



UNIVERSIDADE ESTADUAL DE CAMPINAS
Instituto de Estudos da Linguagem

FERNANDO LUIS SABATINI

**ACESSO LEXICAL: ATIVAÇÃO E INIBIÇÃO DE PALAVRAS EMBUTIDAS EM
POSIÇÃO FINAL**

Campinas
2024

FERNANDO LUIS SABATINI
ACESSO LEXICAL: ATIVAÇÃO E INIBIÇÃO DE PALAVRAS EMBUTIDAS EM
POSIÇÃO FINAL

Dissertação apresentada ao Instituto de Estudos da Linguagem (IEL) da Universidade Estadual de Campinas (Unicamp) como parte dos requisitos exigidos para obtenção do título de Mestre em Linguística.

Orientador: Thiago Oliveira da Motta Sampaio.

ESTE TRABALHO CORRESPONDE À
VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELO ALUNO
FERNANDO LUIS SABATINI
ORIENTADO PELO PROFESSOR
DOUTOR THIAGO OLIVEIRA DA
MOTTA SAMPAIO

Campinas

2024

Ficha catalográfica

Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Estudos da Linguagem
Ana Lucia Siqueira Silva - CRB 8/7956

Sabatini, Fernando Luis, 1990-
Sa47a Acesso lexical : ativação e inibição de palavras embutidas em posição final
 / Fernando Luis Sabatini. – Campinas, SP : [s.n.], 2024.

Orientador: Thiago Oliveira da Motta Sampaio.
Dissertação (mestrado) – Universidade Estadual de Campinas
(UNICAMP), Instituto de Estudos da Linguagem.

1. Linguística. 2. Psicolinguística. 3. Cognição. 4. Linguagem. I. Sampaio,
Thiago Oliveira da Motta, 1984-. II. Universidade Estadual de Campinas
(UNICAMP). Instituto de Estudos da Linguagem. III. Título.

Informações complementares

Título em outro idioma: Lexical access : activation and deactivation of final
embedded
words

Palavras-chave em inglês:

Linguistics
Psycholinguistics
Cognition
Language

Área de concentração: Linguística

Titulação: Mestre em Linguística

Banca examinadora:

Thiago Oliveira da Motta Sampaio [Orientador]
Marije Soto
Mahayana Cristina Godoy

Data de defesa: 17-12-2024

Programa de Pós-Graduação: Linguística

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0001-8533-375X>
- Currículo Lattes do autor: <http://lattes.cnpq.br/1533749356421694>



BANCA EXAMINADORA:

Thiago Oliveira da Motta Sampaio

Marije Soto

Mahayana Cristina Godoy

IEL/UNICAMP

2024

Ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós Graduação do IEL.

AGRADECIMENTOS

Este trabalho foi concebido e desenvolvido, em boa parte, enquanto atravessávamos a pandemia da COVID-19, e, por isso, foi profundamente condicionado pelo contexto social e político desse período. Enfim, chego à sua conclusão, muito graças ao privilégio de ter seguido vivo, com saúde e condições materiais de realizar as atividades que embasam essa dissertação, mas também graças ao apoio, à colaboração e a toda ajuda de colegas de trabalho, amigos e familiares — e, sobremaneira, apesar dos esforços do então governo de minar as condições de vida do povo trabalhador e deixar morrer parte dele, no que, em centenas de milhares de casos, infelizmente teve êxito. Por isso, agradeço aos trabalhadores da Unicamp, em especial os da área da Saúde — que, atravessando esse momento tão violento da nossa história, salvaram tantas vidas, muitas vezes ao custo da própria — e os do Instituto de Estudos da Linguagem (IEL), que proporcionaram as condições necessárias para a realização deste trabalho e das atividades do instituto como um todo. A vocês, sou profundamente grato.

Em especial, agradeço à minha esposa e eterna companheira, Sara, que me incentivou, motivou, acompanhou, ensinou e inspirou em cada passo. Obrigado pela Maya, esse universo que você gestou com tanto amor e que agora cresce com a gente. Obrigado por nós três, por todas as nossas vidas juntos, pela nossa família, pelos nossos sonhos compartilhados. Pela antropologia, pelos livros, pela terceira margem do rio. Por dar vida à nossa casa. Obrigado por ser você, exatamente como você é. E sendo, entre outras coisas, Doutora Sara, obrigado por mostrar que é possível transformar o mundo.

À Maya. Por cada sorriso, cada passo, cada choro, cada abraço. Obrigado por ser possível. Obrigado por ser.

Aos meus pais, Ednéia e Varley, e a toda a minha família: Gilberto, Tereza, Marley, Felipe, Suzana, Vitor, Gabriel, Mariana, Anna, Clara, Valentim, Catarina, Gabriela, Eliane, Rolando, Leo, Gi, Beto, Val, Henrique, Wesley, Carol, Helô. Obrigado pela vida, pelo amor e pelo apoio incondicional. Obrigado pela paciência com a minha ausência. À tia Maurícia (in memoriam), cujo amor nos acompanha todos os dias. A Nilson, Marcia e toda a família — em especial, Caio, Tamires, Nicolas, Ubaldo e Elsa (in memoriam) —, pelo apoio e pelo afeto incondicionais. Amo todos vocês.

Aos professores do IEL, pelo papel que desempenharam na minha formação e por toda a ajuda durante esses 10 (!) anos de Unicamp.

Às professoras Marije Soto e Mahayana Godoy, que compuseram a banca de defesa e contribuíram muito não só com este trabalho, mas com minha formação como pesquisador.

Ao professor Alex de Carvalho, da Université de Paris/Sorbonne, pela disponibilidade, pela recepção e pelas contribuições valiosas para a minha pesquisa.

Ao meu orientador e amigo, Thiago Motta, por tantos anos de trabalho, atenção e companheirismo, por me incentivar, por ter paciência para discutir cada uma das minhas ideias mirabolantes e por tantas vezes me ajudar a dosar as expectativas para tentar dar um passo mais ou menos do tamanho da perna. Obrigado pelos áudios de WhatsApp (os curtos e os longos), pelos cafés, pelos almoços, por encorajar minha escrita. Obrigado por estar disponível — e obrigado por não estar quando era preciso cuidar da sua saúde! Obrigado por acreditar em mim. Obrigado por me ensinar. Em tempo, obrigado por me apresentar tantas pessoas maravilhosas, algumas das quais eu menciono aqui.

À Josie Siman, pelas conversas, pela disponibilidade, por tanta ajuda, pela excelência em tudo o que faz e por me apresentar as palavras dentro de palavras.

Aos colegas e amigos da UFRJ, que me receberam com tanta hospitalidade no laboratório, na Faculdade de Letras e no Rio de Janeiro — em especial, Aniela e Marije, pela recepção, por tanto aprendizado que me proporcionaram e por serem inspiração.

À Julia Cataldo, mestrona e amiga: obrigado por me acolher no Rio, pelo companheirismo nessas estradas árduas do acesso lexical, do mestrado, do InCognitus. Por tanto aprendizado e tantas trocas, pelo karaokê, pelo Fla-Flu, por me apresentar sua família incrível — e obrigado a eles pelo carinho! —, pelo seu trabalho incrível, por ser incrível. Voe alto! Maya e Zezinho aguardam sua próxima visita.

A todos os membros, atuais e antigos, do InCognitus e do LAPROS, a quem não sou capaz de agradecer o suficiente por cada momento compartilhado e pelo privilégio de crescer com vocês. Ainda assim, faço algumas menções específicas e, às vezes, duplicadas.

À Maria Eugênia, grande amiga, primeira professora da Maya, linguista de verdade, Guardiã do 2 de Setembro; pela amizade, por manter vivo o InCognitus, pelos chafés que foram e os que estão por vir. Por tanto capricho em simplesmente tudo o que faz. Por me apresentar o Daniel. Por compartilhar histórias. Pelo Inner Vision, essa preciosidade. Elen síla lúmenn' omentielvo.

À Alice, pela ajuda valiosa com a análise estatística, com minhas apresentações, com o planejamento e a rotina de trabalho. Por compartilhar o sentimento de que a beleza do léxico está na semântica. Por me ensinar palavras que não existem. *bafo*. Pelos cafés em copo retornável. Pela amizade inesperada e sincera. Merci!

À Natalia Araboni, amiga, semanticista, apreciadora da matematização do sujeito; por compartilhar a jornada desses últimos anos, pelo seu belíssimo conjunto de xícaras de café e por ter um coração maior que Minas Gerais. Obrigado pelo privilégio da sua amizade.

À Larissa Chiodetto, por todas as conversas e risadas, pela sinceridade e por toda a ajuda durante a fase final da pesquisa, sobretudo com a análise estatística.

Aos amigos da Liga das Humanas da Unicamp (LHU), em especial: Edson, Romano, Pedrosa, Caio, Renan e Pedro Zan. A lista é bem maior; se ela se resumisse apenas a vocês, eu já seria uma pessoa de muita sorte.

Aos amigos: Renato Lopes, Matheus Gustavo e Felipe Chow. Muito obrigado por tudo, principalmente pelos conselhos e pela paciência.

À Fernanda Luz, amiga querida, pelo companheirismo nesses últimos anos, por tantos conselhos e tantos cafés, mesmo quando mal tínhamos tempo de conversar entre as reuniões. Estou esperando o próximo.

To the Fellowship of Umeå: Ann, Inês, Meis, Pedro, Pim, Silke, Niklas, and Seyda, my dear friends!, for you were light in (literally) dark places. Ann, Silke: thank you so much for Berlin!

To Henrik, my friend, my brother, wherever you are in the forest. Tack så mycket.

A todas as pessoas que deram suporte, cuidado e afeto à Sara, à Maya e ao Zezinho durante as semanas em que estive viajando e concluindo a dissertação: Ednéia, Marcia, Maria Eugênia, Rafaele, Ana, Sandra, José e Luiza. Faltam palavras pra descrever o quanto sou grato a vocês. Muito obrigado!

Por fim, a todas as pessoas que aceitaram o convite e participaram do experimento. Muito obrigado!

Esta pesquisa foi realizada com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil (processo nº 2021/00377-3) e, em sua fase final, da Unicamp, via Funcamp (Auxílio Ponte FAEPEX: 3445/24). As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do autor e não necessariamente refletem a visão da FAPESP.

Le Brésil s'esquissait dans mon imagination comme des gerbes de palmiers contournés, dissimulant des architectures bizarres, le tout baigné dans une odeur de cassolette, détail olfactif introduit subrepticement, semble-t-il, par l'homophonie inconsciemment perçue des mots « Brésil » et « grésiller », mais qui, plus que toute expérience acquise, explique qu'aujourd'hui encore je pense d'abord au Brésil comme à un parfum brûlé.

Levi-Strauss. *Tristes Tropiques*, 1955.

(...) there is nothing in the physical world by itself that would predict a priori the response of the brain to a stimulus. It is often largely the state of the brain that determines the behavioral outcome.

György Buzsáki. *Rhythms of the Brain*, 2006.

RESUMO

Na Psicolinguística, os estudos em torno do acesso lexical buscam explicar a ativação de palavras na memória durante o processamento de linguagem. O termo “palavra embutida” ou “palavra dentro de palavra” se refere a um segmento espúrio que constitui parte de uma palavra maior e que coincide com a forma — fonológica ou ortográfica — de outra palavra (e.g. ‘fé’ em *café*). Esse epifenômeno inspira questões linguísticas relevantes sobre, por exemplo, como ocorre a segmentação da fala contínua durante a interpretação de uma sentença; além disso, nos permite testar a eficiência de modelos de percepção/reconhecimento da fala, como o *TRACE* (McClelland; Elman, 1986) e as diferentes versões do *Shortlist* (Norris, 1994; Norris; McQueen; Cutler, 1995) e do *Cohort* (Marslen-Wilson; Welsh, 1978; Marslen-Wilson, 1987, 1993). Afinal, palavras embutidas são ativadas no léxico mental quando ouvidas/lidas? Elas entram na concorrência pelo acesso lexical junto à portadora, que por hipótese é a mais adequada para vencer a concorrência e ser integrada à representação mental da interpretação? Na literatura, os resultados são inconsistentes: são reportados efeitos de *priming* facilitador em ao menos algumas situações (Shillcock, 1990; Vroomen; De Gelder, 1997; Luce; Cluff, 1998; Isel; Bacri, 1999; Shatzman, 2006; Bowers *et al.*, 2009; Van Alphen; Van Berkum, 2010, 2012; Zhang; Samuel, 2015), inibidor em outras (Shatzman, 2006; Marslen-Wilson *et al.*, 1994), e até mesmo ausência de efeito (Gow; Gordon, 1995; Macizo; Van Petten; O’rouke, 2012; Marslen-Wilson *et al.*, 1994; Norris *et al.*, 2006). Esta dissertação detalha um estudo experimental de *priming* transmodal em Português Brasileiro com *prime* auditivo (sentença) e *target* visual (palavra), cuja hipótese é a de que o efeito de *priming* — interpretado como ativação das palavras embutidas durante o processamento — pode ser observado quando o *Stimulus Onset Asynchrony* (SOA) é igual a 0 ms, enquanto um SOA de 300 ms não gera *priming*, por haver tempo suficiente para que a palavra embutida, mesmo que ativada, já tenha sido inibida. Os resultados não endossam a hipótese; ainda assim, eu argumento que parte dos dados fornece indícios de ativação de palavras embutidas finais e discuto aspectos metodológicos, como o controle da relação entre *prime* e *target*, que podem ser aprimorados para estudos posteriores.

Palavras-chave: *embedded words*; palavras embutidas; acesso lexical; *crossmodal priming*; processamento de linguagem.

ABSTRACT

In psycholinguistics, studies on lexical access aim at explaining the activation of words in memory during language processing. The term “embedded word” or “words within words” refers to a spurious segment that is part of a larger word and coincides with the phonological or orthographic form of another word, e.g. 'fé' ('faith') in *café* (coffee). The epiphenomenon inspires relevant linguistic questions, such as how segmentation of continuous speech occurs during the interpretation of a sentence; moreover, it allows us to test the efficiency of word processing models, such as TRACE (McClelland; Elman, 1986) and the different versions of Shortlist (Norris, 1994; Norris; McQueen; Cutler, 1995) and Cohort (Marslen-Wilson; Welsh, 1978; Marslen-Wilson, 1987, 1993). After all, are embedded words activated in the mental lexicon when heard/read? Do they compete for lexical access with the carrier, which, by hypothesis, is the most suitable to win the competition and be integrated into the mental representation of the interpretation? Previous results are inconsistent: reports include facilitatory priming effects in at least some situations (Shillcock, 1990; Vroomen; De Gelder, 1997; Luce; Cluff, 1998; Isel; Bacri, 1999; Shatzman, 2006; Bowers et al, 2009; Van Alphen; Van Berkum, 2010, 2012; Zhang; Samuel, 2015), inhibitory in others (Shatzman, 2006; Marslen-Wilson et al., 1994), and even absence of effect (Gow; Gordon, 1995; Macizo; Van Petten; O'rouke, 2012; Marslen-Wilson et al., 1994; Norris et al., 2006). This dissertation details an experimental study with crossmodal priming in Brazilian Portuguese with auditory prime (sentence) and visual target (word), whose hypothesis is that the priming effect - interpreted as activation of the embedded words during processing - can be observed when the Stimulus Onset Asynchrony (SOA) is equal to 0 ms, while a 300 ms SOA does not elicitate priming, since there is enough time for the embedded word, even if activated, to have already been inhibited. The results do not endorse the hypothesis; nevertheless, I argue that part of the data provides evidence of activation of final embedded words and discuss methodological aspects, such as the control of the prime-target relationship, which can be improved for further studies.

Keywords: embedded words; lexical access; crossmodal priming; language processing.

SUMÁRIO

1 INTRODUÇÃO.....	13
2 PALAVRAS EMBUTIDAS.....	16
2.1 DEFINIÇÕES DO TERMO E DESCRIÇÕES DO FENÔMENO.....	17
2.2 ESCRITA E ORALIDADE: A QUESTÃO FONOLÓGICA.....	20
3 ACESSO LEXICAL E PROCESSAMENTO DE PALAVRAS.....	24
3.1 MEMÓRIA E PALAVRA.....	24
3.2 ALGUNS MODELOS DE RECONHECIMENTO DE PALAVRAS FALADAS.....	27
3.3 ATIVAÇÃO LEXICAL DE PALAVRAS EMBUTIDAS FINAIS.....	29
4 O EXPERIMENTO.....	44
4.1 HIPÓTESE E EXPECTATIVAS.....	44
4.2 MÉTODO.....	45
4.2.1 PARTICIPANTES.....	45
4.2.2 MATERIAIS.....	45
4.3 RESULTADOS.....	51
5 DISCUSSÃO GERAL.....	55
6 CONSIDERAÇÕES FINAIS.....	58
REFERÊNCIAS.....	61
APÊNDICE A – LISTA DE ESTÍMULOS.....	65
APÊNDICE B – ANÁLISE DO PRÉ-TESTE DE CLOZE.....	76
ANEXO 1: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – CLOZE.....	77
ANEXO 2: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – EXPERIMENTO PRINCIPAL.....	79

1 INTRODUÇÃO

Do ponto de vista fonológico, as línguas faladas são compostas por um certo número de elementos básicos, os fonemas. Adotando uma perspectiva computacional, podemos definir a língua como um sistema capaz de gerar, a partir da aplicação de determinadas regras a um conjunto finito de elementos¹, um número potencialmente infinito de expressões (sentenças). Deixando de lado implicações mais complexas desse tipo de abordagem, qualquer sistema que combine um número finito de elementos para formar outros elementos implica necessariamente a aparição repetida dos mesmos elementos em diferentes situações/contextos; conseqüentemente, é provável que certas combinações de dois ou mais elementos — neste caso, fonemas — se repitam ao acaso.

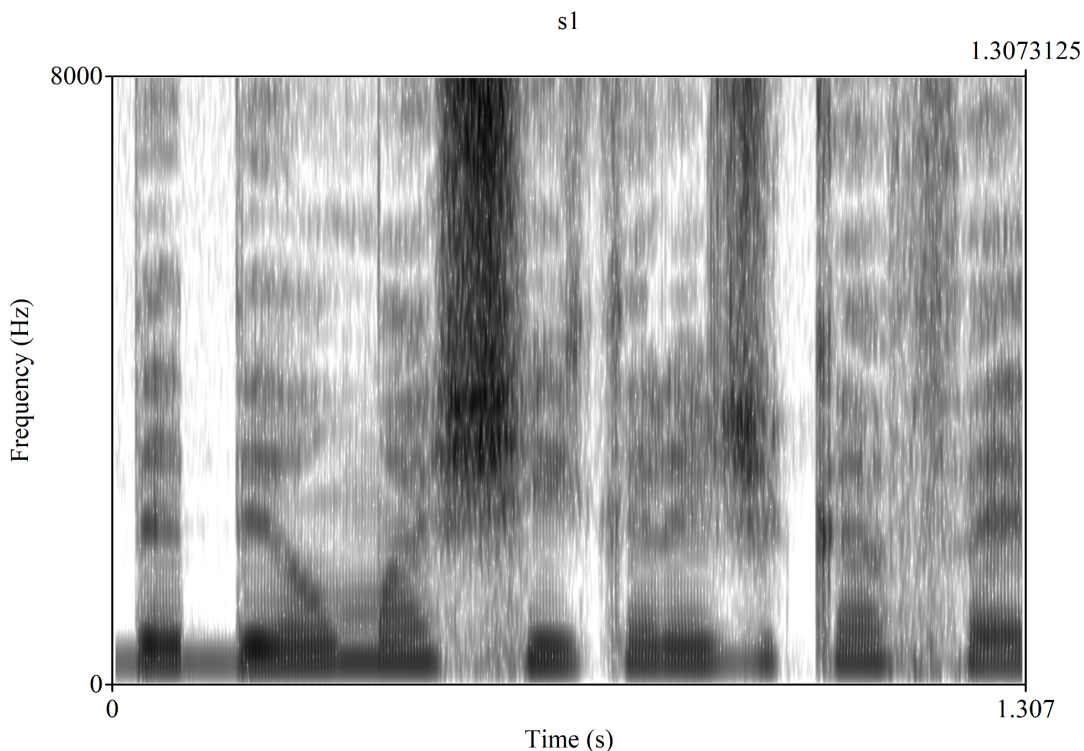
Vamos tomar como exemplo a língua escrita e, para fim de simplificação, adotar como elementos básicos as letras. Seguindo o raciocínio, podemos formar um grande número de palavras e, hipoteticamente, um número infinito de sentenças, a partir de um número finito de letras. Note, por exemplo, que a combinação de ‘a’ e ‘r’ é suficiente — e necessária — para formarmos a palavra ‘ar’. Contudo, a arbitrariedade no uso dos mesmos elementos dessa língua pode gerar uma palavra do tipo ‘luar’, assim como de um bom número de verbos no infinitivo (e.g. ‘casar’, ‘morar’, ‘pintar’) e diversas outras palavras (e.g. ‘armário’, ‘amargo’, etc.) que trazem embutido o mesmo segmento ‘ar’.

