavra escrita corresponde à cor que ela denota também é testado (vermelho escrito em vermelho). Para um melhor detalhamento da técnica, cf. *Stroop*, 1935. No teste reportado foi feita uma adaptação do efeito *stroop* aplicada à fronteira de morfemas

as palavras multimorfêmicas tiveram mais aceitação na tarefa de decisão lexical, quando comparadas às palavras e pseudopalavras monomorfêmicas. Não descartaram, no entanto, a possibilidade da existência de um processamento dupla rota.

Alguns outros estudos também apostaram nas pseudopalavras como estímulos ideais para avaliar essa questão. Kazanina *et al.* (2008), por exemplo, realizaram em um estudo em russo cujos resultados os fizeram propor que houvesse decomposição, não obrigatoriamente governada por transparência semântica, até que as menores unidades morfológicas fossem alcançadas. Isso porque encontraram facilitação para alvos transparentes (flexionados no diminutivo em relação ao *prime*) e pseudoderivados, mas não para alvos que compartilhavam apenas forma (ortografia/fonologia) com seus *primes* (i.e., que com eles não compartilhavam qualquer relação morfológica).

No holandês, Roelofs e Baayen (2002) utilizaram-se de uma tarefa cognitiva diferente da decisão lexical – a escolha de uma resposta dentre várias disponíveis – para estimular a produção de linguagem (até então apresentamos apenas trabalhos com compreensão). No entanto, também concluíram seu estudo tendendo para a autonomia da morfologia em relação à semântica no processamento linguístico de palavras complexas, quando comparadas a palavras simples (não complexas).

Contudo, Longtin, Segui e Hallé (2003) demonstraram que, no francês, o *priming* multimodal em pseudoderivação é um fator que gera facilitação apenas para pares semanticamente interpretáveis, ao contrário do que acontece na modalidade encoberta. Diependaele, Sandra e Grainger (2005) utilizaram o mesmo método em um estudo com francês e holandês, mas obtiveram resultados ligeiramente diferentes para cada uma das línguas: enquanto no holandês a transparência semântica foi um fator de facilitação, independente de os alvos serem apresentados visual ou auditivamente, no francês ambos os alvos transparentes e opacos facilitaram o processamento auditivo e visual de suas pseudoraízes. As autoras sugerem, então, que ocorram dois processos no reconhecimento inicial de palavras: um morfo-ortográfico e um morfossemântico.

Ao variar o fator transparência semântica, esses estudos estão lidando com a questão da interface sintaxe-semântica nos pontos de negociação intra e extra lexicais. Meunier e Longtin (2007), seguindo esse caminho, aplicaram um *priming* multimodal com decisão lexical em francês e perceberam facilitação apenas para pseudopalavras morfológicamente complexas que fossem interpretáveis, de forma que fatores semânticos seriam considerados quando o *prime*



fosse aberto. As autoras propõem, então, que exista tanto decomposição morfológica baseada em propriedades formais quanto integração semântica baseada na compatibilidade semântica entre morfemas.

Já Feldman, O'Connor e Moscoso del Prado Martín (2009) controlaram propriedades da raiz – como frequência, tamanho da vizinhança e sobreposição de letras entre *prime* e alvo – em um *priming* encoberto no inglês. Encontraram facilitações maiores para pares semanticamente transparentes quando comparados a opacos, de maneira que a classificação do parentesco entre *prime* e alvo foi um fator de predição da magnitude da facilitação. Os autores criticaram, ainda, o *null finding*, isto é, a interpretação de que a ausência de diferença estatisticamente relevante entre a facilitação de pares semanticamente transparentes e opacos signifique que o momento inicial do processamento de palavras é governado exclusivamente por fatores morfo-ortográficos.

Marelli *et al.* (2013) encontraram resultados similares, mas operando um *eyetracking* com *priming* encoberto em inglês. Detectaram tempos de fixação de olho menores para pares transparentes do que para opacos ou puramente ortográficos, sugerindo que o acesso ao nível morfo-ortográfico nem sempre seja necessário para a identificação lexical.

Beyersmann, Castles e Coltheart (2011) investigaram a questão tratada extensivamente nesse capítulo, mas usando não-palavras (sequências de letras das quais não é possível depreender sentido). Aplicaram, também, o método de troca de letras (*transposed letter*) no radical. O experimento possuía condições de não-palavras monomorfêmicas, com sufixos existentes e com finais não-morfêmicos (i.e., não-sufixais). Por terem encontrado ativação na condição de não-palavras sufixadas, concluíram que o *affix-stripping* ocorre nos estágios mais iniciais do reconhecimento visual de palavras e que ele opera mesmo quando elas são opacas e apresentam letras trocadas.

Lázaro, García e Illera (2021) realizaram dois experimentos com estímulos em espanhol, sendo um deles dividido em duas fases, completadas pelos mesmos participantes, mas com uma semana de espaço temporal entre elas. Os autores não encontraram diferenças significativas de *priming* entre condições morfológicas semanticamente transparentes e opacas, mas ambas geraram menos ativação que condições apenas ortográficas, o que significa que o processamento dessas últimas foi mais custoso.

Dois outros estudos que também propõem efeitos de *priming* equivalentes para pares semanticamente transparentes e opacos foram feitos em alemão, uma língua morfológicamente

rica. Baayen e Smolka (2020) utilizaram *priming* visual aberto com SOA bem alto, que convida à reflexão, de 500ms, e Smolka, Preller e Eulitz (2014) utilizaram tanto *priming* visual quanto auditivo, verificando se a modalidade influenciaria os resultados, o que não aconteceu. Os resultados do trio foram, porém, inconstantes no que diz respeito a facilitação semântica (comparada à morfológica) e a inibição de forma. Vale ressaltar, também, a importância dada pelos autores a estudos translinguísticos, uma vez que seus resultados se distanciaram dos encontrados em muitos estudos com línguas indo-europeias.

Por fim, Järvikivi e Pykkönen (2011), em um estudo de decisão lexical e *priming* encoberto e multimodal em finlandês, propuseram que ambos os estágios iniciais e finais sejam sensíveis a informações *bottom-up* e *top-down* (que nos estágios iniciais talvez fossem originárias de conexões semânticas supralexicais entre os membros da família morfológica de uma palavra e, nos finais, constituiriam interpretabilidade semântica). A informação morfológica seria, então, representada em dois níveis interativos – sua forma sublexical (ortografia) e sua forma supralexical (semântica) – mediados pelo nível lexical. Por fim, os autores sugerem que o estágio inicial de processamento morfológico não seja apenas sensível à possibilidade de a palavra ser segmentada em raízes e afixos existentes, mas também a se a sua forma total é uma palavra existente cujo sentido difere do sentido da raiz.

Resolvemos, então, testar justamente essa proposição de Jarvikivi e Pykkonen (2011), mas no PB e inovando na escolha de estímulos. Nossos materiais experimentais são as palavras como *refrigerante*, que são sincronicamente opacas, mas cujo sentido total é diacronicamente relacionado ao sentido de sua raiz. Acreditamos que o processamento inicial seja guiado por fatores morfo-ortográficos, enquanto fatores morfossemânticos seriam importantes em um segundo estágio do processamento, quando percebemos que a forma total da palavra *refrigerante*, por exemplo, tem sentido diferente do da raiz de sentido diferente do da raiz *refrig-*, em uma espécie de Garden Path lexical. Isso porque começaríamos a processar *refrigerante* pela raiz *refrig-* mas, conforme déssemos continuidade ao processamento, perceberíamos que não conseguiríamos processar como afixos os segmentos *-e-*, *-rar*, *-te*. Eles fariam, na realidade, parte de uma nova raiz, de maneira que precisaríamos voltar à Enciclopédia para recuperar seu sentido.

## 2.4 Stimulus Onset Asynchrony (SOA)

*Stimulus onset asynchrony* (doravante SOA) é a diferença de tempo entre a apresentação do onset do *prime* e a apresentação do onset do *target*. Ela é seguida pelo tempo de resposta, que geralmente é contado a partir do onset do *target*. Uma distinção muito importante é entre SOA e intervalo entre *trials* (do inglês *intertrial interval*, ISI). ISI considera apenas o tempo entre o fim da apresentação do *prime* e (assim como o SOA) o início da apresentação do *target* (figura 3).

Assim como a variação no tipo de apresentação do *prime* (encoberto ou aberto) parece influenciar os resultados experimentais, o mesmo acontece com a variação no tempo de SOA. Heyer e Kornishova (2018) realizaram um estudo de decisão lexical em russo e em inglês, utilizando um paradigma de *priming* encoberto que, de acordo com a literatura, é o que parece facilitar um processamento morfológico para alvos semanticamente opacos. As autoras só encontraram uma diferença grande de efeito de *priming* morfológico entre palavras semântica-

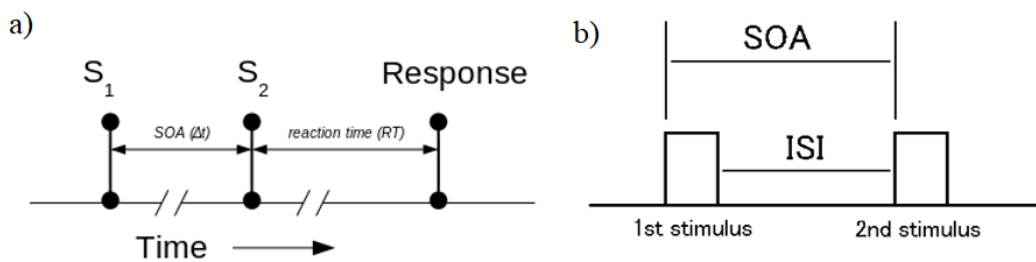


Figura 3. a) Diagrama exemplificando SOA. Fonte: <https://commons.wikimedia.org/wiki/File:SOA-diagram.svg>. Nenhuma modificação feita, acessado em: 31/07/2023. b) Diagrama exemplificando a diferença entre SOA e ISI. Fonte: <http://www.eegget-it.nl/erp.html#>. Nenhuma modificação feita, acessado em: 31/07/2023.

mente transparentes e opacas quando o SOA era alto, isto é, quando o participante tinha mais tempo para pensar ativamente sobre o que estava processando. Assim, elas ratificaram a proposta de modelos decomposicionais, sugerindo que a fase inicial da decomposição fosse cega à semântica e motivada apenas por fatores morfo-ortográficos.

Resultados similares foram encontrados por Feldman *et al.* (2004), que aplicaram um *priming* multimodal com alvos visuais no inglês. Eles ainda sugeriram que o nível de facilitação que ambos os alvos semanticamente transparentes e opacos recebem fosse escalado e potencializado pelo tamanho da família morfológica da raiz. Quanto maior a família, maior

seria a facilitação. Resultados como esse parecem ser diretamente correlacionados com a frequência de uso da raiz.

Feldman, Barac-Cikoja e Kostic (2002) realizaram uma manipulação experimental valiosíssima ao replicarem esse estudo em sérvio, mas utilizando materiais experimentais escritos em dois diferentes alfabetos: o latino e o cirílico. Encontraram o mesmo padrão: facilitação de processamento morfológico quando SOA é mais baixo, quando comparado ao mais alto, e influência do tamanho da família morfológica. Isso os fez questionar modelos que não preveem mudança de resultados em função do paradigma experimental.

Já Rastle *et al.* (2000) variaram o SOA (em 43, 72 e 230 ms) de estímulos em inglês, ambos na modalidade encoberta de *priming*. Chegaram à conclusão de que a decomposição morfológica acontece num estágio inicial do reconhecimento visual de palavras e de que ela é independente de semântica. Propuseram, mais especificamente, que a facilitação gerada por condições morfológicas em testes de *priming* não derivaria simplesmente de uma combinação de características ortográficas e semânticas compartilhadas por seus *primes* e alvos.

Por fim, Marslen-Wilson, Bozic e Randall (2008) variaram o SOA entre 36, 48 e 72ms em um estudo de *priming* encoberto no inglês. Em nenhum SOA encontraram efeito positivo no relacionamento ortográfico, nem decomposição morfológica quando os pares apresentavam apenas relacionamento semântico – apresentamos resultados similares em Cataldo e França (2021) para a condição de relacionamento semântico. Por outro lado, os autores observaram, nos três SOAs, decomposição morfológica para palavras complexas existentes e para pseudosufixadas, mesmo quando havia inversão da ordem de *prime* e alvo (e.g. palavra derivada-raiz vs. raiz-palavra derivada), uma manipulação que estamos realizando no presente estudo.

Portanto, o aumento do SOA parece ter uma relação inversamente proporcional com a facilidade de acesso a palavras semanticamente opacas: quanto maior o SOA, maior a dificuldade em acessar a morfologia dentro delas. Dessa maneira, escolhemos utilizar um SOA altíssimo no presente trabalho para verificar se, mesmo nas condições mais adversas, um processamento decomposicional ainda é verificado.

### 3. Neurofisiologia das palavras

---

A principal tarefa dos sistemas neuronais é construir na mente uma representação fidedigna do mundo captado pelos sentidos, para que esta representação possa ser manipulada pelos processos cognitivos. A primeira etapa desta tarefa é traduzir o estímulo físico na linguagem dos neurônios (FRANÇA, 2005:75)

A neurofisiologia da linguagem surgiu com a descoberta e o estudo de afasias, que são um tipo de disfunção, com vários subtipos, relacionado à linguagem. Em 1861, o médico francês Pierre Broca recebeu, no hospital Bicêtre, em Paris, um paciente chamado Monsieur Leborgne, que havia sido acometido por um acidente vascular cerebral (AVC) e, apesar de entender perfeitamente tudo que lhe era dito e ter sua prosódia preservada na produção de linguagem, não conseguia produzir sentenças, pois todas as suas palavras eram trocadas por “tan”. Quando eventualmente o paciente veio a óbito, Broca fez uma necropsia no cérebro dele e encontrou uma lesão na terceira circunvolução do giro frontal inferior esquerdo, endereço cerebral que ficou conhecido como Área de Broca e a doença como Afasia de Broca (FRANÇA, 2022).

Alguns anos depois, o neurologista polonês radicado na Alemanha Karl Wernicke analisou alguns pacientes que conseguiam articular sentenças, mas não entendiam o que lhes era dito. Eram disfunções diferentes das de Broca. Wernicke e seus discípulos realizaram, então, exames post-mortem nesses pacientes e puderam associar essa afasia ao lobo temporal superior esquerdo do cérebro, sítio que passa a ser denominado Área de Wernicke, e que começa a ser relacionado com a recepção da linguagem. Na sequência das investigações, Wernicke e seu assistente, Lichtheim, então, construíram um modelo, conhecido como A Casa de Lichtheim, que articula esses três pontos – produção de linguagem, percepção e conhecimento de mundo. Wernicke propôs ainda a existência de um feixe de nervos, fascículo arqueado, que conectaria Broca a Wernicke, embora este feixe só tivesse sido visualizado muitos anos depois. No local para o armazenamento de conhecimento do mundo estariam guardadas as representações conceituais (FRANÇA, 2011; CATALDO LOPES; FRANÇA, 2015) (figura 4).

A Casa de Lichtheim é um importante esforço de retomar, com alto rigor científico da época, o modularismo que havia sido completamente desacreditado com o advento da Frenologia, proposto por Franz Gall no final do século XVIII/início do século XIX. A Frenologia propunha que o cérebro seria dividido em pequenos módulos relacionados a características psíquicas. As irregularidades que temos na calota craniana, ora protuberâncias,

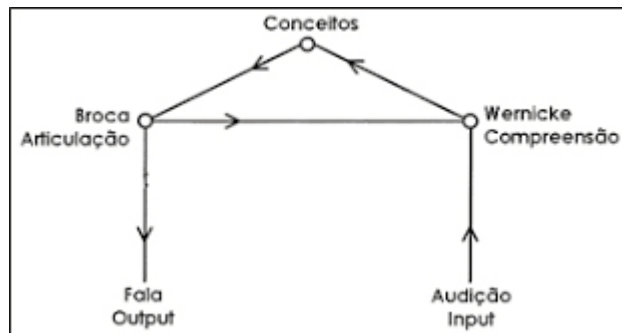


Figura 4. A Casa de Lichtheim de 1874. Fonte: França, 2011:II.

ora depressões, seriam uma medida do funcionamento de cada módulo psíquico em dada pessoa: uma protuberância no módulo da bondade significaria bastante disponibilidade para o bem, enquanto que uma depressão sugeriria falta ou ausência de bondade.

Apesar de, felizmente, ter sido desacreditada como ciência de qualidade, a Frenologia explorou uma ideia muito importante de que o cérebro dispõe de áreas específicas para processar diferentes tipos de capacidades (emoções, linguagem, pensamentos etc.). Hipóteses contrárias à modularidade são chamadas de interativistas e propõem que o cérebro possua uma unidade de domínios informacionais não hierarquizados, que funcionariam em rede (FRANÇA, 2005). Paul Broca, então, com sua pesquisa, pode reafirmar a realidade modular cerebral ao vincular a disfunção do M. Leborgne à já mencionada área de Broca.

Com o tempo, a tecnologia nos permitiu entender melhor esse funcionamento cerebral e descobrir particularidades desse e até muitas outras áreas cerebrais envolvidas nesse processo. Friederici *et al.* (2013) mostraram que as áreas de Broca e de Wernicke são ligadas principalmente por dois fascículos (coleções de fibras longas de associação): o uncinado (via ventral), que seria a via do léxico, não subordinada ao período crítico para a linguagem, uma vez que aprendemos palavras novas durante toda a vida; e o arqueado (via dorsal), já mencionado por Wernicke, que seria responsável pela sintaxe da linguagem humana, ou seja, seria o local onde a hierarquia estrutural estaria ancorada, subordinado ao desenvolvimento durante o período crítico.

Mais recentemente, a partir da primeira década dos anos 2000, diversas outras áreas cerebrais vêm sendo relacionadas a outras capacidades linguísticas e suas interrelações vêm sendo investigadas. A defesa da existência de redes neurais é a cada vez mais forte, mas não em

detrimento da existência de módulos cerebrais, especializados em uma ou algumas atividades. Algumas áreas às quais já foi atribuído algum tipo de especialidade para a linguagem são: o AG (giro angular), sensível a estrutura argumental, de aproximadamente 180ms a 250ms após o onset de um estímulo lexical; o LATL (lobo temporal anterior esquerdo), ativado durante combinações conceituais, mais ou menos aos 200ms a 280ms; o PTL (lobo temporal posterior), sensível a efeitos combinatoriais envolvendo, possivelmente, sintaxe, dos 200ms ao 250ms; o LIFG (giro frontal inferior esquerdo), que tem sido relacionado a dependências de longa distância; e o vmPFC (córtex pré-frontal ventromedial), que também parece estar relacionado a efeitos combinatoriais, mas dessa vez sem envolver sintaxe (Pylkkänen, 2019).

Como no presente estudo avaliamos o processamento visual de palavras soltas, nos dedicamos na próxima seção a detalhar um pouco mais as particularidades desse processo.

### 3.1 *Processamento visual de palavras*

A linguagem é uma habilidade naturalmente falada ou sinalizada. A escrita foi uma invenção humana, posterior ao aparecimento da habilidade para a linguagem na espécie. Isso é ratificado pelo fato de não termos um substrato neural nato para a escrita, como temos para a linguagem falada/sinalizada. Contudo, uma vez ultrapassado o *primeiro* momento de detecção e processamento da modalidade do estímulo recebido do ambiente, o processamento do core da linguagem é ancorado nas mesmas regiões cerebrais, conforme ressalta Brennan (2022:88-89):

Up to this point, the pathway for spoken words and written words has been very distinct. Keep in mind also that written words are presented to the visual system all at once, in contrast to spoken (and signed) words, which unfold over time. Despite these striking differences, their paths begin to merge within just 200 milliseconds in the posterior temporal lobe, especially the pMTG. After about 300 milliseconds, activation continues to spread to anterior temporal and inferior frontal regions; these temporal and frontal regions are involved in high-level aspects of word processing regardless of whether the input was written or spoken (BRENNAN, 2022:88-89).

O autor também ressalta uma característica importante do processamento visual: palavras são lidas de uma só vez, ao contrário de quando são enunciadas. Conforme visto, o sistema visual é o *primeiro* a ser engajado e, mais ou menos aos 200ms, o lobo temporal posterior começa a ser mais ativado, passando, a partir dos 300ms, para regiões temporais

anteriores e frontais inferiores, envolvidas no processamento de alto-nível da linguagem. Vale ressaltar que Brennan está mencionando intervalos de tempos advindos de pesquisas com MEG, enquanto o experimento do presente projeto será com EEG. Há um leve atraso, no EEG, quanto aos tempos encontrados no MEG. Isso não significa, no entanto, que esse atraso é cognitivo. Pesquisas ainda estão sendo feitas para determinar a natureza dessa diferença de tempo.

Quando lemos uma palavra, então, o *primeiro* sistema cognitivo ativado é a visão. Como resposta cognitiva normal à estimulação visual, há o um ERP denominado N170 (ou N1), por ter polaridade Negativa e latência de aproximadamente 170ms após a apresentação do estímulo visual. Tal ERP é relacionado ao processamento grafêmico. Soto *et al.* (2018), em um estudo de EEG com falantes de português brasileiro em idade escolar, demonstrou uma especificidade desse ERP para o processamento de palavras e pseudopalavras no hemisfério esquerdo, quando comparado a símbolos e figuras. Isso acontece porque, como o cérebro não é nativamente adaptado para a leitura, há uma reciclagem neuronal durante o processo de alfabetização. O cérebro, que possui áreas para identificação de objetos e faces em ambos os hemisférios, na região occipital, adapta tais conexões do hemisfério esquerdo para reconhecimento de letras. Dessa maneira, o hemisfério esquerdo fica mais especializado em processamento de leitura, enquanto o direito permanece especializado em processamento de faces e objetos (DEHAENE, 2009; ANDRADE, 2019). Essa área que se torna especializada em processamento de palavras recebeu o apelido de *visual word form area* (VWFA) e é localizada, mais especificamente, em uma região ventral na fronteira entre os lobos temporal e occipital (BRENNAN, 2022:88). O N170 é, então, um dos medidores do nível de alfabetização de crianças e adultos e demonstra uma das maiores mudanças fisiológicas propiciadas por um processo cultural.

Alguns autores apontam ainda um outro componente ligado ao reconhecimento visual de palavras. Trata-se do N250, um ERP negativo aproximadamente aos 250ms de latência. De acordo com Eddy *et al.* (2016:1776), esse componente tem uma distribuição posterior no escalpo, consistente com o sistema visual ventral e é “[a] o componente de latência média, que supostamente reflete o sistema neural que mapeia representações sublexicais ortográficas e fonológicas em representações de palavras inteiras”. Dessa maneira, a sintaxe encontrada dentro das palavras (i.e., a morfologia) estaria refletida nesse componente. No entanto, alguns autores mencionam que *prime* e *target* precisam ser muito próximos temporalmente para um efeito de N250 seja captado (MORRIS *et al.*, 2007). Isso porque talvez a amplitude registrada



para tal assinatura neural reflita o nível de incompatibilidade entre as representações ativadas pelo *prime* e as ativadas pelo *target* (Holcomb, Grainger, 2006).

Morris *et al.* (2007), que realizaram um estudo de EEG com palavras semanticamente transparentes e opacas, explicam como os diferentes modelos de processamento preveem a atuação do N250:

The supralexical model therefore predicts that semantic transparency should not influence the amplitude of the N250. However, given the importance of morphological structure for prelexical processing in the sublexical model, according to this account we ought to observe stronger *priming* effects from morphological *primes* independently of semantic transparency in the N250 ERP (MORIS *et al.*, 2007:6).

Dessa maneira, caso tivéssemos analisado as amplitudes do N250 no presente estudo, esperaríamos valores mais salientes (mais negativos) para *targets* da condição *refrigerar-refrigerante* caso essas palavras não fossem mais relacionadas atualmente (ou seja, caso possuíssem entradas lexicais diferentes). No entanto, valores estatisticamente irrelevantes entre os pares *refrigerar-refrigeração* e *refrigerar-refrigerante* levariam a uma análise de que essas palavras são morfologicamente relacionadas em nosso léxico mental, apesar de atualmente não serem mais semanticamente relacionadas.

Morris *et al.* (2007) analisaram, ainda o componente N400, o mais estudado dos componentes relacionados à linguagem. Sua descoberta surgiu quando as pesquisadoras Kutas e Hillyard (1980) aplicaram o paradigma de oddball à estímulos linguísticos. Oddball é um paradigma em que, seguido à apresentação de diversões estímulos iguais ou similares, apresenta-se um muito destoante. Auditoriamente, pode-se apresentar diversos [p] seguidos e apenas um [t] no meio deles. Esse fonema [t] seria, então o oddball, que geraria um ERP distinto. Ao aplicar esse paradigma com a linguagem, Kutas e Hillyard (1980) criaram sentenças como “he spread the warm bread with **socks**” (em vez de manteiga, por exemplo). Encontraram, para essas palavras, amplitudes mais negativas que para sentenças controle bem formadas, aproximadamente aos 400ms após a apresentação da palavra oddball (figura 7). Assim o N400 foi introduzido nas pesquisas psicolinguísticas e, desde então, muito foi descoberto sobre ele.

Em 2008, Lau, Phillips e Poeppel (2008) realizaram uma revisão da literatura do que havia sido descoberto até então sobre esse componente. Em seguida, Lau, Almeida, Hines e Poeppel (2009) propuseram uma associação desse componente com efeitos de contexto lexical. Já Lau, Holcomb e Kuperberg (2013) realizaram uma dissociação entre efeitos de predição e de associação, através do N400, em contextos de palavra única. Lau, Namyst, Fogel e Delgado

(2016), por sua vez, realizaram uma comparação entre predição e incongruência em combinações adjetivo+nome. Em 219, Delaney-Busch, Morgan, Lau e Kuperberg (2019) propuseram um modelo Bayesiano de predições semânticas no N400. Por fim, através do comportamento da assinatura N400, Lago, Namyst, Jäger e Lau (2019) estudaram os mecanismos de acesso a antecedentes durante o processamento de pronomes e Lau e Namyst (2019) investigaram os efeitos de predição que são independentes da presença de congruência.

É importante, também, ressaltar que, em 2011, Kutas e Friedermeier publicaram um trabalho sobre os achados dos 30 anos precedentes de pesquisa sobre esse componente. O artigo conta com um grande resumo visual dos esforços da comunidade científica durante esse tempo e demonstra que a repetição de palavras gera um N400 baixo (i.e., amplitude mais positiva) e a apresentação de palavras associadas prevê um N400 com amplitude relativamente menor, quando comparado à apresentação de palavras não associadas.

Morris *et al.* (2007), em um estudo com palavras semanticamente transparentes e opaca com EEG, encontraram amplitudes maiores de N400 para palavras semanticamente opacas, quando comparadas a palavras semanticamente transparentes, o que denota uma facilitação no processamento de palavras semanticamente transparentes. A condição semanticamente opaca apresentava relacionamento morfológico apenas aparente, como no par *corner-corn*, do inglês esquina e milho (cf. RASTLE, BORIS, NEW, 2004). Pares opacos foram, no entanto, mais facilitados que pares apenas ortograficamente relacionados. A tarefa experimental dada aos participantes foi a decisão lexical – com a resposta “sim” para a mão dominante e “não” para a mão não dominante, assim como fizemos em ambos os experimentos apresentados nessa dissertação.

Jared, Jouravlev e Joanisse (2017) também propuseram que a semântica influencia o processamento de palavras desde o seu início, defendendo um processamento paralelo distribucional (PDP), em uma abordagem conexionista. Analisando dados do inglês, os pesquisadores utilizaram decisão lexical e decisão semântica (a última com manipulação de limite de cor – *color boundary manipulation*) para, através tanto de *priming* encoberto quanto de decisões com palavras isoladas (*nonpriming*), colher dados comportamentais e de EEG. Observaram maior efeito de facilitação para palavras transparentes, seguidas de quase transparentes e, por fim, de opacas, tendo a transparência semântica, então, gerado um efeito gradual.

Por outro lado, (Lavric, Clapp e Rastle, 2007) encontraram, no EEG, resultados correspondentes aos comportamentais de Rastle, Boris e New (2004). Analisando as amplitudes do N400, reportaram efeitos de facilitação de *priming* tanto para condições de relacionamento morfológico quanto para de relacionamento apenas aparentemente morfológico (e sem relação semântica), sendo ambos maiores que o efeito de condições não morfológicas. Defenderam, assim, que os estágios mais iniciais do processamento de palavras visualmente apresentadas fossem operados unicamente pela segmentação da estrutura morfêmica.

Bozic *et al.* (2007) acharam resultados similares, mas com MEG. Perceberam que ambos os pares semanticamente transparentes e opacos tinham ativação facilitada (um N400 reduzido) nas regiões frontais esquerdas do cérebro, o que não aconteceu com pares com relação unicamente de forma, unicamente de sentido e unicamente de identidade (*prime* e alvo sendo a mesma palavra, como a condição M11 do presente trabalho). Sugeriram, então, que a segmentação de palavras complexas fosse realizada pela presença de uma estrutura morfológica de superfície e de maneira independente de semântica e forma.

Resultados de estudos de processamento lexical que não manipulam a transparência semântica também se mostram relevantes de consideração. Além do projeto já reportado de França *et al.* (2008) com português brasileiro, um outro estudo neurofisiológico cujos resultados indicam a existência de um processamento *bottom-up* é o conduzido por Domínguez, Vega e Barber (2004), com espanhol. Os pesquisadores sugeriram uma facilitação na condição morfológica, advinda de tendência à positividade do N400 para tais estímulos. Enquanto isso, o processamento de sinônimos apresentou latências maiores e um pico mais positivo no ERP P600 – geralmente associado a reanálises sintáticas (BRENNAN, 2022). A condição ortográfica apresentou efeito similar ao de pares não relacionados, i.e., nenhuma atenuação na amplitude do N400, e a condição homográfica (palavras com raiz superficialmente similar, mas sem relação morfológica ou semântica), gerou uma atenuação inicial similar à de pares morfológicos, mas que se transformou logo em um N400 tardio, pela impossibilidade de integração.

Estudos neurofisiológicos, mas que utilizam o MEG como técnica, também reportaram resultados similares. Nos resultados obtidos através dessa técnica, o acesso lexical (isto é, o pareamento de forma e sentido) geralmente é captado 350ms após a apresentação do estímulo de interesse, gerando a assinatura M350 – comparável ao N400 em estudos com EEG. Usando essa técnica, Stockall e Marantz (2006) conduziram dois experimentos de *priming* com para

testar alomorfes regulares e irregulares do inglês. Seus resultados permitiram uma distinção entre, de um lado, relacionamento morfológico e, de outro, similaridade semântica e fonológica. Os autores encontraram ativação da raiz antes do acesso lexical para ambos os tipos de alomorfes, medindo, no entanto, latências diferentes de *priming* na assinatura neurofisiológica M350 entre a condição morfológica (com ambos os alomorfes) e as semântica e fonológica. Advogaram, assim, por modelos *bottom-up*, dizendo serem seus resultados incompatíveis tanto com modelos conexionistas, que tratam relacionamento morfológico como similaridade, quanto com modelos mistos, para os quais apenas formas regulares passam por processo de composição.

