



## **Não chame de erro o que a linguística chama variação: processamento de variação linguística e de agramaticalidade no âmbito da concordância verbal variável**

### ***Do Not Call It an Error if Linguistics Has Called It Variation: Processing Linguistic Variation and Ungrammaticality in the Scope of Variable Subject-Verb Agreement***

Wellington Couto de Almeida

Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ / Brasil

wellingtonalmeida@letras.ufrj.br

<https://orcid.org/0000-0001-6813-3801>

**Resumo:** Em uma articulação entre a psicolinguística e a sociolinguística, e a partir de uma visão mais formal de linguagem, este trabalho discute eventuais efeitos de adaptabilidade via *priming* no processamento de variação linguística e como esses efeitos impactam custos de processamento associados à concordância verbal variável de terceira pessoa do plural e a construções agramaticais. Para tanto, realizamos um experimento de leitura automonitorada que contou não só com construções atestadas na língua para a concordância verbal variável, mas também com instâncias agramaticais, com o objetivo de verificar diferenças na proporcionalidade dos custos entre instâncias gramaticais variáveis e agramaticais. O experimento também contou com a manipulação de estímulos longos, o que permitiu que fossem verificados efeitos de adaptabilidade como fruto de *priming* operante. Esse tipo de manipulação, até onde sabemos, é pioneira nos estudos da interface psico-sociolinguística no Brasil. Os resultados sugeriram custos de processamento e efeitos de adaptabilidade diferentes associados a condições gramaticais (concordâncias redundante e não redundante) e à condição agramatical. Em outras palavras, os resultados mostram que processar instâncias gramaticais variáveis é diferente de processar instâncias agramaticais. Este estudo, portanto, lança luz a questões relativas ao processamento *online* de variação linguística, tópico cada vez mais produtivo na ciência linguística.

**Palavras-chave:** processamento; variação linguística; concordância verbal; *priming*; agramaticalidade.

**Abstract:** In an articulation between psycholinguistics and sociolinguistics, and from a formalist-like approach of language, this paper discusses possible adaptative effects caused by operant priming effects in language variation processing, and how these effects impact processing costs associated with third-person plural variable subject-verb agreement. To do so, we ran a self-paced reading experiment that included not only constructions attested for variable verb agreement but also ungrammatical instances, aiming to check for differences in cost proportionality between variable (then grammatical) and ungrammatical instances. The experiment also included the manipulation of long stimuli, which allowed adaptive effects to be verified as a result of an operant priming effect. This type of manipulation, as far as we know, is unprecedented in studies of the psycho-sociolinguistic interface in Brazil. The results suggested different costs and adaptative effects associated with grammatical conditions (redundant and non-redundant agreements) that were not attested in the ungrammatical condition. In other words: processing variable grammatical instances is different from processing ungrammatical ones. This study, therefore, shed light on questions concerning the online processing of linguistic variation, an increasingly productive topic in linguistics.

**Keywords:** processing; linguistic variation; subject-verb agreement; priming; ungrammaticality.

Recebido em 02 de setembro de 2022

Aceito em 19 de novembro de 2022

## 1 Introdução

Diante de enunciados linguísticos, ouvintes precisam dar conta de uma miríade de materiais intervenientes e variáveis, envolvendo processos cognitivos que vão desde ignorar ruídos externos à fala a ele direcionada até, por exemplo, enquadrar fonemas eventualmente produzidos de maneiras diferentes do esperado em suas representações mentais prototípicas. Também reportado na literatura como o problema da falta de invariabilidade (HEALD; KLOS; NUSBAUM, 2006), olhar para essas situações nos ajuda a dimensionar o engajamento cognitivo dispensado na compreensão linguística, principalmente quando assumimos a língua como uma realidade, que, embora ordenada, é heterogênea, i.e., repleta de variabilidade (WEINREICH; LABOV; HERZOG, 2006 [1968]).