A diferença desse segmento espúrio em relação a uma raiz é que não há qualquer relação morfológica entre a palavra embutida e sua portadora — ou seja, o conteúdo semântico do substantivo *ar* não desempenha qualquer papel na formação de *luar*, tampouco na derivação das formas infinitivas dos verbos acima. Dito de outra forma, estamos diante de uma aparente coincidência de formas, o que, como apontei anteriormente, é previsível, dada a finitude dos elementos do sistema. Somado a isso, considere que a fala contínua não traz pistas fonológicas inequívocas, como os espaços em branco

¹ Na modalidade oral, fonemas; na escrita, letras. Essa distinção tem implicações importantes para nossa discussão.

na escrita, para determinar o momento específico no qual tem início uma palavra (Shillcock, 1990, p. 24). Segue um exemplo:

Figura 1 – Espectrograma de trecho de uma sentença



Fonte: Software PRAAT (Boersma; Weenink, 2025), elaborada pelo autor. A figura mostra o espectrograma do trecho “bebeu uma xícara de café”, de uma das sentenças do experimento descrito neste trabalho a fim de ilustrar a continuidade do sinal acústico — que, por sua vez, demonstra que um parser de processamento linguístico eficiente deve ser capaz de segmentar a sentença nas partes que a compõem, seja identificando o início (*onset*) das palavras, seja adotando outra estratégia.

Aqui, vale destacar que o termo “palavra embutida” pode gerar uma interpretação equivocada, já que o fenômeno que abordarei aqui não constitui exatamente uma *palavra*: a coincidência está **na forma** — seja ela escrita, seja falada². Para que ‘ar’ em ‘armário’ caracterizasse uma palavra, seria preciso assumir que o segmento carregasse consigo, **nesse contexto interno à palavra portadora**, seu próprio significado, o que é falso. Como detalharei, é uma coincidência exclusivamente da forma — e nem sempre com tanta fidelidade, como fica visível se o segmento (sílabas) for escrutinado sob as lentes criteriosas de uma análise prosódica. É seguro dizer que o ‘ar’ de

² Não abordarei neste trabalho eventuais palavras embutidas em línguas sinalizadas.

‘armário’ não se refere à substância gasosa contida na mobília vazia ou na sua composição material.

Na psicolinguística, esse fenômeno suscita perguntas interessantes sobre o processamento de linguagem e dados experimentais contribuem para a discussão sobre a eficiência de modelos teóricos. Afinal,

- ‘fé’ é ativada no léxico mental quando ouvimos a palavra ‘café’?
- qual é a dinâmica de competição pelo acesso lexical?
- palavras embutidas no final de outras palavras também são ativadas durante a competição?
- fatores extralexicais (*top-down*), como efeitos de contexto sentencial, podem impor restrições a essa ativação?
- e fatores cognitivos, como a memória de trabalho?
- como é a dinâmica de segmentação de palavras na fala contínua? Esse processo ocorre antes da identificação das palavras? Paralelamente?

De fato, essas e outras perguntas vêm sendo discutidas a partir de estudos experimentais há mais de 30 anos. A ativação de ‘fé’ em ‘café’, por exemplo, implicaria que o reconhecimento de palavras envolve múltiplas ativações a partir de diversos pontos do estímulo fonológico, não só do *onset*. Do ponto de vista do processamento linguístico, isso implica uma estratégia que não necessariamente segmenta a sentença em palavras antes de iniciar a busca pelo melhor *match* entre o estímulo percebido e a representação mental da palavra. Ainda assim, a literatura apresenta resultados inconsistentes entre si — mas interpretá-los como contraditórios seria equivocado; afinal, temos uma produção científica com línguas variadas, métodos distintos e objetivos também diversos, fatores esses que condicionam os resultados.

Tendo isso em vista, busco neste trabalho desembaraçar parte dessa questão. Assim, a seção 2, a seguir, traz uma definição do fenômeno e distinções relevantes entre a ocorrência na forma oral e a escrita. A seção 3 busca situar o estudo no campo do acesso lexical ao resgatar brevemente a discussão atual sobre a natureza e os elementos básicos do léxico mental, descrever alguns modelos psicolinguísticos de reconhecimento de palavras

faladas (*spoken word recognition*³), apresentar as implicações e contribuições dos estudos experimentais sobre palavras embutidas nesse contexto e, finalmente, detalhar a inconsistência na literatura e uma explicação possível para dados tão diversos — aproveitando para listar meus objetivos específicos.

A seção 4 traz informações detalhadas do experimento e os resultados registrados; a seção 5 traz uma discussão geral, tendo em vista que a hipótese não foi confirmada, e a seção 6 fecha o trabalho com algumas ressalvas sobre fatores que limitaram o design experimental e comentários sobre caminhos que ficam em aberto para investigação futura.

2 PALAVRAS EMBUTIDAS

No campo da psicolinguística, o fenômeno das “palavras embutidas” (doravante PEs) recebeu ainda ao menos um outro nome: “palavras dentro de palavras”⁴ (Salgado, 2014). Acontece que, intencionalmente inseridas ou não, palavras menores podem ser encontradas *dentro* de outras palavras. Nesse sentido, seja por introspecção, seja por um estudo etimológico/diacrônico de uma palavra, é possível afirmar que não há qualquer relação semântica ou associativa⁵ que possa atribuir ao substantivo ‘ar’ uma participação na formação da palavra ‘luar’ — ao contrário, por exemplo, do que observamos entre raízes e suas derivações, ou mesmo em composições (e.g. ‘chuva’ em ‘guarda-chuva’). Afinal, como questiona Salgado (2014) no título de sua dissertação de mestrado, *Existe fim em marfim?*⁶. A pergunta, de fato, suscita reflexões interessantes. Antes de trazê-las à discussão, entretanto, quero expandir brevemente minhas considerações sobre a terminologia adotada, para situar o presente trabalho no cenário da psicolinguística, e, depois,

³ O termo costuma ser usado de forma intercambiável com outro, *speech perception*, ainda que o escopo dos modelos de processamento seja maior, abarcando níveis de análise que vão além do reconhecimento de palavra.

⁴ *Words within words* para Van Alphen e Van Berkum (2010), entre outros; *embedded words* para Vroomen e De Gelder (1997), Shatzman (2006), Zhang e Samuel (2015), entre outros. Opto pelo segundo termo por considerar o primeiro intrinsecamente contraditório.

⁵ Ao longo deste trabalho, adotarei a noção de “associação” de Fodor (1983), aplicável à relação entre significados que costumam ocorrer em um mesmo contexto (e.g. ‘médico’ e ‘hospital’), diferentemente das relações propriamente semânticas, que envolvem uma mesma raiz (e.g. ‘medicamento’ e ‘medicar’).

⁶ O título completo é “*Existe fim em marfim? Um estudo da relação de palavras dentro de palavras no acesso lexical*”.

elencar algumas questões pertinentes que vêm sendo investigadas a partir das tais *palavras embutidas*.

O ponto é que a aparente simplicidade do fenômeno pode mascarar a riqueza de suas implicações para os estudos linguísticos. Em paralelo, parece nutrir descrições polissêmicas, talvez até ambíguas ou equivocadas, na literatura. Afinal, é adequado afirmar que estamos diante de palavras *dentro* de outras palavras? Primeiro, vejamos como o termo vem sendo abordado na literatura. Partirei desse ponto para expor algumas definições que assumo neste trabalho.

2.1 DEFINIÇÕES DO TERMO E DESCRIÇÕES DO FENÔMENO

Em alguns trabalhos experimentais, o termo “palavras embutidas” (tradução minha) se refere a segmentos que podem ser encontrados dentro de palavras maiores e que coincidem com a forma de uma outra palavra⁷ (e.g. ‘fé’ em *café*; cf. Shillcock, 1990; Isel; Bacri, 1999; Van Alphen; Van Berkum, 2010; entre outros). No entanto, um segmento desse tipo é espúrio — ou seja, não tem significado pretendido pelo falante. Não entrarei aqui em uma discussão mais aprofundada sobre a definição do conceito de palavra, mas é necessário destacar que a leitura acima implica chamar de *palavra* embutida apenas a contraparte relativa à *forma* (ortográfica ou acústica) do fenômeno, ignorando o *lado semântico* da moeda saussuriana. Como apontei anteriormente, considero essa interpretação equivocada. Assim, retomando o exemplo acima, descarto a possibilidade de que a palavra que se refere ao sistema de crença — isto é, *fé* — cumpra alguma função semântica na composição do significado da sua portadora, *café*, seja a fruta, o estabelecimento comercial, a bebida etc.; objetivamente, ‘fé’ em *café* é pura coincidência.

Uma palavra embutida (doravante PE), neste trabalho, é estritamente um epifenômeno. Por isso, daqui em diante, utilizarei o termo apenas como referência ao **segmento**, escrito ou falado, contido em uma palavra — essa sim, com significado eliciado intencionalmente — e que

⁷ Note que eu diferencio o **segmento** ‘fé’ da **palavra** *fé*, considerando-o apenas uma coincidência da forma (escrita ou falada). Por isso, a partir deste ponto do texto, adotarei aspas simples quando me referir ao segmento e itálico para me referir à palavra.

coincide com a forma ortográfica e/ou fonológica de outra palavra. O segmento espúrio, portanto, não contribui com a composição semântica do significado da portadora, seja na sua raiz, seja como flexão, derivação ou outro processo morfológico. É coincidência de formas, apenas.

Pode-se argumentar que a discussão terminológica seja apenas um pormenor, que o fenômeno em questão é óbvio: se a palavra maior (doravante *portadora*)⁸, estando em forma escrita, é mantida na nossa frente, lá está a palavra embutida, *CAFÉ*, para que nossos olhos possam ver e rever, mas que no processo de compreensão da linguagem (*i.e.* quando ouvimos, lemos uma única vez ou vemos⁹ a palavra em uma sentença), o significado que integramos à interpretação da sentença seja somente o de *café* e, portanto, não há motivo para nos determos em uma parte da sua forma. Aí está o engano, pois o próprio processamento linguístico parece *enxergar* esses segmentos, como pretendo mostrar adiante. Já em relação à terminologia, é importante ter em mente que nem todos os autores adotam os mesmos critérios. Aqui entra a polissemia.

Em um artigo, Zhou *et al.* (2018) detalham um estudo experimental inovador e interessante sobre o curso temporal do processamento do que chamam de “palavras embutidas” durante a leitura em língua chinesa. Por se tratar de um paradigma psicolinguístico que aborda a leitura de um sistema de escrita logográfico, os autores adotam outro conceito de palavra dentro de palavra: aqui, os significados das supostas palavras embutidas são **transparentes** e **necessários** para a construção do significado das portadoras; ou seja, eles participam da composição do significado. Por exemplo, a palavra 专科学士 (*estudantes universitários*) pode conter três palavras embutidas: 专科 (*universidade*), 科学 (*ciência*) e 学生 (*estudantes*)¹⁰. O estudo tem como objetivo justamente entender a dinâmica do processamento; os autores questionam, por exemplo, se essas palavras embutidas são ativadas no léxico mental durante a leitura da portadora. O artigo dialoga com uma literatura que, na mesma linha, aborda palavras

⁸ *Carrier* para Shillcock (1990), Isel e Bacri (1999), Van Alphen e Van Berkum (2010), entre outros.

⁹ Me refiro às línguas sinalizadas. Desconheço, porém, estudos com palavras embutidas que abordem essa modalidade.

¹⁰ As traduções são minhas e foram feitas a partir dos exemplos dados em inglês.

formadas por outros constituintes morfológicos cujo conteúdo semântico tem influência no significado da palavra portadora — seja uma palavra embutida, seja uma raiz lexical, um afixo etc.

Note que a discussão pode se tornar mais complexa se considerarmos o fator **opacidade**. Em composições, por exemplo, os segmentos internos podem operar como elementos que constituem o significado desejado (e.g. *pé de moleque*). O significado de cada termo, nesse caso, é totalmente transparente. Mais complicada é a derivação a partir de raízes. Por um lado, a palavra *cortante* parece deixar rastros visíveis à análise por intuição: há uma relação com *cortar* e/ou *corte*. Por outro lado, a relação entre *restaura* e *restaurante* parece, no mínimo, opaca¹¹ — ou seja, não identificamos de imediato uma raiz comum em ambas as palavras. Entre uma PE estritamente *invisível* e uma raiz legítima, portanto, há um espectro de possibilidades. Seja qual for a resposta para a questão, este fenômeno não é estritamente o mesmo que abordo neste trabalho; no primeiro caso, as palavras embutidas são transparentes e pode haver uma metáfora em jogo, mesmo que já consolidada; no segundo, as raízes passam por processos morfofonológicos para formarem palavras e, ao final, frequentemente escapam da coincidência total das formas¹². Não é meu objetivo abordar esses fenômenos; como registrei, o foco deste trabalho é a PE no sentido estrito, isto é, cuja ocorrência é pura coincidência.

Meu intuito acima é apontar que a definição do termo varia na literatura e ele é utilizado, ao menos, de duas maneiras diferentes: para se referir (i) a um segmento fonológico/ortográfico espúrio de uma palavra que coincide ou se assemelha¹³ com outra palavra, ou seja, um epifenômeno; e (ii) a uma palavra cujos aspectos fonológicos (ou ortográficos) e semânticos são constituintes legítimos de outra palavra. Neste trabalho, me dedico ao primeiro caso.

¹¹ Cf. Lopes (2020), que aborda experimentalmente transparência e opacidade no acesso lexical.

¹² A ocorrência desse processo pressupõe que o uso da linguagem envolve composição morfológica, ou seja, processos de derivação e flexão a partir de elementos menores que palavras (raízes e afixos); uma alternativa seria a seleção das palavras já prontas de uma extensa lista. Essa é uma questão que modelos de acesso lexical tentam responder.

¹³ Diferenciarei coincidência e semelhança após descrever outro grau de complexidade do fenômeno (ver seção 2.2 “Escrita e oralidade: a questão fonológica”).

Somado a isso, as palavras embutidas podem estar situadas em diferentes partes das suas portadoras. Aproveito para apresentar mais dois itens da terminologia que uso daqui em diante:

- **palavra embutida inicial (PEI):** segmento cuja ocorrência coincide com o início (*onset*) da palavra portadora (e.g. ‘mar’ em *marfim*; *initial embedding*¹⁴);
- **palavra embutida final (PEF):** segmento cuja ocorrência coincide com o término (*offset*) da palavra portadora (e.g. ‘mar’ em *pomar*; *final embedding*¹⁵).

Duas outras possibilidades são: a ocorrência no meio de uma palavra portadora (e.g. ‘aranha’ em *emaranhado*); e a ocorrência que não coincide exatamente com o interior de uma palavra, mas se constitui no limite entre duas (e.g. ‘below’ em *terrible omen*¹⁶). Este último caso — que representa o local de ocorrência das cacofonias — tem sido abordado em estudos acerca da estratégia de segmentação das palavras em sentenças (e.g. Zhou *et al.*, 2018; Norris *et al.*, 2006), um tópico de nosso interesse neste trabalho.

Voltaremos às palavras embutidas mais à frente para discutir algumas pesquisas sobre o fenômeno e as respectivas implicações para os modelos teóricos de processamento de palavras, que é o foco principal deste trabalho. Estabelecidas as diferenças entre descrições oferecidas pela literatura, podemos avançar para um segundo fator relevante: as particularidades inerentes à modalidade.

2.2 ESCRITA E ORALIDADE: A QUESTÃO FONOLÓGICA

O fenômeno das palavras embutidas é matéria-prima valiosa para a psicolinguística. Contudo, antes de abordar os objetivos de outros estudos e apresentar os meus, preciso enfatizar que — e explicar os motivos pelos quais — o fenômeno tem particularidades em cada modalidade. Tendo em vista que

¹⁴ Van Alphen; Van Berkum (2010).

¹⁵ *Ibid.*

¹⁶ Norris *et al.* (2006).

desconheço trabalhos semelhantes com línguas de sinais, refiro-me aqui especificamente às modalidades oral e escrita.

No exemplo da língua escrita, simplifiquei a relação entre letra e significado, abstraindo uma característica importante do sistema: a multiplicidade das combinações possíveis a partir de (ao menos) dois elementos. O ponto é que a escrita, por si só, não abarca a totalidade das características fonético-fonológicas da fala. Por exemplo, (i) algumas letras representam mais de um som (e.g. ‘c’ em carro e cedo); (ii) alguns sons são representados ora por uma letra, ora por outra (e.g. [s] em cedo e seno), ora por mais de uma letra ou pela repetição de uma mesma (e.g. excesso). Indo além, onde não há distinção fonológica para um som, pode haver uma distinção fonética relevante na modalidade oral, como a que ocorre entre sílabas tônicas e átonas (e.g. as ocorrências de /a/ em *batata*). Ou seja, mantém-se a representação mental de um fonema, digamos /a/, enquanto a realização acústica (física) do som, [‘a], [a] ou [ə], varia de acordo com outros fatores, como a tonicidade da sílaba, a posição em relação a ela — isto é, pré ou pós-tônica — ou o próprio acento frasal. A título de exemplo, Shatzman (2006, p. 57-72) mostra que a manipulação da duração de um segmento fonológico em 30 ms já é capaz de modular a ativação lexical. Utilizando rastreamento ocular, a autora variou o tempo de realização da sílaba e observou que os participantes olhavam por mais tempo para as figuras de objetos cujo nome (e.g. *spijker* [unha]) sobrepunha parcialmente o da palavra testada (e.g. *pijp* [cachimbo]) quando a consoante inicial [s] tinha uma duração mais longa, enquanto a fixação era maior nas palavras com plosiva inicial [p] antecederida por um [s] curto. A interpretação é a de que a consoante [s] mais curta tende a ocorrer no meio da palavra, sendo sua realização longa mais típica do *onset* e, assim, serve como referência para identificar uma nova palavra.

E por que isso é relevante para a nossa discussão? Porque uma pequena variação pode gerar implicações decisivas para a eventual (não) ativação semântica da suposta palavra no léxico mental durante o processamento da linguagem. Vamos por partes.

Para começar, se falamos na coincidência de um segmento equivalente a um elemento completo (uma palavra) com apenas parte de outro

(ou seja, uma PE em uma palavra portadora), a modalidade em questão influencia o pareamento das formas. Veja abaixo dois exemplos que ilustram a questão:

(1) ‘cobra’ em cobrador;

(2) /dente/ em identidade.

Em (1), a forma escrita de *cobra* caracteriza uma palavra embutida inicial em *cobrador*, mas sua realização na modalidade oral¹⁷ traz duas distinções: a abertura da vogal /o/, realizada como [ɔ] em vez de [o], e a localização da sílaba tônica, justamente [kɔ], enquanto na portadora a tônica é [dor]. Cabe salientar que a tonicidade silábica, tendo suas especificidades prosódicas (*i.e.* duração mais longa da vogal, *pitch* mais alto etc.), é um fator relevante para os estudos de palavras embutidas, como abordarei mais à frente.

Já em (2), a forma escrita da PE não coincide inteiramente com parte da portadora devido à realização da vogal pretônica [i], enquanto na modalidade oral essa é justamente uma das possíveis realizações de /e/ em sílaba final átona, resultando na coincidência das formas¹⁸. Ademais, o ponto de tonicidade da PE também coincide com o da portadora.

Em suma, como apontam Marslen-Wilson *et al.* (1994), um estudo sobre a questão deve levar em conta a modalidade — “*must be modality sensitive*” (Marslen-Wilson *et al.*, 1994, p. 4). Do contrário, corremos o risco de jogar no mesmo balaio fenômenos que envolvem processos psicolinguísticos distintos. Portanto, retomando uma distinção da seção anterior, eu chamo de **coincidência** a ocorrência do segmento completo e relativamente preservado dentro de outra palavra, seja qual for a modalidade. Chamo de **semelhanças** as situações nas quais o segmento não corresponde totalmente à forma da palavra. Logo: em (1), ‘cobra’ coincide na forma escrita, mas é apenas

¹⁷ Uso como referência minha variedade do PB, a da região de Campinas, no interior de São Paulo.

¹⁸ O exemplo (2) assume a coincidência oral ao abstrair de questões fonológicas mais complexas, como a duração (em milissegundos) do segmento na sua realização acústica.

semelhante na forma oral; em (2), ‘dente’ é semelhante na forma escrita e coincide na forma oral¹⁹.

Para concluir esse tópico, duas pontuações importantes. A primeira é que, na fala, apontar a coincidência da PE é um processo complicado e que tende a demandar alguma dose de relativismo, já que a produção está sujeita a variações, seja entre falantes, seja por adequação fonológica ao contexto sentencial. Assim, julgo ser equivocado afirmar que há um *match* de 100% entre os estímulos (e.g. *dente* em *identidade*) sem uma análise acústica mais acurada que comparasse, entre os estímulos, a duração de sílabas e fonemas, o acento frasal e outros fatores — o que, de fato, não é feito neste nem nos trabalhos com os quais eu dialogo aqui. Essa é uma limitação importante a ser considerada na nossa discussão.