Kastner, Pylkkänen e Marantz (2018) aplicaram, também com MEG, um *priming* encoberto em hebraico, uma língua de morfologia não concatenativa. Também apostaram em uma interpretação *bottom-up*, mesmo nos casos em que o conteúdo é abstrato, sem marcações ortográficas ou fonológicas aparentes. Chegaram a essa conclusão demonstrando que um forma verbal que é ortográfica e fonologicamente indistinguível de formas não verbais é primada por outros verbos que usam o mesmo modelo derivativo, mas não por nomes e adjetivos similares às formas não verbais. Assim, haveria sempre representação e acesso a conteúdo morfológico.

Por fim, na mesma linha de raciocínio, Gwilliams e Marantz (2018), usando MEG em inglês, defenderam um processamento composicional em que o conhecimento de regras morfossintáticas pode ser usado para formar representações morfológicas de palavras lidas, mesmo sem uma experiência prévia com todos os seus morfemas constituintes. Nas palavras de interesse para essa dissertação, como *refrigerante*, nós inclusive temos experiência com seus morfemas constituintes. No entanto, a semântica que eles veiculam atualmente é diferente da semântica veiculada pela palavra inteira. Dessa maneira, nosso objetivo é verificar se há diferenças entre o acesso de palavras como *refrigerar*, cujos morfemas veiculam a mesma semântica que a palavra inteira, e palavras como *refrigerante*, cujas representações sub-lexicais não parecem mais ser captadas semanticamente. Para tanto, analisaremos o comportamento da assinatura neurofisiológica N400 que, conforme discutido, reflete o momento do acesso lexical em estudos de EEG.

### 3.2 Hipótese da ativação contínua

Sabendo que há um componente neurofisiológico associado ao momento do acesso lexical, resta, no entanto, descobrir como acontece a ativação das palavras de modo a gerar tal componente. Nesse sentido, França (2005:60) pontua que “é consenso entre os modelos modularistas de acesso lexical que, para acessarmos a representação mental correspondente a uma palavra ouvida ou lida, acabamos por ativar milhares de representações mentais de palavras que estavam desativadas na mente”. Por exemplo, ao começar a ouvir a sílaba [da], representações de diversas palavras que começam com esses mesmos fonemas são ativadas na mente: dado, dádiva, damasco, data, etc. A semelhança fonológica é, portanto, um critério para escolhermos quais palavras serão ativadas no início do processamento de uma palavra (FRANÇA, 2005).

Como visto, entretanto, muitos são os competidores acessados que não correspondem à representação correta que iremos acessar. É preciso, portanto, saber como ativá-los e desativá-los (isto é, a maneira mais eficaz de realizar tal processo), durante os milissegundos do acesso a uma palavra. A Hipótese da Ativação Contínua ou Competição se opõe à Hipótese da Inibição na tentativa de explicação desse processo (figura 5). Ainda nas palavras de França (2005:61):

A Hipótese da Competição defende que palavras ativadas competem até que uma consegue um nível mais alto de ativação, se sobressai e é reconhecida. A Hipótese da Inibição propõe que quando uma palavra que estava sendo ativada deixa de se parecer com o input, ela é imediatamente desativada. Assim, ao final do processo de inibição de competidores, resta apenas a representação que é igual ao input (FRANÇA, 2005:61).

Estudos experimentais foram realizados com o intuito de verificar qual dessas hipóteses se aproximava mais do processo que acontece no cérebro humano (ISEL, BACRI, 1999; PYLKKÄNEN et al, 2002; PYLKKÄNEN *et al.*, 2003; STOCKALL, MARANTZ, 2006). Eles descobriram que há dois momentos de ativação de competidores por semelhança fonológica: um no início da palavra e uma no meio. Em relação ao início da palavra, acessar o alvo do par *teach-tea* (ensinar-chá) é mais custoso do que o alvo *reach* (alcançar) em *teach-reach*, com o qual o alvo não tem nenhum relacionamento fonológico. Isso porque *tea* foi desativado (inibido) depois que *teach* ganhou de seus competidores. Além disso, *reach* (e também *each*, do inglês cada), têm semelhança fonológica com a parte medial da palavra *teach*, comprovando que a

semelhança pelo meio é feita – inclusive como uma segurança do sistema de reconhecimento de palavras, visto que frequentemente precisamos lidar com condições não ideais de escuta de um interlocutor (PYLKKÄNEN, 2002).

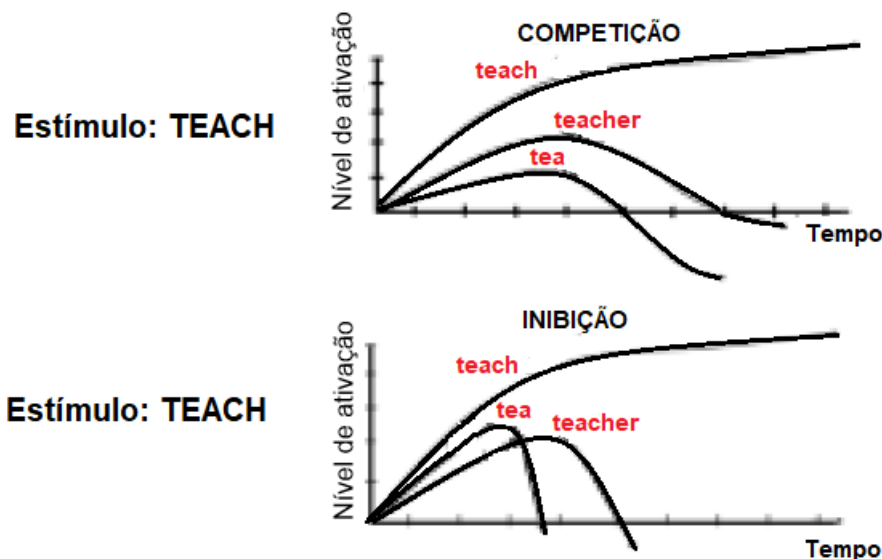


Figura 5. Hipótese de Competição (acima) e de Inibição (abaixo) para eliminação de representações concorrentes durante o acesso lexical de *teach*, do inglês ensinar. Fonte: Adaptado de França, 2005:62.

Além disso, acessar o alvo *teach* no par *teacher-teach* é mais rápido do que acessar o alvo *tea* no par *teacher-tea*. Assim como acontece com o alvo *tea*, *teach* deveria ser inibido uma vez que o processamento de *teacher* é completado. No entanto, acessar *teach* depois de *teacher* é tão eficaz quanto acessar *teach* após *teach* (*priming* de repetição, a identificação máxima entre *prime* e alvo: identidade). De acordo com a Morfologia Distribuída, esse resultado aparentemente paradoxal faz, na realidade, bastante sentido, *teacher* passaria pelas três listas de derivação de uma palavra. Na última delas, a Enciclopédia, a [raiz + primeiro morfema categorização] *teach-* (de *teacher*) é pareada a um significado. Essa raiz está, portanto, acessada quando o processamento do *prime* é completado. Ao ver/ouvir um alvo com a mesma raiz, não é preciso retornar à Enciclopédia para se realizar uma nova associação de significante e significado (forma e sentido), faz-se apenas a adição composicional de um afixo àquela [raiz + morfema categorizador] já acessados e ainda ativos na mente. À nível de derivação e representação, *teacher* e *teach* seriam, portanto, idênticas: durante a derivação de *teacher*, há

um momento em que ele é efetivamente *teach*. O momento do acesso lexical seria justamente o momento em que esse pareamento de forma e significado para a [raiz + morfema categorizador] é realizado, não importando quantas camadas morfológicas (layers, em inglês) são adicionados à palavra posteriormente através de merge. França *et al.* (2008) encontraram exatamente esse resultado em um estudo com português brasileiro, explicado nos capítulos anteriores. A Teoria da Ativação Contínua discrimina, portanto, semelhança fonológica de identidade morfológica.

Esse ponto é de suma importância para o presente trabalho. Abdicamos da comparação entre morfologia e fonologia, mas buscamos justamente descobrir se há uma relação de identidade morfológica entre *refrigerar* e *refrigerante* ou se, durante a ativação de *refrigerante*, a quebra da negociação semântica faria com que *refrigerante* depois de processos diacrônicos, ganhasse opacidade de raiz que tornaria, o relacionamento com *refrigerar* meramente fonológico. Em nada adiantaria ter-se acabado de ativar *refrigerar*, uma vez que elas não são mais reconhecidas como possuindo a mesma [raiz + morfema categorizador], isto é, como a mesma palavra. Neste caso, ao acessar-se *refrigerante* após *refrigerar*, seria necessário realizar um novo acesso lexical completo para uma outra raiz, passando-se novamente pela 3ª lista da Morfologia Distribuída, a Enciclopédia, para que uma nova negociação semântica seja nela feita: o pareamento de [refrigerant- + morfema categorização] com o sentido de local onde se come. Assim, acessar *refrigerante* após *refrigerar* seria, na realidade, custoso (i.e., tempos de resposta maiores e amplitudes mais negativas, quando comparado ao acesso de *refrigeração*, após *refrigerar*). É um caso semelhante ao que já se estuda muito no processamento de sentença, com efeitos de Garden path em que se aposta em um sentido que ainda precisa ser inibido em prol de outra negociação. Esperamos avaliar esses possíveis caminhos com nossos experimentos eletrofisiológicos. Contudo, antes de entrarmos na parte experimental, procuraremos sucintamente estabelecer as bases da fisiologia cerebral aferida pelo eletroencefalograma para que o leitor possa melhor acompanhar os resultados que obtivemos.

### 3.3 Fisiologia cerebral e medidas do EEG

Neurônios, as células do cérebro, são compostas da soma, dos dendritos e do axônio. A soma da atividade elétrica gerada pelos dendritos de um neurônio gera um potencial de ação, que passa pelos seus axônios até que seja transmitido a um outro neurônio, gerando uma sinapse. Isso

acontece através de uma distribuição diferencial de íons no lado interno e externo da membrana celular dos neurônios, caracterizada por um influxo de  $\text{Na}^+$  e um efluxo de  $\text{K}^+$ . Tal distribuição diferencial de íons gera um potencial de membrana, registrando uma atividade elétrica que vai sendo deslocada na célula. Os neurônios se comunicam entre si, então, através de sinais elétricos.

No entanto, para que esse potencial de ação seja criado, é preciso que a soma da atividade elétrica dos dendritos de um neurônio ultrapasse um dado limiar de eletricidade. Ou seja, se a soma das atividades dos dendritos de um neurônio for menor que esse limiar de eletricidade, um potencial de ação não será gerado e a onda terá apenas uma crista; se for maior, ele será gerado e haverá, assim, um pico de eletricidade. Poderíamos dizer que os potenciais de ação funcionam, então, no estilo tudo ou nada (figura 6).

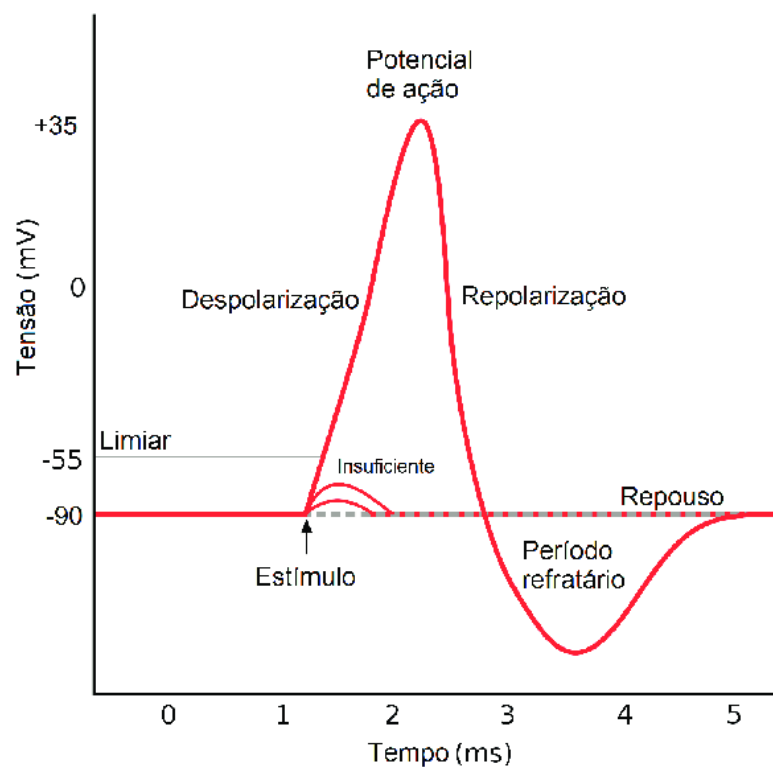


Figura 6. Esquematização do limiar para que um potencial de ação seja gerado e do período refratário sofrido por um neurônio após tal potencial de ação. Fonte: Freitas, 2017:15.

Uma vez que um pico de eletricidade é registrado, o neurônio passa por um período de refração, para que seu estado de coisas volte ao normal (ao repouso). Em uma representação gráfica do sinal elétrico cerebral, esse é o momento em que se registra uma queda do pico, i.e.,



um vale. Assim vai-se formando o padrão gráfico das ondas elétricas, com seus “*ups and downs*”. Cada neurônio, no entanto, tem uma frequência de disparo particular, de acordo com quanto longo é seu período de refração (figura 6).

Estamos fazendo alusão ao processo sofrido por um único neurônio apenas a título de explicação, porque o sinal captado em um EEG vem sempre de atividade elétrica de uma população de neurônios, nunca apenas de um. Isso dificulta a localização da fonte do sinal do EEG – o que é chamado de *problema inverso* –, mas não impacta na grande acurácia temporal, de milissegundos, que essa técnica possui. Além disso, o potencial captado pelo EEG tende a ser pós-sináptico (LUCK, 2014).

Uma outra questão de suma importância se refere a ondas diferenciais. Sinais elétricos são gerados por todas as funções fisiológicas rotineiras performadas pelo corpo humano, das mais concretas, como mover um membro do corpo, piscar os olhos e movimentar a boca para falar, até as que cujo resultado não é físico, como realizar um julgamento ou um raciocínio. Como, então, saber que a onda que estamos analisando reflete exatamente o processamento da linguagem? O procedimento realizado para lidar com essa questão é subtrair as ondas geradas logo após a apresentação de um estímulo das ondas geradas quando nenhum estímulo está sendo apresentado, que é o momento em que todas as outras atividades vitais do corpo humano continuam sendo performadas. Esses sinais são tratados como ruído para as ondas de interesse e são filtrados.

Como mesmo realizando essa rotina é possível ter influência de outras questões fisiológicas, uma prática no método de EEG para pesquisa é apresentar estímulos de uma mesma condição várias vezes, em vários *trials* diferentes, e fazer uma média do sinal obtido em todos os *trials* (figura 7). Assim, chegamos aos ERPs que costumamos ver na literatura e que serão apresentados nos resultados do experimento neurofisiológico realizado no âmbito do presente projeto.

Entender o funcionamento do cérebro e o que os amplificadores de EEG captam nos permite compreender por que analisar o processamento de linguagem por uma medida neurofisiológica é diferente de o analisar apenas por uma medida comportamental, como tempo de resposta. Enquanto a medida comportamental nos mostra apenas o *resultado motor* do processamento gerado por um estímulo linguístico, o sinal gerado pelo EEG é uma medida *contínua de atividade cerebral* em cada momento de tempo, nos permitindo medir os sinais do

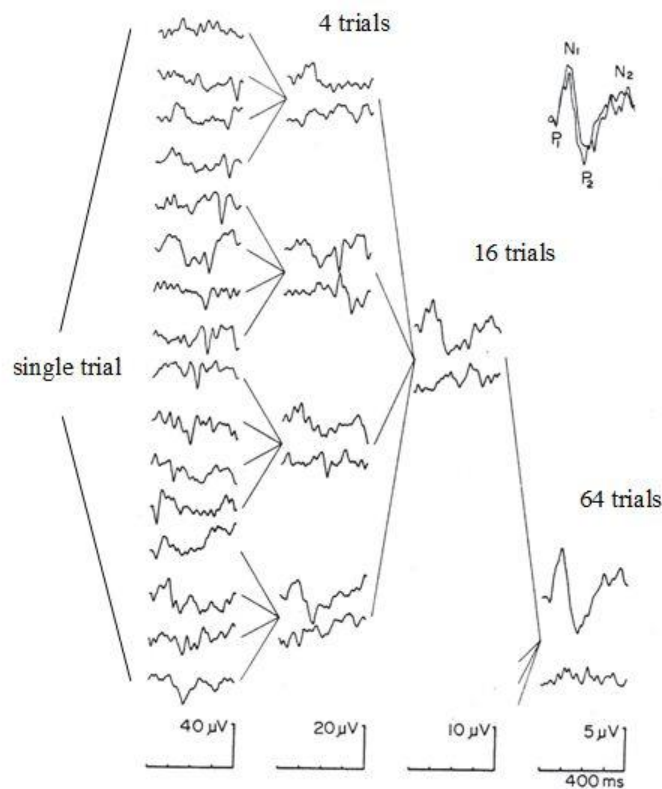


Figura 7. Geração de um ERP através de média de todos os potenciais de ação gerados por um dado estímulo, ao longo de vários *trials* (apresentações ao participante). Em cada par de ondas da imagem, a superior representa o ERP analisado e, a inferior, a onda diferencial com o ruído filtrado. Fonte: <http://www.eegget-it.nl/erp.html#>. Nenhuma modificação feita, acessado em: 31/07/2023.

processamento cerebral que ocorrem *entre* o estímulo e a resposta comportamental (Luck *et al.*, 2009). Além disso, ele não conta com a interferência do delay causado pela resposta motora, o que é uma influência incontornável nos resultados comportamentais. Utilizar medidas de EEG para analisar a linguagem são também uma prova de como ela é um processo com substrato biológico no corpo humano.

Dessa maneira, veremos nos próximos capítulos como essa medida pode nos ser útil no presente estudo, para que possamos medir biologicamente o processamento de palavras no cérebro humano. Primeiro realizamos um estudo comportamental (capítulo 4) para, dentre outros objetivos, validar o design experimental para o experimento de EEG (capítulo 5).

## 4. Experimento comportamental

---

### 4.1 *Justificativa, objetivo e variáveis*

Uma vez que diferentes teorias linguísticas e teorias de processamento propõem hipóteses diferentes sobre o processamento de palavras opacas, e que observamos questões a serem ajustadas em Cataldo e França (2021), acreditamos ser justificável um novo experimento comportamental. Além disso, quisemos validar o novo design experimental e os novos materiais antes de aplicá-los à avaliação neurofisiológica (EEG-ERP), cuja implementação é muito mais custosa.

O objetivo principal desse experimento comportamental foi verificar o processamento da relação entre alvos de pares como *restaurar-restaurante* e *refrigerar-refrigerante*, que são sincronicamente opacos, mas que mantêm as peças morfológicas das palavras que os originaram, respectivamente *restaurar* e *refrigerar*. O objetivo secundário foi verificar se o processamento segue um curso decomposicional.

Como variáveis dependentes, tivemos o tempo de resposta (response time, em inglês – doravante RT) e a acurácia (taxa de acerto e erro) da decisão lexical.

Como variáveis independentes, tivemos:

a) *O tipo de relacionamento entre prime e target:*

- Morfolologicamente intacto (MI):

*Restaurar-restaurar* (caso especial com *prime* idêntico a *target*),  
*restaurar-restauração*, *refrigerar-refrigeração*, *animal-animalesco*;

- Morfolologicamente relacionado, mas hoje semanticamente obscuro (MO):

*Restaurar-restaurante*, *refrigerar-refrigerante*, *animal-animado*;

- Sem nenhum tipo de relacionamento (morfológico, semântico, fonológico ou ortográfico – SR):

*Restaurar-praia*, *refrigerar-caneta*, *animal-sorvete*;

b) *A quantidade de camadas morfológicas da palavra derivada:*

- 1 camada morfológica (ou layer, em inglês) (MI1, MO1, SR1):

*Animal-animado* (MO1)

- 2 camadas morfológicas (ou layers, em inglês) (MI2, MO2, SR2):

*Animal-animadão* (MO2)

Tivemos um design 3x2 e as variáveis, com seus materiais, tiveram uma distribuição *within-subjects*: a distribuição dos materiais nas versões experimentais foi feita através de um quadrado latino. As versões experimentais foram atribuídas aos sujeitos conforme eles chegavam para fazer o teste. Nenhuma versão específica foi atribuída a qualquer participante específico.

Por fim, um controle importante em nosso experimento foi o SOA, conforme discutido previamente no capítulo 2. Escolhemos um SOA longo para ver se, mesmo nas condições mais desvantajosas, um processamento decomposicional seria atestado. Vale mencionar que um SOA longo já é uma consequência direta de um *prime* auditivo, em que o participante necessariamente precisa ouvir o áudio todo, do início ao fim, e não simplesmente bater o olho em uma palavra e lê-la, utilizando-se de seu processamento visual livre em tempo. No entanto, além disso, configuramos 600ms de tela preta entre o fim do *prime* auditivo e o *onset* do *target* escrito.

## 4.2 *Materiais e métodos*

O experimento teve a forma de um *priming* multimodal, sendo o *prime* auditivo e o alvo visual, numa tentativa de tentar diminuir ou mesmo eliminar um possível viés ortográfico no tempo de reconhecimento do alvo. Diversos estudos apontam que a ortografia não interfere significativamente num paradigma de *priming* (DOMÍNGUEZ; de la VEJA; BARBER, 2004; LÁZARO; GARCÍA; ILLERA, 2021; MORRIS *et al.*, 2007; RASTLE *et al.*, 2000), mas, como era possível realizar esse controle, achamos melhor assim proceder.

A tarefa experimental foi a decisão lexical, isto é, julgar se uma palavra existe ou não no português. A tarefa foi explicada aos participantes da seguinte forma: “você deverá focar na palavra que você lê, para julgar se ela existe ou se é uma palavra inventada. A *primeira* palavra, que você ouvirá, sempre existirá no português”. Esse protocolo experimental não só incita o acesso lexical, como, ao apresentar o participante a várias pseudopalavras e várias palavras que existem, faz com que os participantes foquem a sua atenção nesta distinção, dificultando que percebam as diferenças de relacionamento existentes dentre as palavras de interesse. Sem essa tarefa, provavelmente haveria o viés metacognitivo impactando negativamente os resultados.

Tivemos 288 estímulos de interesse, distribuídos na modalidade *within-factors* por 6 versões experimentais, através de um quadrado latino, de maneira que um mesmo participante

nunca visse mais de uma vez o mesmo estímulo. Cada versão experimental teve 48 estímulos de interesse e 48 pseudopalavras distratoras, das quais 24 apresentavam similaridade fonológica entre *prime* e *target* e 24 não apresentavam. Achemos importante realizar esse controle porque, dentre as condições de interesse, MI1, MI2, MO1 e MO2 apresentavam similaridade fonológica entre *prime* e *target*, enquanto SR1 e SR2 não apresentavam. Portanto, fazer a mesma manipulação com as pseudopalavras nos pareceu uma maneira de deixar os estímulos de interesse mais equilibrados com os estímulos distratores, dificultando inclusive uma reflexão do participante sobre o objetivo do estudo.

Assim, cada participante julgou 48 pares de estímulos de interesse e 48 pares de *targets* pseudopalavras. Dessa maneira, cada um deles foi exposto a 8 pares de palavras de cada condição de interesse e, no total, a 96 pares. Além disso, precedendo o experimento, havia uma seção de prática com 10 pares de palavras, nos quais 5 alvos eram pseudopalavras. A ordem dos pares de palavras foi pseudorandomizada para cada versão experimental, de maneira que não houvesse mais de 2 ocorrências seguidas de uma mesma condição, exceto as pseudopalavras, que puderam ter até 5 ocorrências seguidas.

Em todo o experimento o *prime* foi mantido o mesmo, enquanto o alvo era a palavra derivada, que variava de acordo com a condição experimental. O objetivo de manter o *prime* constante em todas as condições de uma mesma célula experimental foi atribuir qualquer diferença nos tempos de resposta e amplitudes à mudança do *target*. Caso isso não fosse feito, teríamos ainda o fator das características individuais de cada palavra nova (como frequência, tamanho da família morfológica e experiência prévia de cada participante com ela) influenciando no resultado. Assim, tendo ‘*animal*’ como *prime*, analisamos o custo (em tempo e em amplitudes) de se chegar a 6 tipos de *targets* diferentes: *animal* (MI1, condição de completa identidade), *animal*+esco (MI2), anim+ado (MO1), anim+ad+ão (MO2), sorvete (S1), jurament+a+do (SR2). As pseudopalavras tinham *primes* diferentes.

Células experimentais				
Célula	Condição		Prime Auditivo	Alvo Visual
1	MI1 ( <i>baseline</i> )	Relacionamento Morfológico Intacto/ Identidade com 1 camada morfológica	animal	animal
	MI2	Relacionamento Morfológico Intacto/ Composicional com 2 camadas morfológicas	animal	animalesco
	MO1	Relacionamento Morfológico Opaco com 1 camada morfológica	animal	animado
	MO2	Relacionamento Morfológico Opaco com 2 camadas morfológicas	animal	animadão
	SR1	Sem (nenhum tipo de) Relacionamento com 1 camada morfológica	animal	sorvete
	SR2	Sem (nenhum tipo de) Relacionamento com 2 camadas morfológicas	animal	juramentado
	DIS/PSEUDO ( <i>baseline</i> )	Distrator Pseudopalavra	bigode	bigueto
2	MI1 ( <i>baseline</i> )	Relacionamento Morfológico Intacto/ Identidade com 1 camada morfológica	oriente	oriente
	MI2	Relacionamento Morfológico Intacto/ Composicional com 2 camadas morfológicas	oriente	oriental
	MO1	Relacionamento Morfológico Opaco com 1 camada morfológica	oriente	orientar
	MO2	Relacionamento Morfológico Opaco com 2 camadas morfológicas	oriente	orientação
	SR1	Sem (nenhum tipo de) Relacionamento com 1 camada morfológica	oriente	lazer
	SR2	Sem (nenhum tipo de) Relacionamento com 2 camadas morfológicas	oriente	ajustado
	DIS/PSEUDO ( <i>baseline</i> )	Distrator Pseudopalavra	contínuo	girodaga
3	MI1 ( <i>baseline</i> )	Relacionamento Morfológico Intacto/ Identidade com 1 camada morfológica	resíduo	resíduo
	MI2	Relacionamento Morfológico Intacto/ Composicional com 2 camadas morfológicas	resíduo	residual
	MO1	Relacionamento Morfológico Opaco com 1 camada morfológica	resíduo	residir
	MO2	Relacionamento Morfológico Opaco com 2 camadas morfológicas	resíduo	residente
	SR1	Sem (nenhum tipo de) Relacionamento com 1 camada morfológica	resíduo	paletó
	SR2	Sem (nenhum tipo de) Relacionamento com 2 camadas morfológicas	resíduo	amabilidade
	DIS/PSEUDO ( <i>baseline</i> )	Distrator Pseudopalavra	aquário	ocrela

Tabela 2. Três células experimentais, com exemplos dos materiais experimentais. *Primes* se mantiveram constantes em todas as condições e eram palavras não-derivadas, enquanto *targets* eram os estímulos de interesse, palavras derivadas, que mudavam de acordo com a condição experimental.

Uma tabela com todos os estímulos experimentais está disponível no anexo (capítulo 8). Uma planilha com todas as versões experimentais, já randomizadas, está disponível no repositório OSF (Open Science Framework) do projeto. Sempre que possível, anotamos a frequência dos estímulos usando os valores zipf encontrados no corpus Léxico do Português Brasileiro<sup>8</sup> (ESTIVALET; MEUNIER, 2015; Van HEUVEN *et al.*, 2014). Esse é o maior corpus de Português Brasileiro disponível, mas nem todos os nossos estímulos puderam ser encontrados nele.

É possível notar que as condições MI1 e PSEUDO foram usadas como baselines opostos. MI1 (*animal-animal*) apresenta o maior nível possível de relação entre *prime* e *target*: a identidade completa. Já PSEUDO tinha como objetivo controlar a atenção do participante durante a tarefa e, conseqüentemente, validar quaisquer dados coletados das outras condições.

Vale ressaltar também, que, entre MI1 e SR2 (MI1 → MI2 → MO1 → MO2 → SR1 → SR2), há uma progressão bastante sutil entre as seis condições experimentais, começando pela plena identidade (*animal-animal*), progredindo para situações em que o conteúdo das palavras do par é menos relacionado (*animal-animalesco*) e finalizando com situações em que nem o conteúdo nem a forma das palavras é relacionado (*animal-sorvete*).

---

<sup>8</sup> Disponível em: <<http://www.lexicodoportugues.com/>>. Acessado em 13/07/2023.