O segundo ponto é que a sensibilidade à modalidade é mais que um simples detalhe. Falar dessa complexidade é relevante, dados os objetivos dos estudos. No caso do processamento de sentenças na modalidade oral, as palavras embutidas nos ajudam a investigar mais a fundo como se dá a segmentação das palavras nas sentenças durante o processamento linguístico. Quando ouvimos algo como “carrossel”, por exemplo, há um ponto no meio da palavra que coincide com uma outra palavra inteira, *carro*. Como se dá o processo linguístico (inconsciente?) de decidir que as informações que continuam chegando fazem parte da mesma palavra e não da seguinte? No cenário oposto, o que ocorre com “Eu vi ela ontem”? O que nos leva a selecionar o verbo e seu complemento, em vez de *viela*? O contexto da sentença? Dicas prosódicas, quando existem? Uma análise probabilística baseada na combinação desses fatores?

Situações como essas são muito frequentes, mas geralmente lidamos com elas de maneira inconsciente e sem muita dificuldade para construir um significado coerente para a sentença. Entender como isso ocorre

¹⁹ Note que a discussão pode se alongar bastante quando consideramos as especificidades prosódicas da oralidade, pois é possível discriminar diferenças pontuais na realização acústica de fonemas em diferentes cenários. A título de exemplo, é possível que percebamos de forma distinta as realizações de /te/ no final de /dente/ e de /ti/ em /identidade/, a depender do contexto sintático no qual cada palavra está inserida. Se assim for, a diferença na percepção acústica pode ser relevante a ponto de descartar a possibilidade de que ‘dente’ seja ativada no léxico mental quando ouvimos *identidade* — diferença essa que um especialista em prosódia poderia captar em uma análise acústica do espectrograma vocal.

é um desafio que inspira muitas pesquisas, mas o quadro que delimitei até aqui é suficiente para meu objetivo geral, que é entender melhor se uma palavra embutida pode entrar na concorrência pelo acesso lexical durante o processamento da linguagem. Essa dinâmica do processamento pode nos dar pistas, por exemplo, sobre como se dá a segmentação das palavras na fala contínua e que modelos de reconhecimento de palavras faladas são mais eficientes.

A seguir, abordarei brevemente este último tópico e detalharei de que forma o estudo de palavras embutidas pode contribuir com a discussão.

3 ACESSO LEXICAL E PROCESSAMENTO DE PALAVRAS

3.1 MEMÓRIA E PALAVRA

Usarei aqui o termo “léxico mental” para me referir à noção de “memória de palavras”. Esse conceito, por si só, é objeto de discussão em áreas como a Psicolinguística e a Neurociência da Linguagem. Não é para menos: se já é desafiador definir o significado do termo “palavra”, mais complicado ainda é dissertar sobre a natureza da nossa memória de palavras e dos processos envolvidos no uso da linguagem.

Os estudos acerca do acesso lexical tentam responder justamente essas duas perguntas: como se organizam e como acessamos os conceitos em nosso léxico mental durante o uso da linguagem. A própria composição do léxico é tema de debate; afinal, que elementos básicos o compõem? Palavras? Morfemas?

Se até aqui eu fui capaz de contornar a discussão sobre o conceito de “palavra” e evitar uma definição, me parece que isso é, agora, inevitável. Falarei, então, sobre o termo de forma mais ampla; depois, sobre como ele é tipicamente abordado na Psicolinguística e, por fim, o que é uma palavra dentro do escopo deste trabalho.

Em diversas áreas da ciência, e sem dúvidas na Linguística, é comum que um termo já existente passe por transformações no meio acadêmico até que seja alcançado um aparente e relativo consenso entre os estudiosos da área quanto à sua definição. Geralmente, o significado se afasta

daquele do senso comum popular, isto é, o que não é do meio acadêmico. Quando isso acontece, temos de lidar com a (muitas vezes incômoda) tarefa de denunciar a polissemia na mesa do bar, ou seja, explicar que o termo científico tem um significado particular e distinto. “Quando nós, linguistas, falamos em *palavra*, estamos falando *disso aqui*, e não *daquilo outro*”. Para o cientista, essa é uma terça-feira qualquer.

Já na academia — ou no laboratório —, não é raro que o mesmo termo se desdobre em subcategorias úteis para finalidades ainda mais específicas, talvez complementares, nenhuma delas abrangente o suficiente para dispensar a necessidade das outras. É o caso de *palavra*, cuja definição não é simples e gerou ramificações. Começemos pelo senso comum.

Estamos tão profundamente habituados à cultura escrita que uma pessoa letrada (mas não linguista) que recebesse a tarefa de responder “o que é uma palavra?” poderia, surpreendendo absolutamente ninguém, recorrer a uma imagem mental de uma sequência de letras delimitada por dois espaços em branco; se for preciso responder, é bem possível que essa pessoa arrisque algo do tipo “uma unidade de algo (um som?) **com significado**”. Do ponto de vista linguístico-acadêmico, ambas as definições parecem insuficientes, talvez equivocadas, já que, primeiro, a palavra precede a escrita e, portanto, não pode ter uma definição estritamente ortográfica; segundo, que é discutível se toda palavra tem significado. Qual é o significado de “*que*”?

No livro *Dicionário de Linguagem e Linguística*²⁰, Trask (2008, p. 218) define palavra como “uma unidade linguística tipicamente maior que o **morfema**, mas menor que um **sintagma**”, havendo ao menos quatro maneiras não equivalentes de defini-la, sendo a primeira a **palavra ortográfica**, conceito que mais se aproxima da primeira ideia acima. A **palavra fonológica**, por sua vez, “se pronuncia como uma única unidade”, enquanto o **item lexical** ou **lexema** é “uma palavra do dicionário [mental], uma unidade para a qual se espera que exista uma entrada própria no dicionário”. Trask (2008) define ainda a **forma gramatical de palavra** ou **palavra morfossintática** como qualquer das formas que o item lexical possa assumir. Do ponto de vista da Linguística como ciência, é possível abordar um aspecto específico dentre esses — e, no caso deste trabalho, o foco é o item lexical, levando em consideração que se

²⁰ *Key Concepts in Language and Linguistics*, traduzido e adaptado para o PB por Rodolfi Ilari.

trata de uma abordagem psicológica, ou seja, que busca descrever a faculdade da linguagem do ponto de vista biopsicológico de um indivíduo (adulto típico), não a língua como fenômeno social.

Na Psicolinguística, é comum nos depararmos com uma distinção ampla entre palavras lexicais e funcionais. As primeiras são aquelas que possuem, por si só, conteúdo semântico próprio: seja uma conexão mental com um conceito, uma imagem concreta/abstrata, etc. Entram nessa categoria, principalmente, verbos e substantivos, mas podemos incluir também os adjetivos e até mesmo os advérbios. Já as palavras funcionais são aquelas que desempenham alguma função gramatical, ou seja, que estabelecem relações entre sintagmas em geral, ou alteram/complementam seu significado. São elas as preposições e os artigos, por exemplo. Dito isso, meu foco aqui são as palavras lexicais, que tratarei dentro da definição bem ampla de “unidade com significado próprio” — definição essa que, curiosamente, não está tão distante daquela do senso comum, mas que justifico. Uma das vantagens dessa abordagem é que, por se tratar de um elemento numeroso nas línguas, a palavra lexical permite certo grau de generalização sobre o processamento linguístico a partir de um paradigma experimental controlado. Além disso, o lastro semântico na memória nos permite mensurar um dado comportamental mais tangível, por meio do fenômeno de *priming* semântico/associativo em um paradigma experimental controlado bem comum na psicolinguística. Por outro lado, uma das desvantagens reside no fato de que abstrair a palavra lexical de eventuais subcategorias pode justamente borrar nuances relevantes e, portanto, é um mais um fator que restringe a análise. Felizmente, estudos como este podem ser complementados por outros capazes de captar aspectos específicos de subcategorias lexicais, como ocorre em trabalhos sobre eventuais diferenças entre o acesso lexical de palavras concretas e abstratas, palavras com emoções de valência positiva e negativa, palavras tabu, entre outras.

Passemos agora à palavra em uso.

3.2 ALGUNS MODELOS DE RECONHECIMENTO DE PALAVRAS FALADAS

Para produzir uma palavra ou sentença falada, realizamos alguns processos cognitivos: é necessário ter a intenção de fazer isso, então selecionar o conteúdo semântico, recuperar sua respectiva forma fonológica e, por fim, expressá-la por meio do nosso aparelho fonador. Boa parte disso ocorre de forma majoritariamente inconsciente, *no piloto automático*. Já a compreensão envolve o caminho inverso: recebemos o estímulo (nesse caso, auditivo), convertemos as ondas acústicas em uma representação mental (forma fonológica) e mapeamos as formas e seus respectivos significados; no caso da sentença, combinamos os itens identificados de maneira a construir uma interpretação geral coerente com as regras sintáticas da língua. A ordem desses processos, por si só, já é tópico de muito debate: há quem defenda que as palavras são identificadas primeiro e a combinação sintática é inteiramente posterior (*bottom-up*); e há quem advogue que a sintaxe desempenha um papel mais amplo, em paralelo à segmentação e identificação das palavras. Sem nos aprofundarmos na questão, a ideia geral é essa — ao menos para teorias representacionalistas²¹.

Modelos teóricos de reconhecimento de palavras faladas buscam descrever essa primeira fase do processamento, na qual a fala contínua é segmentada em palavras que, uma vez **reconhecidas**, integrarão uma representação mental do significado da sentença²². Aqui, vale destacar que me refiro especificamente a esse processo e não ao acesso lexical como um todo, já que os modelos de acesso lexical buscam, por sua vez, descrever a dinâmica de ativação dos itens no léxico mental em função da sua **organização** — que pode ser em lista de palavras (*Full Listing Hypothesis*²³)

²¹ São exemplos não-representacionalistas algumas abordagens da Teoria dos Sistemas Complexos (e.g. Spivey, 2007) e a proposta de Libben (2017), que adota como metáfora noções de Física Quântica.

²² É possível encontrar referências à suposta *decodificação* do significado da sentença, termo que entendo ser impreciso pois sugere que o receptor (ouvinte) estaria necessariamente captando o mesmo significado intencionado pelo emissor (falante).

²³ Cf. Butterworth (1983); Hay e Baayen (2005); Jackendoff (1983).

ou morfemas (*Full Decomposition Hypothesis*²⁴) ou ainda em uma mistura de ambas as estratégias, como preveem os Modelos de Dupla Rota (*Dual-Route Models*²⁵). Entender de que forma a memória de palavras — ou itens menores — se organiza exige entender a dinâmica de *ativação* desses elementos durante o uso da linguagem, daí o interesse no processamento de palavras quando o objetivo é descrever o acesso lexical. Já os modelos que discuto a seguir foram elaborados exclusivamente para explicar como ocorre a seleção de uma palavra específica em vez de outra, durante uma etapa pré-lexical de competição, e de que forma identificamos os limites entre as palavras para segmentar a sentença. O acesso lexical, portanto, se dá quando a concorrente mais adequada é selecionada para integrar o significado da representação da sentença que se constrói.

Em geral, as teorias de reconhecimento de palavras faladas preveem uma competição entre diversas palavras candidatas que podem corresponder em parte ou totalmente ao *input* fonológico. Shatzman (2006, p. 74, tradução própria, grifo próprio) explica que “o nível de ativação de uma candidata em particular é uma função do seu **nível de correspondência**²⁶ com o sinal corrente, o número de outras candidatas lexicais ativas e seus níveis de correspondência com o sinal”. Um exemplo popular desses modelos é a versão mais antiga do Cohort (Marslen-Wilson; Welsh, 1978). No início (*onset*) do sinal acústico, as palavras cujo *onset* coincidem com o segmento fonológico captado são ativadas, formando o que o modelo chama de *cohort*, isto é, uma espécie de conjunto (*cluster*) de palavras, do qual apenas uma continuará ativada. Ao ouvir /pa/, por exemplo, cria-se um cohort com *pato*, *palha*, *pássaro* etc. Em um determinado “ponto de exclusividade” (*uniqueness point*), as palavras que não correspondem exatamente ao código fonológico são inibidas pela que alcança o melhor *match*. A versão posterior do Cohort (Marslen-Wilson, 1993), por sua vez,

Outros modelos, no entanto, propõem uma espécie de ativação contínua, se opondo à ideia de uma ativação linear que se desenrolaria

²⁴ Cf. Taft (1979); Taft e Forster (1975, 1976); Forster (1979); Stockall e Marantz (2006); Pykkänen *et al.* (2000, [2003?]); Pykkänen, Stringfellow e Marantz (2002); Embick *et al.* (2001); Pykkänen e Marantz (2003).

²⁵ Cf. Caramazza, Laudanna, Romani (1988); Pinker (1991); Schreuder, Baayen (1995).

²⁶ O termo original adotado em inglês é “*goodness of fit*”.

exclusivamente a partir do *onset* da palavra. Tanto o TRACE (McClelland; Elman, 1986) quanto as versões posteriores do Cohort (Marslen-Wilson, 1987, 1993) e modelos de *Shortlist A* (Norris, 1994; Norris; McQueen; Cutler, 1995) preveem o recebimento de *inputs bottom-up* continuamente, ou seja, em qualquer ponto do código fonológico. Assim, mesmo que uma palavra no léxico mental não tenha um *onset* que corresponda exatamente ao *onset* do estímulo recebido, ela pode ser ativada em algum nível e, talvez, participar da competição pela ativação. É aqui que entram as palavras embutidas.

Um modelo como o Cohort (Marslen-Wilson; Welsh, 1978) prevê a ativação de palavras embutidas no início das portadoras, como *lar* em *largo*. Assim, no exemplo, *lar* integraria o cohort até o momento em que o *input* /g/ fosse recebido, deixando de ser uma concorrente adequada e, conseqüentemente, sendo inibida. O que esse modelo não prevê é a ativação, mesmo que momentânea, de uma palavra embutida final, como ‘fé’ em *café* — cenário compatível com as previsões do TRACE e do *Shortlist*.

Isso nos leva às implicações para as discussões teóricas e, naturalmente, à literatura experimental existente.

3.3 ATIVAÇÃO LEXICAL DE PALAVRAS EMBUTIDAS FINAIS

Se uma PEF é ativada, mesmo que momentaneamente, e compõe o *cohort* que compete pelo acesso lexical, podemos inferir mais que a simples falha do modelo *Cohort* em descrever a dinâmica do processamento de palavras; afinal, uma das implicações é que o processamento de sentenças não ocorre de maneira exclusivamente linear, no sentido de que pistas fonológicas imediatas seriam suficientes para determinar o início (*onset*) de cada palavra e somente depois disso aconteceria a identificação da palavra. Na prática, é comum que as informações que chegam tragam consigo uma possível ambigüidade: pode ser o início de uma palavra, pode ser a continuidade da palavra anterior. Se o *parser* está recebendo informação fonológica para uma palavra que condiz com o *onset* da informação recebida, ele também considera uma nova palavra competindo pela ativação se ela começa mais à frente (e.g. ‘mar’ em *pomar*)?

Vroomen e De Gelder (1997) detalham que os demais modelos preveem a possibilidade, mas o comportamento pode variar em cada um deles. No exemplo que trazem, *bone* (osso, em inglês) poderia estar embutida em *trombone*. Contudo, em um paradigma experimental de *priming*, o TRACE, por exemplo, prevê diferentes cenários. Uma possibilidade é que “não haveria facilitação de *bone* quando *trombone* é ouvida porque, quando *bone* recebe *input*, *trombone* está tão ativa que inibe *bone* por inibição lateral” (Vroomen; De Gelder, 1997, p. 711, tradução própria); é possível ainda que haja ativação residual da PEF, caso a primeira sílaba seja muito curta, sua frequência seja muito baixa ou a frequência da própria PE seja muito alta (Vroomen; De Gelder, 1997, p. 713-714). O *Shortlist A* (Norris, 1994; Norris; McQueen; Cutler, 1995), por sua vez, incorpora a Estratégia de Segmentação Métrica (*Metric Segmentation Strategy – MMS*), que prevê, ao menos para a língua inglesa, um papel especial das sílabas fortes no fornecimento de pistas fonológicas para a demarcação do *onset* das palavras. Mas esse não é um fator absolutamente decisivo: essa informação pré-lexical ajudaria a localizar os limites entre as palavras, mas também possibilitaria que *bone* embutida em *trombone* fosse ativada por receber um “*input extra*” que permitiria uma curta ativação.

As palavras embutidas finais parecem ter uma desvantagem em relação às iniciais. Afinal, há material fonético suficiente antes do seu *onset* para torná-las candidatas demasiadamente fracas. Levar isso em consideração é importante por que, caso as PEFs ainda assim sejam ativadas, dois pontos merecem atenção. Primeiro: temos um fenômeno que desafia os modelos mais rígidos de processamento linear e, portanto, demanda explicação. Segundo: mesmo que a ativação seja supérflua e o reconhecimento da palavra ocorra com base em outros critérios, como eventuais pistas fonológicas, a velocidade do processamento linguístico e a quantidade existente de palavras embutidas implicam que elas estão constantemente, e em bom número, sendo ativadas em algum nível, ainda que de forma tão breve e fraca que, majoritariamente, não sejam percebidas. Na língua inglesa, por exemplo, cerca de 84% dos polissílabos contêm pelo menos uma palavra embutida, segundo McQueen *et al.* (1995).

Se as previsões são variadas, estudos experimentais podem nos ajudar a aprimorar a descrição do fenômeno para, a partir disso, refinarmos as hipóteses e as teorias psicolinguísticas de acesso lexical. Contudo, a literatura está marcada por resultados aparentemente inconsistentes entre si, abrindo espaço para uma discussão ainda mais ampla que envolve os modelos teóricos, as especificidades de cada língua e algumas diferenças metodológicas. De forma geral, minha hipótese busca acomodar parte dos resultados encontrados anteriormente, explicar a origem das inconsistências (*spoiler*: questões metodológicas) e propor uma descrição para o processamento de palavras embutidas finais em PB que contribua para a explicação do fenômeno também em outras línguas.

Alguns estudos apontam para a ativação de PEFs (Shillcock, 1990; Vroomen; De Gelder, 1997; Luce; Cluff, 1998; Isel; Bacri, 1999; Shatzman, 2006; Bowers *et al.*, 2009; Van Alphen; Van Berkum, 2010, 2012). Mais especificamente, Shillcock (1990) reportou *priming* facilitatório de PEFs em língua inglesa, como *bone* em *trombone*. O paradigma adotado foi o *priming* associativo; os estímulos foram utilizados como *prime* para um alvo²⁷ semanticamente relacionado — nesse exemplo, *bone* (osso) teria causado *priming* em RIB (costela) mas não em BUN, a palavra-alvo semanticamente não relacionada²⁸ usada como base de comparação. Note que estamos falando de alvos distintos e um mesmo *prime* em vez de um mesmo alvo cujos tempos de resposta seriam registrados em dois cenários diferentes (com dois estímulos de *prime* diferentes). Voltaremos a isso.

Luce e Cluff (1998) também reportaram *priming* facilitatório em um estudo com palavras embutidas em Inglês (e.g. ‘*lock*’ em *hemlock*) e palavra-alvo semanticamente relacionada (e.g. *key*). A particularidade aqui é que as portadoras eram palavras cuja primeira parte também caracterizaria uma PE (e.g. ‘*hem*’ em *hemlock*). Em Holandês, Vroomen e De Gelder (1997) compararam PEFs em situação de correspondência ou não com a sílaba final da palavra portadora, registrando *priming* facilitatório quando há correspondência (e.g. ‘*boos*’, que significa ‘nervoso’, em *framboos*, que

²⁷ Neste trabalho, “alvo”, “palavra-alvo” e *target* são sinônimos. Uso os termos alternadamente apenas para evitar repetição excessiva.

²⁸ Seguindo a tradição da literatura, as palavras-alvo estão grafadas em caixa alta.

significa ‘framboesa’). Os autores reportam, no entanto, ausência de *priming* quando a PE começa em um ponto fora do *onset* da sílaba na portadora (e.g. ‘*wijn*’ em *zwijn*), ressaltando o papel das pistas fonológicas — no caso, fronteiras silábicas — para indicar um possível início de palavra.

Isel e Bacri (1999), por sua vez, reportaram um estudo de *priming* transmodal (*crossmodal priming*) em Francês, indicando *priming* facilitatório eliciado pela PE final em um *target* visual semanticamente relacionado (e.g. ‘*car*’ em *brancard*²⁹; o respectivo alvo é AUTOBUS). Já a dissertação de Shatzman (2006) nos traz, entre outros, um estudo experimental com rastreamento ocular no qual os participantes ouviam sentenças curtas (em holandês) e viam quatro figuras, sendo instruídos a clicarem no objeto mencionado na sentença; nos itens da condição de interesse, havia uma palavra na sentença contendo uma PEF e imagens das duas palavras (a portadora e a PE). Os resultados foram interessantes e surpreendentes: ativação de PEFs — nesse caso, era mensurado o tempo em que o participante olhava para a figura correspondente — quando coincidem com uma sílaba átona da palavra portadora, mas não com sílabas tônicas. Também usando o *eye tracker*, Weingartner, Juhasz e Rayner (2012) registraram tempos mais longos a partir de dados que incluem palavras embutidas iniciais e finais. O subconjunto das *final embeddings*, no entanto, era pequeno demais para uma análise isolada.

Mais resultados a favor da ativação foram reportados por Van Alphen e Van Berkum (2010, 2012). Nos experimentos reportados, o objetivo era responder se as palavras embutidas eram levadas em consideração durante a interpretação de sentenças, o que seria uma evidência de ativação, ao menos momentânea, durante a competição pelo acesso lexical. Utilizando EEG, os autores observaram efeito de N400 alinhados ao *input* da PE, tanto em sentenças com viés semântico favorável à sua interpretação, quanto em cenários nos quais o contexto sentencial produzia uma expectativa pela palavra portadora — e, portanto, uma provável pré-ativação dessa palavra, com potencial para inibir a PE e outras concorrentes mais fracas.

²⁹ O estudo usou *prime* auditivo; note que a forma ortográfica *brancard* traz um segmento final /d/ que não é realizado na forma oral e, por isso, a palavra embutida *car* parece coincidir totalmente com a sílaba.

Já o estudo de Bowers *et al.* (2009) adota um paradigma de “interferência imagem-palavra” (*Picture-Word Interference – PWI*), no qual os autores avaliam se uma palavra ouvida durante a nomeação de imagens é capaz de interferir no desempenho da tarefa. Os autores manipulam intencionalmente o SOA, isto é, o alinhamento entre o início dos dois estímulos, para que a palavra fosse ouvida:

- antes da imagem ser exibida (100-150 ms);
- no exato mesmo momento (SOA = 0 ms); ou
- depois da aparição da imagem (150 ms).

É interessante notar que efeito de PWI por palavras embutidas finais só foi obtido com SOA negativo, ou seja, quando a imagem foi exibida 100 ms depois do *onset* da palavra portadora, sobrepondo o momento em que a PE era realizada (Bowers *et al.*, 2009, p. 1589). Além disso, o experimento envolveu uma tarefa de **produção** de linguagem e, portanto, não exclui a possibilidade de que eventuais variações nos tempos de resposta estejam ligadas a outros processos linguísticos que não especificamente reconhecimento de palavras e competição lexical.

Para concluir esse levantamento de estudos com resultados favoráveis à ativação de palavras embutidas, temos o trabalho de Zhang e Samuel (2015), cuja investigação abrange ainda a possível influência do aumento da “carga cognitiva” (*cognitive load*) no processamento linguístico, mais especificamente na competição lexical. Usando *priming* auditivo-auditivo de palavras isoladas, os autores testaram a ativação de palavras embutidas e inseriram uma tarefa de memória como forma de sobrecarregar cognitivamente o participante; o intuito era observar se isso impediria a ativação e, por consequência, a ocorrência de *priming*. Não foi registrado efeito facilitatório a partir de PEFs curtas (monossílabas), mas sim nas condições em que elas eram mais longas (multissilábicas) e **sem** *cognitive load*; na presença de uma tarefa de memória adicional, as palavras embutidas não produziram efeito de *priming*. A interpretação dos resultados foi a de que palavras espúrias podem, sim, ser ativadas em algumas situações, e que essa ativação pode ser restringida por fatores extralexicaís ou mesmo extralinguísticos, como é o caso

do *cognitive load*. Cabe destacar que os autores utilizaram **intervalo entre estímulos** (*Interstimulus Interval* – ISI)³⁰ de 500 ms. Esse controle, similar ao SOA, é central para a minha argumentação, pois o tempo entre os estímulos pode ser longo o suficiente para que a PE, se inicialmente ativada, já tenha sido rapidamente inibida por não ser a candidata mais adequada; logo, se considerarmos a hipótese de Isel e Bacri (1999) de que isso pode acontecer em menos de 250 ms, um ISI de 500 ms não seria capaz de registrar esse *priming*. Retornamos a essa discussão mais à frente.

Agora, vejamos alguns resultados contrários à ativação.

A contrapartida está em estudos que reportaram *priming* inibitório (Shatzman, 2006³¹; Marslen-Wilson *et al.*, 1994³²) ou mesmo a ausência de qualquer efeito de *priming* (Gow; Gordon, 1995; Marslen-Wilson *et al.*, 1994; Norris *et al.*, 2006).

Essa aparente inconsistência se desdobra em alguns aspectos. Para começar, ainda que Shillcock (1990) tenha adotado um paradigma de *priming* transmodal associativo, voltemos à escolha experimental: a comparação foi feita entre os tempos de resposta de diferentes palavras-alvo (*i.e.* RIB e BUN). Para Vroomen e De Gelder (1997), esse “problema metodológico” abre espaço para questionamentos sobre a interpretação dos dados, já que eventuais diferenças nos tempos de resposta, argumentam, poderiam ser atribuídos a diferenças intralexicais entre os próprios *targets*. Em outras palavras, pode ser que os tempos de resposta sejam diferentes não por efeito de *priming*, mas devido a eventuais diferenças entre as frequências das palavras-alvo, por exemplo, algo que Shillcock não declara se adotou como critério de controle na seleção de palavras para o experimento.

Somado a isso, trata-se de um estudo que observou estímulos de *prime* em um contexto sentencial, enquanto outros (*e.g.* Marslen-Wilson *et al.*, 1994) adotaram um paradigma que compara a leitura de palavras isoladas. Nesse sentido, é possível que a ocorrência do *priming* possa ser modulada por fatores específicos do contexto em questão. Na sentença, por exemplo, é

³⁰ Sobre definições de SOA e ISI, ver Kirman (1974).

³¹ Shatzman (2006) reporta efeito positivo em um experimento de rastreamento ocular (Shatzman, 2006, p. 73-96) e *priming* inibitório em outro experimento, de *priming* de identidade (*identity priming*) (Shatzman, 2006, p. 97-121).

³² O efeito de *priming* inibitório reportado por Marslen-Wilson *et al.* (1994) — de, em média, 12 ms — não é estatisticamente relevante.

possível que a palavra receba um nível maior de atenção do receptor (ouvinte) devido à sua importância para a construção do sentido da sentença, como sugerem Norris *et al.* (2006), aumentando a possibilidade de um efeito de *priming* a partir da portadora ou da própria PE; por outro lado, é possível que a sentença favoreça a ativação da portadora por meio de efeito de contexto sentencial (e.g. predição lexical³³), a ponto de que sua ativação seja tão forte que iniba a ativação, mesmo que breve, da PE — possibilidade que inspira os experimentos de Van Alphen e Van Berkum (2010; 2012). Se Zhang e Samuel (2005) estiverem corretos quanto à possibilidade de influência por fatores extralexicais, é plausível que outros efeitos estejam em jogo, tornando ainda mais importante o compartilhamento de dados detalhados sobre o método (*i.e.* materiais, design, procedimentos etc.).

Há também críticas aos estudos que não registraram efeito de *priming*. Quanto aos experimentos de Marslen-Wilson *et al.* (1994), como mencionamos, a diferença é metodológica: o paradigma usado comparou palavras isoladas. Gow e Gordon (1995), por sua vez, testaram palavras em contexto sentencial neutro; contudo, as palavras embutidas não coincidiam com a sílaba forte, que vinha antes (e.g. *lips* em *tulips*), algo que outros estudos concluem que seja relevante (cf. Shatzman, 2006). Já Norris *et al.* (2006), além de não encontrarem efeito facilitatório, mencionam um *priming* inibitório substancial no experimento de *priming* de identidade (*Identity priming*) com palavras em contexto sentencial. A interpretação é de que “as palavras embutidas são inicialmente ativadas, mas depois devem ser suprimidas, e isso resulta em *priming* negativo” (Shatzman, 2006, p. 166).

Por fim, Shatzman (2006) reporta ativação surpreendente de palavras embutidas finais apenas em sílabas átonas. Em contrapartida, seu experimento de *priming* de identidade reporta efeito inibitório. Uma hipótese interessante compartilhada pela autora e por outros (e.g. Norris *et al.*, 2006) para explicar essa inconsistência é a de que o *priming* inibitório sugere **ativação da representação fonológica** da palavra embutida, mas não necessariamente acompanhada da ativação da **representação semântica**, resultando na possibilidade de haver *priming* por identidade (pela *forma* da

³³ *Predictability*; cf. a dissertação de mestrado de Souza Filho (2021) para um estudo sobre a predição lexical como efeito de contexto sentencial.

palavra) sem que haja *priming* associativo (pelo *significado*) O método em questão permitiria dissociar os dois processos e, portanto, traz uma possível explicação para a ausência de *priming* semântico em outros experimentos.

Aqui, vale destacar que a ausência de dados detalhados dificulta e pode comprometer a análise de alguns desses trabalhos. Um caso bastante relevante é o de Shillcock (1990), estudo que tem destaque na literatura por ser um dos primeiros a reportar de forma bem categórica a ativação de palavras embutidas em posição final (e.g. *bone* em *trombone*). Contudo, ao conferir o texto original – um capítulo intitulado *Lexical hypotheses in continuous speech*, publicado em *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (Shillcock, 1990, p. 24-49), de G. T. M. Altmann (ed.), não se encontra uma descrição detalhada do experimento, mais especificamente os materiais. Tudo o que temos é a sentença do exemplo usado no texto. Em vez disso, outro artigo do autor, do mesmo ano, é citado como referência para esses dados – no original: “*Fuller experimental details are given in Shillcock 1990*” (Shillcock, 1990, p. 32). Acontece que este segundo texto aparece nas referências como **submetido**, mas não encontrei o artigo de fato publicado, nem qualquer outra citação a ele em outros trabalhos. Ao que parece, não foi publicado, qualquer que tenha sido o motivo. Assim, diversos outros estudos partem de Shillcock (1990) como referência de ativação de embutidas finais – e Marslen-Wilson *et al.* (1994) como evidência inicial contrária – sem que haja uma base sólida para checagem e replicação do experimento. Nesse caso, só nos resta traçar hipóteses a partir de um exemplo único trazido no texto, e.g. “*He carefully placed the trombone on the table*” versus “*He carefully placed the bone on the table*” (Shillcock, 1990, p. 32); não há margem para exploração e crítica das palavras e sentenças selecionadas. Não descarto a legitimidade dos resultados reportados, e outros experimentos reforçam a hipótese do autor. Contudo, para a discussão que faço sobre a influência de fatores metodológicos específicos na ativação de embutidas finais, não há dados suficientes para usar esse trabalho como baliza, já que não há registro, por exemplo, do SOA ou ISI utilizado, tampouco há uma lista

de palavras embutidas para exploração, ou mesmo informações sobre a gravação dos áudios³⁴.

Nos demais estudos publicados, o registro de certos parâmetros metodológicos varia bastante. Em geral, quanto mais recente é a publicação, mais rica é a descrição de materiais e métodos (ver Quadro 1).

Somado a isso, outro obstáculo nessa discussão é o fato de que os trabalhos citados não abordam necessariamente a mesma questão de maneira frontal; na prática, as palavras embutidas são muitas vezes um fenômeno lateral que entra em uma discussão mais ampla. Isso é relevante porque significa que alguns experimentos não foram elaborados para abordar a questão em primeiro plano e, naturalmente, a comparação dos resultados não é tão objetiva quanto entre experimentos replicados. A título de exemplo, Marslen-Wilson *et al.* (1994) apresentam seis experimentos que se somam em uma abordagem do acesso lexical de palavras morfologicamente complexas, enquanto Gow e Gordon (1995) focam primariamente na segmentação de palavras. Isso ajuda a compreender as escolhas metodológicas e a aparente **inconsistência** dos resultados – *inconsistência*, e não contradição, já que, como eu argumento aqui, acredito que seja possível acomodar os resultados elencando fatores que podem influenciar a (não) ativação.

Como ponto de partida, um ponto que parece ter sido pacificado é que a ativação de PEFs, em si, é plausível, dado que diversos estudos registraram o efeito em ao menos alguns cenários específicos. Posto isso, podemos avançar e investigar mais a fundo quais são os fatores necessários e/ou as restrições para que a ativação aconteça. Aqui estão algumas possibilidades:

- ativação da representação semântica independente da representação fonológica;
- processamento de palavra isolada *versus* palavra em contexto sentencial;

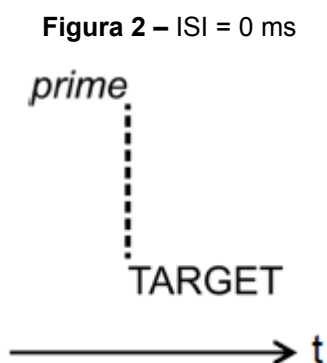
³⁴ Segundo Norris *et al.*, “Shillcock (comunicação pessoal) apontou que as sentenças foram lidas de forma vagarosa e cuidadosa” (2006, p. 174, tradução própria); no original: “*Shillcock (personal communication) pointed out that the materials in Shillcock (1990) were spoken in a slow and careful manner*”.

- relação temporal entre estímulos *prime* e *target*, isto é, SOA ou ISI.

Cada um desses tópicos pode ser explorado individualmente em um experimento. Nesse sentido, o terceiro ponto, que se refere ao **intervalo entre os estímulos**, ou à sobreposição entre eles, é uma das bases do meu argumento.

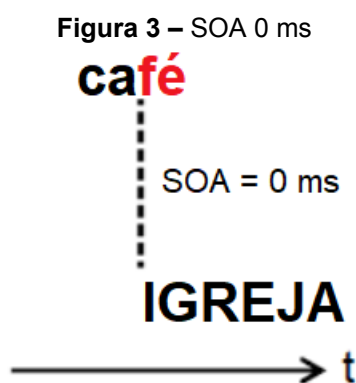
Comecemos por uma hipótese simples: a de que a palavra embutida final é ativada e, por isso, seria passível de eliciar efeito de *priming* associativo facilitatório em uma palavra-alvo relacionada. O experimento que descrevo aqui testa essa hipótese em PB.

Dando um passo além, devemos levar em conta que a literatura levanta a possibilidade de uma inibição rápida da PEF, já que ela perde a competição para a palavra portadora (Isel e Bacri, 1999, p. 66). Se for o caso, uma possível explicação para a inconsistência entre os resultados dos experimentos é: pode ser que o método esteja observando o tempo de resposta de um *target* após um intervalo muito longo em relação à realização do estímulo *prime*, talvez tempo suficiente para que a PE já tenha sido ativada e rapidamente inibida. Isel e Bacri, com base em seus resultados, sugerem que isso pode acontecer em menos de 250 ms. Se voltarmos aos experimentos para observar esse ponto, o panorama é variado: quando reportam ter controlado o intervalo entre estímulos, alguns autores (ver Quadro 1) o fazem considerando o ISI, geralmente em 0 ms — ou seja, o início do *target* está cravado junto ao final do *prime*:



Fonte: Salgado (2014, p. 20).

Se a hipótese estiver correta, uma escolha mais eficiente para registrar o suposto efeito de priming seria SOA abaixo de 250 ms, o que significa alinhar o início do *target* ao **início** do *prime* — no caso, a palavra embutida —, como no exemplo abaixo:



Fonte: Elaborada pelo autor.

Nota: Exemplo de SOA de 0 ms entre o *prime* 'fé' (palavra embutida final) e o *target* IGREJA para um experimento de *priming* associativo. Nesse caso, a palavra-alvo é exibida na tela no momento em que o participante ouve o **início** da última sílaba da palavra portadora 'café'.

A literatura, como aponte, nem sempre é clara quanto a essa escolha metodológica. O Quadro 1 a seguir ilustra isso:

Quadro 1 – Alguns parâmetros experimentais na literatura de palavras embutidas

Referência	Efeito de <i>Priming</i>	Ativ. PEF	ISI (ms)	SOA (ms)	<i>prime</i>	Observações
Shillcock (1990)	Sim (facilitatório)	Sim	-	-	Sentença	<i>Priming</i> transmodal
Marslen-Wilson <i>et al.</i> (1994)	Não	Não	0	-	Palavra isolada	<i>Priming</i> transmodal identidade
Isel e Bacri (1999)	Sim (facilitatório)	Sim	0	-	Palavra isolada	<i>Priming</i> transmodal
Gow e Gordon (1995)	Não	Não	0	-	sentença	<i>Priming</i> transmodal
Vroomen e De Gelder (1997)	Sim	Sim	0	-	Palavra isolada	<i>Priming</i> transmodal
Luce e Cluff (1998)	Sim	Sim	0	-	palavra isolada	<i>Priming</i> transmodal
Shatzman (2006)	Não (Exp. 2)	Sim (Exp. 1)	0 (Exp. 2)	Início da sentença (Exp. 1)	-	Exp. 1: eye tracker; Exp. 2: <i>pr.</i> identidade
Norris <i>et al.</i> (2006)	Inibitório (não relevante)	Não	0 e 500	-	Sentença + pal. isolada	Exp. 1: <i>pr.</i> associativo; Exp. 2: <i>pr.</i> identidade
Bowers <i>et al.</i> (2009)	Sim (SOA -150 ms)	Sim	-	-150, 0, +150	Imagem	<i>Picture-Word Interference</i>
Van Alphen e Van Berkum (2010)	-	Sim	NA	NA	Sentença	EEG/N400
Van Alphen e Van Berkum (2012)	-	Sim	NA	NA	Sentença	EEG/N400
Macizo, Van Petten e O'rouke (2012)	Não	Não	100	-	Palavra isolada	EEG/N400 + <i>priming</i>
Zhang e Samuel (2015)	Não (1 sílaba); Sim (≥ 2 sílabas)	Sim (longas)	500	-	Palavra isolada	<i>Priming</i> audit. → audit.

Fonte: Elaborada pelo autor.

Nota: A tabela aponta, para cada estudo listado, se foi reportado *efeito de priming*; se houve **ativação da palavra embutida final** (Ativ. PEF); o ISI do experimento em questão, se reportado; o SOA, se reportado; o tipo de *prime* utilizado; e informações específicas sobre o método utilizado. No caso dos estudos com EEG/N400, ISI e SOA são Não Aplicáveis (NA), por se tratar de um paradigma com tarefa passiva na qual os participantes apenas ouvem sentenças e não há um estímulo *target*.

Os dados relevantes para essa primeira discussão estão destacados nas colunas SOA e ISI. Recapitulando: há uma hipótese de que palavras embutidas finais são ativadas no léxico mental e participam da competição pelo acesso lexical, mas que, perdendo a competição, são rapidamente inibidas (em

menos de 250 ms). Havendo uma inconsistência na literatura quanto aos dados reportados, podemos começar observando que tempo os autores utilizam entre *prime* e *target*. Shillcock (1990), que é nosso ponto de partida, não oferece essa informação. O Quadro 1 nos ajuda a listar aqueles que não registraram resultados favoráveis à ativação de palavras embutidas: Marslen-Wilson *et al.* (1994), Gow e Gordon (1995), Shatzman (2006) no Experimento 2 do seu capítulo sobre *embedded words*, Norris *et al.* (2006), Macizo, Van Petten e O'rouke (2012) e Zhang e Samuel (2015). Se a hipótese estiver correta, é possível explicar esses resultados pela escolha metodológica, já que todos utilizam ISI 0 ms ou mais, e não SOA 0 ms³⁵. Isso significa que o *target* é exibido algum tempo depois do *onset* da PE, tempo esse que será ao menos a duração dessa sílaba final. Não me aprofundei nessa análise para verificar se o tempo médio da sílaba final, quando informado, é próximo dos 250 ms, pois considero que o resultado seria frágil devido à falta de controle desse parâmetro nos experimentos.

A pesquisa de Beatriz Salgado (2014), especificamente, não consta no quadro para comparação, ainda que seja precursora do assunto em Português Brasileiro e uma das inspirações para este trabalho, por sugerir a execução de um estudo que investigasse a ativação de PEFs com SOAs distintos como meio possível para identificar a ativação e subsequente desativação da PEF por não ser a hipótese lexical ótima (p. 44). Em sua dissertação, Salgado descreve dois experimentos de priming associativo transmodal de palavras embutidas em posições final e inicial, respectivamente. A palavra contendo a PE servia de priming para diferentes *targets*, que poderiam ser relacionados à portadora ou à PE, além de um *target* controle

³⁵ Marslen-Wilson *et al.* (1994) informam que o SOA seria “o mais curto possível”, mas cabe notar que a passagem é um tanto confusa e, seja por equívoco terminológico ou não, o procedimento adotado foi ISI 0 ms e não SOA 0 ms, ou seja, o alinhamento é *offset-onset*. No original: “An important consideration in priming tasks, **especially those using immediately adjacent primes and targets**, is to deter the subjects from developing strategies based on expectations about likely relations between primes and targets. One way of combating this, **which we followed here, is to keep the stimulus onset asynchrony (SOA) between prime and target as short as possible**” (Marslen-Wilson *et al.*, 1994, p. 9, grifo próprio). Na mesma página, no entanto, a seção *Design and Procedure* registra o uso de ISI = 0 ms: “The prime was presented binaurally to the subject and, **immediately at the offset of the prime, the target word was displayed on a CRT screen in front of the subject.**” (Marslen-Wilson *et al.*, 1994, p. 9, grifo próprio).

para referência e um *target* não-palavra (pseudopalavra) para balancear as respostas ao teste de decisão lexical.