### 4.3 *Desenho experimental e procedimentos*

O experimento foi composto de uma fase treino, com 10 pares de palavra, e de 3 blocos experimentais, cada um com 32 pares experimentais. Entre o treino e o início do experimento havia uma pausa, assim como entre cada bloco. Portanto, havia 3 pausas no total. O participante poderia descansar pelo tempo que quisesse e, quando estivesse pronto, apertava um botão para continuar. Cada *trial* era composto de, na seguinte ordem (Figura 8):

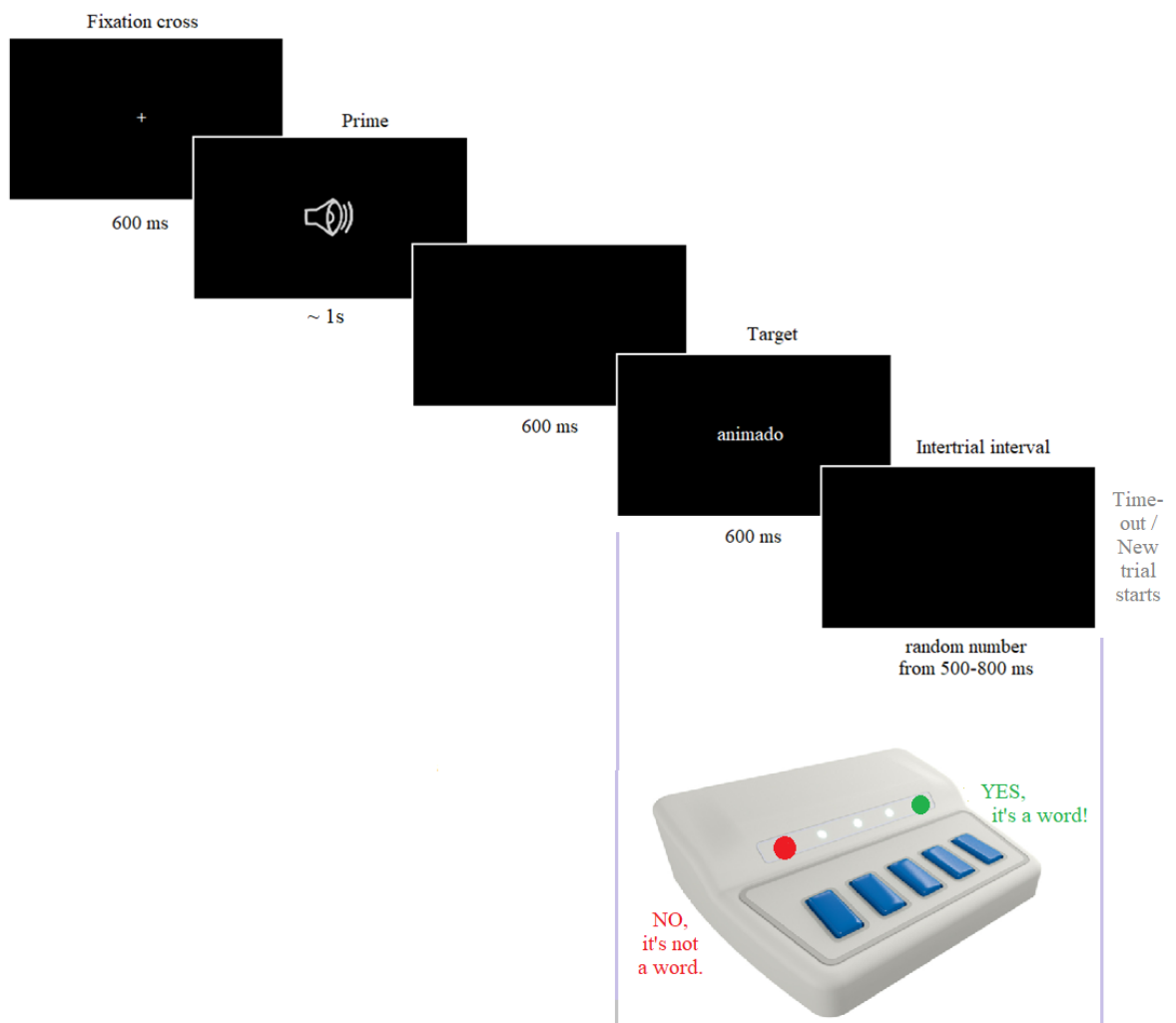


Figura 8. Protocolo de apresentação do experimento na tela do computador, em linha do tempo. Fonte: autora.



O experimento aconteceu no Laboratório de Acesso Sintático (Acesin), localizado na sala D103 do prédio da Faculdade de Letras da UFRJ. A sala apresenta iluminação natural e artificial simultaneamente. A lista de eventos experimentais tinha a seguinte forma:

- a) Após ser recebido, o participante lia, concordava e assinava o TCLE (Termo de Consentimento Livre e Esclarecido), aprovado pelo CEP responsável pela análise desse projeto.
- b) Em seguida, ele recebia instruções gerais e se sentava em uma cadeira confortável, a 90 cm de distância de um monitor de 19 polegadas, onde seriam apresentados os estímulos do experimento através da plataforma experimental *E-Prime 3.0* (Psychology Software Tools, USA). Em frente ao monitor, o participante poderia ver a caixa de botões do Chronos da Brain Products (Figura 14), através da qual o participante registrava suas respostas ao longo do experimento. Esse dispositivo foi configurado como periférico do *E-Prime 3.0* e foi conectado ao computador em que os estímulos seriam apresentados.
- c) Uma vez bem sentado e instruído, o participante era solicitado pelo pesquisador a ajustar o headphone profissional Audio Technica ATH-M40x, com cancelamento de ruído.
- d) Depois de completamente acomodado, o participante respondia a indagação do pesquisador sobre se o experimento poderia ser iniciado.
- e) Caso a resposta fosse positiva, o pesquisador iniciava o experimento e, no centro da tela do monitor, aparecia uma cruz de fixação durante 600ms.
- f) Em seguida, o participante ouvia o *prime*, que durava aproximadamente 1s, dependendo da locução palavra. Durante esse tempo de escuta, a tela permanecia preta.
- g) Depois da locução da palavra, a tela continuava preta antes de o *target* (algo) escrito ser apresentado. O tempo de duração dessa segunda tela preta era de 600ms, constituindo um SOA bem longo, de aproximadamente 1600ms. A fonte utilizada foi Times New Roman, no tamanho 32 e centralizada na tela.
- h) Em seguida vinha o *target* escrito, que permanecia na tela durante 600ms. A partir do onset do *target*, o participante já poderia responder, apertando um botão vermelho de Chronos para o caso de a palavra *target* não existir no português e um botão verde para o caso de ela existir. O botão verde estava na extrema direita do Chronos e, o vermelho, na extrema esquerda, para que a mão dominante – mais rápida – respondesse quando a resposta fosse positiva, i.e., quando se tratava de nossos estímulos de interesse.

- i) Havia, então, uma tela preta, que permanecia na tela durante uma quantidade aleatória de tempo entre 500ms e 800ms. Essa tela era um intervalo entre *trials*. Seu valor aleatório, *jitter*, teve como objetivo não permitir que o participante se acostumassem com o tempo inter-*trials*, o que poderia gerar uma habituação e, por conseguinte, uma resposta mecânica em algum momento, mesmo que a instrução fosse a de que ele deveria dar uma resposta mais reflexa, respondendo o mais rápido possível.
- j) O Chronos armazenava toda a cronologia do experimento de forma que os tempos de resposta puderam ser submetidos a tratamento estatístico dentro de cada condição para serem devidamente comparados.
- k) Durante esse passo, a resposta dada ainda era computada. Dessa maneira, os participantes tinham, no total, entre 1100ms e 1900ms após o *onset* da palavra *target* para responder.

Se o participante respondesse em menos de 600ms, o experimento pulava a tela do *target* (item h) para a tela do intervalo entre *trials* (item i). Se o participante respondesse apenas na tela do intervalo entre *trials*, um novo *trial* começava, assim que sua resposta era detectada. Caso o participante não respondesse até o final da tela do intervalo entre *trials*, o experimento seguia em frente e o Chronos computava a resposta não dada como um erro.

#### 4.4 Hipóteses e previsões

Tivemos como hipóteses e previsões:

- a) A existência de entradas lexicais diferentes para palavras sincronicamente transparentes e opacas, como *refrigerar-refrigeração* e *refrigerar-refrigerante*, respectivamente. Para tanto, prevemos tempos de resposta diferentes para as condições MI1 e MO1, conforme encontramos em Cataldo e França (2021).
- b) Um curso decomposicional de processamento para todas as palavras, tanto antes quanto após o acesso lexical (respectivamente, decomposição para retirada das partes constituintes da palavra e posterior composicionalidade do sentido de cada morfema adicionado à [raiz+morfema categorizador]), conforme previsto pela MD (HALLE; MARANTZ, 1997). Para tanto, prevemos tempos de resposta maiores para as condições

com 2 camadas morfológicas (MI2, MO2), em comparação com as de 1 camada morfológica (MI1, MO1), como manifestação da pequena adição computacional sofrida. Por exemplo, o par *refrigerar-refrigerantezinho* teria um custo adicional em relação a *refrigerar-refrigerante*, por conta da adição de uma camada morfológica, conforme previsto pela hipótese do Affix Stripping (TAFT; FOSTER, 1975) e encontrado nos resultados de Cataldo e França (2021).

- c) Ausência de facilitação (*priming*) para as condições em que *prime* e alvo não têm nenhum tipo de relacionamento entre si, como *refrigerar-caderno*. Para tanto, prevemos tempos de resposta e amplitudes maiores para os estímulos dessas condições, quando comparados com estímulos com a mesma quantidade de camadas morfológicas, mas com algum tipo de relacionamento entre *prime* e alvo (morfológico intacto ou morfológico semanticamente opaco): SR1 vs. MI1 e MO1; e SR2 vs. MI2 e MO2.

## 4.5 Participantes

Participaram deste experimento 47 pessoas, das quais 1 foi retirada por não atingir a acurácia mínima de 70% na tarefa de decisão lexical e 2 por terem mais de 10% de suas observações retiradas após limpeza de outliers (foram considerados outliers tempos de resposta que fossem <200ms ou >1500ms).

Assim, entraram para a análise os dados de 45 participantes (M=33, H=12), com média de idade 23,85 anos e mediana 24 anos. Todos possuíam visão normal ou corrigida, não tinham problemas auditivos, eram destros e eram estudantes do *primeiro* e *segundo* ano do curso de Letras na Faculdade de Letras da UFRJ.

## 4.6 Resultados

As análises estatísticas foram feitas no software R (R Core Team, 2013), usando os pacotes *ggplot2*, *ggsignif*, *dplyr*, *tidyr*, *plotrix*, *tibble*, *stringr*, *writex*, *lme4*, *lmerTest*, *fpp* e *emmeans*. Para todos os resultados comportamentais, ajustamos modelos mistos generalizados lineares (*glmer*), utilizando a família gaussiana invertida (LO; ANDREWS, 2016). Em seguida, realizamos comparações aninhadas entre modelos mais simples e mais complexos, e dos

modelos escolhidos com seus respectivos modelos nulos, para determinar a significância de cada um. Finalmente, realizamos comparações post-hoc entre pares, para observar diferenças estatísticas entre as médias de tempo de resposta – doravante RT, do inglês response time (GODOY; NUNES, 2020).

Considerando todos os sujeitos e todas as condições, a acurácia foi de 88.91% (SE=1.62). Na tabela 4 encontra-se acurácias detalhadas por condição. Conforme esperado, as pseudopalavras distratoras foram as menos acertadas, com 86.19% de acurácia (SE=1.75), seguidas da condição MO2 (87.80%, SE=2.30), SR2 (89.29%, SE=2.26), SR1 (90.63%, SE=1.78), MO1 (92.21%, SE=1.74), MI2 (93.20%, SE=1.79) e MI1 (93.31%, SE=1.12).

**Acurácia por condição experimental, em %**

condition	media	sdev	n	se	CImin	CImax
<b>DIS</b>	86.1864	11.62485	44	1.752513	82.75148	89.62133
<b>MI1</b>	96.30682	7.421853	44	1.118886	94.1138	98.49984
<b>MI2</b>	93.20211	11.90319	44	1.794474	89.68494	96.71928
<b>MO1</b>	92.20779	11.54125	44	1.739909	88.79757	95.61801
<b>MO2</b>	87.80438	15.23763	44	2.297159	83.30195	92.30681
<b>SR1</b>	90.625	11.78939	44	1.777318	87.14146	94.10854
<b>SR2</b>	89.28571	14.99486	44	2.26056	84.85502	93.71641

Tabela 3. Acurácia da decisão lexical por condição experimental, no teste comportamental.

Ajustamos um modelo de RT em função da condição experimental e com *intercepts* aleatórios para participante e *target* ( $RT \sim \text{condition} + (1|\text{Subject}) + (1|\text{target})$ ) que convergiu e foi significativamente mais verossímil que o modelo nulo ( $\chi^2(6)=53.032$ ,  $p=1.156e-09$  \*\*\*). RT por condição experimental podem ser conferidos na tabela 5 e no gráfico 1. Distratores pseudopalavras apresentaram o maior tempo de resposta: 649.20ms (SE=3.65) e foram significativamente diferentes de todas as outras condições: SR2 ( $\beta= -42.56$ , SE=9.53,  $t=4.468$ ,  $p=0.0002$ ), MO2 ( $\beta= -55.46$ , SE=9.99,  $t=5.551$ ,  $p<.0001$ ), MI2 ( $\beta= -87.63$ , SE=9.31,  $t=9.410$ ,  $p<.0001$ ), MO1 ( $\beta= -70.27$ , SE=9.17,  $t=7.663$ ,  $p<.0001$ ), SR1 ( $\beta= -62.42$ , SE=9.19,  $t=6.791$ ,  $p<.0001$ ), MI1 ( $\beta= -92.84$ , SE=8.85,  $t=10.492$ ,  $p<.0001$ ).

Em pares post-hoc ajustados por Tukey, MI1 não foi significativamente diferente de MO1 ( $\beta= -22.58$ , SE=8.54,  $t= -2.643$ ,  $p=0.1132$ ), mas MI2 foi estatisticamente diferente de MO2 ( $\beta= -32.17$ , SE=9.56,  $t= -3.365$ ,  $p=0.0135$ ) e MI1 foi diferente de MO2 ( $\beta=-37.38$ , SE=9.38,  $t=-3.984$ ,  $p=0.0013$ ). Contudo, em uma análise post-hoc não ajustada, a diferença

entre MI1 e MO1 foi significativa ( $\beta = -22.58$ ,  $SE = 8.54$ ,  $t = -2.643$ ,  $p = 0.0082$ ) e a entre MI2 e o MO2 foi ainda mais significativa que a ajustada por Tukey ( $\beta = -32.17$ ,  $SE = 9.56$ ,  $t = -3.365$ ,  $p = 0.0008$ ).

**RT por condição experimental, em ms**

condition	media	sdev	n	se	CImin	CImax	lim_min
<b>DIS</b>	649.1992	167.8219	2113	3.65089	642.0435	656.355	145.7337
<b>MI1</b>	515.5306	149.1831	360	7.86264	500.1198	530.9413	67.98125
<b>MI2</b>	560.5726	151.6498	358	8.014942	544.8633	576.2819	105.6232
<b>MO1</b>	559.1369	156.3885	358	8.265391	542.9367	575.337	89.97127
<b>MO2</b>	592.5767	162.6274	352	8.668072	575.5873	609.5661	104.6944
<b>SR1</b>	550.6389	128.1882	360	6.756113	537.3969	563.8809	166.0742
<b>SR2</b>	595.5112	146.8996	358	7.763885	580.294	610.7284	154.8124

Tabela 4. RT da decisão lexical por condição experimental, no teste comportamental.

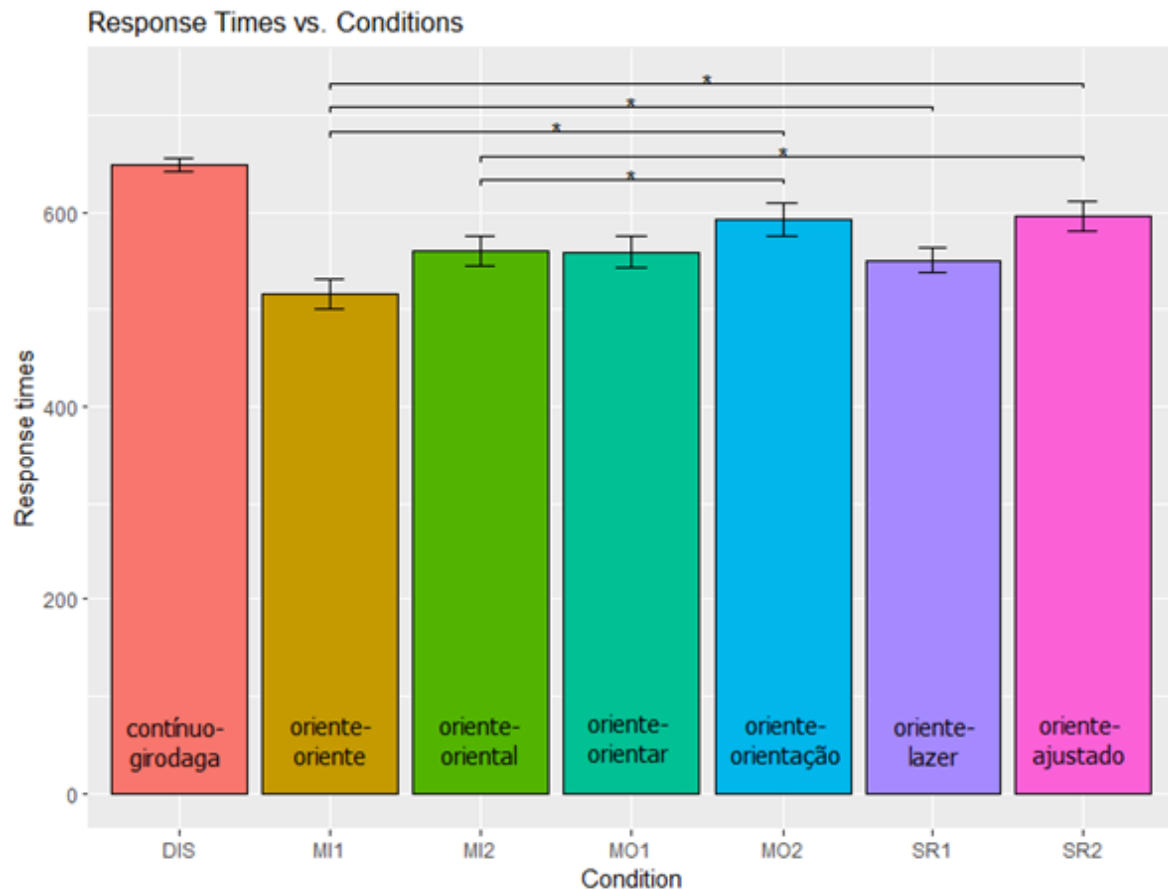


Gráfico 1. RT da decisão lexical por condição experimental, no teste comportamental.

Ademais, em pares post-hoc ajustados por Tukey, nenhuma comparação entre estímulos com diferentes camadas morfológicas mas de mesmo tipo de relacionamento foi significativa: MI1 x MI2:  $\beta = -5.21$ ,  $SE = 8.95$ ,  $t = -0.582$ ,  $p = 0.9973$ ; MO1 x MO2:  $\beta = -14.81$ ,  $SE = 10.84$ ,  $t = -1.366$ ,  $p = 0.8197$ ; SR1 x SR2:  $\beta = -19.86$ ,  $SE = 10.66$ ,  $t = -1.863$ ,  $p = 0.5052$ . Nenhuma mudança substancial foi observada nesses resultados quando uma análise pós-hoc não ajustada foi realizada: MI1 x MI2:  $\beta = -5.21$ ,  $SE = 8.95$ ,  $t = -0.582$ ,  $p = 0.5606$ ; MI2 x MO2:  $\beta = -32.17$ ,  $SE = 9.56$ ,  $t = -3.365$ ,  $p = 0.0008$ ; SR1 x SR2:  $\beta = -19.86$ ,  $SE = 10.66$ ,  $t = -1.863$ ,  $p = 0.0625$ .

Além disso, em pares post-hoc ajustados por Tukey, SR1 foi diferente de MI1 ( $\beta = 30.42$ ,  $SE = 8.44$ ,  $t = 3.606$ ,  $p = 0.0058$ ), mas não de MO1 ( $\beta = 7.84$ ,  $SE = 8.78$ ,  $t = 0.893$ ,  $p = 0.9737$ ). Algo similar aconteceu com SR2, que foi diferente de MI2 ( $\beta = 45.07$ ,  $SE = 9.54$ ,  $t = 4.723$ ,  $p < 0.0001$ ), mas não de MO2 ( $\beta = 12.90$ ,  $SE = 10.00$ ,  $t = 1.289$ ,  $p = 0.8572$ ). SR2 foi, ainda diferente de MI1 ( $\beta = 45.07$ ,  $SE = 9.54$ ,  $t = 4.723$ ,  $p < 0.0001$ ).

Outro modelo ajustado tomou RT como variável resposta, a interação entre tipo de relacionamento (MI, MO, SR) e quantidade de camadas morfológicas (1, 2) como variável preditora e *intercepts* aleatórios para participante e *target*. Ele, no entanto, não foi mais verossimilhante que um modelo sem a interação entre os fatos preditores ( $RT \sim rel\_type * layers + (1|Subject) + (1|target)$ ) vs.  $RT \sim rel\_type + layers + (1|Subject) + (1|target)$ :  $\chi^2(2) = 0.242$ ,  $p = 0.886$ ). Tendo sempre como objetivo a parcimônia, escolhemos o último modelo, sem interação, que tem RT como variável resposta, tipo de relacionamento e camadas morfológicas como variáveis preditoras, e *intercepts* aleatórios para participante e *target* ( $RT \sim rel\_type + layers + (1|Subject) + (1|target)$ ). Uma comparação com modelos aninhados indicou que o fator tipo de relacionamento contribuía significativamente para o modelo ( $\chi^2(2) = 47.883$ ,  $p = 4.003e-11$  \*\*\*), assim como aconteceu com o fator camadas morfológicas ( $\chi^2(1) = 10.922$ ,  $p = 0.0009502$  \*\*\*) e que a verossimilhança desse modelo era melhor que a de um modelo nulo ( $\chi^2(3) = 65.868$ ,  $p = 3.271e-14$  \*\*\*).

No gráfico 2 e na tabela 6, estão disponíveis os RT de acordo com o tipo de relacionamento (MI, MO, SR). Comparações post-hoc por pares ajustadas por Tukey determinaram que MI é significativamente diferente de MO ( $\beta = -31.4$ ,  $SE = 6.36$ ,  $t = -4.939$ ,  $p < 0.0001$ ) e a SR ( $\beta = -44.0$ ,  $SE = 6.33$ ,  $t = -6.962$ ,  $p < 0.0001$ ). No entanto, a diferença entre MO e SR não é significativa ( $\beta = -12.7$ ,  $SE = 6.75$ ,  $t = -1.875$ ,  $p = 0.1461$ ).

### RT por condição experimental, em ms

rel_type	media	sdev	n	se	CImin	CImax	lim_min
<b>MI</b>	537.9889	151.9932	718	5.672337	526.8711	549.1066	82.00917
<b>MO</b>	575.7155	160.2752	710	6.015021	563.9261	587.5049	94.8899
<b>SR</b>	573.0125	139.5573	718	5.208233	562.8044	583.2207	154.3405

Tabela 5. RT da decisão lexical por tipo de relacionamento, no teste comportamental.

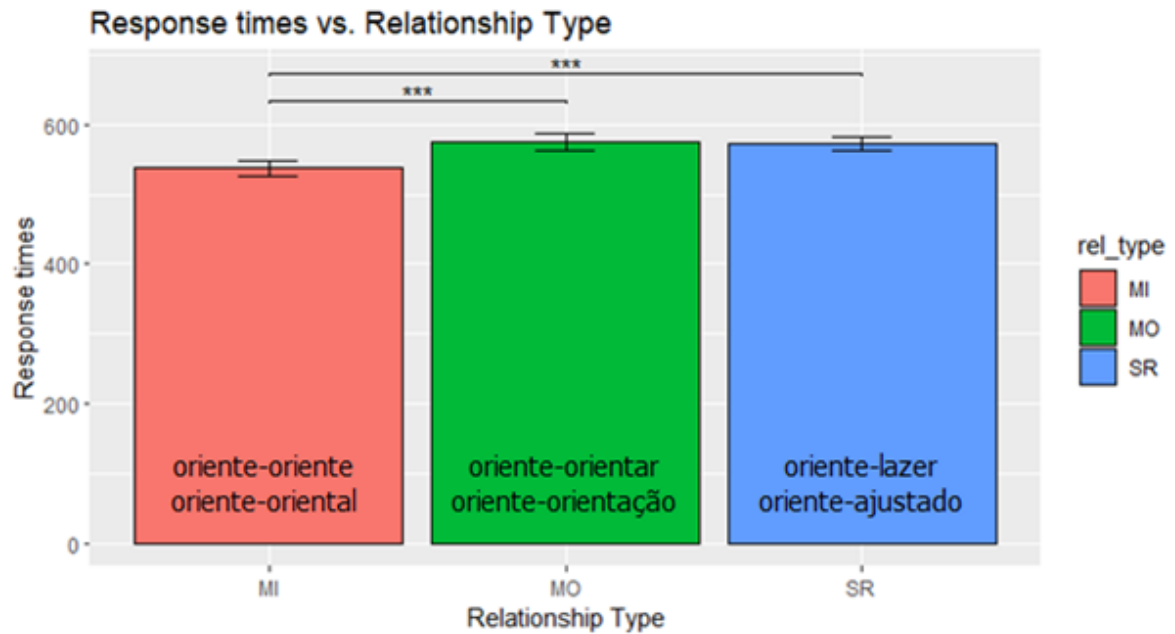


Gráfico 2. RT da decisão lexical por tipo de relacionamento, no teste comportamental.

### RT por quantidade de camadas morfológicas, em ms

layers	media	sdev	n	se	CImin	CImax	lim_min
1	541.7365	146.1495	1078	4.45131	533.012	550.4611	103.2881
2	582.8324	154.4878	1068	4.727251	573.567	592.0978	119.3689

Tabela 6. RT da decisão lexical por quantidade de camadas morfológicas, no teste comportamental.

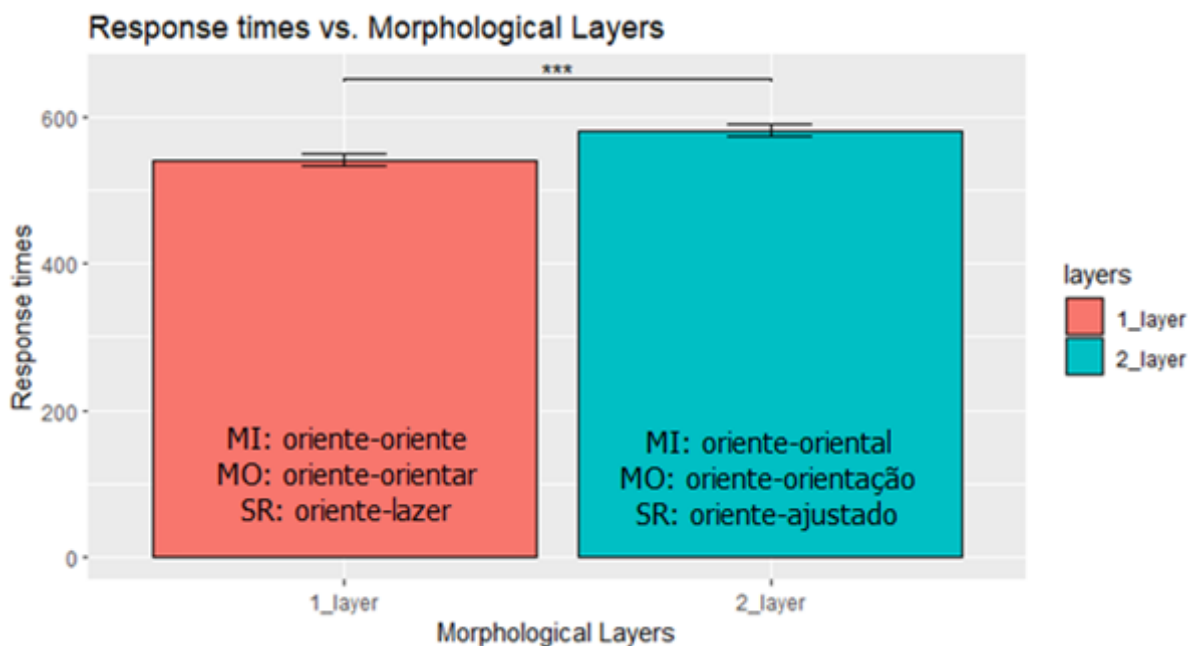


Gráfico 3. RT da decisão lexical por quantidade de camadas morfológicas, no teste comportamental.

Em relação às camadas morfológicas, tomando como base o mesmo modelo, comparações post-hoc por pares ajustadas por Tukey determinaram que estímulos com 1 camada geraram resultados significativamente diferentes que estímulos com 2 camadas ( $\beta = -30.3$ ,  $SE = 8.07$ ,  $t = -3.753$ ,  $p = 0.0002$ ), conforme disponível no gráfico 3 e na tabela 7.



## 4.7 Discussão

Nossa primeira hipótese foi a da existência de entradas lexicais diferentes para palavras sincronicamente transparentes e opacas, como, respectivamente, *restaurar-restauração* e *refrigerar-refrigerante*. Para tanto, previmos tempos de resposta estatisticamente diferentes para as condições MI1 (*animal-animal*) e MO1 (*animal-animado*), conforme encontramos em Cataldo e França (2021). A comparação entre essas condições não foi significativa ( $p=0.1132$ ), mas MI2 foi estatisticamente diferente de MO2 ( $p=0.0135$ ) e MI1 também o foi de MO2 ( $p=0.0013$ ). Além disso, no modelo que teve tipo de relacionamento (MI vs. MO vs. SR) MI foi significativamente diferente de MO ( $p<.0001$ ).

Junto à significância estatística dessa diferença, a observação de que os tempos de acesso a MO são sempre maiores que os a MI nos leva a concluir que **(i) acessar *refrigerar* antes de *refrigerante* inibe a ativação de *refrigerante***. Ou seja, a herança trazida pelo *prime refrigerar* na realidade dificulta o acesso ao *target refrigerante*. Se ambos possuísem identidade morfológica entre si (i.e., possuísem a mesma raiz), essa dificuldade não seria esperada e eles apresentariam RT estatisticamente similares (Pylkkänen *et al.*, 2002; 2003; França, 2005; Stockall, Marantz, 2006). Isso é o que acontece com os alvos de MI: acessar *animal* (MI1) e *animalesco* (MI2) depois de *animal* é simples, uma vez que esses alvos recebem uma herança positiva de ativação do *prime*, i.e., a mesma raiz havia acabado de ser ativada durante o acesso lexical do *prime* e, portanto, ela é recuperada rapidamente quando uma segunda palavra que também a carrega é apresentada (ambas têm identidade morfológica entre si). Com os alvos de MO, no entanto, a rota é diferente: se é mais custoso acessar *animado* (MO1) e *animadão* (MO2) como alvos de *animal*, quando comparados a *animal* (MI1) e *animalesco* (MI2), isso significa que a negociação semântica realizada na Enciclopédia durante o processamento de *animal* não pode ser aproveitada para o processamento de *animado* e *animadão*. Um novo acesso à Enciclopédia precisa ser feito, para gerar um novo pareamento de forma e significado em cima de uma nova raiz *animad-* e, portanto, é necessário que seja realizado o acesso a uma nova entrada lexical. **Dessa maneira, *animado* e *animadão* não apresentam mais identidade morfológica com *animal*. Trata-se de entradas lexicais diferentes em nosso léxico mental.**

Nossa segunda hipótese previa um curso decomposicional de processamento para todas as palavras, tanto antes quanto após o acesso lexical (respectivamente, decomposição para retirada das partes constituintes da palavra e posterior composicionalidade do sentido de cada

morfema adicionado à [raiz+morfema categorizador]), conforme previsto pela MD (HALLE; MARANTZ, 1997). Porém, seguindo os achados de Cataldo e França (2021) e conforme previsto pela hipótese do *Affix Stripping* (TAFT; FOSTER, 1975), esperávamos encontrar tempos de resposta maiores para as condições com 2 camadas morfológicas (MI2, MO2), em comparação com as de 1 camada morfológica (MI1, MO1), como manifestação da pequena adição computacional sofrida. Por exemplo, o par *animal-animadão* teria um custo adicional em relação a *animal-animado*, por conta da adição de uma camada morfológica.