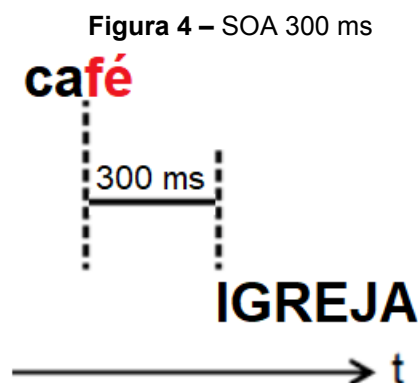
Minha justificativa para não comparar diretamente os resultados que trago com os da autora reside em dois aspectos do experimento da autora.

O primeiro ponto, e mais importante, é que a análise dos resultados foi feita com **as médias de RTs divididas pelo número de sílabas das palavras-alvo**, com o objetivo de “normalizar” os dados, já que os *targets* tinham tamanhos (em sílabas) variados e, portanto, uma palavra mais longa “demoraria mais tempo para ser lida”, tendo “resposta mais demorada em um teste de decisão lexical” (p. 20-21). Contudo, esse recurso da divisão do RT pelo número de sílabas só permite generalizações se assumirmos que o tempo de resposta das palavras é proporcional ao número de sílabas, afirmação que carece de evidências. Em suma, uma palavra com mais sílabas não tem um RT **proporcionalmente** maior, ainda que haja, sim, alguma influência desse fator, assim como há do número de letras (Estivalet, 2020, p. 273). Indo além, outros fatores, como a frequência lexical e a vizinhança ortográfica, também influenciam o RT, e precisam ser levados em consideração em um cálculo de RT por sílaba. Assim, entendo que as diferenças entre RTs na análise são insuficientes para interpretar a ocorrência ou não de *priming*.

O segundo ponto se refere ao SOA. No Experimento 2 de Salgado (2014), referente às PEFs, o *onset* da palavra-alvo está alinhado ao *offset* da palavra *prime*, equivalendo a um ISI 0 ms (ver Figura 2). Logo, ainda que eu buscasse uma comparação apenas da ocorrência ou não de efeito de *priming*, são cenários distintos e a configuração coloca o onset dos estímulos em um momento impreciso entre minhas condições — SOA 0 ms e 300 ms.

Outros autores, como Shatzman (2006) e Bowers *et al.* (2009), observaram justamente efeitos de *priming* em determinada configuração experimental, mas não em outra, reforçando minha hipótese de que fatores como o SOA podem influenciar os resultados.

Tendo em vista essa inconsistência, o experimento que apresento traz duas condições — SOA 0 ms e SOA 300 ms — voltadas a observar a possibilidade de ativação de palavras embutidas finais e posterior inibição rápida dessas mesmas palavras.



Fonte: Elaborada pelo autor.

Nota: Exemplo de SOA de 300 ms entre o *prime* 'fé' (palavra embutida final) e o *target* IGREJA para um experimento de *priming* associativo. Nesse caso, a palavra-alvo é exibida na tela 300 ms depois que o participante ouve o **início** da última sílaba da palavra portadora 'café'.

Outro ponto que eu abordo é a escolha metodológica do estímulo *prime*. Há especificidades na performance durante o processamento de palavras isoladas em relação ao processamento de sentenças — e essa é uma das diferenças entre os experimentos de Marslen-Wilson *et al.* (1994) e Shillcock (1990). Uma hipótese é a de que o contexto sentencial pode criar uma **expectativa** quanto às palavras que completam a sentença, sendo esse efeito capaz de favorecer a palavra portadora a ponto de inibir a ativação e concorrência da PEF. Assim, jogando contra a minha hipótese, eu adoto sentenças com contexto favorável à portadora, o que, seguindo essa lógica, representaria um fator contrário à ativação da PEF, ainda que minha expectativa seja a de que isso não será suficiente para inibir completamente a ativação das palavras espúrias — ou seja, o experimento deve registrar *priming* facilitatório eliciado por palavras embutidas com SOA 0 ms, mas não com SOA 300 ms, independente do contexto sentencial. As sentenças foram elaboradas de maneira a favorecer um efeito de predição lexical (Souza Filho, 2021) em favor da palavra portadora, no intuito de gerar uma restrição adicional à ativação da PE. Esta não é uma variável de controle do experimento, pois não era meu objetivo comparar a eventual ativação das palavras embutidas em um contexto sentencial desfavorável e em um contexto neutro, já que a adição

dessa condição tornaria o experimento muito longo e com muitos fatores para controlar, podendo comprometer sua viabilidade. Optei por utilizar as sentenças favoráveis à portadora neste primeiro momento, deixando em aberto a possibilidade de um segundo experimento para compará-las a um contexto neutro, caso o paradigma que detalho a seguir resulte em ativação lexical das embutidas. Ao final, trarei uma problematização sobre a escolha por esse design e possíveis influências nos resultados.

4 O EXPERIMENTO

Descrevo a seguir um experimento de *priming* transmodal (*prime* auditório de sentenças que contêm palavra portadora; *target* visual/ortográfico de palavras isoladas) para examinar a ativação de PEFs em contexto sentencial favorável à portadora em SOA 0 ms e 300 ms³⁶. Os participantes ouvem sentenças em PB e leem uma sequência de letras na tela do computador no momento em que a palavra *prime* é falada. A tarefa é a decisão lexical: os voluntários são instruídos a responder se a sequência de letras na tela é ou não uma palavra do PB. Nas condições críticas, uma palavra da sentença ouvida contém uma PEF. O *target* é exibido na tela alinhado ao *onset* da palavra embutida ou 300 ms depois desse ponto (ver figuras 3 e 4).

O objetivo é mensurar os tempos de resposta (RT) e compará-los para saber se, em média, os participantes respondem mais rapidamente aos alvos precedidos de palavra portadora com PE relacionada semanticamente. A comparação é feita com os mesmos alvos precedidos de palavras que não contêm PE semanticamente relacionada.

4.1 HIPÓTESE E EXPECTATIVAS

O experimento foi elaborado para testar a hipótese de que palavras embutidas em posição final podem ser ativadas quando a palavra portadora é ouvida em um contexto sentencial. A expectativa é de que os tempos de resposta da condição *prim00* sejam menores que os eliciados pelas condições

³⁶ Para simplificação, eu chamo aqui de “sentença” o estímulo auditivo que carrega o *prime*, ainda que esse estímulo seja composto por **duas** sentenças.

nprim e *prim03*, o que demonstraria que não só a PEF é ativada (causando *priming*), como ela é rapidamente inibida, por perder a competição pelo acesso lexical, não havendo, então, *priming* após 300 ms.

4.2 MÉTODO

4.2.1 PARTICIPANTES

Foram coletados dados de 49 participantes brasileiros com idade entre 18 e 33 anos (média 23; DP 3.42), destros, cuja língua materna é o PB, com visão normal ou corrigida. Todos cursavam ou já haviam concluído o ensino superior.

4.2.2 MATERIAIS

Palavras embutidas. Foram selecionadas 27 monossílabas do corpus *LexPorBr Infantil*³⁷ (Estivalet *et al.*, 2019) para serem usadas como palavras embutidas:

- em posição final (e.g. 'fé' em *café*);
- que coincidem totalmente com a sílaba final da portadora;
- sem relação semântica com a portadora;
- de ortografia não necessariamente idêntica ao segmento da palavra portadora (dois exemplos de não coincidência são 'fio' em *perfil* e 'som' em *garçom*);
- de frequência média ou alta (mínima de 3,73 e máxima de 6,11; média 4,6).

Além disso, adotei alguns critérios para aumentar o controle de fatores adversos, tendo como limite as restrições do próprio paradigma experimental. Especificamente, destaco que:

³⁷ Disponível em: <https://www.lexicodoportugues.com/infantil>. Acesso em: 10 jul. 2024.

- usei majoritariamente substantivos (são 24, além de 2 verbos e um adjetivo);
- usei majoritariamente palavras de duas ou três letras (à exceção de 'grau' em *degrau*);
- usei palavras no singular;
- na impossibilidade de replicar a mesma sentença alternando apenas a palavra *prime*, elaborei sentenças com estruturas semelhantes.

A lista total de estímulos está disponível no Apêndice A.

Palavras portadoras. Os critérios para escolha das portadoras, por ordem de importância, foram:

1. Presença da palavra embutida em posição final;
2. Coincidência total com a sílaba final, que é tônica;
3. Frequência em escala Zipf (*Zipf scale*) aproximada entre portadora e condição controle;
4. Substantivo singular.

A lista total de estímulos está disponível no Apêndice A.

Sentenças. Cada estímulo é composto por duas sentenças, tendo a primeira a função de criar contexto com efeito de previsibilidade em favor da portadora nas condições críticas para a análise. A segunda sentença contém a palavra portadora. Para evitar efeito de *spillover* ou *wrap-up*, palavras adicionais foram inseridas depois da portadora, de forma que ela não esteja em posição final. O quadro abaixo exemplifica as condições:

Quadro 2 – Sentenças experimentais

Sentença	<i>Prime</i>	<i>Target</i>
Lucas saiu para trabalhar com muito sono. Por isso, parou na padaria e bebeu uma xícara de café bem forte.	<u>café</u>	igreja
Lucas saiu para trabalhar com muita preguiça. Por isso, comprou uma lata de coca para tomar no caminho.	<u>coca</u>	igreja

Fonte: Elaborada pelo autor.

Nota: Na coluna '*Prime*', o trecho sublinhado é o *prime* em sentido estrito, ou seja, o trecho que estará sendo ouvido quando o *target* aparece na tela.

Foram elaboradas:

- 27 sentenças com PEF;
- 27 sentenças de controle (baseline);
- 27 sentenças distratoras, sendo 9 usadas com *targets* distratores (palavras) e 18 com *targets* do tipo “não-palavra”, para balancear as respostas.

A lista total de estímulos está disponível no Apêndice A. Parte das sentenças distratoras são as mesmas ou versões adaptadas das sentenças de Souza Filho (2021).

Foi realizado um pré-teste de CLOZE, via Google Forms, para avaliação do efeito de predição lexical gerado pelo contexto. Os participantes foram instruídos a completarem a lacuna da sentença com a primeira palavra que viesse à mente, podendo ainda escrever uma segunda e uma terceira opções. Posteriormente, adicionei algumas palavras ao final de cada sentença selecionada, a fim de evitar efeitos de *spillover* e *wrap-up*, como mencionei anteriormente. Foram selecionadas as palavras portadoras cujos índices de resposta foram de 72% ou mais. A análise do pré-teste está disponível no Apêndice B.

Figura 5 – Teste de CLOZE

O elevador do prédio estava quebrado. Joana decidiu usar a escada, mas torceu o pé no primeiro... *

Texto de resposta curta

O filme que assistimos contava a luta entre duas personagens num reino encantado. A princesa era a heroína, enquanto uma bruxa malvada era a grande... *

Texto de resposta curta

Fonte: Captura de tela registrada pelo autor.

Na condição crítica — isto é, onde pode haver efeito de *priming* associativo eliciado pela PE, seja com SOA 0 ms, seja com SOA 300 ms —, todas as sentenças possuem efeito favorável à palavra **portadora**. Para não comprometer a viabilidade do experimento — isto é, para que a tarefa experimental não ficasse longa demais (mais de 30 minutos) e para trabalhar com menos condições, buscando evitar um cruzamento excessivo de fatores que poderia tornar a análise estatisticamente irrelevante —, optei por usar **apenas** sentenças *enviesadas* a favor da palavra portadora, o que não caracteriza uma variável para o estudo, ao menos até aqui. Assim, não há uma comparação objetiva com uma outra condição “sem efeito de predição lexical” ou “efeito desfavorável à portadora”, que serviria para investigar se o efeito de contexto sentencial é capaz de ativar a palavra portadora com força suficiente a ponto de inibir completamente a ativação da PE, excluindo-a da competição. Essa decisão foi motivada pela limitação prática do estudo, mas traz um benefício: se não for registrado *priming* dessas palavras embutidas **nessas sentenças**, um segundo experimento pode ser feito para observar especificamente se o mesmo acontece em sentenças de contexto neutro ou desfavorável.

Por fim, foram elaborados cinco pares de *prime-target* para serem usados como treino no início do experimento (três com *target* ‘palavra’, duas com ‘não-palavra’ e SOAs de 0 ms ou 300 ms).

Design e procedimentos. O experimento foi elaborado usando o PsychoPy3 (v2023.2.3). Três versões diferentes foram elaboradas, de maneira que cada grupo de participantes seja exposto a:

- 9 sentenças da condição *priming* + SOA 0 ms;
- 9 sentenças da condição *priming* + SOA 300 ms;
- 9 sentenças da condição *não-priming*;
- 9 sentenças distratoras (*target* = palavra do PB);
- 18 sentenças de balanceamento de respostas (*target* = não-palavra);
- 5 sentenças de treino.

Em cada versão, os pares *prime-target* são distribuídos alternadamente da seguinte maneira: $\frac{1}{3}$ na condição *prime* + SOA 0 ms; $\frac{1}{3}$ na condição *prime* + SOA 300 ms; $\frac{1}{3}$ na condição *não-priming*. O objetivo é justamente que cada *target* seja executado três vezes: duas para testar o *priming* com SOA 0 ms e 300 ms, mais uma para a comparação (*não-priming*). Os demais pares (distratoras e balanceamento) são os mesmos nas três versões.

As sentenças foram gravadas por uma profissional do audiovisual que desconhecia os objetivos e o método do experimento. A orientação foi para que as sentenças fossem lidas com naturalidade. Outras informações sobre a gravação:

- local: sala com isolamento acústico (LAPROS³⁸);
- microfone dinâmico Shure SM58;
- software Audacity (versão 3.4.2);
- taxa padrão (*sample rate*): 48 KHz com resolução de 16 bits;
- os áudios foram comprimidos e normalizados para remoção de clipes e eventuais artefatos acústicos, além de equalização do volume;
- na persistência de alguns artefatos, algumas sentenças foram regravadas;

³⁸ Laboratório de Aquisição de linguagem, Processamento e Sintaxe (LAPROS) do IEL/UNICAMP.

- as sentenças gravadas têm duração média de 9,88 segundos.

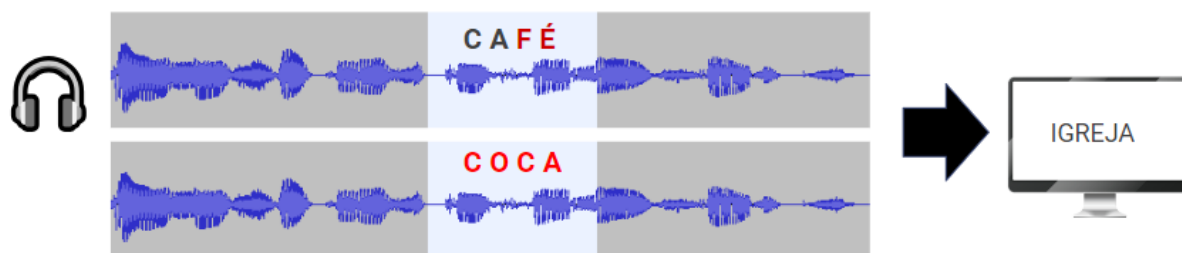
Cada *trial* do experimento ocorre da seguinte maneira:

1. O participante lê as instruções e aperta um botão no teclado para iniciar o treino;
2. Uma ##### de fixação é exibida no centro da tela por 1 segundo como *Inter-Trial Interval* (ITI);
3. A sentença é ouvida nos fones de ouvido em um volume confortável;
4. Um ##### de fixação é exibido novamente no centro da tela, três segundos antes do *target*;
5. No início da palavra crítica (em vermelho na Figura 6, abaixo), a palavra-alvo é exibida no centro da tela por quatro segundos;
6. O participante tem esses quatro segundos para responder SIM ou NÃO utilizando duas teclas do teclado³⁹;
7. A palavra 'OK' na cor verde aparece na tela pouco acima do target no momento em que uma resposta é fornecida, se mantendo ali até o final do intervalo, quando o ciclo se reinicia com a exibição da ##### de fixação.

Os estímulos foram randomizados; o texto foi exibido em fonte Open Sans (tamanho: 0.05; contraste: 1), centralizado na tela (posição: 0,0), em cor branca sobre um fundo cinza. A figura abaixo ilustra essa dinâmica:

³⁹ SIM = 'a' e NÃO = 'l'. Os participantes, portanto, devem usar a mão esquerda para a resposta positiva e a direita para a negativa, além de serem instruídos a manterem as mãos sobre os botões o tempo todo, a fim de evitar que desviem o olhar da tela no momento crítico.

Figura 6 – Representação visual do experimento de *priming* transmodal



Fonte: Elaborada pelo autor.

Nota: Acima, a condição crítica, na qual espera-se que haja *priming* eliciado pela palavra embutida (*fé*), que tem relação semântica com o *target* ('igreja'). Na condição '*não-priming*', a palavra-alvo é exibida no início do *prime* 'coca'.

A participação no experimento foi individual. O experimento foi realizado em um laboratório silencioso, sem a presença de terceiros. Os participantes eram instruídos a utilizar o fone de ouvido durante todo o experimento, a ouvirem atentamente a sentença e responder, o mais rápido possível, quando a sequência de letras aparecesse na tela, se aquela era ou não uma palavra da sua língua. Em média, a duração foi de 15 a 20 minutos.

Os arquivos referentes às três versões, incluindo as planilhas com os respectivos estímulos organizados, estão disponíveis na plataforma Open Science Framework (OSF)⁴⁰ e no repositório institucional da Unicamp, assim como todos os dados coletados⁴¹.

4.3 RESULTADOS

Foram coletados dados de 49 participantes. Destes, 40 eram de SP, 3 de MG, 2 do RS, e um participante do MS, CE, PE e PB. Cada uma das três versões do experimento contava com uma divisão de um terço dos estímulos para cada condição — ou seja, 9 para ***priming* com SOA 0 ms** (doravante *prim00*), 9 para ***priming* com SOA 300 ms** (doravante *prim03*) e 9 para ***não-priming*** (doravante *nprim*), de forma que todos os participantes foram expostos a 27 sentenças de interesse com *prime* ou condição de controle, além

⁴⁰ Disponível em: https://osf.io/czytq/?view_only=5ef2900180b74a1cb4ebd6b6372b6b3f.

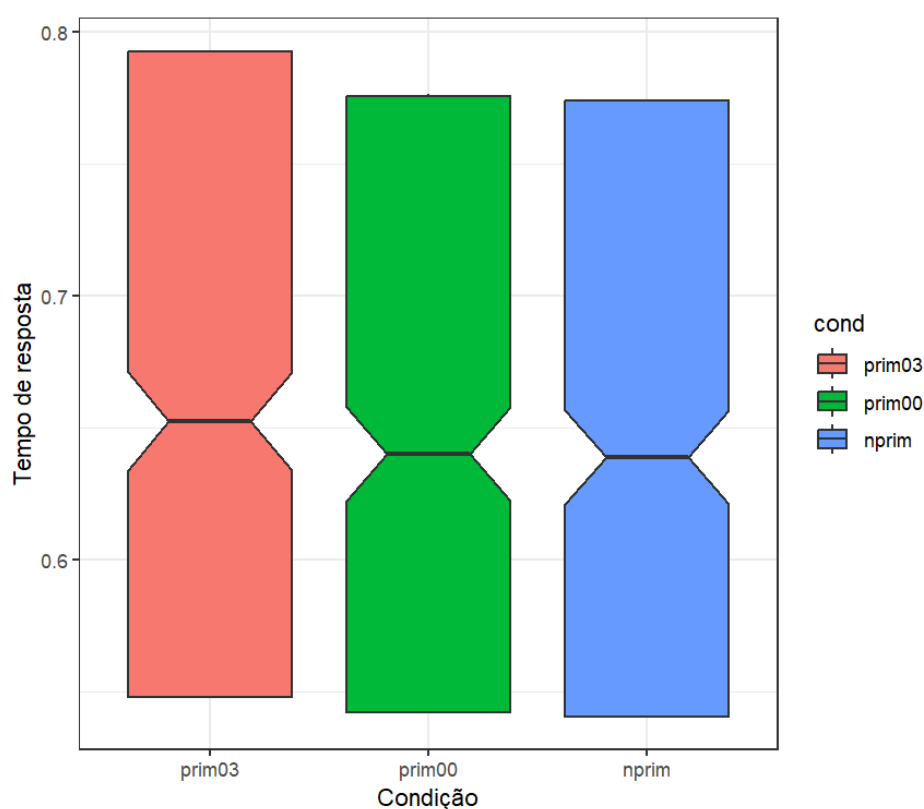
⁴¹ Exceto a identidade dos participantes, que não é associada aos arquivos de dados coletados e publicados.

das sentenças com *target* não-palavras, utilizadas para balancear as respostas. A análise foi feita em R, por meio do software RStudio.

Um dos participantes teve índice de acertos inferior a 70% na tarefa de decisão lexical; seus dados foram descartados. Ademais, nenhum dos demais voluntários teve mais de um erro nas respostas.

Os *outliers* foram tratados por *winsorização* (Turkey, 1962; Hoaglin, Iglewicz, Turkey, 1986), ou seja, foram substituídos pelo valor mais alto/baixo dentro da margem estabelecida para separar *outlier* — que, nesse caso, foram os limites do primeiro e terceiro quartis (Q1 e Q3). Apenas os RTs para as respostas corretas foram utilizados na análise; esse processo excluiu 17 (0,91%) dos 1872 dados. O gráfico e a tabela abaixo mostram as médias de RTs por condição com os *outliers* já ajustados.

Figura 7 — Média de RTs por condição

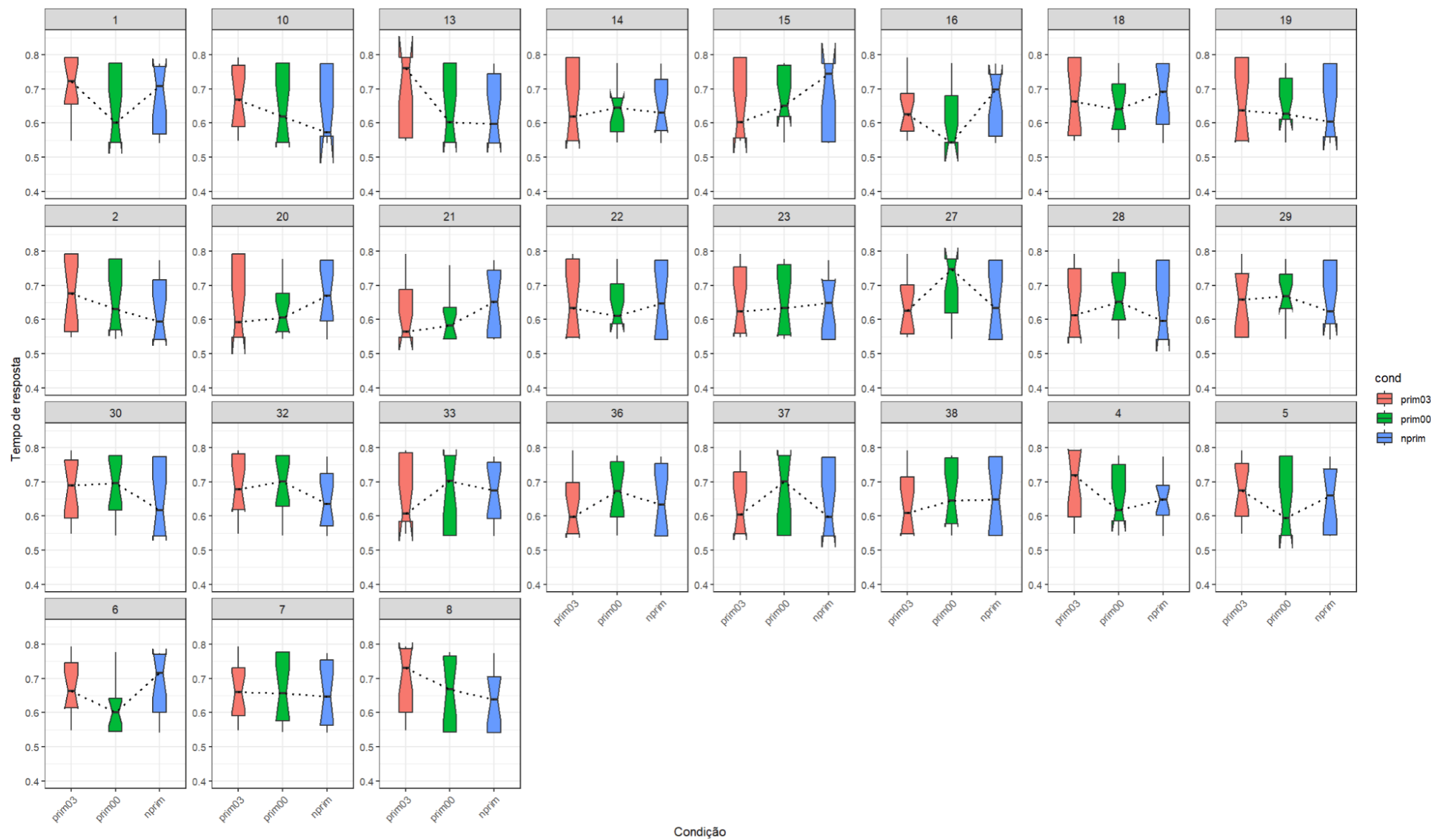


Fonte: elaborada pelo autor.

Nota: Desvio Padrão por condição: **prim00** = 0,283; **prim03** = 0,291; **nprim** = 0,230. Não há diferença significativa entre as três condições, considerando apenas as médias de RT.

Foi realizado um teste de Kruskal-Wallis para identificar eventuais diferenças significativas entre as condições; o resultado foi negativo ($H = 2$, $df = 2$, $p = 0,368$). Para uma análise mais detalhada, o mesmo teste foi realizado em cada item (sentença), comparando as condições (*prim00*, *prim03*, e *nprim*); essa análise identificou duas sentenças específicas com diferenças significativas, além de outras duas com diferença marginal. A figura abaixo ilustra os resultados gerais.

Figura 8 — Médias de RTs de condições por item



Interessante notar que o item 1, que traz a sentença que ilustra os exemplos deste trabalho⁴², apresentou resultados coerentes com a hipótese: o RT da condição *prim00* (*priming* com SOA 0 ms) foi significativamente mais baixo que as duas outras condições, *prim03* (*priming* com SOA 300 ms e *não-priming*; $p = 0,048$). Já o item 13 apresenta tempo médio de resposta significativamente mais lento para a condição *nprim* em relação às outras duas ($p = 0,043$), que não diferem significativamente entre si. Outra distinção está no item 8, com o item de *priming* e SOA 300 ms eliciando tempo médio de respostas mais lento que os demais; porém, o resultado tem significância marginal ($p = 0,054$). O item 2, por sua vez, também gerou RTs médios marginalmente mais longos para a condição com SOA 300 ms, se comparada à condição *não-priming* ($p = 0,075$).

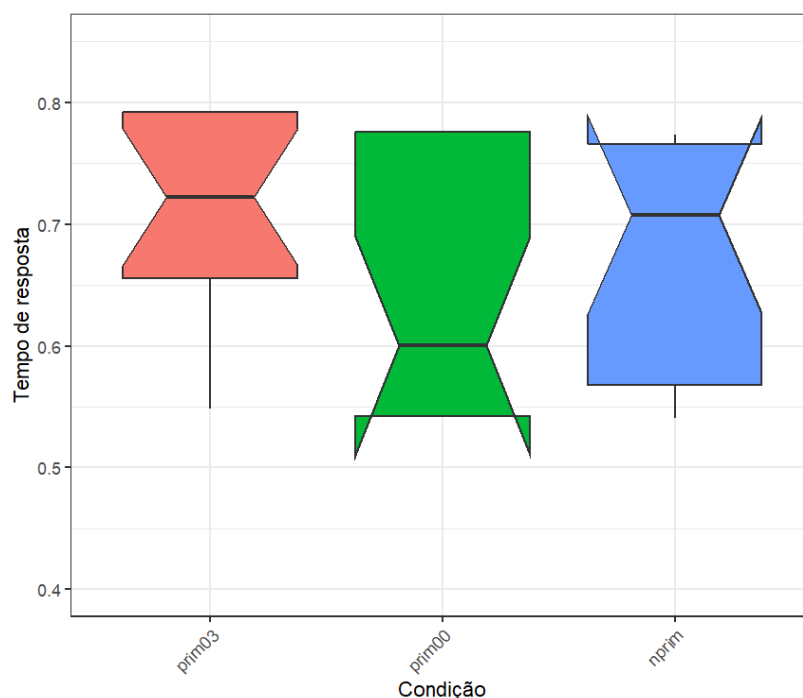
As demais condições apresentaram resultados variados e nenhuma diferença significativa no teste de Kruskal-Wallis, o que representa um desafio para a hipótese e o próprio design experimental. Abordarei essa questão com mais profundidade na seção a seguir.

5 DISCUSSÃO GERAL

O experimento foi elaborado para testar se PEFs podem ser ativadas durante a competição pelo acesso lexical em um estímulo auditivo em contexto sentencial. É notável que o conjunto total de dados traz dados que desafiam uma análise categórica do processo, isto é, resultados consistentes entre os itens experimentais que, associados, permitissem a interpretação de confirmação ou rejeição da hipótese. O que os resultados mostram é que a grande maioria dos itens não apresentou tempos de resposta com alguma correlação significativa às variáveis controladas. Em outras palavras, são RTs médios que, na maioria, não refletem a previsão de que há efeito de *priming* com SOA 0 ms a partir das PEFs, especificamente nesses contextos.

Ainda assim, devido à ocorrência de resultados significativos em dois itens e marginais em outros dois, minha interpretação é a de que a hipótese não pode ser descartada, mas o método precisa ser refinado. O *boxplot* abaixo ilustra os resultados registrados a partir do item 1:

⁴² A sentença das duas condições *priming* é “Lucas saiu para trabalhar com muito sono. Por isso, parou na padaria e bebeu uma xícara de café bem forte.”; a palavra embutida é *fé* em ‘café’, enquanto o *target* semanticamente associado é ‘igreja’ e a palavra controle é ‘coca’ em “Lucas saiu para trabalhar com muita preguiça. Por isso, comprou uma lata de coca para tomar no caminho”.

Figura 9 — Médias de RTs da condição 1

Na condição crítica de interesse (*prim00*), os tempos de resposta foram significativamente mais baixos que as demais condições. Conforme previa a hipótese alternativa, a resposta foi mais rápida quando o *target* é visualizado junto ao *onset* da PE; já com o alvo exibido 300 ms depois, o RT médio não foi significativamente diferente do RT para a condição de controle. Naturalmente, este item sozinho não é suficiente para sustentar a argumentação acerca da hipótese.

As diferenças de RT, significativas ou marginais, encontradas em outros itens, não corroboram exatamente a hipótese. Portanto, quaisquer propostas explicativas não seriam produtivas, já que novos argumentos exigiriam condições experimentais controladas que o sustentassem. Então, não entrarei nessa discussão.

Temos, então, a possibilidade de que a hipótese nula seja, de fato, mais adequada que a alternativa; se for o caso, as evidências indicam que o processamento do Português Brasileiro não envolve múltiplas ativações em diversos pontos das palavras, havendo algum mecanismo que inibe a ativação desses segmentos quando não coincidem com o *onset* da palavra portadora — mas há também espaço para uma discussão sobre desafios metodológicos que podem ter

influenciado os resultados, sobretudo se considerarmos que alguns itens apresentaram diferenças de RT.

Experimentos psicolinguísticos são frequentemente difíceis de controlar, sobretudo quando o *design* é relativamente inovador, como é o caso. Nesse sentido, um ponto central parece ser a relação semântica/associativa entre *priming* e *target*. Para testar essa hipótese, seria necessário executar uma nova versão do experimento, selecionando novos *targets* com base em um pré-teste, como o CLOZE, para avaliar quais são as palavras que, em média, são as primeiras associações feitas pelas pessoas ao serem expostas à palavra embutida.

Outros dois pontos que merecem atenção são a frequência lexical e eventuais interações com outras faculdades cognitivas, como a atenção e a memória de trabalho. Quanto à frequência, cabe salientar que, em uma configuração experimental como esta, equilibrar as frequências lexicais de todas as palavras envolvidas em cada comparação é um desafio enorme. A hipótese seria a de que palavras com frequência maior podem inibir com mais facilidade outras candidatas na competição pelo acesso lexical. Infelizmente, uma análise *post-hoc* para investigar a influência da frequência nos tempos de resposta não seria confiável, já que, apesar de ser uma variável controlada até certo ponto, este não foi um fator previsto no design experimental, carecendo então de um equilíbrio na distribuição dos itens que seria necessário para permitir comparar se um grupo de palavras portadoras com frequência mais alta, por exemplo, resulta em menos efeito de *priming*, enquanto palavras portadoras com frequência mais baixa são mais propensas a permitir a ativação de diversas concorrentes, favorecendo o *priming* a partir da PE.

O segundo fator é a memória de trabalho. O estudo de Zhang e Samuel (2015), por exemplo, sugere a influência do *cognitive load* na possível ativação de palavras embutidas. Tendo em vista que a duração do experimento foi de cerca de 15 minutos e que alguns participantes relataram certo cansaço “na segunda metade” ou mais próximo do final, é possível que haja uma diferença de performance devido à queda no engajamento e, conseqüentemente, uma limitação na memória de trabalho que seja suficiente para reduzir o *cohort* de competidoras ativadas durante o processamento. Mais uma vez, este é um fator que exige um controle experimental prévio, não sendo adequada a análise *post-hoc* de performance entre a primeira e a segunda metades do experimento.

Um terceiro ponto que merece rápida menção é a acurácia da execução fonológica dessas palavras. Como detalha Shatzman (2006), variações mínimas na duração dos segmentos fonológicos (e.g. consoante longa vs. consoante curta) podem funcionar como um critério fonológico adicional para limitar a ativação de competidoras. Afinal, se essa diferença mínima for suficiente para que o *parser* identifique se aquela realização específica da sílaba é a que comumente ocorre no *onset* de uma palavra, isso pode ser suficiente para inibir a ativação da PE quando esse mesmo segmento ocorre com duração mais curta, estando em sílaba final de outra palavra.

Em suma, a interpretação geral aqui é de que os resultados não são suficientes para negar a hipótese nula de que não é possível haver ativação de palavras embutidas finais alocadas em portadoras que fazem parte de um contexto sentencial. Ainda assim, há indícios que sugerem essa possibilidade e, portanto, há também motivação para a elaboração de novos experimentos que controlem de forma mais eficiente fatores como os abordados acima.

6 CONSIDERAÇÕES FINAIS

Este trabalho traz uma revisão geral da discussão acerca da ativação de palavras embutidas e propõe um experimento de *priming* transmodal com tarefa de decisão lexical. A literatura é marcada por inconsistências nos resultados de experimentos anteriores e um objetivo mais amplo aqui é avaliar se diferenças metodológicas podem ser as responsáveis por isso, nomeadamente o uso de SOA mais longo (acima de 250 ms) nos trabalhos que não registraram ativação de palavras embutidas.

É importante ressaltar que essa hipótese não aborda, neste primeiro momento, eventuais especificidades do processamento de cada língua. Quanto a isso, o estudo é relevante no sentido de levantar dados referentes ao PB, provendo mais material para a discussão. Além disso, o design experimental deixa em aberto a possibilidade de investigação futura acerca de outra escolha metodológica que varia na literatura e pode ter um papel decisivo na eventual ativação de palavras embutidas: o processamento de palavras soltas *versus* em contexto sentencial.

As previsões iniciais aqui eram:

1. Serão registrados RTs menores na condição SOA 0 ms, mesmo com contexto sentencial favorável à palavra portadora;
2. Não serão registrados RTs menores para os mesmos *targets* quando as mesmas sentenças de *prime* na condição SOA 300 ms.

Não houve confirmação dessas hipóteses, mas não descarto a possível ativação dessas palavras embutidas. Fica em aberto, então, a realização de novos testes com controles ajustados aos pontos levantados. Dito isso, é notável que a ativação de palavras embutidas não parece ser fruto de uma regra categórica que pode prever especificamente a ativação em qualquer cenário. Em sintonia com os demais estudos, é possível que o fenômeno da ativação seja um subproduto da estratégia linguística para lidar com a competição lexical e que esteja condicionado a diversos fatores. Cabe elucidar, então, até que ponto é possível controlar experimentalmente este cenário para mapear esses fatores e as condições necessárias e suficientes para a ativação das palavras embutidas finais.

Um ponto relacionado ao design experimental que merece atenção é o controle de relação associativa entre *prime* e *target*. A seleção de estímulos foi feita por mim, o que deixa em aberto a possibilidade de que os participantes não façam a mesma relação associativa no léxico, sobretudo de forma inconsciente. É plausível, então, que um teste de CLOZE seja realizado para buscar palavras-alvo possivelmente mais eficientes que as adotadas aqui.

Ademais, há a questão do viés das sentenças. A decisão de elaborar conjuntos de sentenças com potencial efeito de predição lexical, controlado por um teste de CLOZE, acrescentava um fator que operava contra a hipótese: se fosse registrado efeito de *priming*, confirmando a hipótese do trabalho, haveria um indício a favor da primazia da ativação das PEFs, sendo a inibição total por efeito de contexto sentencial uma questão secundária ou inexistente. Um experimento complementar que tratasse do efeito como variável controlada poderia elucidar a questão. Contudo, essa configuração se mostrou uma falha no design experimental. Com os dados inconclusivos, o efeito se tornou um fator adicional entre os que podem ter impedido a ativação e o *priming*. Logo, sem uma outra variável para comparação (*i.e.* sentenças sem efeito de predição), tornou-se inviável apontar com mais precisão a(s) causa(s) da ausência de *priming*. Em outras palavras, o efeito de

predição pode ter impedido a ativação, algo que exige um segundo experimento para averiguação; outra explicação plausível é a descrita pela hipótese nula — de que as palavras embutidas finais não são ativadas.

Por fim, não podemos descartar, é claro, outras possíveis limitações do design experimental, como o nível de controle das características fonológicas das palavras. Como mostram estudos anteriores (e.g. Marslen-Wilson, 1993; Shatzman, 2006), a sensibilidade à incompatibilidade (*mismatch*) é alta: pequenas variações nas características fonológicas da sílaba que coincide com a palavra embutida — e.g. duração de uma vogal — podem ser suficientes para inibir a ativação; não tendo adotado um controle desse tipo, considero que esse fator tenha um potencial disruptivo para o experimento. Se uma palavra como ‘céu’ em *píncel* tem a realização do [s] minimamente diferente em cada situação — uma duração mais curta em *píncel*, por exemplo —, é possível que a ativação da PEF não ocorra. Assim, um experimento com esse controle fonológico⁴³ seria produtivo para investigar o nível de sensibilidade fonológica para ativação de palavras concorrentes em Português Brasileiro e, a partir disso, estabelecer critérios mais palpáveis para um estudo sobre a ativação de palavras embutidas e a competição pelo acesso lexical.

⁴³ Isto é, comparando o comportamento dos participantes diante de uma mesma palavra, ora com a consoante que compõe o *onset* da sílaba final tendo realização *natural* (digamos, mais curta se comparada à realização no *onset* da palavra monossílaba), ora com uma realização modificada artificialmente (mais longa, como na monossílaba), ambas servindo de *prime* para a subsequente visualização de imagens na tela, nos mesmos moldes do estudo de rastreamento ocular de Shatzman (2006, p. 57-72), para compararmos os tempos de fixação nas imagens após cada tipo de estímulo. Isso exige, é claro, uma descrição mais detalhada das características fonológicas dessas palavras. Seria interessante, inicialmente, focar na questão fonológica por meio de um estudo com palavras soltas; se os resultados forem sólidos, a favor ou contra a hipótese do alto nível de sensibilidade fonológica, seria possível expandir o estudo para os níveis da sentença e do discurso a fim de investigar possíveis influências de outros níveis de processamento (sintático, pragmático etc.) e de fatores extralinguísticos (atenção, memória de trabalho etc.) no acesso lexical.

REFERÊNCIAS

BOERSMA, P.; WEENINK, D.. Praat: doing phonetics by computer. Versão 6.4.27, 27 jan 2025. Disponível em <http://www.praat.org/>

BOWERS, J. S.; COLIN, J. D.; MATTYS, S. L.; DAMIAN, M. F.; HANLEY, D. The activation of embedded words in spoken word identification is robust but constrained: Evidence from the Picture-Word Interface paradigm. **Journal of Experimental Psychology: Human Perception and Performance**, v. 35, n. 5, p. 1585-1597, 2009.

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

CARAMAZZA, A.; LAUDANNA, A.; ROMANI, C. Lexical Access and Inflectional Morphology. **Cognition**, v. 28, n. 3, p. 297-332, abr. 1988.

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. 345-348, jan. 2001.

ESTIVALET, GUSTAVO LOPEZ. Variáveis lexicais e ortográficas no acesso lexical das palavras do português brasileiro. **REVISTA LINGÜÍSTICA**, v. 16, p. 264-277, 2020.

ESTIVALET, G. L.; HARTMANN, N. S.; MARQUIAFÁVEL, V.; LUKASOVA, K.; CARTHERY-GOULART, M. T.; ALUÍSIO, S. M. LexPorBR Infantil: uma base lexical tripartida e com interface Web de textos ouvidos, produzidos, e lidos por crianças. *In*: SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 12., 2019, Salvador. **Anais [...]**. Porto Alegre: SBC, 2019. p. 190-199.