Os tempos de resposta das condições com 2 camadas morfológicas foram maiores que os das condições com 1 camada morfológica em todos os tipos de relacionamento (MI, MO, SR). Todavia, não encontramos significância estatística entre as comparações MI1 vs. MI2 ( $p=0.9973$ ), MO1 vs. MO2 ( $p=0.8197$ ) e SR1 vs. SR2 ( $p=0.5052$ ). Esse resultado foi diferente do encontrado pelo modelo que continha camadas morfológicas como variável explicativa: a diferença entre [estímulos de MI, MO e SR com 1 camada morfológica] e [estímulos de MI, MO e SR com 2 camadas morfológicas] foi significativa ( $p=0.0002$ ). A diferença entre esses modelos pode se dar por conta de o segundo considerar mais estímulos em seus cálculos, apresentando, portanto, maior possibilidade de gerar diferença estatística. Assim, um estudo com mais materiais experimentais seria necessário para uma comparação mais verossímil entre MI1 x MI2, e MO1 x MO2.

Ainda, esses resultados se distanciam um pouco dos encontrados pelo experimento comportamental de Cataldo e França (2021), em que as autoras reportaram a possibilidade de a frequência dos itens ter influenciado especialmente os RT da condição MI2. Como a frequência foi melhor controlada dessa vez e possuíamos o dobro de estímulos experimentais, acreditamos que os presentes resultados reflitam melhor a realidade.

Estes resultados parecem mostrar, então, que **(ii) o processamento morfológico é decomposicional e sofre efeito de tamanho de palavra, de maneira que palavras com mais sílabas (mais material fonético) levam mais tempo para serem processadas que palavras com menos sílabas (menos material fonético)**. Esse mesmo resultado foi reportado diversas vezes na literatura de experimentos comportamentais. Os resultados de Taft e Foster (1975; 1976), por exemplo, os levaram a propor a já mencionada hipótese do *affix stripping*, segundo a qual o processamento é decomposicional: a palavra é decomposta, sua raiz é acessada e, em seguida, cada afixo é adicionado (sintática e semanticamente). Dessa maneira, é natural que testes comportamentais como os desses autores e o nosso reportassem um efeito de tamanho de

palavra, de maneira a prever uma diferença entre o processamento de palavras com mais ou menos camadas morfológicas/material fonético.

Por fim, nossa terceira hipótese versava sobre as condições sem relacionamento SR: condições cujos *primes* não forneciam nenhum tipo de facilitação para o processamento do alvo. Ao ler o *target*, o participante recomeçava do zero, pois nenhum tipo de herança era recebida do *prime*. Nossa hipótese não previa facilitação para os alvos dessas condições sem relacionamento (SR), com a consequente previsão de RT maiores para elas, quando comparados com alvos que possuíam a mesma quantidade de camadas morfológicas, mas outros tipos de relacionamento com seus *primes*: SR1 vs. MI1 e MO1; SR2 vs. MI2 e MO2.

SR1 foi estatisticamente diferente de MI1 ( $p=0.0058$ ), mas não de MO1 ( $p=0.9737$ ). O mesmo aconteceu com SR2, ou seja, ele foi estatisticamente diferente de MI2 ( $p<.0001$ ), mas não de MO2 ( $p=0.8572$ ). Esses resultados estão de acordo com os obtidos no modelo com tipo de relacionamento como variável explicativa: SR foi significativamente diferente de MI ( $p<.0001$ ), mas não de MO ( $p=0.1461$ ). Os tempos de resposta de MO (575.72 ms) foram, ainda, ligeiramente maiores que os de SR (573.01 ms). Isso provavelmente se deu pelos tempos de SR1 (550.64 ms), que foram os segundos mais rápidos, atrás apenas de MI1 (515.5306ms). SR2 (595.51 ms), no entanto, só não foi mais custoso que os distratores (649.20 ms). Em ordem crescente de tempo de resposta, tivemos: MI1 < SR1 < MO1 < MI2 < MO2 < SR2 < DIS.

Considerando, no entanto, a acurácia das respostas, a ordem crescente de condições experimentais foi: DIS < MO2 < SR2 < SR1 < MO1 < MI2 < MI1. Apesar de não transparecerem exatamente a nossa previsão, esses números demonstram, no geral, que **(iii) o processamento de SR foi tão custoso quanto o de MO**, que precisou realizar uma nova negociação semântica na Enciclopédia, **e mais custoso que o de MI, cuja derivação foi a mais facilitada possível**, pelo efeito mais forte de identidade do *priming* morfológico, já reportado diversas vezes na literatura (STOCKALL; MARANTZ, 2006; FRANÇA et al., 2008; GARCIA, 2009). Verificamos, portanto, uma escala de facilitação semântica e morfológica entre esses três tipos de condições.

## 5. Experimento neurofisiológico (EEG)

---

### 5.1 *Justificativa, objetivo e variáveis*

Os objetivos principal e secundário desse experimento são os mesmos do comportamental apresentado no capítulo anterior. No entanto, diferentemente dos testes comportamentais cuja reação se dá ao *final* da apreensão visual do *target* (um efeito *wrap-up*), a medida neurofisiológica proporcionada pelo EEG poderá nos trazer informações relevantes sobre computações internas à palavra, como por exemplo o momento da arbitrariedade saussuriana, ou seja, da negociação entre a raiz e o *primeiro* morfema categorizador, que se dá *durante* o processamento da palavra, dentro dela (HALLE; MARANTZ, 1993; HARLEY; NOYER, 1999). Esta resposta bioelétrica não dependerá de nenhuma resposta ou expressão comportamental do participante, ao apertar um botão dizendo se a palavra que lê existe ou não. Ela será uma consequência direta e *primeira* do impacto dos estímulos experimentais no cérebro do participante, decidindo a base semântica do composto sendo processado.

Como variáveis dependentes tivemos o tempo de resposta comportamental (RT), a acurácia comportamental e as amplitudes das ondas elétricas.

As variáveis independentes foram as mesmas do teste comportamental apresentado no capítulo anterior: (a) o tipo de relacionamento entre *prime* e *target*: MI, MO, SR; (b) a quantidade de camadas morfológicas da palavra derivada: 1 ou 2.

### 5.2 *Materiais e métodos*

Os materiais utilizados foram idênticos aos do teste comportamental, apresentados no item 4.2 e disponíveis no anexo (capítulo 8). Os métodos também foram iguais, sem exceção.

### 5.3 *Desenho experimental e procedimentos*

O desenho experimental e os procedimentos foram similares aos do teste comportamental, descritos na seção 4.3, com duas exceções: (i) o fone utilizado: para não gerar nenhum tipo de interferência no sinal e na conectividade dos eletrodos, os participantes não vestiram um headphone e sim um fone intra-auricular JBL Quantum 50, com cancelamento de ruído; (ii) a touca utilizada com 32 eletrodos através dos quais os sinais eletrofisiológicos foram coletados.

O procedimento técnico relacionado à coleta dos dados neurofisiológicos (ilustrado na figura 9) será melhor detalhado na seção 5.3.1.



Figura 9. No canto superior esquerdo, a pesquisadora está inserindo gel condutor de eletricidade entre o eletrodo e o escalpo do participante. No canto superior direito, vê-se o participante com a touca e o conjunto de eletrodos conectados, de um lado, no escalpo do participante e, de outro, no amplificador. Na tela do computador, vê-se ondas elétricas captadas do escapo do participante e amplificadas. No canto inferior, vê-se uma participante lendo a tela de instruções do experimento, pronta para inicia-lo. Suas mãos e dedos já estão posicionados nos botões do dispositivo Chronos que usará para registrar suas respostas.

### 5.3.1 Coleta dos dados neurofisiológicos

Os participantes tiveram seu crânio medido e sua testa e mastoides limpos com um lenço umedecido contendo álcool, numa tentativa de eliminar a oleosidade natural presente nessas áreas, que atrapalha uma interface de contato entre eletrodo e pele. Em seguida, uma touca *actiCap insert* ajustada ao tamanho do crânio foi encaixada, de maneira que o centro da cabeça estivesse alinhado com o eletrodo 24, correspondente à área Cz (Centro).

A atividade neurofisiológica foi gravada através de 32 eletrodos ativos, todos importados: Fp1, Fp2, Fz, F3, F7, F4, F8, FC1, FC5, FT9, FC2, FC6, FT10, Cz, C3, T7, C4, T8, CP1, CP5, TP9, CP2, CP6, TP10, Pz, P3, P7, P4, P8, Oz, O1, O2. Os eletrodos eram compostos de Ag/AgCl (prata e prata com cloro) e foram dispostos no escalpo dos participantes através do sistema internacional 10-20 (Figura 10).

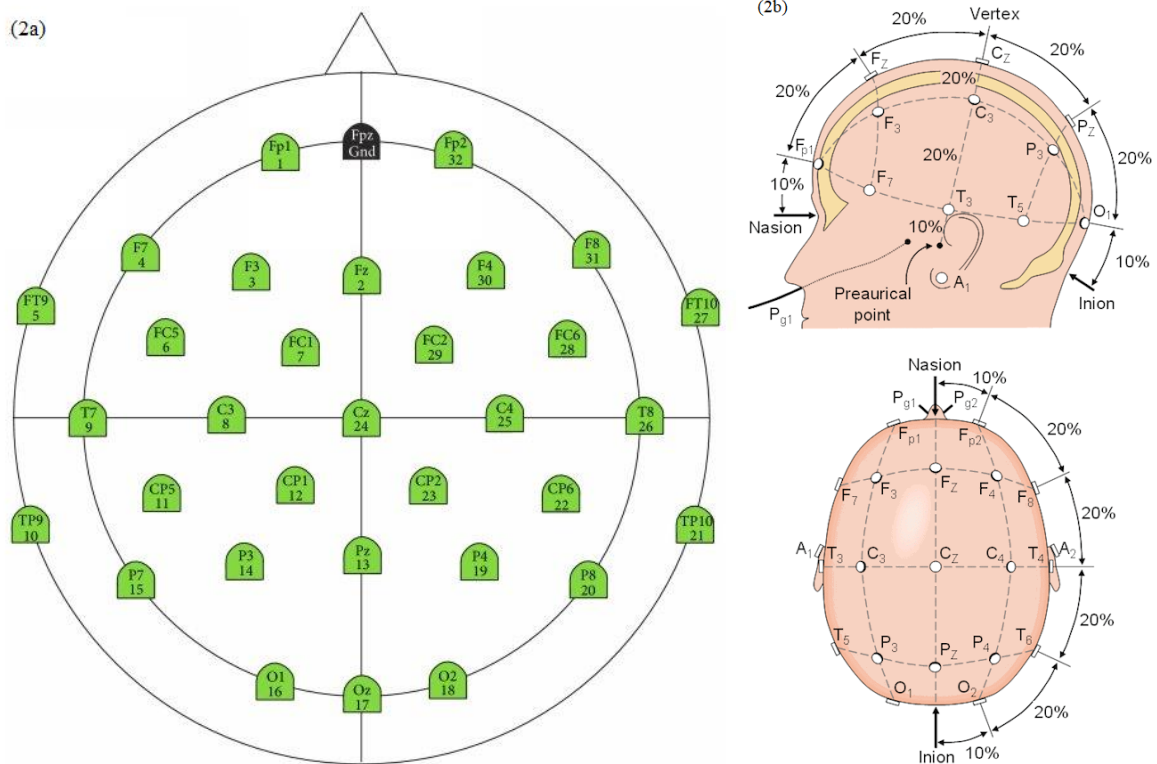


Figura 10. (2a) Sistema internacional 10-20 de posicionamento de eletrodos com 32 canais na touca *actiCap* (BrainProducts). Fonte: Bastos, Adamatti e Billa, 2016:3. (2b) Lógica e métrica utilizada para divisão dos eletrodos do sistema 10-20 pelo escalpo. Fonte: <<https://info.tmsi.com/blog/the-10-20-system-for-eeeg>>, visitado em 13/07/2023.

Um gel condutor contendo uma solução salina foi aplicado no sítio de cada um dos eletrodos, para permitir o fluxo de dados. Todos os eletrodos estavam conectados a um amplificador actiCHamp (de Brain Products, GmbH). Durante a gravação, os mastoides (FT10 e FT9) foram definidos, online, como a referência e a taxa de amostragem foi de 500 Hz. A impedância dos eletrodos foi mantida abaixo de 50 k $\Omega$ . O programa PyCorder foi usado para registrar os sinais amplificados pelo EEG, depois de terem sido filtrados online, pelo mesmo programa, com um filtro notch de 60 Hz (correspondente à frequência elétrica do Rio de Janeiro), um filtro passa-baixo (highpass) de 100Hz e um filtro passa-alto (lowpass) de 0,01Hz. Triggers relativos a cada uma das condições experimentais foram codificados no software *E-Prime 3.0*, o mesmo que apresentava os estímulos experimentais. A acurácia e tempo de resposta comportamental foram gravados por um dispositivo Chronos (de BrainProducts), conectado ao computador que passava os experimentos codificados no *E-Prime* e ao amplificador do sinal neurofisiológico.

#### 5.4 Hipóteses e previsões

Tivemos como hipóteses e previsões:

- a) A existência de entradas lexicais diferentes para palavras sincronicamente transparentes e opacas, como *refrigerar-refrigeração* e *refrigerar-refrigerante*, respectivamente (CATALDO; FRANÇA, 2021). Para tanto, prevemos tempos de resposta e **amplitudes** diferentes para as condições MI1 e MO1, sendo MO1 maior que MI1, por ser mais custoso (a herança do *prime* falha e uma nova entrada na Enciclopédia se faz necessária);
- b) Um curso decomposicional de processamento para todas as palavras, tanto antes quanto após o acesso lexical (respectivamente, decomposição para retirada das partes constituintes da palavra e posterior composicionalidade do sentido de cada morfema adicionado à [raiz+morfema categorizador]), conforme previsto pela MD (HALLE; MARANTZ, 1997). Resolvemos ser conversadores e manter as mesmas previsões do experimento comportamental realizado anteriormente com esse mesmo design. Assim, previmos tempos de resposta e **amplitudes** maiores para as condições com 2 camadas morfológicas (MI2, MO2), em comparação com as de 1 camada morfológica (MI1,

MO1), como manifestação da pequena adição computacional sofrida. Por exemplo, o par *refrigerar-refrigerantezinho* teria um custo adicional em relação a *refrigerar-refrigerante*, por conta da adição de uma camada morfológica, conforme previsto pela hipótese do Affix Stripping (TAFT; FOSTER, 1975) e encontrado nos resultados de Cataldo e França (2021).

- c) Ausência de facilitação (*priming*) para as condições em que *prime* e alvo não têm nenhum tipo de relacionamento entre si, como *refrigerar-caderno*. Para tanto, prevemos tempos de resposta e amplitudes maiores para os estímulos dessas condições, quando comparados com estímulos com a mesma quantidade de camadas morfológicas, mas com algum tipo de relacionamento entre *prime* e alvo (morfológico intacto ou morfológico semanticamente opaco): SR1 vs. MI1 e MO1; e SR2 vs. MI2 e MO2.

## 5.5 *Participantes*

Participaram desse experimento 36 voluntários que tinham entre 18 e 28 anos. Todos possuíam visão normal ou corrigida, não tinham problemas auditivos, eram destros e eram estudantes do primeiro, segundo e terceiro anos do curso de Letras na Faculdade de Letras da UFRJ.

Foram retirados da análise estatística dos resultados comportamentais 6 participantes: um afirmou que não conseguiu ler as palavras, fator essencial para que a decisão lexical pudesse ser feita; 3 não atingiram um mínimo de 70% de acurácia na decisão lexical (DIEPENDAELE; SANDRA; GRAINGER, 2005; FELDMAN *et al.*, 2012; WILDER; DAVIES; EMBICK, 2019) e 2 tiveram mais de 10% de perda de observações após a retirada de outliers, sendo que um deles teve 33% de sua perda nas condições de interesse. Foram considerados outliers tempos de resposta menores que 200ms e maiores que 1500ms. Dessa maneira, os números e análises apresentados na seção de resultados comportamentais foram provenientes de dados coletados de 30 participantes (M=18, H=12), com média de idade 23.5 anos e mediana 20 anos.

Para a análise neurofisiológica, à da rotina de limpeza de dados feita para a análise comportamental foi adicionada uma rotina relativa aos sinais elétricos captados. Dessa maneira, mais três participantes precisaram ser retirados da análise por problemas técnicos durante a coleta de dados e dois por ter apresentado amplitudes de onda consideradas visualmente como



outliers. Dessa maneira, utilizamos dados de 25 voluntários nas análises (M=15, H=10), com média de idade 23,5 anos e mediana 21 anos.

## 5.6 *Resultados*

Nesse capítulo, serão apresentados: os resultados comportamentais do teste (tempo de resposta de botão e acurácia da resposta – seção 5.6.1), o pré-processamento dos dados neurofisiológicos (seção 5.6.2) e os resultados neurofisiológicos (amplitudes de ondas elétricas, seção 5.6.3). Foram feitas análises separadas para os resultados comportamentais e para os resultados neurofisiológicos, de maneira que os dados retirados da amostra neurofisiológica por problemas técnicos durante a coleta não foram retirados da análise comportamental.

### 5.6.1 Resultados comportamentais do EEG (RT de aperto de botão e ACC)

As análises estatísticas foram feitas no software R (R Core Team, 2013), usando os pacotes ggplot2, ggsignif, dplyr, tidyr, plotrix, tibble, stringr, writex, lme4, lmerTest, fpp e emmeans. Para todos os dados comportamentais mostrados nesta seção, ajustamos modelos mistos generalizados lineares (glmer), utilizando a família gaussiana invertida (LO; ANDREWS, 2016). Em seguida, realizamos comparações aninhadas entre modelos mais simples e mais complexos, e dos modelos escolhidos com seus respectivos modelos nulos, para determinar a significância de cada um. Finalmente, realizamos comparações post-hoc entre pares, para observar diferenças estatísticas entre as médias de tempo de resposta – doravante RT, do inglês response time (GODOY; NUNES, 2020).

A acurácia total foi de 92% e a acurácia por condição pode ser verificada na tabela 8. MI2 foi a condição de maior acurácia com 97,5% (DV: 6,05%), seguida de MO1 (96,61%, DV: 6,6%), MI1 (96,25%, DV: 9,37%), DIstratores i.e., as pseudopalavras (91,90%, DV: 8,51%), SR2 (91,60%, DV: 9,51%) e, finalmente, SR1 com 91,25% (DV: 9,93%).

Para todas as análises de resultados comportamentais, ajustamos modelos mistos generalizados lineares, utilizando a família gaussiana invertida. Em seguida, fizemos comparações aninhadas entre modelos mais simples e mais complexos, e dos modelos escolhidos com seus respectivos modelos nulos, para determinar a significância de cada modelo.

**Acurácia por condição experimental, em %**

<b>condition</b>	<b>media</b>	<b>sdev</b>	<b>n</b>	<b>se</b>	<b>CImin</b>	<b>CImax</b>
<b>DIS</b>	91.90134	8.507121	30	1.553181	88.8571	94.94557
<b>MI1</b>	96.25	9.371407	30	1.710977	92.89648	99.60352
<b>MI2</b>	97.5	6.052927	30	1.105108	95.33399	99.66601
<b>MO1</b>	96.60714	6.604563	30	1.205823	94.24373	98.97056
<b>MO2</b>	90.59524	15.75659	30	2.876746	84.95682	96.23366
<b>SR1</b>	91.25	9.92971	30	1.812909	87.6967	94.8033
<b>SR2</b>	91.60714	9.508853	30	1.736071	88.20444	95.00984

Tabela 7. Acurácia da decisão lexical por condição experimental, na análise comportamental do EEG.

Por fim, realizamos comparações post-hoc para delimitar as diferenças estatísticas entre as médias de tempo de resposta (doravante RT, de *reaction or response time*).

Para esses dados, ajustamos um modelo de RT em função da condição experimental e considerando participantes e *targets* como fatores aleatórios ( $RT \sim condition + (1|Subject) + (1|target)$ ). Considerar *target* como fator aleatório teve como objetivo dar conta de fatores que mudam de acordo com cada palavra, como frequência, quantidade de letras e de sílabas. Esse modelo foi consideravelmente diferente de um modelo nulo ( $\chi^2(6) = 94.189, p < 2.2e-16$  \*\*\*), o que significa que ele explica melhor os dados do que um modelo que analisa os tempos de resposta, mas não em função da condição experimental. Em outras palavras, a condição experimental é um fator relevante para os valores dos RT.

Assim, RT por condição podem ser verificados na tabela 9 e também no gráfico 4. Os distratores pseudopalavras apresentaram o maior RT, como esperado para esse tipo de teste, com média de 698.13ms. Eles foram significativamente maior que os tempos de todas as outras condições; por ordem decrescente: MO2 ( $\beta = -46.71, SE = 11.26, t = 4.148, p = 0.0007$ ), SR2 ( $\beta = -53.79, SE = 10.83, t = 8.287, p < .0001$ ), SR1 ( $\beta = -89.77, SE = 10.86, t = 8.287, p < .0001$ ), MO1 ( $\beta = -98.57, SE = 11.01, t = 8.950, p < .0001$ ), MI2 ( $\beta = -103.80, SE = 10.53, t = 9.861, p < .0001$ ) e MI1 ( $\beta = -126.23, SE = 10.50, t = 12.020, p < .0001$ ).

Na análise post-hoc ajustada por Tukey, a diferença entre MI1 e MO1 tendeu à significância ( $\beta = -27.66, SE = 9.71, t = -2.849, p = 0.0661$ ) e a entre MI2 e MO2 foi significativa ( $\beta = -57.09, SE = 10.85, t = -5.263, p < .0001$ ). Já numa análise post-hoc não ajustada, a comparação MI1 vs. MO1 foi significativa ( $\beta = -27.66, SE = 9.71, t = -2.849, p = 0.0044$ ) e a entre MI2 e MO2 permaneceu idêntica, ratificando sua prévia significância.

### RT por condição experimental, em ms

condition	media	sdev	n	se	CImin	CImax
<b>DIS</b>	667.8491	154.5698	1418	4.104748	659.8038	675.8944
<b>MI1</b>	534.9916	137.2616	239	8.878715	517.5894	552.3939
<b>MI2</b>	562.1975	136.7085	238	8.861494	544.829	579.566
<b>MO1</b>	565.3234	136.2443	235	8.887596	547.9037	582.7431
<b>MO2</b>	629.0043	176.9648	233	11.59335	606.2813	651.7273
<b>SR1</b>	588.636	168.4812	239	10.89814	567.2756	609.9963
<b>SR2</b>	616.4641	151.7103	237	9.854641	597.149	635.7792

Tabela 8. Tempos de resposta da decisão lexical por condição experimental, na análise comportamental do EEG.

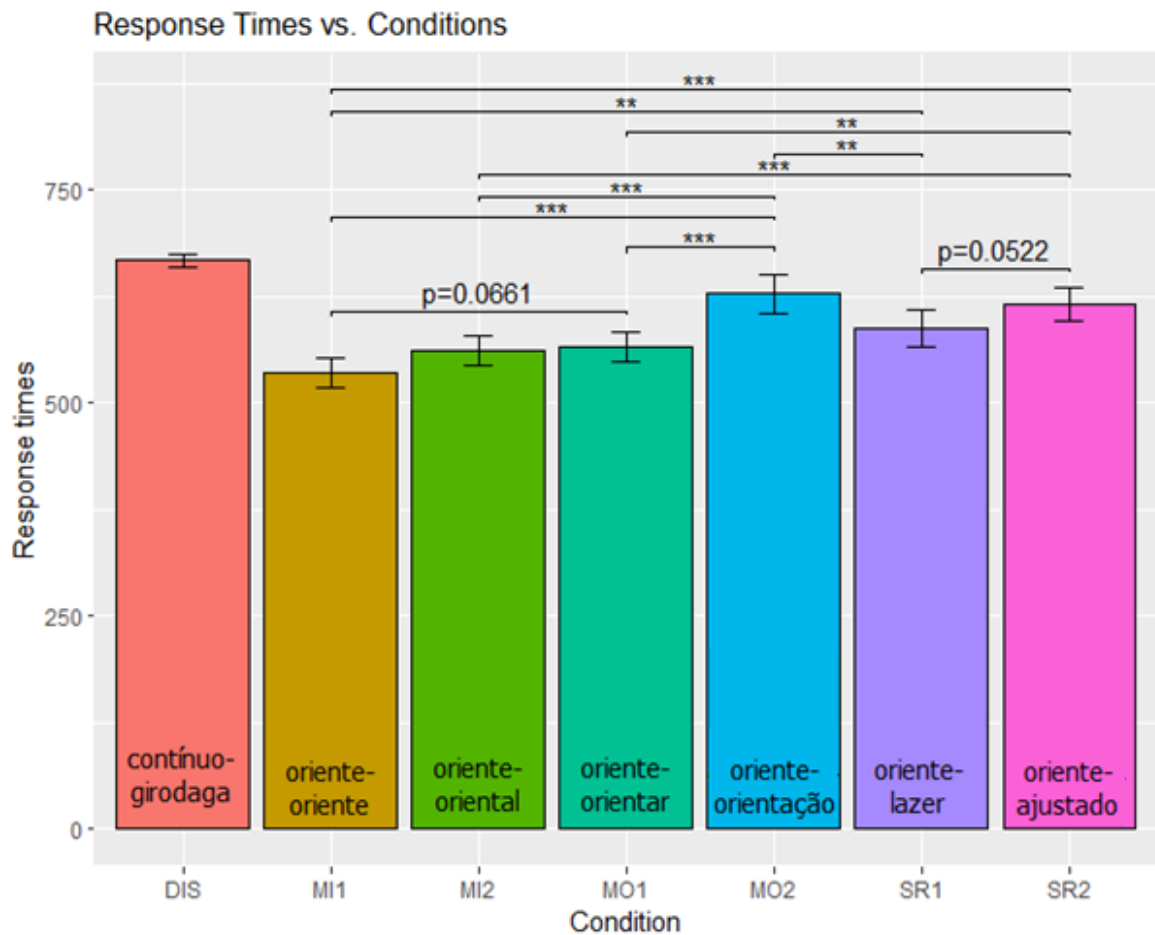


Gráfico 4. Tempos de resposta por condição experimental, na análise comportamental do EEG.

Além disso, ajustando os valores por Tukey, MI1 não apresentou diferença significativa com MI2 ( $\beta = -22.43$ ,  $SE = 10.09$ ,  $t = -2.223$ ,  $p = 0.2830$ ), mas MO1 apresentou em relação a MO2 ( $\beta = -51.86$ ,  $SE = 12.30$ ,  $t = -4.217$ ,  $p = 0.0005$ ) e SR1 apresentou uma tendência de significância com relação a SR2 ( $\beta = -35.98$ ,  $SE = 12.27$ ,  $t = -2.933$ ,  $p = 0.0522$ ). Ademais, MI1 foi estatisticamente diferente de MO2 ( $\beta = -79.52$ ,  $SE = 10.94$ ,  $t = -7.272$ ,  $p < 0.0001$ ). Num modelo não ajustado, porém, todas essas comparações foram estatisticamente significativas: MI1 x MI2:  $\beta = -22.43$ ,  $SE = 10.09$ ,  $t = -2.223$ ,  $p = 0.0262$ ; MO1 x MO2:  $\beta = -51.86$ ,  $SE = 12.30$ ,  $t = -4.217$ ,  $p < 0.0001$ ; SR1 x SR2:  $\beta = -35.98$ ,  $SE = 12.27$ ,  $t = -2.933$ ,  $p = 0.0034$ ; e MI1 x MO2 obteve exatamente o mesmo resultado.

Além disso, em pares post-hoc ajustados por Tukey, SR1 foi diferente de MI1 ( $\beta = 36.47$ ,  $SE = 9.84$ ,  $t = 3.706$ ,  $p = 0.0040$ ), mas não de MO1 ( $\beta = 7.84$ ,  $SE = 8.78$ ,  $t = 0.893$ ,  $p = 0.9737$ ). Algo similar aconteceu com SR2, que foi diferente de MI2 ( $\beta = 50.01$ ,  $SE = 10.49$ ,  $t = 4.769$ ,  $p < 0.0001$ ), mas não de MO2 ( $\beta = -7.08$ ,  $SE = 11.38$ ,  $t = -0.622$ ,  $p = 0.9961$ ). SR2 foi, ainda, diferente de MI1 ( $\beta = 72.44$ ,  $SE = 10.65$ ,  $t = 6.803$ ,  $p < 0.0001$ ) e MO1 ( $\beta = 44.79$ ,  $SE = 12.11$ ,  $t = 3.699$ ,  $p = 0.0041$ ). Já SR1 foi diferente de MO2 ( $\beta = -43.06$ ,  $SE = 12.30$ ,  $t = -3.501$ ,  $p = 0.0084$ ).

Ajustamos, também, um modelo com RT em função da interação do tipo de relacionamento (MI, MO, SR) com a quantidade de camadas morfológicas (1 ou 2) e considerando participante e *target* como efeitos aleatórios mas, após comparação por modelos aninhados, ele não foi mais significativo do que um modelo que não continha a interação entre tipo de relacionamento e quantidade de camadas ( $RT \sim rel\_type * layers + (1|Subject) + (1|target)$ ) vs.  $RT \sim rel\_type + layers + (1|Subject) + (1|target)$ :  $\chi^2 =$ ,  $p = 0.1584$ ). Tendo sempre como objetivo a parcimônia, escolhemos o último modelo, sem interação, que tem RT como variável resposta, tipo de relacionamento e camadas morfológicas como variáveis preditoras, e *intercepts* aleatórios para participante e *target* ( $RT \sim rel\_type + layers + (1|Subject) + (1|target)$ ). Uma comparação com modelos aninhados indicou que o fator camadas morfológicas contribuía significativamente para o modelo ( $\chi^2(1) = 9.8585$ ,  $p = 0.00169$  \*\*), assim como aconteceu com o fator tipo de relacionamento ( $\chi^2(2) = 35.949$ ,  $p = 1.562e-08$  \*\*\*), e que a sua verossimilhança era maior que a de um modelo nulo ( $\chi^2(3) = 50.4$ ,  $p = 6.568e-11$  \*\*\*).

No gráfico 5 e na tabela 10, é possível ver os RT de acordo com o tipo de relacionamento. Comparações post-hoc ajustadas por Tukey determinaram que MI é significativo quando comparado a MO ( $\beta = -39.88$ ,  $SE = 7.68$ ,  $t = -5.194$ ,  $p < 0.0001$ ) e a SR ( $\beta = -43.05$ ,  $SE = 7.73$ ,  $t = -$

5.573,  $p < .0001$ ). No entanto, a diferença entre MO e SR não é significativa ( $\beta = -3.17$ ,  $SE = 8.26$ ,  $t = -0.384$ ,  $p = 0.9218$ ).

### RT por tipo de relacionamento, em ms

rel_type	media	sdev	n	se	CImin	CImax	lim_min
MI	548.566	137.5178	477	6.296507	536.2249	560.9072	136.0127
MO	597.0278	160.8564	468	7.43559	582.454	611.6015	114.4586
SR	602.4916	160.7861	476	7.369618	588.0471	616.936	120.1333

Tabela 9. Tempos de resposta da decisão lexical por tipo de relacionamento, na análise comportamental do EEG.

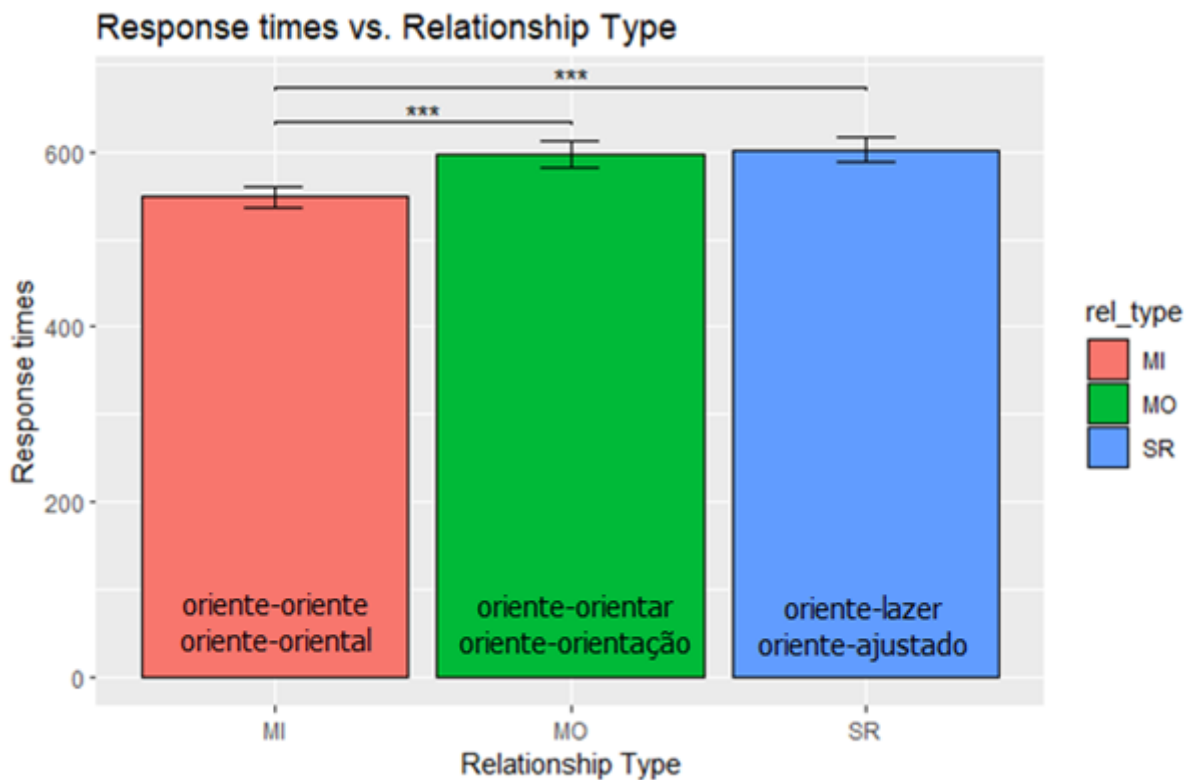


Gráfico 5. Tempos de resposta da decisão lexical por tipo de relacionamento, na análise comportamental do EEG.

### RT por quantidade de camadas morfológicas, em ms

Layers	media	sdev	n	se	CImin	CImax	lim_min
1_layer	562.9705	149.5752	713	5.601637	551.9913	573.9498	114.2448
2_layer	602.3489	158.3285	708	5.95035	590.6862	614.0116	127.3633

Tabela 10. Tempos de resposta da decisão lexical por quantidade de camadas morfológicas, na análise comportamental do EEG.

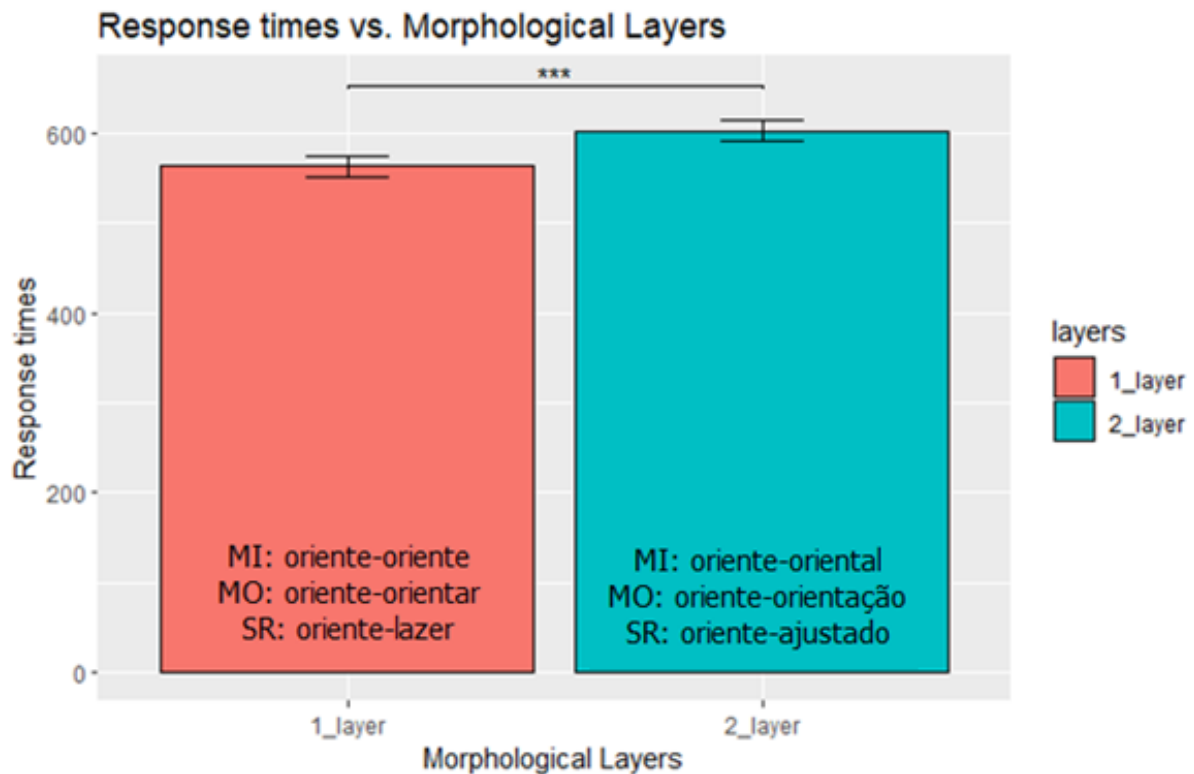


Gráfico 6. Tempos de resposta da decisão lexical por quantidade de camadas morfológicas, na análise comportamental do EEG.

Uma comparação post-hoc do mesmo modelo e também ajustada por Tukey, mas focada na quantidade de camadas morfológicas determinou que estímulos com 1 camada geraram resultados significativamente diferentes de estímulos com 2 camadas ( $\beta = -30.7$ ,  $SE = 9.08$ ,  $t = -3.388$ ,  $p = 0.0007$ ), conforme disponível no gráfico 6 e na tabela 11.

É possível perceber, portanto, que, apesar de em um modelo que considere todas as condições experimentais não ter significância, nesses há. Acreditamos que isso pode acontecer pela quantidade maior de estímulos considerados na análise estatística, conferindo robustez à mesma.

### 5.6.2 Pré-processamento dos dados neurofisiológicos

Os dados neurofisiológicos foram pré-processados no programa Analyzer (BrainProducts), versão 2.2, de acordo com a seguinte ordem de operações:

- i) Canais muito ruidosos foram interpolados por 2 a 4 canais vizinhos, com a função *pooling*;
- ii) Todos os canais foram re-referenciados em relação aos mastoides;
- iii) O sinal foi filtrado com um passa-alto (*lowpass*) de 30 Hz (Butterworth filter, cf. BUTTERWORTH, 1930);
- iv) O sinal contínuo de EEG foi inspecionado para artefatos, como piscada de olho, franzimento de testa, deglutição de saliva e possíveis outros sinais distorcidos (devido à perda de eletrodos, diafonia etc.). Isso foi feito usando a função de inspeção semiautomática do Analyzer, com critério de seleção, como atividade superior ou inferior a 100  $\mu$ V. Segmentos afetados foram eliminados. Nenhum participante obteve mais do que 15% de perda de dados, de maneira que nenhum foi eliminado por esse motivo. Um total de 1,67% da amostra foi retirado, permanecendo dentro do limite de boas práticas estipulado por Luck (LUCK, 2014);
- v) Uma vez inspecionado para artefatos, o sinal contínuo de EEG foi segmentado de 200 ms antes do *onset* do estímulo *target* até 1000ms depois dele, usando a função *epoching* do Analyzer. Cada segmentação, portanto, tinha o tamanho total de 1200 ms;
- vi) *Baseline correction* foi aplicada a cada segmento, de -200 ms a 0 ms antes do *onset* da palavra *target*. ERPs foram sempre time-locked para o onset do *target*;
- vii) Um média de todos os segmentos de uma mesma condição (MI1, MI2, MO1, MO2, SR1, SR2 e as PSEUDOpalavras distratoras) foi feita por sujeito, usando a função *averaging*. Em seguida, usando a função *Grand Average*, foram feitas médias gerais para cada uma dessas condições, além de para cada tipo de relacionamento (MI, MO, SR).

### 5.6.3 Resultados neurofisiológicos do EEG (amplitudes de ondas)

As análises estatísticas foram feitas no software R (R Core Team, 2013), usando os pacotes ggplot2, ggsignif, dplyr, tidyr, plotrix, tibble, stringr, writex, lme4, lmerTest, fpp e emmeans. Para todos os dados comportamentais mostrados nesta seção, ajustamos modelos lineares mistos (lmer), uma vez que a função glmer, utilizada para as análises comportamentais, não lida bem com números negativos. Em seguida, realizamos comparações aninhadas entre modelos mais simples e mais complexos, e dos modelos escolhidos com seus respectivos modelos nulos, para determinar a significância de cada um. Finalmente, realizamos comparações post-hoc entre pares, para observar diferenças estatísticas entre as médias de amplitude (GODOY; NUNES, 2020).

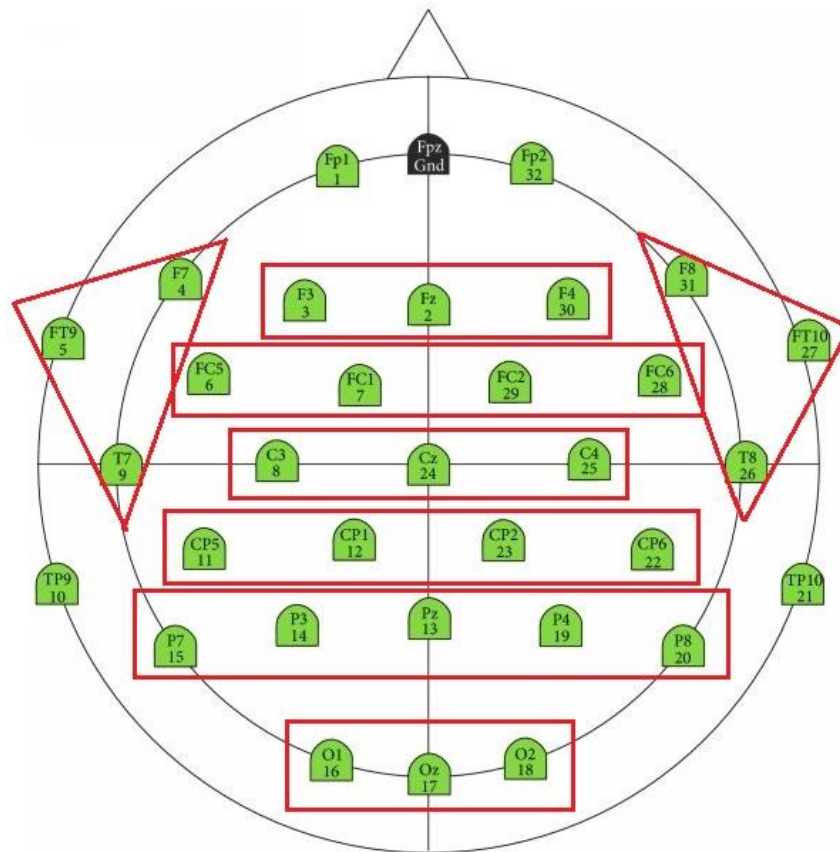


Figura 11. Mapa de delimitação das regiões cerebrais de interesse (ROIs), com seus respectivos canais (eletrodos): ROI Frontal com eletrodos F3, Fz, F4; ROI Fronto-temporal esquerda com eletrodos F7, FT9, T7; ROI Fronto-temporal direita com eletrodos F8, FT10, T8; ROI Fronto-central com eletrodos FC5, FC1, FC2, FC6; ROI Central com eletrodos C3, Cz, C4; ROI Centro-parietal com eletrodos CP5, CP1, CP6, CP2; ROI Parietal com eletrodos P3, P7, P4, P8; e ROI Occipital com eletrodos Oz, O1, O2. Fonte: Silva-Nasser, 2022:76.



Os dados de amplitude analisados nessa seção foram gerados pelo programa Analyzer considerando os sinais gravados no intervalo de 300ms a 500ms após o onset do *target*, em uma busca pelo efeito do componente neurofisiológico N400 (MORRIS; STOCKALL, 2012).

As regiões de interesse (ROI, do inglês *region of interest*) foram geradas agrupando os seguintes canais: frontal (canais F3, Fz, F4), fronto-temporal esquerda (F7, FT9, T7), fronto-temporal direita (F8, FT10, T8), fronto-central (FC5, FC1, FC2, FC6), central (C3, Cz, C4), centro-parietal (CP5, CP1, CP6, CP2), parietal (P3, P7, P4, P8), occipital (O1, Oz, O2) (cf. figura 11).

Assim como feito para os dados comportamentais de RT, queríamos analisar respostas neurofisiológicas de amplitudes de onda em relação a três fatores:

- A condição experimental (MI1, MI2, MO1, MO2, SR1, SR2, DIS):

Para tanto, ajustamos um modelo (Mod1) com amplitudes como variável resposta, a interação entre ROI e condição experimental como variável preditora e *intercept* aleatório para participante (meanamp~ROI\*condition + (1|File)). No entanto, em uma comparação aninhada, esse modelo não foi mais significativo do que um sem a interação entre ROI e condição experimental (vs. meanamp~ROI+condition + (1|File):  $\chi^2(42)=23.325$ ,  $p=0.9913$ ). Escolhemos, então, o último modelo, que foi significativamente mais verossímil que um modelo nulo:  $\chi^2(13)=186$ ,  $p< 2.2e-16$  \*\*\*. O fator condição experimental se mostrou relevante para tal modelo ( $\chi^2(6)=67.623$ ,  $p=1.255e-12$  \*\*\*), assim como o fator ROI ( $\chi^2(7)=124.08$ ,  $p< 2.2e-16$  \*\*\*). O gráfico 7 demonstra amplitudes plotadas por condição experimental, em cada ROI.

- O tipo de relacionamento (MI, MO, SR):

Para analisa-lo, ajustamos um modelo (Mod2) com amplitudes como variável resposta, a interação entre ROI e tipo de relacionamento como variáveis preditoras e *intercept* aleatório para participante. Esse modelo, no entanto, não foi mais significativo que um modelo similar mas não considerativo da interação entre ROI e tipo de relacionamento: [meanamp ~ ROI \* rel\_type + (1 | File)] vs. [meanamp ~ ROI + rel\_type + (1 | File)]:  $\chi^2(21)=13.451$ ,  $p=0.892$ . Escolhemos, então, o último modelo, que foi significativo quando comparado ao modelo nulo:  $\chi^2(10)=177.85$ ,  $p< 2.2e-16$  \*\*\*. O fator tipo de relacionamento se mostrou relevante nesse modelo ( $\chi^2(3)=59.475$ ,  $p=7.61e-13$  \*\*\*), assim como o fator ROI ( $\chi^2(7)=123.38$ ,  $p< 2.2e-16$  \*\*\*). O gráfico 8 demonstra amplitudes plotadas por tipo de relacionamento, em cada ROI.

- A quantidade de camadas morfológicas (1 camada, 2 camadas):

Assim, ajustamos um modelo (Mod3) com amplitudes como variável resposta, a interação entre ROI e quantidade de camadas morfológicas como variáveis preditoras e *intercept* aleatório para participante. Esse modelo, no entanto, não foi mais significativo que um modelo similar mas não considerativo da interação entre ROI e quantidade de layers:  $\chi^2(14)=10.137$ ,  $p=0.7521$ . Escolhemos, então, o último modelo, que foi significativamente diferente do modelo nulo:  $\text{meanamp} \sim \text{ROI} + \text{layers} + (1 | \text{File})$  vs.  $\text{meanamp} \sim 1 + (1 | \text{File})$ :  $\chi^2(23)=181.89$ ,  $p < 2.2e-16$  \*\*\*. O fator quantidade de camadas morfológicas se mostrou relevante para esse modelo ( $\chi^2(2)=53.377$ ,  $p=2.567e-12$  \*\*\*), assim como o fator ROI ( $\chi^2(7)=122.86$ ,  $p < 2.2e-16$  \*\*\*). O gráfico 9 demonstra amplitudes plotadas por quantidade de camadas morfológicas, em cada ROI.

Como ROI se mostrou um fator relevante em todos esses modelos, decidimos analisar esses três fatores em cada ROI, ou seja, aplicar cada um desses 3 modelos a sub-dataframes que continham amplitudes correspondentes a apenas uma ROI. Dessa maneira, para os dados de cada ROI, aplicamos modelos tipo 1 (Mod1), que analisam as amplitudes em função das condições experimentais, modelos tipo 2 (Mod2), que analisam as amplitudes em função dos tipos de relacionamento, e modelos tipo 3 (Mod3), que analisam as amplitudes em função da quantidade de camadas morfológicas. Um resumo da comparação de cada um desses modelos com modelos nulos, em cada ROI, está disponível na tabela 12.

No gráfico 10, estão plotadas as amplitudes para todas as condições experimentais, em todos os ROIs. Lançando mão do mod1, como primeira análise é possível verificar que os distratores PSEUDopalavras foram diferentes das outras condições, especialmente de MI1, uma condição de identidade que, conforme previsto, apresentou os menores valores de amplitude aproximadamente aos 400ms após o onset do *target*. A diferença MI1 (representada em preto no gráfico 10) x PSEUD (verde claro no gráfico 10) foi estatisticamente significativa em 6 dos 8 ROIs, tendendo à significância nesses 2: Frontal:  $\beta = -2.5331$ ,  $SE=1.1122$ ,  $t=-2.278$ ,  $p=0.0242^*$ ; FT\_Esq:  $\beta = -1.7696$ ,  $SE=0.9366$ ,  $t = -1.889$ ,  $p=0.0608$ . ; FT\_Dir:  $\beta = -1.4835$ ,  $SE=0.7555$ ,  $t = -1.963$ ,  $p=0.0514$  .; FC:  $\beta = -2.8509$ ,  $SE=0.8098$ ,  $t = -3.521$ ,  $p=0.000571$  \*\*\*; Central:  $\beta = -4.562$ ,  $SE=0.824$ ,  $t = -5.536$ ,  $p=1.35e-07$  \*\*\*; CP:  $\beta = -4.8758$ ,  $SE=0.7796$ ,  $t = -6.255$ ,  $p=3.95e-09$  \*\*\*; Parietal:  $\beta = -3.5475$ ,  $SE=0.6903$ ,  $t = -5.139$ ,  $p=8.47e-07$  \*\*\*; Occipital:  $\beta = -3.1799$ ,  $SE=0.6992$ ,  $t = -4.548$ ,  $p=1.11e-05$  \*\*\*.

<b>Comparação com um modelo nulo nas seguintes ROIs:</b>	<b>Modelo1</b> Por condição experimental (meanAmp~condition + (1 File))	<b>Modelo2</b> por tipo de relacionamento (meanAmp~type_rel + (1 File))	<b>Modelo 3</b> Por quantidade de camadas (meanAmp~layers + (1 File))
<b>Frontal</b>	$\chi^2(6)=6.6983$ , p=0.3497	$\chi^2(3)=5.5555$ , p=0.1354	$\chi^2(2)=5.0222$ , p=0.08118 .
<b>Fronto-temporal Esquerdo</b>	$\chi^2(6)=10.067$ , p=0.1219	$\chi^2(3)=6.3377$ , p=0.09629 .	$\chi^2(2)=4.4642$ , p=0.1073
<b>Fronto-temporal Direito</b>	$\chi^2(6)=12.727$ , p=0.04757 *	$\chi^2(3)=10.581$ , p=0.01422 *	$\chi^2(2)=10.68$ , p= 0.004795 **
<b>Fronto-central</b>	$\chi^2(6)=13.445$ , p=0.03649 *	$\chi^2(3)=11.582$ , p=0.008963 **	$\chi^2(2)=10.547$ , p=0.005126 **
<b>Central</b>	$\chi^2(6)=29.19$ , p=5.599e-05 ***	$\chi^2(3)=25.891$ , p=1.005e-05 ***	$\chi^2(2)=21.857$ , p=1.794e-05 ***
<b>Centro-parietal</b>	$\chi^2(6)=36.38$ , p=2.325e-06 ***	$\chi^2(3)=30.41$ , p=1.131e-06 ***	$\chi^2(2)=23.313$ , p=8.661e-06 ***
<b>Parietal</b>	$\chi^2(6)=25.986$ , p=0.000224 ***	$\chi^2(3)=20.203$ , p=0.000154 ***	$\chi^2(2)=18.939$ , p=7.715e-05 ***
<b>Occipital</b>	$\chi^2(6)=22.485$ , p=0.0009886 ***	$\chi^2(3)=13.333$ , p=0.00397 **	$\chi^2(2)=13.711$ , p=0.001053 **

Tabela 11. Comparações dos modelos 1, 2 e 3 com modelos nulos, em cada ROI. Células coloridas de acordo com a intensidade da significância estatística.

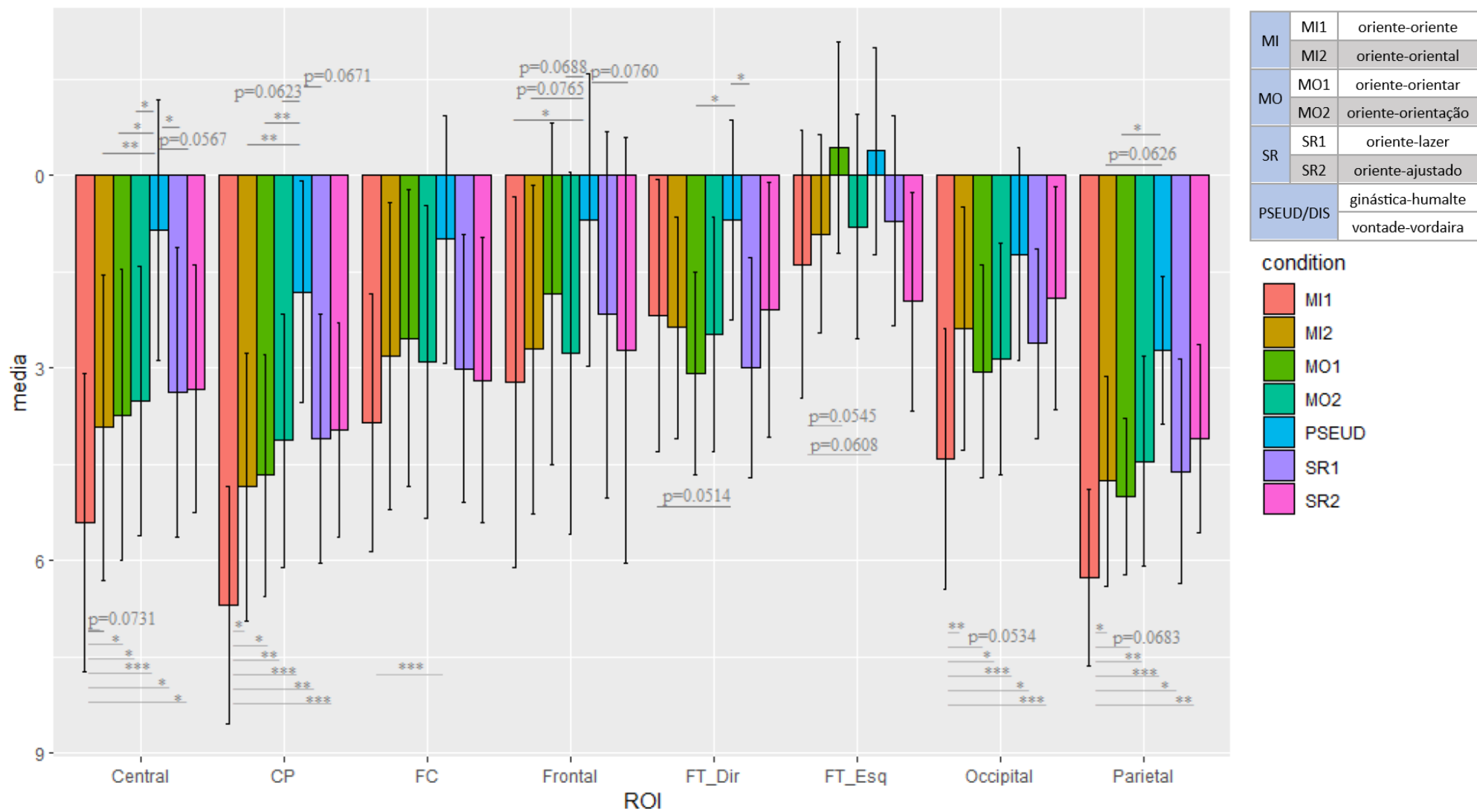


Gráfico 7. Amplitudes médias (em µV) por condição experimental, em cada ROI.

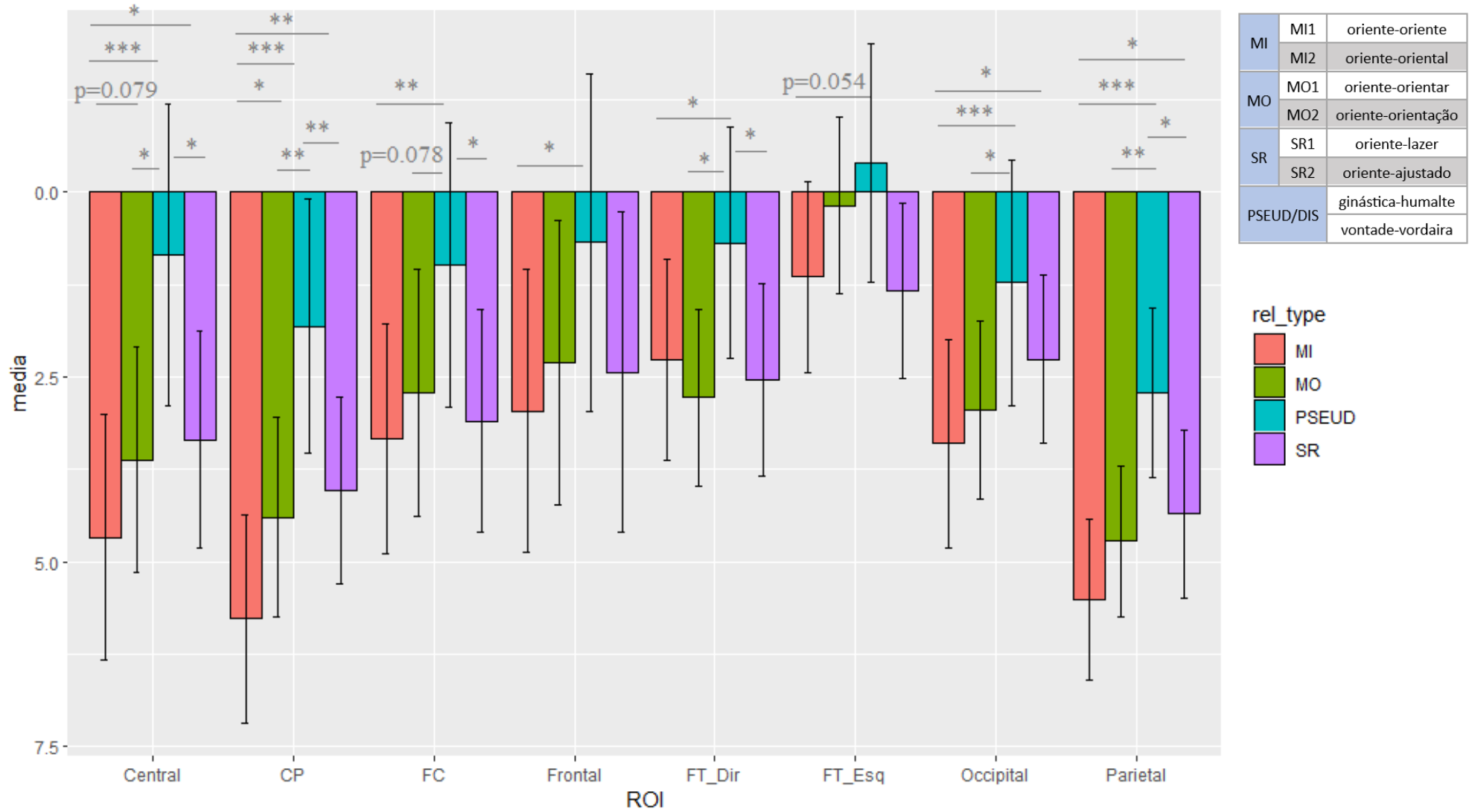


Gráfico 8. Amplitudes médias (em  $\mu\text{V}$ ) por tipo de relacionamento, em cada ROI.