FODOR, J. A. **Modularity of mind**. Cambridge: MIT Press, 1983.

FORSTER, K. I. **Levels of processing and the structure of the language processor**. *In*: COOPER, W. E.; WALKER, E. (ed.). Hillsdale: Erlbaum, 1979.

GOW, D. W.; GORDON, P. C. Lexical and prelexical influences on word segmentation: Evidence from priming. **Journal of Experimental Psychology-Human Perception and Performance**, v. 21, n. 2, p. 344-359, abr. 1995.

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

HOAGLIN, D.; IGLEWICZ, B.; TURKEY, J. W. Performance of some resistant rules for outlier labeling. **Journal of American Statistical Association**, v. 81, p. 991-999, 1986.

ISEL, F.; BACRI, N. Spoken-word recognition: The access to embedded words. **Brain and Language**, v. 68, n. 1-2, p. 61-67, jun. 1999.

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

KIRMAN, J. H. Tactile apparent movement: The effects of number of stimulators. **Journal of Experimental Psychology**, v. 103, n. 6, p. 1175-1180, 1974.

LIBBEN, G. The quantum metaphor and the organization of words in the mind. **Journal of Cultural Cognitive Science**, v. 1, p. 49-55, mar. 2017.

LOPES, Julia Cataldo. **Acesso lexical**: avaliando as especificidades das rotas semântica e morfológica. 2020. 61 f. Monografia (Licenciatura em Letras: Português-Francês), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2020.

LUCE, P. A.; CLUFF, M. S. Delayed commitment in spoken word recognition: Evidence from cross-modal priming. **Perception & Psychophysics**, v. 60, p. 484-490, jan. 1998.

MACIZO, P.; VAN PETTEN, C.; O'ROURKE, P. L. Semantic access to embedded words? Electrophysiological and behavioral evidence from Spanish and English. **Brain and Language**, v. 123, n. 2, p. 113-124, nov. 2012.

MARSLEN-WILSON, W. D. Functional parallelism in spoken word-recognition. **Cognition**, v. 25, n. 1-2, p. 71-102, mar. 1987.

MARSLEN-WILSON, W. Issues of process and representation in lexical access. In G. T. M. Altmann & R. Shillcock (Eds.), **Cognitive models of speech processing: The Second Sperlonga Meeting**, Lawrence Erlbaum Associates Publishers, p. 187-210, 1993.

MARSLEN-WILSON, W. Issues of process and representation in lexical access. In: ALTMANN, G. T. M.; SHILLCOCK, R. (ed.). **Cognitive models of speech processing: The Second Sperlonga Meeting**. Londres: Psychology Press, 1993. p. 187-210.

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.

MARSLEN-WILSON, W. D.; WELSH, A. Processing interactions and lexical access during word recognition in continuous speech. **Cognitive Psychology**, v. 10, n. 1, p. 29-63, jan. 1978.

MCCLELLAND, J. L.; ELMAN, J. L. The TRACE model of speech perception. **Cognitive Psychology**, v. 18, n. 1, p. 1-86, jan. 1986.

MCQUEEN, J. M.; CUTLER, A.; BRISCOE, T.; NORRIS, D. Models of continuous speech recognition and the contents of the vocabulary. **Language and Cognitive Processes**, v. 10, n. 3-4, p. 309-331, 1995.

NORRIS, D. Shortlist: A connectionist model of continuous speech recognition. **Cognition**, v. 52, n. 3, p. 189-234, 1994.

NORRIS, D.; CUTLER, A.; McQUEEN, J. M.; BUTTERFIELD, S. Phonological and conceptual activation in speech comprehension. **Cognitive Psychology**, v. 53, n. 2, p. 146-193, set. 2006.

NORRIS, D., McQUEEN, J. M.; CUTLER, A. Competition and segmentation in spoken-word recognition. **Journal of Experimental Psychology: Learning, Memory, and Cognition**, v. 21, n. 5, p. 1209-1228, set. 1995.

OPEN SCIENCE TOOLS LTD. **Pavlov** (R), [s. l.], 2020. Disponível em: <https://pavlov.org>. Acesso em: 6 ago. 2024.

OPEN SCIENCE TOOLS LTD. **PsychoPy** (R): versão 2020.2.3, [s. l.], 2020. Aplicativo de computador. Disponível em: <https://psychopy.org/>. Acesso em: 6 ago. 2024.

PINKER, S. Rules of Language. **Science**, v. 253, n. 5019, p. 530-535, ago. 1991.

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

PYLKKÄNEN, L.; MARANTZ, A. Tracking the time course of word recognition with MEG. **Trends in Cognitive Sciences**, v. 7, n. 5, p. 187-189, maio 2003.

PYLKKÄNEN, L.; STRINGFELLOW, A.; FLAGG, E.; MARANTZ, A. A neural response sensitive to repetition and phonotactic probability: MEG investigations of lexical access. *In*: INTERNATIONAL CONFERENCE BIOMAGNETISM, 12., 2000, Espoo. **Anais** [...]. Espoo: Helsinki University of Technology, 2000. p. 363-367.

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, n. 1-3, p. 666-78, abr./jun. 2002.

SALGADO, B. O. **Existe fim em marfim?** Um estudo da relação de palavras dentro de palavras no acesso lexical. 2014. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2014.

SCHREUDER, R.; BAAYEN, R. H. Modeling morphological processing. *In*: FELDMAN, L. B. (ed.). **Morphological aspects of language processing**. Hillsdale: Lawrence Erlbaum Associates, 1995. p. 131-154.

SHATZMAN, K. B. **Sensitivity to detailed acoustic information in word recognition**. 2006. Dissertação (Doutorado em Psicolinguística) – Universidade de Nijmegen, Nijmegen, 2006.

SHILLCOCK, R. C. Lexical hypotheses in continuous speech. *In*: ALTMANN, G. T. M. (ed.). **Cognitive models of speech processing**: Psycholinguistic and computational perspectives. Cambridge: MIT Press., 1990. p. 24-49.

SOUZA FILHO, N. S. **A small but significant effect**: lexical prediction in a self-paced reading study. 2021. 106 f. Dissertação (Mestrado em Estudos da Linguagem) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal do Rio Grande do Norte, Natal, 2021.

SPIVEY, M. **The continuity of mind**: Oxford Psychology Series. Oxônia: Oxford University Press, 2007.

STOCKALL, L.; MARANTZ, A. A single route, full decomposition model of morphological complexity: MEG evidence. **The Mental Lexicon**, v. 1, p. 85-123, 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, fev. 1979.

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

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.

TRASK, R. L. **Dicionário de Linguagem e Linguística**. Tradução de Rodolfo Ilari. Revisão técnica de Ingedore Koch, Thaís Cristófaros Silva. 2. ed. São Paulo: Contexto, 2008. p. 218.

TURKEY, J. W.. The future of data analysis. **The Annual of Mathematical Statistics**, v. 33, n. 1, p. 1-67, 1962

VAN ALPHEN, P. M.; VAN BERKUM, J. J. A. Is there pain in champagne? Semantic involvement of words within words during sense-making. **Journal of Cognitive Neuroscience**, v. 22, n. 11, p. 2618-2626, nov. 2010.

VAN ALPHEN, P. M.; VAN BERKUM, J. J. A. Semantic involvement of initial and final lexical embeddings during sense-making: The advantage of starting late. **Frontiers in Psychology**, v. 3, jun. 2012.

VROOMEN, J.; DE GELDER, B. Activation of embedded words in spoken word recognition. **Journal of Experimental Psychology: Human Perception and Performance**, v. 23, n. 3, p. 710-720, 1997.

WEINGARTNER, K. M.; JUHASZ, B. J.; RAYNER, K. Lexical embeddings produce interference when they are morphologically unrelated to the words in which they are

contained: Evidence from eye movements. **Journal of Cognitive Psychology (Hove)**, v. 24, n. 2, Hove, p. 179-188, 2012.

ZHANG, X.; SAMUEL, A. G. The activation of embedded words in spoken word recognition. **Journal of Memory and Language**, v. 79, p. 53-75, fev./abr. 2015.

ZHOU, J.; MA, G.; LI, X.; TAFT, M. The time course of incremental word processing during Chinese reading. **Reading and Writing**, v. 31, p. 607-625, 2018.

APÊNDICE A – LISTA DE ESTÍMULOS

Código	Sentença (prime)	PRIME	FreqZ S EW	FreqZS carrier	target
s1	Lucas saiu para trabalhar com muito sono. Por isso, parou na padaria e bebeu uma xícara de café bem forte.	caFÉ	4,5148	5,3645	igreja
n1	Lucas saiu para trabalhar com muita preguiça. Por isso, comprou uma lata de coca para tomar no caminho.	COCA	NA	4,0317	igreja
s2	O casal achou o atendimento do restaurante muito bom. Na hora de pagar a conta, lembraram de dar uma gorjeta generosa ao garçom que tinha sido muito eficiente.	garÇOM	4,808	4,3204	rádio
n2	O casal achou o atendimento do restaurante muito bom. Na hora de ir embora, lembraram de perguntar ao dono os ingredientes do prato.	DONO	NA	4,6935	rádio
s4	As crianças aguardavam ansiosas pela chegada do Papai Noel. Ao anoitecer, elas olharam para o céu procurando ver as renas puxando o seu trenó em direção à chaminé.	treNÓ	3,8627	3,786	corda
n4	As crianças aguardavam ansiosas pela chegada do Papai Noel. Ao amanhecer, elas correram para a árvore enfeitada e cada uma abriu seu embrulho para ver o presente.	EMBRULHO	NA	3,3952	corda
s5	Depois de anos de uso, a cama de Jorge já não estava tão confortável. Para evitar um problema nas costas, ele foi até uma loja comprar um novo colchão mais macio e anatômico.	colCHÃO	4,9955	3,9855	piso
n5	Depois de anos de uso, a cama de Jorge já não estava tão confortável. Para evitar gastar à toa, ele decidiu fazer um novo teste e comprou outro travesseiro.	TESTE	NA	4,9846	piso

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
s6	O eletricista da companhia de luz sempre trabalha em lugares altos. É importante que ele use o cinto amarrado ao poste para não correr o risco de cair daqueles lugares altos.	caIR	6,1141	4,9796	volta
n6	O eletricista da companhia de luz sempre trabalha em lugares altos. É importante que ele use o cinto amarrado ao poste para não correr o risco de levar choque.	LEVAR	NA	5,4931	volta
s7	Gabriel gostava dos sabores mais azedos. Assim que chegou na praia, ele pediu uma deliciosa caipirinha de limão no quiosque mais próximo.	liMÃO	5,3587	4,137	dedo
n7	Gabriel gostava dos sabores mais azedos. Assim que chegou em casa, ele pegou uma jarra de água para fazer um suco de maracujá.	ÁGUA	NA	5,3412	dedo
s8	Depois de horas produzindo o retrato, o artista finalmente terminou o esboço feito a lápis. O próximo passo era pegar a paleta de tintas para usar com o pincel na finalização da obra.	pinCEL	4,9211	3,4047	chuva
n8	Depois de horas produzindo o retrato, o artista finalmente terminou o esboço feito a lápis. O próximo passo era começar o desenho do bebê que brincava com um peixe no canto da imagem.	PEIXE	NA	4,8579	chuva
s10	A amiga de Marcos disse que, para ele arrumar uma namorada, deveria atualizar suas redes sociais. Então, ele entrou no Facebook e escolheu uma foto mais bonita para colocar no seu perfil de uso pessoal.	perFIL	4,3937	4,144	poste
n10	A amiga de Marcos disse que, pra ele arrumar uma namorada, deveria atualizar suas redes sociais. Então, ele tirou uma foto mais bonita para colocar no seu Facebook e no Instagram.	FACEBOOK	NA	4,1635	poste
s13	A fazenda já estava com a terra arada e adubada. Assim que as sementes chegaram, os agricultores foram ao campo e começaram o plantio das mudinhas da estação.	planTIO	5,2319	2,546	família

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
n13	A fazenda já estava com a terra arada e adubada. Assim que as sementes foram plantadas, eles se reuniram e relembrou o plano para irrigar a área.	PLANO	NA	5,2768	família
s14	A psicanalista sempre aguardava Juliano na porta do consultório. Depois se sentava e aguardava em silêncio enquanto ele se deitava no divã para começarem a sessão.	divÃ	4,7398	3,0803	carro
n14	A psicanalista sempre aguardava Juliano na porta do consultório. Depois aguardava em silêncio enquanto ele se sentava no sofá para começarem a sessão.	SOFÁ	NA	4,7031	carro
s15	Meu primo me ensinou a jogar xadrez e depois de dois meses eu consegui dar um xeque-mate nele. Mas percebemos que eu tinha feito um lance proibido, por isso a jogada não valeu como eu imaginava.	vaLEU	4,5052	4,7831	livro
n15	Meu primo me ensinou a jogar xadrez e depois de dois meses eu consegui dar um xeque-mate nele. Mas percebemos que eu tinha feito um lance proibido, então tive que aceitar não vencer naquele dia.	VENCER	NA	4,674	livro
s16	Quando assinei o canal de filmes, precisava escolher entre pagar a cada trinta dias ou adiantar o valor do ano inteiro. Como eu não tinha dinheiro para um pagamento anual, optei pelo pagamento mensal que cabia no meu bolso.	menSAL	4,4375	3,3911	cozinha
n16	Quando assinei o canal de filmes, precisava escolher entre pagar a cada trinta dias ou adiantar o valor do ano inteiro. Como eu não tinha dinheiro para um pagamento anual, optei por um plano familiar para dividir com meus pais.	FAMILIAR	NA	4,5533	cozinha
s18	Durante sua estadia em Manaus, João fotografou diversos animais silvestres. A sua preferida é de um macaco pendurado em algo que parecia uma corda, mas na verdade era um cipó que vinha lá do alto até o chão.	ciPÓ	4,3511	2,574	sujeira
n18	Durante sua estadia em Manaus, João fotografou diversos animais silvestres. A sua preferida é de um macaco pendurado em algo que parecia uma rede, mas na verdade era uma árvore cheia de galhos secos.	ÁRVORE	NA	4,8132	sujeira

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
s19	O elevador do prédio estava quebrado. Joana decidiu usar a escada, mas torceu o pé no primeiro degrau logo que abriu a porta.	deGRAU	4,1676	3,6942	óculos
n19	O elevador do prédio estava quebrado. Joana subiu pela escada, mas esqueceu a bolsa no primeiro andar enquanto falava com o porteiro.	ANDAR	NA	4,6767	óculos
s20	O filme que assistimos contava a luta entre duas personagens num reino encantado. A princesa era a heroína, enquanto uma bruxa malvada era a grande vilã que tentava atormentá-la.	viLÃ	3,7246	3,1916	ovelha
n20	O filme que assistimos contava a luta entre duas personagens num reino encantado. A corajosa princesa procurava uma grande arca do tesouro em uma ilha mágica.	ARCA	NA	3,7416	ovelha
s21	A erupção pegou a população toda de surpresa. Quando o alarme soou pela cidade, as pessoas foram pra rua olhar se realmente havia lava saindo do vulcão lá depois da floresta.	vulCÃO	4,8988	3,8242	gato
n21	A erupção pegou a população toda de surpresa. Quando o alarme soou pela cidade, as pessoas foram pra rua olhar se havia carros saindo da estrada para buscar abrigo.	ESTRADA	NA	4,8698	gato
s22	Bárbara comprou um videogame pela internet e o vendedor prometeu entregá-lo no dia seguinte. Mesmo assim, ela precisou esperar um mês e só ontem o produto chegou pra ela jogar com os amigos.	cheGOU	4,0484	5,2482	bola
n22	Bárbara comprou um videogame pela internet e o vendedor prometeu entregá-lo no dia seguinte. Mesmo assim, ele demorou mais de um mês e a garota desistiu da compra.	DESISTIU	NA	4,1412	bola
s23	Durante as férias, visitei a famosa exposição de joias antigas. No centro da exposição, uma estátua que representava uma antiga rainha ostentava pendurado no pescoço um valioso colar cheio de pedras preciosas.	coLAR (n)	4,6316	4,4235	casa

Código	Sentença (prime)	PRIME	FreqZ S EW	FreqZS carrier	target
n23	Durante as férias, visitei a famosa exposição de joias antigas. No centro da exposição, uma estátua que representava uma antiga rainha ostentava na cabeça uma valiosa coroa cheia de pedras preciosas.	COROA	NA	4,3005	casa
s27	No filme de ação, o herói correu no meio do tiroteio para salvar o amigo. Um dos vilões atirou nele à queima roupa, mas por muita sorte a bala passou de raspão sem atingi-lo em cheio.	rasPÃO	4,6627	2,8884	manteiga
n27	No filme de ação, o herói correu no meio do tiroteio para salvar o amigo. Um dos vilões conseguiu escapar, mas por muita sorte o herói o viu de perto e conseguiu identificar seu rosto.	PERTO	NA	5,4691	manteiga
s28	Quando tirou o maço de cigarro do bolso, Rogério não tinha visto o aviso na parede do quarto. Outro hóspede do hotel notou a fumaça no corredor e foi avisá-lo que lá era proibido fumar dentro do prédio sem ser autorizado.	fuMAR	4,8245	4,5849	praia
n28	Quando tirou o violão da bolsa, Rogério não tinha visto o aviso na parede do quarto. Outro hóspede do hotel notou a música no corredor e foi avisá-lo que lá era proibido tocar dentro do prédio sem ser autorizado.	TOCAR	NA	5,02	praia
s29	Laura sabia que não devia odiar o amigo por ter esquecido do seu aniversário. Por mais que ele tivesse pisado na bola, devemos aprender a não guardar rancor das pessoas que gostamos.	ranCOR	4,8007	3,7501	azul
n29	Laura sabia que não devia odiar o amigo por ter esquecido do seu aniversário. Por mais que ele tivesse pisado na bola, devemos aprender a não semear ódio entre as pessoas e em nós mesmos.	ÓDIO	NA	4,2644	azul
s30	Camila achou o professor novo muito estranho. Mas suas amigas disseram que não havia nada de diferente, ele parecia completamente normal para o padrão da escola.	norMAL	5,6106	5,0976	ruim
n30	Camila achou o professor novo muito estranho. Mas suas amigas disseram que não havia nada de diferente, inclusive ele parecia incrivelmente legal para o padrão da escola.	LEGAL	NA	5,7443	ruim

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
s32	Quando eu terminei a prova, tive certeza que acertei 8 das 10 questões. As outras duas eu acho que errei por algum detalhe.	eRREI	5,0593	4,0047	coroa
n32	Quando eu terminei a prova, tive certeza que acertei 8 das 10 questões. As outras duas eu até que tentei mas não sabia a resposta.	TENTEI	NA	4,9778	coroa
s33	Matheus esqueceu de levar roupa de cama para seu chalé na praia. Precisou usar o travesseiro sem fronha e o colchão sem lençol por falta de opção.	lenÇOL	4,9955	3,6636	nuvem
n33	Matheus esqueceu de levar roupa de cama para seu chalé na praia. Além disso, a sala estava sem lâmpadas e o banheiro sem papel por falta de reposição.	PAPEL	NA	5,0809	nuvem
s36	O filme em cartaz no cinema era sobre um famoso ladrão de joias raras. Porém, dessa vez era uma obra de arte o item que ele planejava roubar de um museu francês.	rouBAR	5,0659	4,8584	cerveja
n36	O filme em cartaz no cinema era sobre um famoso ladrão de jóias raras. Porém, dessa vez ele encontrou um policial que planejava prender o protagonista antes que ele cometesse o crime.	PRENDER	NA	4,4521	cerveja
s37	Armando sempre ia na casa de repouso para velhinhos com sua mãe, para visitar os pais dela. O rapaz já conhecia as enfermeiras do lugar, então sabia que estavam cuidando bem dos seus avós e dos demais idosos.	aVÓS	5,0388	3,8849	falar
n37	Armando sempre ia na casa de repouso para velhinhos com sua mãe, para visitar os pais dela. O rapaz conhecia muito bem seu avô, então sabia que ele estaria contando suas piadas aos outros idosos.	PIADAS	NA	4,5043	falar
s38	Felipe disse à sua dentista que estava com a sensação de ter mal hálito. Ela sugeriu que além da escova e do fio dental, ele poderia usar o enxaguante bucal da Colgate.	buCAL	3,9865	3,2024	branco
n38	Felipe disse à sua dentista que estava com a sensação de ter mal hálito. Ela disse que poderia ser algo no estômago, então o encaminhou para uma consulta com o clínico geral do hospital.	GERAL	NA	4,5835	branco