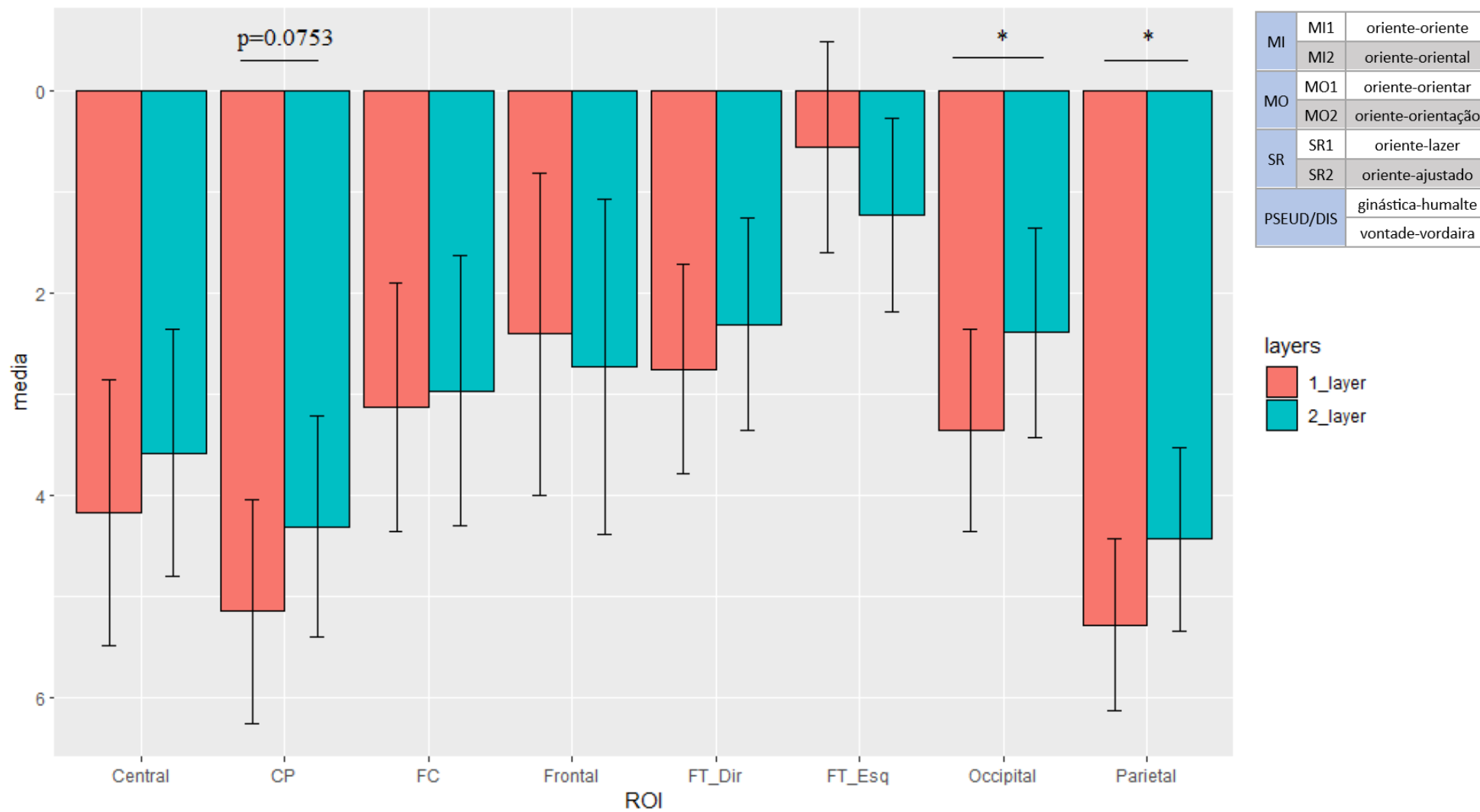


Gráfico 9. Amplitudes médias (em  $\mu\text{V}$ ) por quantidade de camadas morfológicas, em cada ROI.

MI	MI1	oriente-orientado
	MI2	oriente-oriental
MO	MO1	oriente-orientar
	MO2	oriente-orientação
SR	SR1	oriente-lazer
	SR2	oriente-ajustado
PSEUD/DIS		ginástica-humalte
		vontade-vordaira

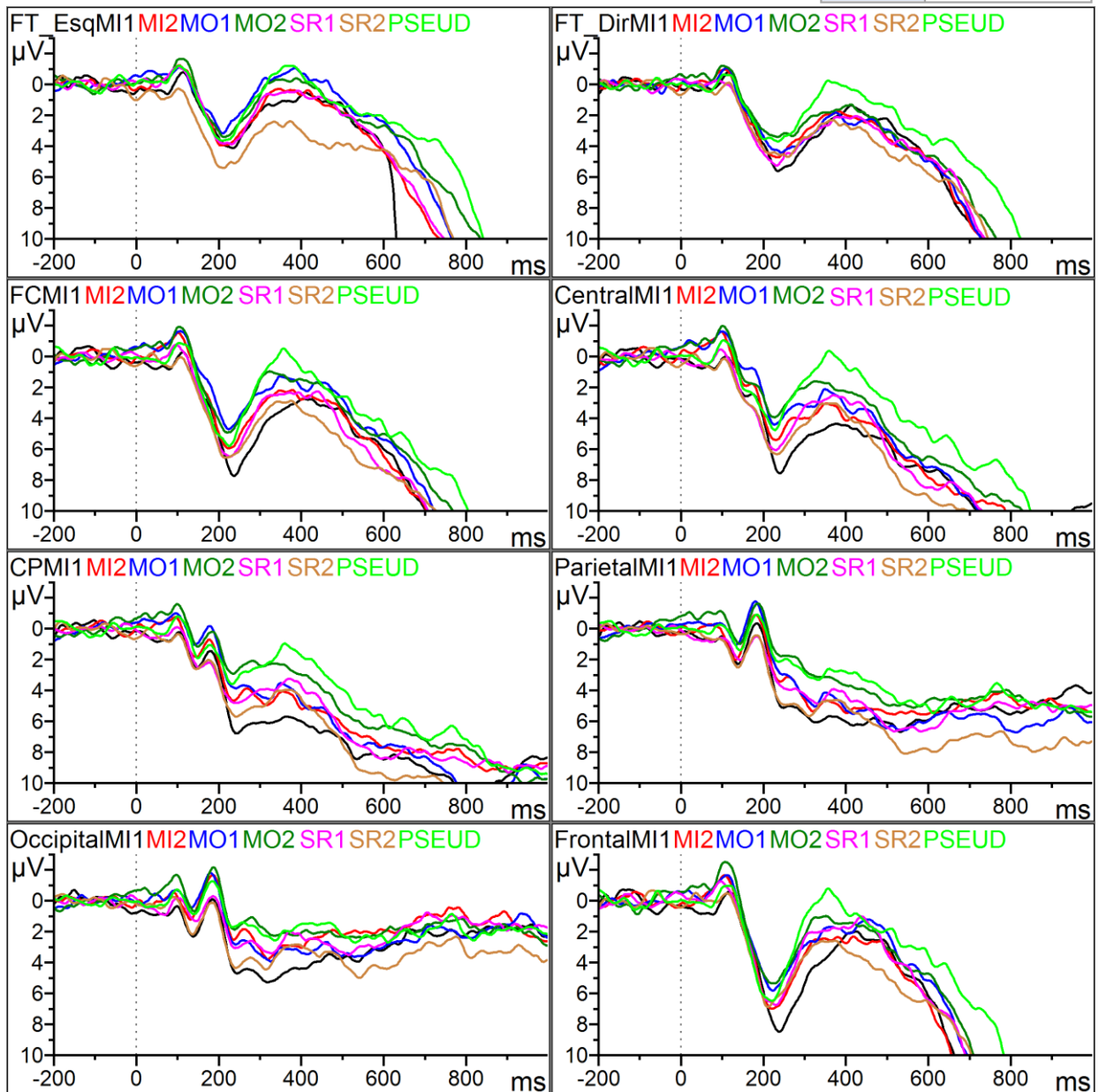


Gráfico 10. Amplitudes (em  $\mu\text{V}$ ), em função do tempo (em ms), plotadas para todas as condições experimentais, em cada e todos os ROI.

O gráfico 10 e o mod1 ainda nos permitem verificar a diferença entre as condições de SR com as de MI. A comparação SR1 (rosa no gráfico 10, azul no gráfico 11) vs. MI1 (preto em ambos os gráficos 10 e 11) foi significativa na ROI Central ( $\beta=2.035$ ,  $SE=0.824$ ,  $t=2.469$ ,  $p=0.0147$  \*) e em CP ( $\beta=2.5862$ ,  $SE=0.7796$ ,  $t=3.318$ ,  $p=0.001140$  \*\*), mas não em FC ( $\beta=0.8408$ ,  $SE=0.8098$ ,  $t=1.038$ ,  $p=0.300786$ ) (cf. também gráfico 11).

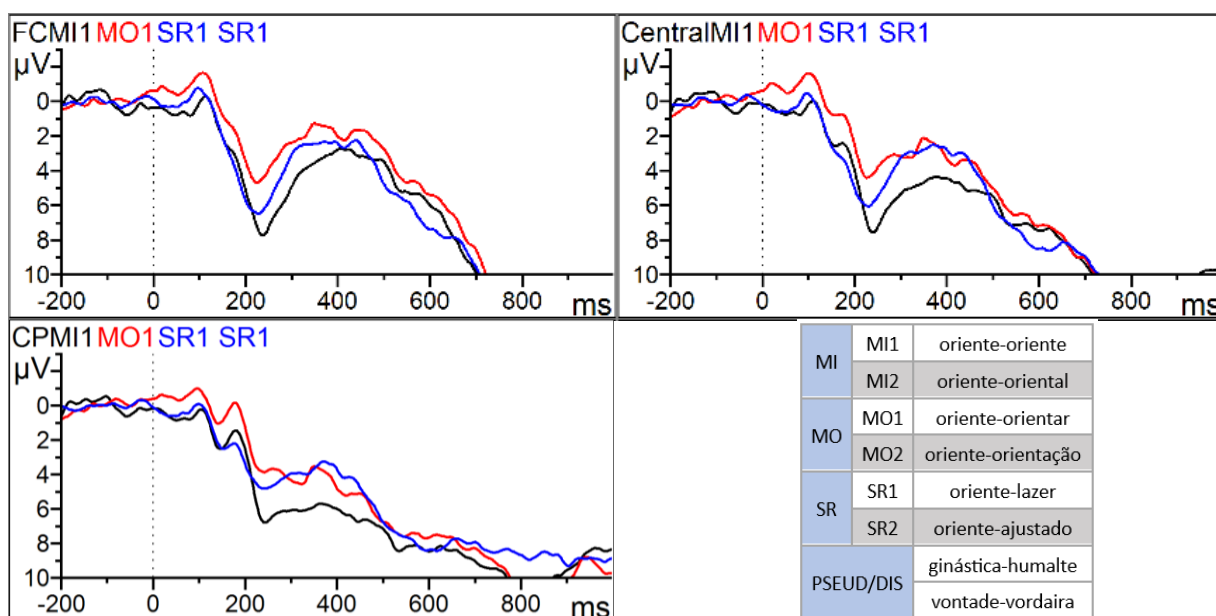


Gráfico 11. Amplitudes (em  $\mu\text{V}$ ) das condições MI1, MO1 e SR1, em função do tempo (em ms), nas ROIs FC, Central e CP.

Já SR2 (laranja no gráfico 10, azul no gráfico 12) não apresentou diferença estatisticamente significativa com MI2 (vermelho no gráfico 10, preto no gráfico 12) em nenhuma dessas três áreas. Em FC:  $\beta=0.3726$ ,  $SE=0.826$ ,  $t=0.451$ ,  $p=0.9993$ ; na ROI Central:  $\beta=-0.601$ ,  $SE=0.841$ ,  $t=-0.715$ ,  $p=0.9915$ ; em CP:  $\beta=-0.8904$ ,  $SE=0.796$ ,  $t=-1.119$ ,  $p=0.9214$ . Em FC, seu processamento foi ainda menos custoso que o de MI2, gerando um N400 menos negativo que MI2. Trata-se de condições diametralmente opostas, uma vez que MI2 têm identidade morfológica entre *prime* e alvo e SR2 não possui nenhum tipo de relacionamento entre *prime* e alvo (cf. também gráfico 12).

Com SR2 não se diferenciando significativamente de MI2, ambos se distanciam de MO2, como é possível verificar no gráfico 12. Trata-se da comparação entre o processamento de MI e MO, o cerne da investigação dessa dissertação. Comparações post-hoc ajustadas por Tukey determinaram que a comparação MI2 x MO2 não é significativa na ROI central



( $\beta=0.414$ ,  $SE=0.841$ ,  $t=0.492$ ,  $p=0.9989$ ), nem na CP ( $\beta=0.7231$ ,  $SE=0.796$ ,  $t=0.909$ ,  $p=0.9707$ ), nem em FC ( $\beta=-0.0852$ ,  $SE=0.826$ ,  $t=-0.103$ ,  $p=1.0000$ ).

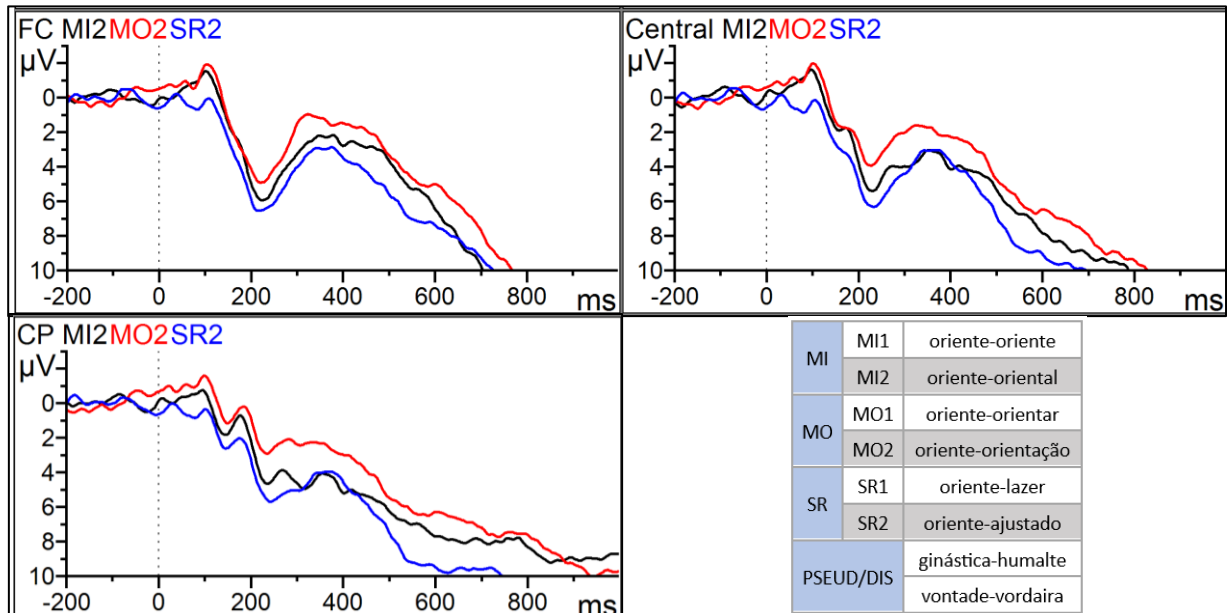


Gráfico 12. Amplitudes (em  $\mu\text{V}$ ) das condições MI2, MO2 e SR2, em função do tempo (em ms), nas ROIs FC, Central e CP.

Entretanto, a comparação de MI1 (preto no gráfico 13) com MO1 (azul no gráfico 13) é significativa na área Central ( $\beta=-1.679$ ,  $SE=0.824$ ,  $t=-2.038$ ,  $p=0.0433$  \*) e em CP ( $\beta=-2.0243$ ,  $SE=0.7796$ ,  $t=-2.597$ ,  $p=0.010349$  \*). Não o é em FC:  $\beta=-1.3118$ ,  $SE=0.8098$ ,  $t=-1.620$ ,  $p=0.107363$  (cf. gráfico 13).

Considerando o mod2, que tem tipo de relacionamento como variável explicativa, podemos avaliar a diferença entre (estímulos com 1 e 2 camadas morfológicas de) MI e (estímulos com 1 e 2 camadas morfológicas de) MO (cf. plotagem de amplitudes do gráfico 14). Na ROI Central, a diferença MI (preto no gráfico 14) vs. MO (vermelho no gráfico 14) é marginalmente significativa ( $\beta=-1.0466$ ,  $SE=0.5891$ ,  $t=-1.777$ ,  $p=0.0777$  .), enquanto na região CP ela se firma como significante ( $\beta=-1.3737$ ,  $SE=0.5623$ ,  $t=-2.443$ ,  $p=0.01573$  \*) e, na região FC, como não significativa ( $\beta=-0.6133$ ,  $SE=0.5762$ ,  $t=-1.064$ ,  $p=0.28885$ ) (cf. gráfico 14).

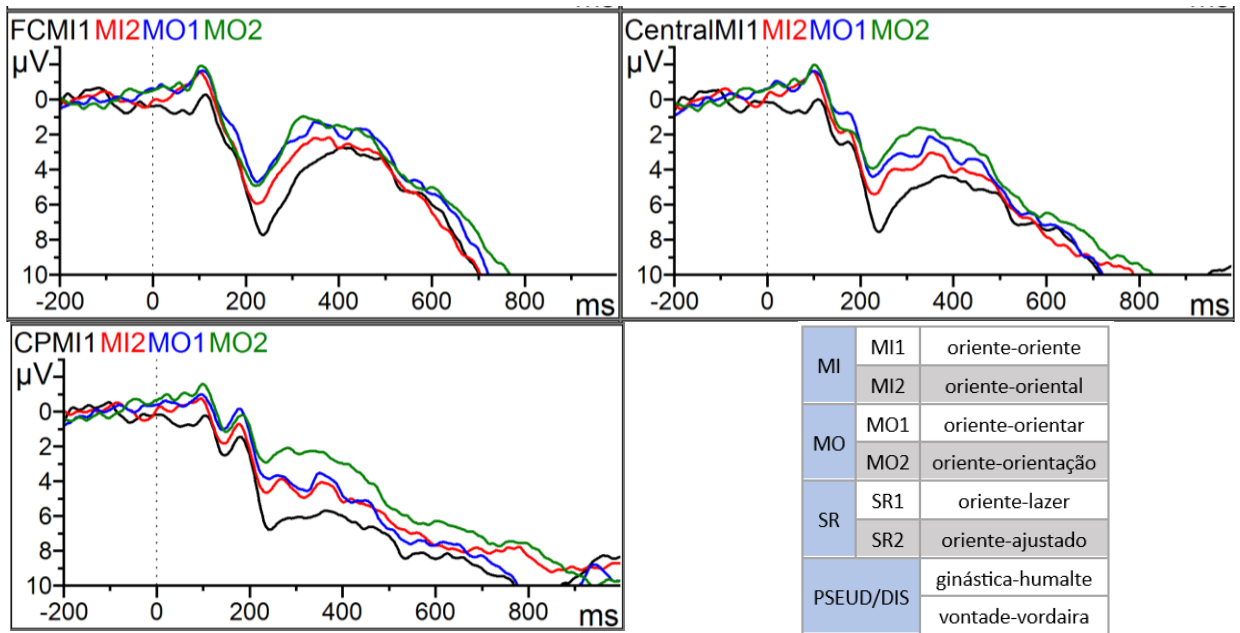


Gráfico 13. Amplitudes (em  $\mu\text{V}$ ) das condições MI1, MI2, MO1 e MO2, em função do tempo (em ms), nas ROIs FC, Central e CP.

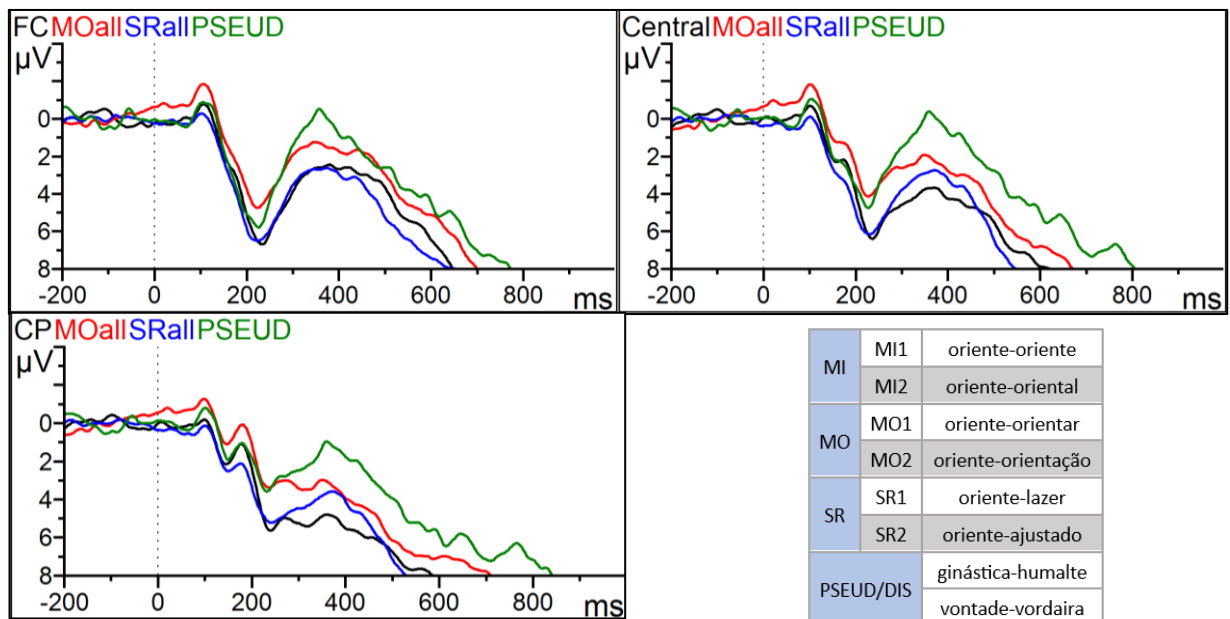


Gráfico 14. Amplitudes (em  $\mu\text{V}$ ) dos tipos de relacionamento MI, MO, SR e PSEUD, em função do tempo (em ms), nas ROIs FC, Central e CP.

Vale também comparar SR com MI e MO. Na ROI Central, SR (azul no gráfico 14) não é significativamente diferente de MI ( $\beta=-1.318$ ,  $SE=0.595$ ,  $t=-2.215$ ,  $p=0.1238$ ) nem de MO

( $\beta=-0.271$ ,  $SE=0.595$ ,  $t=-0.456$ ,  $p=0.9683$ ). Em CP, SR também não é significativamente diferente de MO ( $\beta=-0.365$ ,  $SE=0.568$ ,  $t=0.642$ ,  $p=0.9182$ ), mas o é de MI ( $\beta=-1.738$ ,  $SE=0.568$ ,  $t=3.060$ ,  $p=0.0138$ ). Por fim, na região FC, SR não é significativamente diferente nem de MI ( $\beta=-0.234$ ,  $SE=0.582$ ,  $t=0.402$ ,  $p=0.9779$ ), nem de MO ( $\beta=0.379$ ,  $SE=0.582$ ,  $t=-0.652$ ,  $p=0.9149$ ).

Por fim, lançamos mão do mod3, que toma a quantidade de camadas morfológicas como variável explicativa, e do gráfico 9. É possível verificar que a diferença de processamento entre estímulos de 1 e 2 camadas morfológicas não foi relevante na região Central ( $\beta=-0.5879$ ,  $SE=0.4875$ ,  $t=-1.206$ ,  $p=0.22971$ ) e em FC ( $\beta=-0.1600$ ,  $SE=0.4721$ ,  $t=-0.339$ ,  $p=0.73508$ ), sendo marginalmente significativa em CP ( $\beta=-0.8419$ ,  $SE=0.4701$ ,  $t=-1.791$ ,  $p=0.0753$ ). Vale ressaltar, ainda, que essa diferença foi apenas estatisticamente relevante nas regiões Parietal e Occipital, onde não houve um desenho de onda de N400 (cf. gráfico 9 e 10).

## 5.7 *Discussão*

Nesse item, apresentaremos a discussão dos resultados comportamentais do teste de EEG (seção 5.7.1) e a discussão dos resultados neurofisiológicos do EEG (seção 5.7.2).

### 5.7.1 Discussão dos resultados comportamentais do EEG

Em primeiro lugar, hipotetizamos a existência de entradas lexicais diferentes para palavras sincronicamente transparentes e opacas, como *refrigerar-refrigeração* e *refrigerar-refrigerante*, respectivamente (CATALDO; FRANÇA, 2021). Para tanto, previmos tempos de resposta diferentes para as condições MI1 (*animal-animal*) e MO1 (*animal-animado*), sendo MO1 maior que MI1, por ser mais custoso (a herança do *prime* falha e uma nova entrada na Enciclopédia se faz necessária).

Assim como aconteceu com o estudo comportamental apresentado no capítulo anterior, em comparações post-hoc ajustadas por Tukey, a diferença entre MI1 e MO1 tendeu à significância ( $p=0.0661$ ), enquanto a entre MI2 (*animal-animalesco*) e MO2 (*animal-animadão*) foi significativa ( $p<.0001$ ), sendo todos os RT de MO maiores que os de MI. MI1 foi, também, significativamente diferente de MO2 ( $p<.0001$ ). Além disso, no modelo ajustado que teve tipo de relacionamento (MI, MO, SR) como variável explicativa, MI foi significativamente diferente de MO ( $p<.0001$ ). Tudo isso nos faz, mais uma vez, indicar que

**i) *animal e animado, refrigerar e refrigerante, líquido e liquidação* possuem entradas lexicais diferentes.** Acessar *animal* logo antes de acessar *animado* não é um fator de facilitação para o processamento da segunda palavra; pelo contrário, ele o atrapalha. Como as palavras de cada par não têm identidade morfológica sincrônica entre si, possuem raízes diferentes, sendo necessária uma nova ida à Enciclopédia para uma renegociação do sentido do conjunto [raiz + morfema categorizador].

Em segundo lugar, hipotetizamos um curso decomposicional de processamento para todas as palavras, tanto antes quanto após o acesso lexical (respectivamente, decomposição para retirada das partes constituintes da palavra e posterior composicionalidade do sentido de cada morfema adicionado à [raiz+morfema categorizador]), conforme previsto pela MD (HALLE; MARANTZ, 1997). Porém, seguindo os achados de Cataldo e França (2021) e conforme previsto pela hipótese do *Affix Stripping* (TAFT; FOSTER, 1975), esperamos encontrar tempos de resposta maiores para as condições com 2 camadas morfológicas (MI2, MO2), em comparação com as de 1 camada morfológica (MI1, MO1), como manifestação da pequena adição computacional sofrida. Por exemplo, o par *animal-animadão* teria um custo adicional em relação a *animal-animado*, por conta da adição de uma camada morfológica.

Assim como no experimento comportamental apresentado no capítulo anterior, os RT de MI1 não apresentaram diferença significativa com MI2 ( $p=0.2830$ ), em uma comparação post-hoc por pares ajustada por Tukey. Porém, diferentemente de tal experimento, MO1 foi significativamente diferente de MO2 ( $p=0.0005$ ) e SR1 apresentou tendência de significância quando comparado com SR2 ( $p=0.0522$ ). Um modelo que teve camadas morfológicas e tipo de relacionamento como variáveis explicativas determinou que a diferença entre estímulos (de MI, MO e SR) com 1 camada morfológica e estímulos (de MI, MO e SR) com duas camadas morfológicas foi um fator determinante para os resultados encontrados ( $p=0.0007$ ). Nossa análise, aqui, é a mesma realizada para o experimento comportamental anterior: acreditamos que **ii) o processamento seja decomposicional, mas haja um efeito de tamanho de palavra, de maneira que ter mais material fônico seja naturalmente um fator de atraso nos RT comportamentais.**

Em terceiro e último lugar, hipotetizamos que não haveria facilitação (*priming*) para os alvos das condições sem relacionamento (SR), quando comparados com alvos que possuíam a mesma quantidade de camadas morfológicas, mas outros tipos de relacionamento com seus primes: SR1 vs. MI1 e MO1; SR2 vs. MI2 e MO2. As condições SR são as cujos primes não

forneciam nenhum tipo de facilitação para o processamento do alvo. Ao ler o *target*, o participante recomeçava do zero, pois nenhum tipo de herança era recebida do *prime*.

SR1 foi estatisticamente diferente de MI1 ( $p= 0.0040$ ), mas não de MO1 ( $p=0.9782$ ). O mesmo aconteceu com SR2, ou seja, ele foi estatisticamente diferente de MI2 ( $p<.0001$ ), mas não de MO2 ( $p= 0.9961$ ). Esses resultados estão de acordo com os obtidos no modelo com tipo de relacionamento como variável explicativa: SR foi significativamente diferente de MI ( $p<.0001$ ), mas não de MO ( $p=0.9233$ ). Os RT de SR (602.49 ms) foram, ainda, ligeiramente maiores que os de MO (597.03 ms). Depois dos distratores, que tiveram 667.85 ms de RT, MO2 apresentaram os maiores RT (629.00 ms), seguido de SR2 (616.46 ms) e SR1 (588.64 ms). Em ordem decrescente de tempo de resposta, tivemos: DIS > MO2 > SR2 > SR1 > MO1 > MI2 > MI1. Considerando, no entanto, a acuraria das respostas, a ordem crescente por condições experimentais foi: MO2 < SR1 < SR2 < DIS < MI1 < MO1 < MI2. Vemos, portanto, que **(iii) o processamento de SR1 e SR2 foi menos custoso que o de MO2**, que precisou realizar uma nova negociação semântica na Enciclopédia (mas não que o de MO1, que fez a mesma renegociação mas possuía uma camada morfológica a menos), **e mais custoso que o de MI1 e MI2, cuja derivação foi a mais facilitada possível**, pelo efeito mais forte de identidade do *priming* morfológico, já reportado diversas vezes na literatura (STOCKALL; MARANTZ, 2006; FRANÇA et al., 2008; GARCIA, 2009). Verificamos, portanto, uma escala de facilitação semântica e morfológica entre esses três tipos de condições, com apenas MI2 saindo da previsão e sendo mais custoso que as condições SR.

### 5.7.2 Discussão dos resultados neurofisiológicos coletados pelo EEG

Nossa *primeira* hipótese foi a da existência de entradas lexicais diferentes para palavras sincronicamente transparentes e opacas, como, respectivamente, *refrigerar-refrigeração* e *refrigerar-refrigerante*; ou *animal-animalesco* e *animal-animado*. Para tanto, previmos amplitudes de onda estatisticamente diferentes para as condições MI1 (*animal-animal*) e MO1 (*animal-animado*) na assinatura N400, que reflete o momento do acesso lexical. MO1, por ser mais custoso, apresentaria amplitudes menores (mais negativas) que MI1. De fato, MO1 foi mais custosa que MI1, sendo essa diferença estatisticamente significativa tanto na ROI Central quanto em CP, duas das regiões elegidas para a análise. Além disso, no mod2, ajustado considerando tipo de relacionamento como variável explicativa, as amplitudes de MI foram

significativamente diferentes das de MO na região CP e apresentaram tendência à significância na região Central.

Junto aos resultados do experimento comportamental e às análises comportamentais do presente experimento de EEG, que também apontaram, no geral, uma significância estatística entre as medidas dessas duas condições, chegamos à conclusão que **(i) palavras sincronicamente opacas não possuem mais identidade morfológica com as que as geraram**, isso é, não são mais percebidas como compartilhando a mesma raiz. Acessar *refrigerante* depois de *refrigerar* é mais custoso que acessar *refrigeração* depois de *refrigerar*, porque *refrigeração* compartilha com *refrigerar* a mesma [raiz + morfema categorizador].

Estudos neurofisiológicos demonstraram que o N400 reflete o momento do acesso lexical como previsto pela MD, isto é, o momento que a negociação semântica da [raiz + morfema categorizador] é feita na Enciclopédia (cf. FRANÇA 2008 para um estudo de EGG com português brasileiro e PYLKKÄNEN *et al.*, 2002; 2003 e STOCKALL, MARANTZ, 2006 para estudos de MEG com inglês). Esses estudos demonstraram que palavras morfológicamente relacionadas entre si possuem uma relação de identidade, de maneira que, ler uma antes da outra facilita muito o acesso à segunda. Isso porque a [raiz + morfema categorizador] da *primeira* palavra teria acabado de ser acessada e, ao se processar a segunda, que possui [raiz + morfema categorizador] idêntica, não seria preciso retornar à Enciclopédia: aquela negociação de sentido já estaria feita e o único trabalho seria adicionar, por merge, um sufixo (como o sufixo -esco, de *animalesco* em MI2).

Registrar esse resultado neurofisiologicamente nos permite propor uma conclusão a mais do que apenas resultados comportamentais permitiam. Isso porque a medida comportamental de apertar de botão é uma decisão posterior ao fim da palavra: o participante já ouviu e processou por completo quando aperta um botão julgando se ela existe ou não em sua língua. É, portanto, um julgamento dos efeitos de ***wrap-up* daquela palavra, i.e., do processamento final dela**. O EEG, no entanto, nos permite olhar para **dentro dessas palavras, i.e., para o “durante” desse processamento no nosso cérebro** (=enquanto ele acontece), não precisando esperar para que a raiz e todos os afixos da palavra sejam completamente acessados<sup>9</sup>.

---

<sup>9</sup> Essa é, ainda, uma diferença entre estudos de EGG que analisam palavras utilizando como material **palavras soltas** e estudos de EEG que analisam palavras, mas inseridas em **sentenças**. No segundo caso, não há como atribuir um sinal unicamente ao processamento interno de uma palavra, mas sim ao *wrap-up* dela e ao *wrap-up* dela com a sentença.

Dessa maneira, pudemos ver, em nossos resultados, a confirmação da previsão de que o N400 reflete o momento da junção da raiz com o morfema categorizador, isto é, o momento da atribuição de sentido à uma forma, que recebe o nome de Enciclopédia da MD. Captar o fim (o *wrap-up*) do processamento a palavra não parece possível por uma análise de amplitudes e latências de ondas. A literatura vem se dedicando a fazer isso através da análise das oscilações das ondas (e.g., PEÑA; MELLONI, 2012)

O fato de termos encontrado N400 mais custosos para palavras semanticamente opacas como *refrigerante* demonstra uma espécie de **Garden Path**<sup>10</sup> **Lexical**. Primeiro ouvimos *refrigerar* e acessamos sua [raiz + morfema categorizador]. Ela permanece, então, ativa em nossa mente. Quando vamos ler *refrigerante*, primeiro tomamos o mesmo ponto de arbitrariedade que tomamos em *refrigerar*, isso é, a raiz *refrig-*. Logo ao continuar o processamento, no entanto, nosso parser percebe que está enganado – pois aquela palavra na realidade não tem (mais) identidade morfológica com *refrigerar* – e precisa retornar à Enciclopédia para renegociar o sentido dessa forma, que agora ele entenderá como correspondente à raiz *refrigerant-*. Em outras palavras, a semântica herdada do *prime* entra em choque com o material morfológico e semântico fornecido pelo *target*.

Nossa segunda hipótese previa um curso decomposicional de processamento para todas as palavras, tanto antes quanto após o acesso lexical (respectivamente, decomposição para retirada das partes constituintes da palavra e posterior composicionalidade do sentido de cada morfema adicionado à [raiz+morfema categorizador]), conforme previsto pela MD (HALLE; MARANTZ, 1997). Porém, tanto os resultados de Cataldo e França (2021) quanto os do experimento comportamental reportado no capítulo anterior apontavam para diferenças entre o processamento de palavras com quantidades diferentes de camadas morfológicas. Dessa maneira, resolvemos ser conservadoras e mantivemos nossas previsões baseadas na hipótese do Affix Stripping (TAFT; FOSTER, 1975): esperávamos encontrar tempos de resposta maiores para as condições com 2 camadas morfológicas (MI2, MO2), em comparação com as de 1 camada morfológica (MI1, MO1), como manifestação da pequena adição computacional sofrida. Por exemplo, o par *animal-animadão* teria um custo adicional em relação a *animal-animado*, por conta da adição de uma camada morfológica. Contudo, não encontramos diferença significativa entre estímulos com uma e duas camadas morfológicas em nenhuma

---

<sup>10</sup> Para uma revisão completa de Garden Path, cf. Frazier e Fodor (1978), Fodor e Inoue (2000) e Ribeiro (2015).

ROI, exceto a occipital e a parietal, que não apresentaram desenho de N400. Isso não nos deixa outra opção a não ser concluir que **ii) o processamento é sim decomposicional, mas o N400 reflete apenas o momento em que uma raiz é concatenada a um morfema categorizador e um acesso lexical é feito na Enciclopédia.**

Esses resultados se distanciam de todos os que já encontramos antes, mas na realidade se aproximam dos apontados pela literatura de acesso lexical com EEG e MEG. Pylkkänen et al. (2002, 2003), Stockall e Marantz (2006) e França et al. (2008) encontraram efeitos de tamanho de palavra apenas para pares em que *primes* eram fonologicamente relacionados ao alvo. Essa diferença não foi encontrada em pares com relacionamento morfológico entre *prime* e alvo. Nesses, independentemente do tamanho da palavra (i.e., do número de camadas morfológicas), o acesso lexical se dava sempre no mesmo momento: o ponto de arbitrariedade em que raiz é juntada ao *primeiro* morfema categorização e a forma recebe uma contraparte semântica. Em França et al. (2008), por exemplo, as amplitudes e latências de onda não foram significativamente diferentes entre os pares de relacionamento morfológico com 3 quantidades diferentes de camadas morfológicas. Porém, elas foram estatisticamente diferentes para estímulos com 3 quantidades diferentes de sílabas, quando os pares de palavras eram apenas fonologicamente relacionados. Foi assim que França et al. captaram, em português brasileiro, a diferença entre identidade morfológica e similaridade fonológica proposta por Stockall e Marantz (2006).

Vemos, portanto, que a literatura aponta uma diferença em amplitudes e latências para a comparação entre relacionamento morfológico e fonológico. Aqui comparamos dois tipos diferentes de relacionamento morfológico (transparente e opaco). Isso acontece porque a morfologia possui uma estrutura hierarquizada, enquanto o relacionamento fonológico entre *prime* e *target* geram uma herança positiva de estrutura, hierarquizada ainda por cima. Para a fonologia, cada sílaba é uma informação a mais que está sendo levada para o parser, linearmente. Esses resultados nos fazem propor que, **para analisar diferentes tipos de relação morfológica entre si, experimentos comportamentais não sejam os mais indicados, porque acabam sendo sensíveis a um efeito de tamanho de palavra que encobre o efeito da hierarquia estrutural da morfossintática.**

Por fim, nossa terceira hipótese versava sobre as condições sem relacionamento SR: condições cujos *primes* não forneciam nenhum tipo de facilitação para o processamento do alvo. Ao ler o *target*, o participante começava do zero, pois nenhum tipo de herança era



recebida do *prime*. Nossa hipótese não previa facilitação para os alvos dessas condições sem relacionamento (SR), com a consequente previsão de amplitudes menores (mais negativas) para eles, quando comparados com alvos que possuíam a mesma quantidade de camadas morfológicas, mas outros tipos de relacionamento com seus *primes*: SR1 vs. MI1 e MO1; SR2 vs. MI2 e MO2.

O mod2, que tinha tipo de relacionamento como variável explicativa, nos mostrou que, nas ROI Central e CP, SR não foi significativamente diferente de nem MI nem de MO. Já em CP, SR não foi significativamente diferente de MO, mas sim de MI. A aproximação do custo de processamento de SR ao de MO é justificável pela dificuldade de processamento de MO, que recebe uma herança negativa de seu *prime* e precisa renegociar o sentido de sua raiz na Enciclopédia. Contudo, SR não ser mais custoso que MI é, a princípio, algo muito controverso: trata-se de condições diametralmente opostas, uma vez que MI2 têm identidade morfológica entre *prime* e alvo e SR2 não possui nenhum tipo de relacionamento entre *prime* e alvo.

Analisando a diferença entre SR1 e MI1, vemos uma significância tanto na ROI Central quanto em CP. Ao nos depararmos, no entanto, com a falta de significância da diferença entre SR2 e MI2, em todas as três áreas analisadas, compreendemos o motivo de SR não ter sido diferente de MI. Vale lembrar que, em FC o processamento de SR2 foi ainda mais custoso que o de MI2. Acreditamos que SR2 fosse tão difícil de processar que tenha acabado sendo identificado como uma não-palavra. Não-palavras costumam ser descartadas muito rápido em testes de decisão lexical. Na visão global da palavra que os participantes possuem nos poucos milissegundos em que são expostos a ela, a linha entre pseudopalavras e não-palavras é tênue. Além disso, os estímulos de SR2 eram muito grandes, possuindo mais sílabas que a o padrão para palavras do português. Considerando todos esses fatores, é razoável propor que os participantes tenham jogado os estímulos de SR logo fora, como se fossem não-palavras. Assim, concluímos que **(iii) o processamento de palavras sem relacionamento é mais custoso que o de palavras com identidade morfológica. Vale trazer mais uma vez a ressalva de que SR2 parece ter sido processada como uma não-palavra – o que a faria ser rapidamente descartada, apresentando, portanto, o processamento muito pouco custoso observado.**

## 6. Considerações finais

---

Depois de apresentarmos nossos resultados, acreditamos que cumprimos o objetivo de avaliar a existência de influência da diacronia no nosso conhecimento sincrônico de forma que o acesso lexical ainda leve em conta formas e sentidos opacos. Em outras palavras, queríamos verificar se, hoje em dia, no curso do processamento de *refrigerante*, por exemplo, os falantes de português ainda percebem a raiz de refrigerar dentro dele sendo que esta relação não pode contar com compartilhamento semântico depreensível.

Visitando a literatura da área, escolhemos o *design* que julgamos mais adequado para essa verificação: um teste de *priming* multimodal, aberto, com SOA longo e que mantinha o *prime* sempre constante. O *target* variava quanto ao tipo de relacionamento que mantinha com o *prime* (morfológico intacto, morfológico opaco e nenhum tipo de relacionamento) e quanto à quantidade de camadas morfológicas que possuía.

Ambos os experimentos comportamental e neurofisiológico nos levaram à conclusão de que *refrigerar* e *refrigerante* não possuem mais identidade morfológica. Ou seja, não vemos mais a raiz refrig- dentro de *refrigerante*, de modo a criarmos uma nova raiz refrigerant- em nosso léxico mental. Em razão de essas palavras possuírem entradas lexicais diferentes, ao processarmos *refrigerante* logo após *refrigerar*, pode ser que passemos por uma espécie de Garden Path lexical: nosso parser começa o processamento considerando a herança positiva recebida do *prime*, mas logo se depara com a impossibilidade de continuar dessa maneira e é forçado a voltar à Lista 2 e depois à Enciclopédia para renegociar o sentido daquela raiz que é fonologicamente parecida com refrig- mas que não é ela, e sim refrigerant-. Note-se que mesmo fora da situação experimental do *priming*, é possível que outros fatores, como a frequência das palavras, façam, por vezes, com que a percepção de uma raiz menos frequente seja primeiro percebida como uma outra, mais frequente, criando esta mesma necessidade de reanálise que demonstramos experimentalmente. Por exemplo, no curso da percepção de *restauração*, em algum ponto intermediário, pode ser que o ouvinte já antecipe o acesso a *restaurante*, mais frequente do que *restauração*. Mas assim que *restaurante* se prova como a palavra errada, é necessária uma volta para a borda da raiz, onde uma nova negociação seria requerida: Garden Path lexical de forma natural, na fala.

Apresentamos teorias de processamento de palavras e enfocamos na Morfologia Distribuída porque ela apresenta explicações muito bem ajustadas para uma dissonância entre

nossos resultados comportamentais e neurofisiológicos. No teste psicolinguístico comportamental, que avalia o processamento final (*wrap-up*) da palavra através do aperto de um botão para o julgamento palavra/não-palavra, observamos um efeito de tamanho de palavra nos RT: quanto mais camadas morfológicas uma palavra possuía, maior era o custo do seu processamento, medido em tempo de resposta no botão. Já no teste de EEG, que avalia o acesso lexical durante o processamento das concatenações dos morfemas dentro da palavra, a quantidade de camadas morfológicas não influenciou o custo do processamento. Isso porque, como previsto pela MD, o ponto do acesso lexical marcado pelo pico da onda colhida no escalpo se dá no momento em que a raiz de uma palavra se junta ao seu primeiro morfema categorizador. Esse seria o momento do esforço cognitivo, o momento da arbitrariedade saussuriana, em que a Enciclopédia é acessada e temos a relação entre forma e conteúdo. Após esse ponto, o restante dos afixos são adicionados composicionalmente, de forma quase sem custo cognitivo, de forma previsível (STOCKALL; MARANTZ, 2006; FRANÇA et al., 2008). Essa computação (adições composicionais), portanto, não é refletida nas amplitudes e latências captadas pelo EEG (cf. estudos com oscilações, que parecem oferecer uma contraparte neurofisiológica para o fim do processamento de uma palavra, seu *wrap-up*).

Vê-se, portanto, que a Morfologia Distribuída se mostrou como o modelo teórico mais adequado para a explicação da discrepância entre as metodologias experimentais e também para a ausência de efeito de *priming* para o acesso lexical do alvo do par *refrigerar-refrigerante*. A MD é ainda compatível com a hipótese que propusemos da existência de um Garden Path lexical no processamento dos estímulos morfológicamente opacos.

Chegamos, assim, à conclusão de que o processamento lexical é composto por um primeiro estágio morfo-ortográfico, em que as raízes e afixos seriam identificados e computados, e um segundo estágio morfossemântico, em que o sentido da raiz seria acessado e quaisquer ajustes seriam feitos, como a renegociação de qual é a raiz da palavra *refrigerante*, junto a uma atribuição de sentido à mesma.

Ao final de um trabalho como este, apesar de ele já vir sendo desenvolvido desde meus primeiros anos de graduação, e agora durante o mestrado, sempre vemos que ainda há muito mais para se pesquisar e aprofundar. De forma mais emergente, um *follow-up* necessário e imediato será revisitar os dados neurofisiológicos já colhidos para esta dissertação, mas desta vez fixando o foco em outras janelas de tempo, por exemplo épocas entre 100ms e 300ms pós-estímulo. A análise de potenciais como o N170, de processamento visual de palavra (e.g., a

conduzida em nosso laboratório por Soto *et al.*, 2018), e o N250, de mapeamento de representações sublexicais ortográficas e fonológicas em palavras inteiras, uma espécie de reconhecimento morfológico (e.g., a de Eddy *et al.*, 2016; e Morris *et al.*, 2007), são parâmetros a serem seguidos. Por refletirem momentos ainda mais iniciais no processamento lexical que o N400, essas observações nos proporcionarão meios de confirmar ou mesmo reavaliar as conclusões aqui propostas. Isso inclui, inclusive, a nossa ilação a respeito de a condição experimental SR2 ter sido talvez processada como uma não-palavra.

Em relação à nossa proposta da existência de um Garden Path Lexical, seria interessantíssimo que estudos futuros fossem feitos com o objetivo de confirmar essa hipótese, ou de apontar para a uma inibição imediata de *refrigerante* durante o processamento de *refrigerar*. Discriminar entre competição (Garden path – processamento serial) e inibição sumária de um entre dois ou mais candidatos (processamento paralelo) lançaria luz nos estágios de processamento. Se um Garden-Path lexical fosse verificado, por exemplo, conseguiríamos até entender melhor como o cérebro se engaja e como se desengaja de apostas em que fatores externos como frequência podem servir para deslanchar um caminho derivacional, mas não necessariamente perduram como base para o fim do reconhecimento da palavra.

Além disso, interessa-nos também verificar o processamento mais sutil, pós acesso lexical, que talvez possa dar conta da monitoração das concatenações pós arbitrariedade saussureana. Como podem recordar, não encontramos no N400 diferenças entre *refrigerar-refrigeração*, porque acreditamos que o N400 reflita o merge interno entre a raiz *refri-* e o morfema categorizador verbal, em ambas palavras, verbo e nome. Essa era uma previsão do modelo e uma replicação dos resultados de França *et al* 2008. Porém, seria, é claro, desejável que conseguíssemos verificar quais manifestações bioelétricas se dão depois deste ponto interno de concatenação. Por exemplo, como perceber essas manifestações sutis da concatenação composicional entre *refrigerar* e o nominalizador que se segue a ele? Uma investigação que parece ser apropriada para este fim é a análise de decomposição de tempo-frequência através de uma *wavelet* de Morlet, usada, por exemplo, em Branzi *et al.* (2023). Eles mediram e compararam atividades oscilatórias neuronais em resposta a palavras cognatas versus não-cognatas. Apesar da ativação de palavras ser facilitada quando a tarefa exigia recuperação explícita de palavras (tarefa de nomeação de figuras), o acesso lexical ocorreu mesmo sem a intenção de nomear o objeto (tarefa não verbal de julgamento de tamanho). A ativação de palavras e a seleção de respostas foram acompanhadas por dessincronização beta (25–35 Hz) e

sincronização teta (3–7 Hz), respectivamente medidas ao longo do processamento. Uma análise semelhante, quando aplicada aos nossos materiais, pode vir a revelar atividades pós-acesso que nos façam reconhecer melhor os diferentes esforços cognitivos suplementares em cada condição experimental – esforços esses que não foram capturados pela análise de ERP que extraiu o N400.

Além disso, depois de um estudo sobre os padrões das radiações pós-N400, seria também muito interessante usarmos nossos materiais para um teste mais à longa distância, pensando em verificar o alcance de diferentes teorias. Nesse sentido, poderíamos testar nossos estímulos, contrastando pares de palavras com relacionamento transparente e opaco, através de modelagem computacional, em uma tentativa de apresentação de outras frentes de investigação acerca do status do morfema no acesso lexical. Aqui fazemos referência a contrastar modelos cujos primitivos são morfemas, como faz a MD, com outros que trabalham com unidades amorfas, para responder à pergunta básica se morfemas são objetos reais representados no cérebro humano e, se assim forem, se são processados linear ou hierarquicamente.

Por exemplo, em uma linha de estudos da morfologia, Fruchter e Marantz (2015) usaram a magnetoencefalografia (MEG) para tentar dissociar estágios de processamento morfológico no espaço tempo. Os resultados dessa investigação sugerem que morfemas são unidades reconhecíveis neurologicamente no cérebro humano.

Por outro lado, fazendo modelagem computacional, Grönroos, Virpioja e Kurimo (2017) propuseram o *Naive Discriminative Learning* (NDL), um modelo conexionista com mapeamentos diretos entre forma indecomponível e significado, para explicar o acesso lexical. Eles usaram o Morfessor, um modelo de estado finito não supervisionado com “morfos” induzidos estatisticamente e conseguiram prever tempos de reação humana e respostas neurais sem que houvesse informação de morfemas definidos linguisticamente.

As previsões que fizemos em relação ao nosso conjunto de estímulos e a própria proposta de um Garden Path lexical poderiam ser bastante aprofundados a partir do status de modelos contrastantes como esses citados.

Por fim, esse tema de estudo ainda se beneficiaria de um experimento com design experimental semelhante, mas desta vez avaliado pelo MEG, que tem melhor resolução espacial. Assim poderíamos ter mais sutileza e precisão para avaliarmos o comportamento das áreas cerebrais e dos componentes neurofisiológicos do MEG já relacionados na literatura ao processamento de palavra, como o N170, mais ligado ao reconhecimento morfêmico. Os sinais

gerados por esse componente nos ajudariam a confirmar os sítios onde acontece a nova negociação semântica que enfim nos permite responder: “Não, não usamos **refrigerante para refrigerar** nada.” *Refrigerante é isso aí*, trata-se de algo gostoso que adoça a boca e faz cosquinha na garganta. Mas longe da inocente ação de *refrigerar*, o *refrigerante* é um líquido super açucarado que, ajudado pelo carbono, se infiltra no cérebro e deflagra o neurotransmissor do prazer, dopamina, que logo se esvai, e nos deixa sempre com o gosto de quero mais. Na realidade, essa é a negociação semântica pela qual passamos e que, com o nosso trabalho, pudemos sugerir que nada tem a ver com *refrigerar*.

## Bibliografia

---

AKHLAGHPOUR, H. An RNA-based theory of natural universal computation. **Journal of Theoretical Biology**, v. 537, p. 110984, mar. 2022.

ANDRADE, I. R. DE; FRANÇA, A. I.; SAMPAIO, T. O. DA M. DINÂMICAS DE INTERAÇÃO NATURE-NURTURE: DO IMPRINTING À RECICLAGEM NEURONAL. **ReVEL**, v. 16, n. 31, p. 38–62, 2018.

ANDRADE, I. R. **Lendo faces e palavras: uma abordagem psicolinguística. Dissertação de mestrado.** Universidade Federal do Rio de Janeiro, 2020.

BAAYEN, H.; MILIN, P.; ĐURĐEVIĆ, D.; HENDRIX, P.; MARELLI, M. An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. **Psychological Review**, v. 118, n. 3, p. 438-481, 2011.

BAAYEN, R. H.; SMOLKA, E. Modeling Morphological *Priming* in German with Naive Discriminative Learning. **Frontiers in Communication**, v. 5, 2020.

BASTOS, N.; ADAMATTI, D.; BILLA, C. Discovering Patterns in Brain Signals Using Decision Trees. **Computational Intelligence and Neuroscience**, v. 2016, p. 1–10, 1 jan. 2016.

BEVER, T; POEPEL, D. Analysis by Synthesis: A (Re-)Emerging Program of Research for Language and Vision. **Biolinguistics**, v.4, p. 174-200, 2010.

BEYERSMANN, E.; CASTLES, A.; COLTHEART, M. Early morphological decomposition during visual word recognition: Evidence from masked transposed-letter *priming*. **Psychonomic Bulletin and Review**, v. 18, n. 5, p. 937–942, out. 2011.

BORER, H. **In Name Only. Structuring Sense, Volume I.** Oxford: Oxford University Press, 2005a.

BORER, H. **The Normal Course of Events. Structuring Sense, Volume II.** Oxford: Oxford University Press, 2005b.

BOZIC, M.; MARSLER-WILSON, W.D.; STAMATAKIS, E.A.; DAVIS, M.H.; TYLER, L.K. Differentiating Morphology, Form, and Meaning: Neural Correlates of Morphological Complexity. **Journal of Cognitive Neuroscience**, v. 19, n. 9, p. 1464–1475, 2007.

BRANZI, F; MARTIN, C. D.; BIAU, E. Activating words without language: beta and theta oscillations reflect lexical access and control processes during verbal and non-verbal object recognition tasks. **Cerebral Cortex**, v. 33, n. 10, p.6228-6240, 15 maio 2023.

BRENNAN, J. R. **Language and the Brain A Slim Guide to Neurolinguistics**. 1. ed. Oxford: Oxford University Press, 2022.

BUTTERWORTH, B. Lexical Representation. In: BUTTERWORTH, B. (Ed.) **Language Production**. London: Academic Press, v. 2, p. 257-294, 1983.

BUTTERWORTH, S. On the Theory of Filter Amplifiers. **Experimental Wireless and the Wireless Engineer**, v. 7, p. 536–541, 1930.

BYBEE, J. Regular morphology and the lexicon. **Language and Cognitive Processes**, v. 10, n. 5, p. 425-455, 1995.

CABRAL, L.; SOTO, M. O processamento do particípio passado de primeira conjugação em teste de *priming* encoberto. **Veredas – Revista de Estudos Linguísticos**, v.27, n.1, 2023 (InPress).

CARAMAZZA, A.; LAUDANNA, A. & ROMANI, C. Lexical Access and Inflectional Morphology. **Cognition**, v. 28, p. 297-332, 1988.

CATALDO LOPES, J. **Acesso lexical: avaliando as especificidades das rotas semântica e morfológica**. Rio de Janeiro: Monografia de graduação, Universidade Federal do Rio de Janeiro, 2020.

CATALDO LOPES, J. C.; FRANCA, A. I. Pesquisas fundamentais em neurociência que impactaram o estado da arte da pesquisa linguística. In: **Anais do Congresso Scientiarum Historia VIII**, Rio de Janeiro, 2015.

CATALDO, J.; FRANÇA, A. Um estranho no ninho. **Cadernos de Linguística**, v. 2, n. 4, p. e499, 14 dez. 2021.

CHOMSKY, N. **Syntactic structures**. The Hague/Paris: Mouton, 1957.

\_\_\_\_\_. **Lectures on Government and Binding**. Holland: Foris Publications, 1981.

\_\_\_\_\_. **Reflections on Language**. Nova York: Pantheon, 1975.

\_\_\_\_\_. **The Minimalist Program**. Cambridge. Mass: MIT Press, 1995.

\_\_\_\_\_. **Topics in the Theory of Generative Grammar**. The Hague: Mouton, 1966.

DEHAENE, S. **Reading in the Brain: The New Science of How We Read**. New York: Viking USA, 2009. 384 p.



DELANEY-BUSCH, N.; MORGAN, E.; LAU, E.; KUPERBERG, G. Neural evidence of rational adaptation of semantic predictions: a Bayesian model of trial-by-trial N400 amplitude during semantic *priming*. **Cognition**, v. 187, p. 10-20, 2019.

DIEPENDAELE, K.; SANDRA, D.; GRAINGER, J. Masked cross-modal morphological *priming*: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. **Language and Cognitive Processes**, v. 20, n. 1–2, p. 75–114, fev. 2005.

DOMÍNGUEZ, A.; VEGA, M. DE; BARBER, H. Event-related Brain Potentials Elicited by Morphological, Homographic, Orthographic, and Semantic *Priming*. **Journal of Cognitive Neuroscience**, v. 16, n. 4, p. 598-608, 2004.

EDDY, M.D.; GRAINGER, J.; HOLCOMB, P.J.; GABRIELI, J.D.E. Orthographic and phonological processing in developing readers revealed by ERPs. **Psychophysiology**, v. 53, n. 12, p. 1776-1783, 2016.

EMBICK, D.; HACKL M.; SCHAEFFER, J.; KELEPIR, M.; MARANTZ, A. A magnetoencephalographic component whose latency reflects lexical frequency. **Cognitive Brain Research**, v. 10, n. 3, p. 345–8, 2001.

ESTIVALET, G. L.; MEUNIER, F. The Brazilian Portuguese Lexicon: An Instrument for Psycholinguistic Research. **PLOS ONE**, v. 10, n. 12, p. e0144016, 2 dez. 2015.

FELDMAN, L. B.; SOLTANO, E. G.; PASTIZZO, M. J.; FRANCIS, S. E. What do graded effects of semantic transparency reveal about morphological processing? **Brain and Language**, v. 90, n. 1-3, p. 17-30, 2004.

FELDMAN, L. B.; BARAC-CIKOJA, D.; KOSTIĆ, A. Semantic aspects of morphological processing: Transparency effects in Serbian. **Memory & Cognition**, v. 30, n. 4, p. 629–636, 2002.

FELDMAN, L. B.; KOSTIĆ, A.; GVOZDENOVIĆ, V.; O'CONNOR, P. A.; MOSCOSO DEL PRADO MARTÍN, F. Semantic similarity influences early morphological *priming* in Serbian: A challenge to form-then-meaning accounts of word recognition. **Psychon Bull Rev**, v. 19, p. 668–676, 2012.

FELDMAN, L. B.; O'CONNOR, P. A.; MOSCOSO DEL PRADO MARTÍN, F. Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. **Psychonomic Bulletin and Review**, v. 16, n. 4, p. 684–691, Agosto de 2009.

FODOR, J.D.; INOUE, A. Garden Path Repair: diagnosis and triage. **Language and Speech**, v. 43, n. 3, p. 261-271, 2000.

FORSTER, K. I. **Levels of processing and the structure of the language processor**. In W. E. Cooper & E. Walker (Eds.), Hillsdale, NJ: Erlbaum, 1979.

FRANÇA, A. I. O léxico mental em ação: muitas tarefas em poucos milissegundos. **Linguística**, v.1, n.1, p. 45-80, junho de 2005.

FRANÇA, A. I. O que é afasia? In: **O que sabemos sobre Linguagem: 51 perguntas e respostas sobre a linguagem humana**. (Orgs.) Gabriel da Ávila Othero e Valdir do Nascimento Flores. Parábola Editorial: São Paulo, v. 1. p. 15-21, 2022.

FRANÇA, A. I.; LEMLE, M.; GESUALDI, A. R.; CAGY, M.; INFANTOSI, A, F. C. A neurofisiologia do acesso lexical: palavras em português. **Revista Veredas**. Juiz de Fora, v. 2, p. 14-39, 2008.