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
dist01	Os piratas desembarcaram numa ilha deserta e seguiram até o ponto indicado no mapa com uma cruz. Nesse lugar, estava enterrado um tesouro em um baú antigo.	tesouro	NA	4,4821	notícia
dist02	Mateus correu para a biblioteca quando percebeu que um de seus empréstimos estava vencido há vários dias. Ele precisava devolver o livro o mais rápido possível.	livro	NA	5,2725	mesa
dist03	A professora demorou mais que o previsto para corrigir as provas de uma de suas turmas. Logo que terminou, ela abriu o sistema e inseriu as notas desses alunos.	notas	NA	4,5349	truque
dist04	Felipe havia sido um bom menino durante o ano e estava doido para saber o que ganharia de Natal. Assim que acordou na manhã de Natal, correu para abrir os presentes embaixo da árvore.	presentes	NA	4,6749	legenda
dist05	Jorge foi mandar uma mensagem para a namorada quando notou que seu bolso estava vazio. Ele praguejou muito quando percebeu que tinha esquecido o celular no trabalho.	celular	NA	4,8969	energia
dist06	Uma tempestade com muito vento começou a cair, molhando todo o apartamento. Hélio pulou do sofá e correu para fechar as janelas da sala.	janelas	NA	4,315	ritmo
dist07	Júlio adorava pescar e toda semana ia ao lago da sua fazenda. Chegando lá, ele colocava a isca no anzol e ficava horas naquele lugar tranquilo.	anzol	NA	3,4656	aura
dist08	Cristina fazia musculação porque tinha o plano de ficar com o corpo em forma para o verão. Por isso, todo dia ela frequentava a academia antes de ir trabalhar.	academia	NA	4,5149	roubo
dist09	Jonas estava cada vez mais levando a sério seu gosto por fotografia, então decidiu investir em um equipamento profissional. Ele juntou dinheiro por alguns meses para comprar uma câmera com algumas lentes e uma bolsa protetora.	câmera	NA	4,7501	herói
dist10	João estava suado e fedido depois de correr por mais de cinco quilômetros. Chegando em casa, ele correu até o banheiro para tomar um banho e ficar cheiroso.	banho	NA	4,9614	dança

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
dist11	Felipe precisava estudar, mas seu vizinho estava fazendo muito barulho. O barulho era tanto que ele só pensava em tapar os ouvidos para não arrumar confusão.	ouvidos	NA	4,4916	torre
dist12	Márcia ia viajar para São Paulo e viu que lá estava ocorrendo um surto de sarampo. Por isso, ela foi ao posto de saúde para tomar uma vacina antes de pegar o avião.	vacina	NA	3,5874	bafo
dist13	O jardineiro sempre prefere começar o trabalho pelo chão e depois cuidar das plantas altas. Por isso, assim que chega, ele logo pega o cortador para aparar a grama para depois molhar os vasos.	grama	NA	4,3337	osso
dist14	Carlos gostou muito de um apartamento, mas desistiu de comprá-lo porque o prédio não tinha elevador. A mãe dele já era idosa e não conseguia subir as escadas até o segundo andar.	escadas	NA	4,3334	trato
dist15	Maíra estava completamente exausta e, quando chegou em casa, ficou na sala, pois nem conseguiu ir para o quarto. Ela largou a bolsa no chão e deitou no sofá sem nem trocar de roupa.	sofá	NA	4,7031	dúvida
dist16	O médico suspeitava que Fátima estava com tuberculose, pois ela estava tossindo muito. Para confirmar, ele solicitou alguns exames e marcou um retorno.	exames	NA	4,5286	sono
dist17	Baltazar decidiu criar gado em sua fazenda, mas os animais estavam fugindo porque o terreno não era delimitado. Por isso, ele comprou madeira e arame para construir uma cerca ao redor do seu terreno.	cerca	NA	4,7547	peso
dist18	O Pontão do Lago Sul, em Brasília, é conhecido por abrigar restaurantes caríssimos. Por isso, os manobristas são muito cuidadosos quando estacionam os carros dos clientes.	carros	NA	4,8388	banda
dist19	No dia da procissão de Aparecida, Isabela percorreu todo o caminho até a igreja de joelhos. Ela fez isso porque seu filho foi curado e era necessário cumprir a promessa que ela tinha feito pra santa.	promessa	NA	4,4131	médio

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
dist20	Gabriel estava organizando as comidas para o aniversário de sua filha. Como não sabia fazer nada usando o forno, decidiu ir a uma confeitaria para encomendar um bolo bem gostoso e colorido.	bolo	NA	4,9251	prisão
dist21	Suzana estava pronta para assinar o contrato, mas percebeu que não tinha trazido nada para escrever. Ela perguntou se alguém na sala poderia lhe emprestar uma caneta por um instante.	caneta	NA	4,3998	solo
dist22	Os lanches noturnos de Paulo estavam ficando cada vez mais frequentes e problemáticos. Na última semana, ele acordou em cinco dias seguidos para assaltar a geladeira no meio da noite.	assaltar	NA	3,4331	vigário
dist23	Mário sabe se virar na cozinha, mas ainda não aprendeu a cortar todos os alimentos corretamente. Por isso, ele sempre pede para o açougueiro cortar sua carne antes de embalar.	carne	NA	5,0029	grupo
dist24	Depois de se ensaboar e se lavar com água quente, Aline precisava se secar. Ela abriu o box e bateu até encontrar a toalha pendurada na parede.	toalha	NA	4,3201	sócio
dist25	Luana estava decidindo o que vestir e notou que não estava usando nada nos dedos. Ela abriu sua caixa de jóias para escolher um anel que combinasse com a roupa.	anel	NA	4,8139	povo
dist26	O professor havia esquecido de imprimir a lista de frequência e pediu para que os alunos fizessem uma. Um dos alunos abriu o caderno e arrancou de lá uma folha para todos assinarem.	folha	NA	4,1949	neto
dist27	Rute adorava música e fazia as próprias composições com seus instrumentos. A menina vivia na varanda dedilhando as cordas de um violão ou de um cavaquinho.	violão	NA	3,942	fralda
dist28	A rotina de Marcelo é a mesma todas as manhãs, e ele já está acostumado com isso. Todos os dias, ele deixa os filhos na escola e vai para o trabalho antes das oito da manhã.	trabalho	NA	5,9049	razão

Código	Sentença (prime)	PRIME	FreqZS EW	FreqZS carrier	target
dist29	Júlia tem um grau de miopia muito alto e não consegue ver nada de longe. Por isso, sempre que acorda, ela logo procura seus óculos ou coloca suas lentes.	óculos	NA	4,6382	juntar
dist30	Sérgio estava vestindo um terno porque tinha uma entrevista numa empresa muito importante. Depois de vestir a calça, ele colocou a meia e sentou para calçar os sapatos mais elegantes que tinha.	sapatos	NA	4,944	honra
dist31	Os assaltantes planejaram o assalto ao banco da cidade durante vários meses. Como eles queriam levar o dinheiro, pediram para um funcionário abrir o cofre que ficava nos fundos.	cofre	NA	4,4138	marinha
dist32	A hora de cantar os parabéns é sempre especial numa festa de aniversário. Todos batem palmas e, no fim da música, o aniversariante assopra as velas antes de cortar o bolo.	velas	NA	4,2883	gene
dist33	Rodrigo estava exausto de tanto trabalho e das horas que passava confinado no escritório. Ele não via a hora de dezembro chegar para curtir as férias na praia com a esposa.	férias	NA	4,8022	gelo
dist34	O chef estava preparando sua famosa macarronada para o almoço de domingo. O primeiro passo era picar os tomates para fazer o molho com sua receita secreta.	molho	NA	4,6237	canção
dist35	Fábio comprou cerveja e amendoim para acompanhar a final da copa do mundo com os amigos. Na hora do jogo, ele sentou no sofá e ligou a tevê no canal do jogo.	sentou	NA	3,8007	paletó

APÊNDICE B – ANÁLISE DO PRÉ-TESTE DE CLOZE

Resposta	Índice de ocorrência (%)	Resposta	Índice de ocorrência (%)
fumar (ali)	99,16%	valeu	82,41%
canal	99,16%	brigou	81,51%
garçom	98,15%	degrau	80,67%
batom	97,22%	pulei	79,63%
vulcão	96,64%	rancor	77,31%
cacau	96,30%	avós	77,31%
perfil	96,30%	plantio (...)	75,93%
torcer	96,30%	cair	75,00%
normal	95,80%	colar	73,15%
carvão	95,80%	divã / divan	72,22%
chegou	94,96%	luar	69,75%
raspão	94,96%	(mal) cheiro	69,44%
errei	94,96%	negou (...)	68,52%
café	94,44%	cipó	68,07%
trenó	94,44%	crachá	63,87%
local (do...)	94,12%	marfim	61,21%
roubar	93,28%	pomar	59,26%
colchão	92,59%	locais (...)	57,98%
vilã	92,44%	traí-lo	53,78%
bucal	92,44%	judeus	52,78%
limão (...)	90,74%	correr	50,00%
pincel	89,81%	atacante (variações)	40,19%
lençol (...)	87,39%	rapaz	33,33%
colar (de...)	84,87%	(lhe) servia	32,77%
mensal	83,33%	galpão	26,85%

ANEXO 1: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – CLOZE

Registro de Consentimento Livre e Esclarecido

Você está sendo convidado a participar como voluntário de uma pesquisa (CAAE: 53249121.2.0000.8142).

Por favor, leia com atenção e calma, aproveitando para esclarecer suas dúvidas. Se houver perguntas antes ou mesmo depois de participar, você poderá esclarecê-las com o pesquisador através do WhatsApp (019 98730-8587) ou do e-mail f167532@dac.unicamp.br. Não haverá nenhum tipo de penalização ou prejuízo se você não aceitar participar ou retirar sua autorização em qualquer momento.

Procedimentos

O Teste de Cloze consiste na apresentação de frases com algumas lacunas a serem preenchidas. O teste não visa a avaliações ou julgamentos, o teste busca compreender características da natureza da linguagem. Portanto, evite pensar muito e responda o mais rapidamente possível com a primeira palavra que lembrar.

A duração prevista para o preenchimento deste é de 10 a 15 minutos e deverá ser realizada apenas 1 vez por cada participante.

Os dados da pesquisa serão armazenados no computador pessoal da pesquisadora para tratamento, desagregação e anonimização, de modo a diminuir ao máximo possível a possibilidade de identificação dos participantes. A guarda dos dados já tratados para retirar qualquer possibilidade de identificação dos participantes será armazenada no Open Science Framework por tempo indeterminado, desde que não seja menor do que 5 anos, como preconizado pelas normas de boas práticas em pesquisa.

Você não deve participar deste estudo se:

- 1) Não for um falante nativo de Português Brasileiro;
- 2) Tiver menos de 18 anos;

Desconfortos e riscos

Não há riscos ou desconfortos previsíveis aos participantes.

Benefícios

Não há previsão de benefícios diretos aos participantes, mas os resultados da pesquisa contribuirão para a área da Psicolinguística, permitindo-nos a compreensão de processos mentais envolvidos no uso da linguagem.

Sigilo e privacidade

Os dados serão tratados de maneira a retirar qualquer vestígio previsível de identidade dos participantes. Você tem a garantia de que os pesquisadores buscarão garantir o sigilo de sua identidade e nenhuma informação identifi cada ou identifi cável será fornecida a outras pessoas que não façam parte da equipe de pesquisadores.

Ressarcimento e indenização

Considerando se tratar de um formulário eletrônico, não são previstos custos. De todo modo, se você tiver qualquer custo causado diretamente por esta pesquisa, entre em contato com os pesquisadores comprovando os custos que você será devidamente ressarcido. Você terá a garantia ao direito a indenização diante de eventuais danos decorrentes da pesquisa quando comprovados nos termos da legislação vigente.

Acompanhamento e assistência

A qualquer momento os participantes poderão entrar em contato com os pesquisadores para esclarecimentos e assistência sobre qualquer aspecto da pesquisa, através dos contatos abaixo. Você receberá assistência integral e imediata, de forma gratuita, pelo tempo que for necessário em caso de danos decorrentes da pesquisa.

Contato

Em caso de dúvidas sobre a pesquisa, se precisar consultar esse registro de consentimento ou quaisquer outras questões, você poderá entrar em contato com Fernando Luis Sabatini, (IEL/UNICAMP) por email (sabatini.nando@gmail.com; 19 98730-8587) ou com Thiago Oliveira da Motta Sampaio (thimotta@unicamp.br; 21 98160-4087).

Em caso de denúncias ou reclamações sobre sua participação e sobre questões éticas do estudo, você poderá entrar em contato com a secretaria do Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP-CHS) da UNICAMP das 08h30 às 11h30 e das 13h00 às 17h00 na Rua Bertrand Russell, 801, Bloco C, 2º piso, sala 05, CEP 13083-865, Campinas – SP; telefone (19) 3521-6836; e-mail: cepchs@unicamp.br.

Havendo a necessidade de intermediação da comunicação em Libras você pode fazer contato com a Central TILS da Unicamp no site <https://www.prg.unicamp.br/tils/>.

O Comitê de Ética em Pesquisa (CEP)

O papel do CEP é avaliar e acompanhar os aspectos éticos de todas as pesquisas envolvendo seres humanos. A Comissão Nacional de Ética em Pesquisa (CONEP), tem por objetivo desenvolver a regulamentação sobre proteção dos seres humanos envolvidos nas pesquisas. Desempenha um papel coordenador da rede de Comitês de Ética em Pesquisa (CEPs) das instituições, além de assumir a função de órgão consultor na área de ética em pesquisas

Responsabilidade do Pesquisador

Asseguro ter cumprido as exigências da resolução 510/2016 CNS/MS e complementares na elaboração do protocolo e na obtenção deste Termo de Consentimento Livre e Esclarecido. Asseguro, também, ter explicado e fornecido uma via deste documento ao participante. Informo que o estudo foi aprovado pelo CEP perante o qual o projeto foi apresentado e pela CONEP, quando pertinente. Comprometo-me a utilizar o material e os dados obtidos nesta pesquisa exclusivamente para as finalidades previstas neste documento ou conforme o consentimento dado pelo participante.

Você concorda com os termos e quer continuar?

Marcar apenas uma oval.

☐ Sim

☐ Não

**ANEXO 2: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO –
EXPERIMENTO PRINCIPAL
EXPERIMENTO DE DECISÃO LEXICAL
FERNANDO LUIS ANTUNES SABATINI E THIAGO MOTTA SAMPAIO
CAAE: 53249121.2.0000.8142**

Você está sendo convidado a participar como voluntário de uma pesquisa. Este documento, chamado Termo de Consentimento Livre e Esclarecido, visa informar seus direitos como participante de pesquisa.

Por favor, leia com atenção e calma, aproveitando para esclarecer suas dúvidas. Se houver perguntas antes ou mesmo depois de concordar ou de responder o formulário, você poderá elucidá-las com os pesquisadores. Não haverá nenhum tipo de penalização ou prejuízo se você não aceitar participar ou retirar sua autorização em qualquer momento.

JUSTIFICATIVAS E OBJETIVOS

Este se trata de um teste linguístico. Por teste, não entenda que estamos testando sua habilidade com a língua. Como uma pesquisa em linguística, o objetivo é compreender exatamente como os falantes usam a língua e, a partir daí, buscar compreender como ela funciona naturalmente, independente do que diz a gramática. Nesse sentido, escute as frases e leia as palavras da forma mais natural possível, e responda com a primeira opção que vier à mente. Não se preocupe, pois ninguém irá julgar suas respostas e não iremos divulgar sua identidade.

Ao final do estudo, poderemos explicar melhor do que se trata essa pesquisa e, se for do seu interesse, podemos compartilhar seus resultados explicando como eles auxiliaram em nossa pesquisa. De todo modo, suas respostas contribuirão para uma melhor compreensão da relação entre mente e linguagem.

Você pode participar se você for:

- brasileiro;
- maior de 18 anos; e
- ter visão normal ou corrigida (óculos ou lentes).

PROCEDIMENTOS

Nessa atividade, você ouvirá sentenças no fone de ouvido e será apresentado a diferentes combinações de letras na tela do computador. Sua tarefa é responder se a letras correspondem a palavras do Português do Brasil ou não. A previsão de duração desse teste é de cerca de 15-20 minutos.

Os dados da pesquisa serão armazenados no servidor "gitlab" durante a aplicação. Posteriormente, os dados serão baixados no computador da pesquisa para tratamento, desagregação e anonimização, de modo a diminuir ao máximo possível a possibilidade de identificação dos participantes. A guarda dos dados já tratados para retirar qualquer possibilidade de identificação dos participantes será realizada através do servidor Open Science Framework por tempo indeterminado, desde que não seja menor do que 5 anos, como preconizado pelas normas de boas práticas em pesquisa.

DESCONFORTOS E RISCOS

Você não deve participar deste estudo se for menor de 18 anos ou maior do que 30 anos,

se tiver problemas de visão não corrigidos por óculos ou lentes de contato. Este teste não apresenta riscos físicos, de saúde ou psicológicos previsíveis. De qualquer forma, sinta-se à vontade para interromper imediatamente o teste em caso de tonturas ou outro desconforto.

BENEFÍCIOS

Embora esta pesquisa não preveja algum benefício direto ao participante, ela contribui para uma gama de estudos sobre a relação entre Psicologia e Linguagem.

SIGILO E PRIVACIDADE

Seus dados serão armazenados no computador de forma anônima, visto que os participantes serão identificados por meio de números. Caso você deseje desistir de sua participação, você poderá a qualquer momento e sem qualquer tipo de justificativa solicitar que seus dados sejam apagados de nossa base, exceto em bases permanentes como o REDU (Unicamp). Após a publicação deste estudo em revistas científicas, os dados serão guardados por pelo menos 5 anos no repositório de dados da UNICAMP (REDU) e no Open Science Framework.

RESSARCIMENTO E INDENIZAÇÃO

Considerando se tratar de um experimento a ser realizado no navegador de seu computador, não são previstos custos. De todo modo, se você tiver qualquer custo causado diretamente por esta pesquisa, entre em contato com os pesquisadores com as devidas comprovações e você será devidamente ressarcido. Você terá a garantia ao direito a indenização diante de eventuais danos decorrentes da pesquisa quando comprovados nos termos da legislação vigente.

ACOMPANHAMENTO E ASSISTÊNCIA

Embora a pesquisa em questão não consiga prever desconfortos causados diretamente pelos testes, a qualquer sinal de desconforto o teste deve ser interrompido. A qualquer momento, os participantes poderão entrar em contato com os pesquisadores para esclarecimentos e assistência sobre qualquer aspecto da pesquisa, através dos contatos abaixo. Você receberá assistência integral e imediata, de forma gratuita, pelo tempo que for necessário, em caso de danos decorrentes da pesquisa.

CONTATO

Em caso de dúvidas sobre a pesquisa, caso deseje informações mais detalhadas após a sua participação no experimento, se precisar consultar esse registro de consentimento ou quaisquer outras questões, você poderá entrar em contato com os pesquisadores:

Fernando Luis Antunes Sabatini, sabatini.nando@gmail.com; (19 98730-8587) ou

Thiago Oliveira da Motta Sampaio, thimotta@unicamp.br.

Instituto de Estudos da Linguagem (IEL) da Unicamp. Sala D111.

Em caso de denúncias ou reclamações sobre sua participação e sobre questões éticas do estudo, você poderá entrar em contato com a secretaria do Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP-CHS) da UNICAMP das 08h30 às 11h30 e das 13h00 às 17h00 na Rua Bertrand Russell, 801, Bloco C, 2º piso, sala 05, CEP 13083-865, Campinas – SP; telefone (19) 3521-6836; e-mail: cepchs@unicamp.br.

Havendo a necessidade de intermediação da comunicação em Libras, você pode fazer

contato com a Central TILS da Unicamp no site <https://www.prg.unicamp.br/tils/>.

O COMITÊ DE ÉTICA EM PESQUISA

O papel do CEP é avaliar e acompanhar os aspectos éticos de todas as pesquisas envolvendo seres humanos. A Comissão Nacional de Ética em Pesquisa (CONEP) tem por objetivo desenvolver a regulamentação sobre proteção dos seres humanos envolvidos nas pesquisas. Desempenha um papel coordenador da rede de Comitês de Ética em Pesquisa (CEPs) das instituições, além de assumir a função de órgão consultor na área de ética em pesquisas.

RESPONSABILIDADE DO PESQUISADOR

Asseguramos ter cumprido as exigências da resolução 510/2016 CNS/MS e complementares na elaboração do protocolo e na obtenção deste Termo de Consentimento Livre e Esclarecido. Asseguramos, também, ter explicado e fornecido uma via deste documento ao participante. Informamos que o estudo foi aprovado pelo CEP, perante o qual o projeto foi apresentado.

Comprometemo-nos a utilizar o material e os dados obtidos nesta pesquisa exclusivamente para as finalidades previstas neste documento ou conforme o consentimento dado pelo participante.

Consentimento livre e esclarecido:

Após ter recebido esclarecimentos sobre a natureza da pesquisa, seus objetivos, métodos, benefícios previstos, potenciais riscos e o incômodo que esta possa acarretar, aceito participar:

Nome do(a) participante: _____.

_____ Data: ____/____/____.

(Assinatura do participante)

Responsabilidade do Pesquisador:

_____ Data: ____/____/____.

(Assinatura do pesquisador)