FRAZIER, L. **On Comprehending Sentences: Syntactic Parsing Strategies**. University of Connecticut, 1979.

FRAZIER, L.; FODOR, J.D. The sausage machine: a new two-stage parsing model. **Cognition**, v. 6, p. 291-325, 1978.

FREITAS, D.R.R. DE. **Plataforma para análise do sinal de EEG aplicado ao ERD/ERS no reconhecimento em tempo real da imaginação do movimento**. Universidade Federal de Pernambuco: Tese de doutorado, 2017.

FRUCHTER, J.; MARANTZ, A. Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. **Brain and Language**, v. 143, p. 81-96, 2015.

GALLISTEL, C. R. **The physical basis of memory**. *Cognition*, v. 213, p. 104533, ago. 2021.

GARCIA, D. C. **Elementos estruturais do acesso lexical: o reconhecimento de palavras multimorfêmicas no português brasileiro**. 88 p. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

GERSHMAN, S. J.; BABLBI, P.E.M.; GALLISTEL, J.G. Reconsidering the evidence for learning in single cells. **eLife**, v. 10, 4 jan. 2021.

GODOY, M.C.; NUNES, M.A. Uma comparação entre ANOVA e modelos lineares mistos para análise de dados de tempo de resposta. **Revista da ABRALIN**, v. 19, n. 1, 2020.

GOMES, J. N. **A Direcionalidade no Relacionamento Semântico: um Estudo de ERP**. Tese de Doutorado: Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

GORRELL, P. **Syntax and Parsing**. Cambridge University Press, 1995.

GRÖNROOS, A.; VIRPIOJA, S.; KURIMO, M. Extending hybrid word-character neural machine translation with multi-task learning of morphological analysis. **Proceedings of the Second Conference on Machine Translation**, p. 296-302.

GWILLIAMS, L.; MARANTZ, A. Morphological representations are extrapolated from morpho-syntactic rules. **Neuropsychologia**, v. 114, p. 77–87, 1 jun. 2018.

HALLE, M.; MARANTZ, A. Distributed Morphology and the pieces of inflection. In: HALE, K. L.; KEYSER, S. J. (Eds.). **The view from building 20**. Cambridge: MIT Press, p. 111–176, 1993.

HARLEY, H.; NOYER, R. **Distributed Morphology**. Glot International, v. 4, n. 4, 1999.

HAY, J. B.; BAAYEN, R. H. Shifting paradigms: gradient structure in morphology. **Trends in Cognitive Sciences**, v. 9, n. 7, p. 342-348, 2005.

HEUVEN, W. J. B. VAN; MANDERA, P.; KEULEERS, E.; BRYBAERT, M. Subtlex-UK: A New and Improved Word Frequency Database for British English. **Quarterly Journal of Experimental Psychology**, v. 67, n. 6, p. 1176–1190, 1 jun. 2014.

HEYER, V.; KORNISHOVA, D. Semantic transparency affects morphological *priming* ... eventually. **Quarterly Journal of Experimental Psychology**, v. 71, n. 5, p. 1112–1124, 2018.

JACKENDOFF, R. **Semantics and Cognition**. Cambridge. Mass: MIT Press, 1983.

JARED, D.; JOURAVLEV, O.; JOANISSE, M. F. The effect of semantic transparency on the processing of morphologically derived words: Evidence from decision latencies and event-related potentials. **Journal of Experimental Psychology: Learning Memory and Cognition**, v. 43, n. 3, p. 422–450, 1 mar. 2017.

JÄRVIKIVI, J.; PYYKKÖNEN, P. Sub-and supralephical information in early phases of lexical access. **Frontiers in Psychology**, v. 2, n. OCT, 2011.

KASTNER, I.; PYLKKÄNEN, L.; MARANTZ, A. The form of morphemes: MEG evidence from masked *priming* of two Hebrew Templates. **Frontiers in Psychology**, v. 9, n. NOV, 12 nov. 2018.

KAZANINA, N.; DUKOVA-ZHELEVA, G.; GEBER, D.; KHARLAMOV, V.; TONCIULESCU, K. Decomposition into multiple morphemes during lexical access: A masked

*priming* study of Russian nouns. **Language and Cognitive Processes**, v. 23, n. 6, p. 800–823, fev. 2008.

KENEDY, E. (org). **Curso básico de linguística gerativa**. São Paulo: Contexto, 2013.

KUTAS, M. FEDERMEIER, K. D. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). **Annual Review of Psychology**, v. 62, p. 621-647, 2011.

KUTAS, M.; HILLYARD, S. A. Reading senseless sentences: Brain potentials reflect semantic incongruity. **Science**, 207(4427), 203–205, 1980.

LAGO, S.; NAMYST, A.; JÄGER, L.; LAU, E. Antecedent access mechanisms in pronoun processing: evidence from the N400. **Language, Cognition, and Neuroscience**, v. 34, n. 5, p. 641-661, 2019.

LAU, E., NAMYST, A., FOGEL, A., DELGADO, T. A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. **Collabra**, v. 2, n. 1, 2016.

LAU, E.; NAMYST, A. fMRI evidence that left posterior temporal cortex contributes to N400 effects of predictability independent of congruity. **Brain and Language**, v. 199, 2019.

LAU, E.F., ALMEIDA, D., HINES, P., POEPPPEL, D. A lexical basis for context effects: evidence from the N400. **Brain and Language**, 2009.

LAU, E.F., GRAMFORT, A., HÄMÄLÄINEN, M.S., KUPERBERG, G. Automatic semantic facilitation in anterior temporal cortex revealed through transmodal neuroimaging. **The Journal of Neuroscience**, v. 33, n. 43, p. 17174-17181, 2013.

LAU, E.F., PHILLIPS, C., POEPPPEL, D. A cortical network for semantics: (de)constructing the N400. **Nature Reviews Neuroscience**, 2008.

LAU, E.F., WEBER, K., GRAMFORT, K., HÄMÄLÄINEN, M.S., KUPERBERG, G. Spatiotemporal Signatures of Lexical-Semantic Prediction. **Cerebral Cortex**, v. 26, p. 1377-1387, 2016.

LAU, E.F.; HOLCOMB, P.J.; KUPERBERG, G.R. Dissociating N400 effects of prediction from association in single word contexts. **Journal of Cognitive Neuroscience**, 2013.

LAVRIC, A.; CLAPP, A.; RASTLE, K. ERP Evidence of Morphological Analysis from Orthography: A Masked *Priming* Study. **Journal of Cognitive Neuroscience**, v. 19, n. 5, p. 866–877, 2007.

LÁZARO, M.; GARCÍA, L.; ILLERA, V. Morpho-orthographic segmentation of opaque and transparent derived words: New evidence for Spanish. **Quarterly Journal of Experimental Psychology**, v. 74, n. 5, p. 944–954, 1 maio 2021.

LEMLE, M.; FRANÇA, A. I. Arbitrariedade Saussureana em Foco. **Revista Letras**. Curitiba, Paraná, v. 69, n. Maio/Ago, p. 269-288, 2006.

LENNEBERG, E. **Biological Foundations of Language**. New York: Wiley, 1967.

LO, S., ANDREWS, S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, v. 6, n. 1171, 2015.

LONGTIN, C. M.; SEGUI, J.; HALLÉ, P. A. Morphological *priming* without morphological relationship. **Language and Cognitive Processes**, v. 18, n. 3, p. 313–334, jun. 2003.

LUCK, S. **Introduction to the Event-Related Potential Technique**. Cambridge: MIT, 2014.

LUCK, S. Is attention filtering impaired in schizophrenia? **Schizophrenia bulletin**, v. 45, n. 5, p. 991-1000, 2019.

MAIA, M.; FRANCHETTO, B.; LEMLE, M.; VIEIRA, M.D. **Línguas indígenas e gramática universal**. São Paulo: Contexto, 2019.

MAIA, M.; LEMLE, M.; FRANÇA, A. Efeito *stroop* e rastreamento ocular no processamento de palavras. **Ciências & Cognição (UFRJ)**, v. 12, p. 02-17, 2007.

MAIA, M.; RIBEIRO, A. Jabuticaba liboramina lê mais fácil do que jornalista norbalense: um estudo de rastreamento ocular de palavras e pseudopalavras mono e polimorfêmicas. In: BUCHWEITZ, A.; MOTTA, M. (Org.). **Linguagem & Cognição**. 1ed. Porto Alegre: EDIPUC-RS, p. 143-154, 2015.

MARANTZ, A. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In: **Proceedings of the 21st Annual Penn Linguistics Colloquium. University of Pennsylvania Working Papers in Linguistics (PWPL)**, Pennsylvania: v. 4, n. 2, p. 201-225, 1997.

MARELLI, M.; AMENTA, S.; MORONE, E.A.; CREPALDI, D. Meaning is in the beholder's eye: Morpho-semantic effects in masked *priming*. **Psychonomic Bulletin and Review**, v. 20, n. 3, p. 534–541, jun. 2013.

MARR, D. **Vision: A Computational Investigation into the Human Representation and Processing of Visual Information**. New York: Freeman, 1982.

MARSLEN-WILSON, W. D.; BOZIC, M.; RANDALL, B. Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. **Language and Cognitive Processes**, v. 23, n. 3, p. 394–421, abr. 2008.

MARSLEN-WILSON, W.; TYLER, L. K.; WAKSLER, R.; OLDER, L. Morphology and Meaning in the English Mental Lexicon. **Psychological Review**, v. 101, n. 1, p. 3-33, 1994.

MCCORMICK, S.F.; BRYSSBAERT, M.; RASTLE, K. Is morphological decomposition limited to low-frequency words? **The Quarterly Journal of Experimental Psychology**, v. 62, n. 9, 2009, p. 1706-1715.

MEUNIER, F.; LONGTIN, C. M. Morphological decomposition and semantic integration in word processing. **Journal of Memory and Language**, v. 56, n. 4, p. 457–471, maio 2007.

MORRIS, J.; FRANK, T.; GRAINGER, J. HOLCOMB, P. Semantic transparency and masked morphological *priming*: An ERP investigation. **Psychophysiology**, v. 44, n. 4, p. 506–521, 2007.

MORRIS, J.; STOCKALL, L. Early, equivalent ERP masked *priming* effects for regular and irregular morphology. **Brain and Language**, v. 123, n. 2, p. 81-93, 2012.

MOTSIOU, E. Lexical Innovations (Neologisms) in Children's Speech: Evidence from Greek. **Съпоставително езикознание / Сопоставительное языкознание**, p. 22–32, 2012.

PEDERNEIRA, I. L.; LEMLE, M. Word-internal syntactic complexity. **Revista da ABRALIN**, v. 14, p. 389-410, 2015.

PEÑA, M.; MELLONI, L. Brain Oscillation during Spoken Sentence Processing. **Journal of Cognitive Neuroscience**, v. 24, n. 5, p. 1149-1164, 2012.

PINKER, S. **Rules of Language**. *Science*, v. 253, p. 530-535, 2 August 1991.

PINKER, S.; PRINCE, A. Regular and irregular morphology and the psychological status of rules of grammar. In: SUTTON, L.A.; JOHNSON, C.; SHIELDS, R. (orgs.) Annual Meeting of The Berkeley Linguistics Society. **Linguistics Society: General Session and Parasession on The Grammar of Event Structure**, p. 230-251, 1991.

PINTO, C.; VILLALVA, A.; COSTA, M.A. Priming Morfológico nos Processos de Leitura. **XXX Encontro Nacional da Associação Portuguesa de Linguística: Textos Seleccionados**, p. 113, 2014.

POIX, C. **Contrastive analysis of lexical neology in a multilingual corpus of children's literature: typology and function of occasionalisms**. Tese de Doutorado: Université Lyon Lumière, 2019.

PULVERMÜLLER, F.; ASSADOLLAHI, R.; ELBERT, T. Neuromagnetic evidence for early semantic access in word recognition. **European Journal of Neuroscience**, v. 13, n. 1, p. 201–205, 2001.

PURVES, D.; AUGUSTINE, G.; FITZPATRICK, D.; HALL, W.; LAMANTIA, A.; WHITE, L.; MOONEY, R.; PLATT, M. Memory. In: **Neuroscience**. Oxford University Press, USA, v. 6 2017.

PYLKKÄNEN, L.; GONNERMAN, L.; STRINGFELLOW, A.; MARANTZ, A. Disambiguating the source of phonological inhibition effects in lexical decision: an MEG study. **ResearchGate**, 2003.

PYLKKÄNEN, L.; FEINTUCH, S.; HOPKINS, E.; MARANTZ, A. Neural correlates of the effects of morphological family frequency and family size: an MEG study. **Cognition**, v. 91, n. 3, p. B35-B45, 2004.

PYLKKÄNEN, L. The neural basis of combinatory syntax and semantics. **Science**, v. 366, p. 62-66, 4 October, 2019.

PYLKKÄNEN, L.; MARANTZ, A. Tracking the time course of word recognition with MEG. **Trends in Cognitive Sciences**, v. 7, p.187–189, 2003.

PYLKKÄNEN, L.; STRINGFELLOW, A.; FLAGG, E.; MARANTZ, A. A neural response sensitive to repetition and phonotactic probability: MEG investigations of lexical access. **Proc. 12th Int. Conf. Biomagnetism**, Helsinki University of Technology, p. 363–367, 2000.

PYLKKÄNEN, L.; STRINGFELLOW, A.; MARANTZ, A. Neuromagnetic evidence for the timing of lexical activation: an MEG component sensitive to phonotactic probability but not to neighborhood density. **Brain and Language**, v. 81(1-3), p. 666–78, 2002.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2013, <<https://www.R-project.org/>>.

RASTLE, K.; DAVIS, M.H.; MARSLEN-WILSON, W.D.; TYLER, L.K. Morphological and semantic effects in visual word recognition: A time-course study. **Language and Cognitive Processes**, v. 15, n. 4–5, p. 507–537, 2000.

RASTLE, K.; DAVIS, M. H.; NEW, B. The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. **Psychonomic Bulletin and Review**, v. 11, n. 6, p. 1090–1098, 2004.

RIBEIRO, A. J. C. A teoria do Garden-path e evidências do português do Brasil. **Revista Prolíngua**, v. 10, n.1, jan/fev de 2015.

ROELOFS, A.; BAAYEN, H. Morphology by itself in planning the production of spoken words. **Psychonomic Bulletin and Review**, v. 9, n. 1, p. 132–138, 2002.

SAGUIE, A.; MAIA, M.; DOS SANTOS, S.L. Decisão lexical e rastreamento ocular na leitura de vocábulos com prefixos, raízes e sufixos com letras transpostas. **Signo**, v. 43, n. 77, p. 87-97, 2018.

SAMPAIO, T. O. DA M. **Coerção aspectual: uma abordagem linguística da percepção do tempo**. Rio de Janeiro: Tese de Doutorado, Universidade Federal do Rio de Janeiro, 2015.

SAUSSURE, Ferdinand de. **Curso de linguística geral**. São Paulo: Parábola, 2021. 392p.

SCHER, A. P.; BASSANI, I. S.; MINUSSI, R. D. Morfologia em Morfologia Distribuída: Morphology in Distributed Morphology. **Estudos linguísticos e literários**, n. 47, p. 9-29, 2013.

SCHREUDER, R.; BAAYEN, H.R. Modeling morphological processing. In: FELDMAN, L.B. (org.) **Morphological Aspects of Language Processing**. Lawrence Erlbaum, New Jersey: Hillsdale, p. 131-154, 1995.

SILVA, M. C. F.; MEDEIROS, A. B. **Para conhecer morfologia**. São Paulo: Contexto, 2016. 160 p.

SILVA-NASSER, C.G.A. DA. **O que tem a ver o c\* com as calças? Um estudo de ERP sobre processamento de palavras tabu, palavras neutras e palavras negativas no português brasileiro**. Dissertação de mestrado. Universidade Federal do Rio de Janeiro, 2022.

SMOLKA, E.; PRELLER, K. H.; EULITZ, C. “Verstehen” (‘understand’) *primes* “stehen” (‘stand’): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. **Journal of Memory and Language**, v. 72, n. 1, p. 16–36, abr. 2014.

SOTO, M. **ERP and fMRI evidence of compositional differences between linguistic computations for words and sentences**. Tese (doutorado) – UFRJ / Faculdade de Letras – Programa de Pós-graduação em Linguística, 2014.



SOTO, M.; GOMES, J. N.; FRANÇA, A. I.; MANHÃES, A. G. **Neurophysiology of grapheme decoding: the N170 as a predictive and descriptive tool**. *Revista da Abralín*, v. 17, n. 1, p. 402-433, 2018

STOCKALL, L.; MARANTZ, A. A single route, full decomposition model of morphological complexity. **The Mental Lexicon**, v. 1, n. 1, p. 85–123, 5 maio 2006.

TAFT, M. Lexical access via an orthographic code: the basic orthographic syllabic structure (BOSS). **Journal of Verbal Learning and Verbal Behavior**, v. 18, n. 1, p. 21-39, 1979.

TAFT, M.; FORSTER, K. I. Lexical storage and retrieval of polymorphemic and polysyllabic words. **Journal of Verbal Learning and Verbal Behavior**, v. 15, n. 6, p. 607-620, 1976.

TAFT, M.; FORSTER, K. I. Lexical storage and retrieval of prefixed words. **Journal of Verbal Learning and Verbal Behavior**, v. 14, n. 6, p. 638-647, 1975.

WERKER, J. F.; LOGAN, J. S. Cross-language evidence for three factors in speech perception. **Perception and Psychophysics**, v. 37, n. 1, p. 35–44, jan. 1985.

WERKER, J. F.; TEES, R. C. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. **Infant Behavior and Development**, v. 7, n. 1, p. 49–63, jan. 1984.

WILDER, R.J.; DAVIES, A.G.; EMBICK, D. Differences between morphological and repetition *priming* in auditory lexical decision: Implications for decompositional models. **Cortex**, v. 116, p. 122-142. 2019.

## ANEXOS

### A. Materiais experimentais

Tabela 12. Todos os materiais experimentais utilizados nos testes comportamental e neurofisiológico reportados.

#	Condition	MI1 Compositional Morphological Relationship with 1 layer	MI2 Compositional Morphological Relationship with 2 layers	MO1 Opaque Morphological Relationship with 1 layer	MO2 Opaque Morphological Relationship with 2 layers	SR1 No relationship with 1 layer	SR2 No relationship with 2 layers	Distractors <i>Primes</i> are existent words, <i>targets</i> are <b>pseudowords</b>	
	<i>Prime</i> in versions 1,3,5 <b>Target</b> in versions 2,4,6	<b>Target</b> in versions 1,3,5 <i>Prime</i> in versions 2,4,6	<b>Target</b> in versions 1,3,5 <i>Prime</i> in versions 2,4,6	<b>Target</b> in versions 1,3,5 <i>Prime</i> in versions 2,4,6	<b>Target</b> in versions 1,3,5 <i>Prime</i> in versions 2,4,6	<b>Target</b> in versions 1,3,5 <i>Prime</i> in versions 2,4,6	<b>Target</b> in versions 1,3,5 <i>Prime</i> in versions 2,4,6	<i>Prime</i> in all versions	<b>Target</b> in all versions
1	Professor (teacher)	Professor (teacher)	Professoral (professorial)	Professar (to profess)	Professado (professed)	Colcha (bedspread)	Colorido (colorful)	Acalmar (to calm down)	Vigulir
2	<i>Animal</i> (an <i>animal</i> )	<i>Animal</i> (an <i>animal</i> )	<i>Animalesco</i> ( <i>animalistic</i> )	<i>Animado</i> (lively)	<i>Animadão</i> (very lively)	Sorvete (ice cream)	Juramentado (sworn)	Término (the end)	Lebernal
3	<i>Líquido</i> (liquid)	<i>Líquido</i> (liquid)	<i>Liquidificar</i> (to liquefy)	<i>Liquidar</i> (to liquidate)	<i>Liquidação</i> (sale)	Neurônio (neuron)	Carinhoso (tender)	Bigode (mustache)	Bigueto
4	Conversa (conversation)	Conversa (a chat)	Conversação (conversation)	Conversão (conversion)	Convertido (converted)	Garrafa (bottle)	Acomodado (quiet)	Montanha (mountain)	Cavona
5	Tecer (to weave)	Tecer (to weave)	Tecido (tissue)	Texto (text)	Textual (textual)	Namoro (roman tic relationship)	Namorado (boyfried)	Jangada (raft)	Juncarda
6	Concreto (concrete)	Concreto (concrete)	Concretar (to put concrete)	Concretizar (to concretize)	Concretização (concretization)	Agrado (tip)	Nomeado (designated)	Lembrete (reminder)	Quembrete
7	Militar (military man)	Militar (military man)	Militarista (militaristic)	Milícia (militia)	Militante (a militant)	Teclado (keyboard)	Confortável (comfortable)	Corte (a cut)	Gorte
8	Correto (correct)	Correto (correct)	Corretamente (correctly)	Corretora (broker agency)	Corretagem (brokerage)	Vidro (glass)	Passividade (passivity)	Terraço (terrace)	Terrenolo

	Condition	MI1 Compositional Morphological Relationship with 1 layer	MI2 Compositional Morphological Relationship with 2 layers	MO1 Opaque Morphological Relationship with 1 layer	MO2 Opaque Morphological Relationship with 2 layers	SR1 No relationship with 1 layer	SR2 No relationship with 2 layers	Distractors <i>Primes are existent words, targets are <b>pseudowords</b></i>	
9	Livro (book)	Livro (book)	Livrinho (little book)	Livrar (to free)	Livrado (freed)	Melodia (melody)	Defendido (defended)	Vontade (will, desire)	Vordaira
10	Palavra (word)	Palavra (word)	Palavrinha (little word)	Palavrão (swearword)	Palavrório (baloney)	Câmera (camera)	Caminhada (walk)	Cadeira (chair)	Vibeção
11	Árvore (tree)	Árvore (tree)	Arvoredo (bushes)	Arvorar (to fancy)	Arvorismo (tree climbing)	Ônibus (bus)	Conduzido (conducted)	Órbita (orbit)	Orjido
12	Revista (magazine)	Revista (magazine)	Revisteiro (magazine rack)	Revisitar (to revisit)	Revisitação (revisitation)	Gelado (cold)	Conquistado (conquered)	Tarifa (fare)	Conjersa
13	Coleta (the act of collecting)	Coleta (the act of collecting)	Coletar (to collect)	Coletivo (collective)	Coletivizar (collectivize)	Natação (swimming)	Dorminhoco (sleeper)	Elefante (elefant)	Eletantanfe
14	Título (title)	Título (title)	Titulado (titled)	Titular (official)	Titularizar (to securitize)	Harmonia (harmony)	Caramelizado (caramelized)	Janela (window)	Jadela
15	Estado (state)	Estado (state)	Estatal (relative to state, state-owned)	Estante (bookcase)	Estatuto (statute)	Camisa (shirt)	Arrumada (tidy)	Escola (school)	Esvole
16	Partícula (particle)	Partícula (particle)	Particulado (particulate)	Particular (particular)	Particularmente (particularly)	Cansado (tired)	Aprendizado (learning)	Martelo (hammer)	Mareselo
17	Dívida (debt)	Dívida (debt)	Dividazinha (small debt)	Dividir (to share)	Dividendo (dividend)	Calçada (sidewalk)	Transportado (transported)	Cidade (city)	Ruvinte
18	Sequência (sequence)	Sequência (sequence)	Sequencial (sequential)	Sequela (sequelae)	Sequelado (one who has sequelae)	Felicidade (happiness)	Camaradagem (camaraderie)	Viagem (trip)	Vilgagem
19	Sereia (mermaid)	Sereia (mermaid)	Sereizinha (a little mermaid)	Sereno (serene)	Serenata (serenate)	Brócolis (broccoli)	Raspado (scraped)	Baleia (whale)	Bemeia
20	Cavalo (horse)	Cavalo (horse)	Cavalgar (to ride a horse)	Cavalete (easel)	Cavalheirismo (chivalry)	Processo (process)	Processado (processed)	Banana (banana)	Galefe
21	Certeza (certainty)	Certeza (certainty)	Certeiro (accurate)	Certificar (to certify)	Certificado (a certificate)	Estrela (star)	Arranjado (arranged)	Pureza (purity)	Asvoeda
22	Cinema (cinema)	Cinema (cinema)	Cinematográfico (cinematographer)	Cinético (kinetic)	Cinemático (kinematic)	Buraco (hole)	Guardada (kept)	Bocado (a little)	Limenga

	Condition	MI1 Compositional Morphological Relationship with 1 layer	MI2 Compositional Morphological Relationship with 2 layers	MO1 Opaque Morphological Relationship with 1 layer	MO2 Opaque Morphological Relationship with 2 layers	SR1 No relationship with 1 layer	SR2 No relationship with 2 layers	Distractors <i>Primes</i> are existent words, <i>targets</i> are <b>pseudowords</b>	
23	Resíduo (residue)	Resíduo (residue)	Residual (residual)	Residir (to reside)	Residente (resident)	Paletó (suit)	Amabilidade (friendliness)	Aquário (tank)	Ocrela
24	Restauo (restoration)	Restauo (restoration)	<i>Restauração</i> (the restoration)	<i>Restaurar</i> (to restore)	<i>Restaurante</i> (restaurant)	Televisão (television)	Televisionar (televise)	Pedido (order)	Vafento
25	Edifício (a building)	Edifício (a building)	Edificação (a construction)	Edificar (to edify)	Edificante (edifying)	Comida (food)	Comilança (gluttony)	Cachorro (dog)	Cachego
26	Legenda (subtitle)	Legenda (subtitle)	Legendado (subtitled)	Legendário (legendary)	Legendariamen te (legendarily)	Planta (plant)	Jardinagem (gardening)	Jasmin (jasmine)	Jasmana
27	Abismo (abyss)	Abismo (abyss)	Abismal (abysmal)	Abismado (astonished)	Abismadíssimo (very astonished)	Festa (party)	Capacidade (capacity)	Marca (brand)	Celtigem
28	Legal (cool, legal)	Legal (cool, legal)	Legalizar (to legalize)	Legado (legacy)	Legatário (legatee)	Campo (field)	Preenchido (filled)	Fungo (fungus)	Fulgem
29	Lixo (trash)	Lixo (trash)	Lixeira (trashbin)	Lixar (to file)	Lixado (filed)	Desenho (a drawing)	Desenhado (drawn)	Coração (heart)	Júntilo
30	Órgão (organ)	Órgão (organ)	Orgânico (organic)	Organizar (to organize)	Organização (organization)	Praia (beach)	Cabeçalho (header)	Valente (intrepid)	Penduba
31	Cronos (chronos)	Cronos (chronos)	Crônico (chronic)	Cronista (chronicler)	Cronístico (writer)	Mochila (backpack)	Verificação (verification)	Correria (rush)	Corruroa
32	Imagem (image)	Imagem (image)	Imagético (imagetic)	Imaginar (to imagine)	Imaginação (imagination)	Luva (glove)	Perdido (lost)	Ventilador (fan)	Vantulidor
33	Autor (author)	Autor (author)	Atoral (authorship)	Autorizar (to authorize)	Autoridade (authority)	Plástico (plastic)	Camuflagem (camouflage)	Computador (computer)	Compatesor
34	Criar (to create)	Criar (to create)	Criatividade (creativity)	Criança (child)	Criançada (kids)	Morango (strawberry)	Fechado (closed)	Variiedade (variety)	Quampo
35	Oriente (east)	Oriente (east)	Oriental (eastern)	Orientar (to guide)	Orientação (guidance)	Lazer (leisure)	Ajustado (fitted)	Carona (ride)	Medeiar
36	Carta (letter)	Carta (letter)	Carteiro (mailman)	Carteira (wallet)	Carteirinha (id card)	Servir (to serve, to fit)	Perecível (perishable)	Retrato (portrait)	Ravrato

	Condition	MI1 Compositional Morphological Relationship with 1 layer	MI2 Compositional Morphological Relationship with 2 layers	MO1 Opaque Morphological Relationship with 1 layer	MO2 Opaque Morphological Relationship with 2 layers	SR1 No relationship with 1 layer	SR2 No relationship with 2 layers	Distractors <i>Primes are existent words, targets are <b>pseudowords</b></i>	
37	Secreto (secret – adj)	Secreto (secret -adj)	Secretamente (secretly)	Secretário (secretary)	Secretariado (secretariat)	Parecer (evaluation)	Complexado (insecure)	Contínuo (continuous)	Girodaga
38	Prato (dish)	Prato (plate)	Pratão (big plate)	Prateleira (shelf)	Prateleirinha (small shelf)	Cadeado (locker)	Confiabilidade (reliability)	Vaga (opening)	Vaguel
39	Arma (weapon)	Arma (weapon)	Armado (armed)	Armário (cupboard)	Armariozinho (little cupboard)	Papel (paper)	Narigudo (nosey)	Tombo (fall)	Penilar
40	Acesso (access)	Acesso (access)	Acessar (to access)	Assessorar (to help)	Assessorado (helped)	Casaco (coat)	Cozinhado (cooked)	Piano (piano)	Pienel
41	Cândido (candid)	Cândido (candid)	Candidamente (candidly)	Candidato (candidate)	Candidatar (to apply for)	Madeira (wood)	Responsabilidad e (responsibility)	Ginástica (gym)	Humalte
42	Funcionar (to function)	Funcionar (to function)	Funcionalidade (functionality)	Funcionário (employee)	Funcionariozinho (small employee)	Postal (postal)	Comprado (bought)	Rude (rude)	Cênilos
43	Casa (house)	Casa (house)	Casebre (shack)	Casar (to marry)	Casamento (marriage)	Vareta (stick)	Pilotado (piloted)	Visão (vision)	Clisão
44	Bola (ball)	Bola (ball)	Bolinha (little ball)	Bolar (to plan)	Bolado (planned)	Caderno (notebook)	Faltoso (faulty)	Vulcão (vulcano)	Fênela
45	Exercer (to perform)	Exercer (to perform)	Exercido (performed)	Exército (army)	Exercitado (exercised)	Taça (coupe)	Iluminação (lighting)	Oceano (ocean)	Estonje
46	Barro (mud)	Barro (mud)	Barrento (muddy)	Barraca (tent)	Barracão (shed)	Fone (earbuds)	Contratado (hired)	Vantagem (advantage)	Ventifagem
47	Mole (soft)	Mole (soft)	Molenga (softy)	Molhar (to wet)	Molhado (wet)	Canto (corner)	Cantado (sung)	Roupa (clothing)	Dasca
48	Circo (circus)	Circo (circus)	Circense (from the circus)	Círculo (circle)	Circular (circular)	Travesseiro (pillow)	Juizado (juvenile or small claims or light offense court)	Lustre (chandelier)	Assumpa