Naturalmente, compreender instâncias linguísticas é, por si só, uma tarefa de engajamento cognitivo. No que diz respeito ao problema da falta de invariabilidade, uma série de estudos se ocupou, por exemplo, do controle inibitório necessário frente a ruídos que ou se sobrepõem ou se confundem à fala de um interlocutor, como o barulho da TV, ou a buzina de um carro (cf. HEALD; KLOS; NUSBAUM, 2006). No âmbito das variáveis linguísticas, a sociolinguística, com áreas e ferramentas em interface, como aquelas advindas da psicolinguística, tem dedicado especial atenção à maneira como indivíduos as processam – fazendo uso, em muitos casos, de técnicas *online* de experimentação.

Especificamente em termos de processamento, estudos desenvolvidos em outros países, como o de Squires (2014; 2016) e o de Thomas (2011), já têm proposto esse diálogo, que apontou para a necessidade de se articularem modelos linguísticos com o conhecimento sociolinguístico, especificamente no que diz respeito a como e se aspectos sociolinguísticos impactam processos cognitivos envolvidos no processamento linguístico. No Brasil, estudos de processamento, principalmente no âmbito morfossintático, têm ganhado corpo, embora ainda sejam pouco comuns, e podemos apontar os de Marcilese *et al.* (2015, 2017) como pioneiros nesse sentido.

E é exatamente entre esses estudos que este trabalho se coloca: os dados aqui apresentados têm por objetivo contribuir com a discussão que é posta no processamento de concordância verbal variável, em especial investigando custos associados não só a itens gramaticais (como *eles gostam* e *eles gosta*), mas também a itens agramaticais (como *\*eles gosto*), em contextos experimentais relativamente mais ecológicos. No escopo dessa discussão, estão efeitos psicológicos associados ao processamento de variação linguística (TAMMINGA; MACKENZIE; EMBICK, 2016). Aqui, o principal deles é o efeito de *priming*, que possibilita, por exemplo, na compreensão, um acesso mais facilitado a formas já processadas anteriormente e, em decorrência disso, eventuais adaptações que têm efeito direto no engajamento cognitivo do processamento linguístico.

Nesse sentido, o presente artigo se divide, a partir daqui, em seis seções: na segunda seção, é feita uma revisão de literatura; na terceira, apresentamos o que entendemos como gramática e como processamento variável. Na quarta seção, reportamos o experimento conduzido e seus principais resultados; na quinta, discutimos os resultados obtidos. Finalmente, na sexta seção, fazemos apontamentos futuros e, em seguida, a conclusão do texto.

## 2 Concordância verbal variável no PB: estado da arte e estudos experimentais

Pode-se entender concordância como um fenômeno que estabelece relações de codependência por meio da covariação da morfologia flexional, ou funcional, entre palavras que estejam relacionadas (MOLINARO; BARBER; CARREIRAS, 2011). Esse é um fenômeno observado em aproximadamente 50% das línguas do mundo (BYBEE, 1985), acontecendo, pelo menos, em termos de gênero, pessoa, caso e número, e em alguns casos acomodando variação linguística, que é o que acontece no Português Brasileiro (PB) no âmbito da concordância de número, seja nominal ou verbal – essa última nosso objeto de interesse.

Desde o pioneiro trabalho de Lemle e Naro (1977) sobre a concordância verbal variável na cidade do Rio de Janeiro, uma série de estudos adicionou camadas de complexidade para o entendimento do fenômeno no âmbito da sociolinguística variacionista. Hoje, a quantidade de trabalhos acumulados na área provê um dimensionamento bastante robusto do fenômeno, permitindo que estudos experimentais possam ser feitos tendo como base o arcabouço construído ao longo de aproximadamente 50 anos. Para os propósitos deste trabalho, trataremos à baila alguns dos aspectos desse arcabouço, mas o leitor pode conferir uma revisão bastante precisa dessa literatura em Molina (2018).

Em termos de descrição desse fenômeno, podemos simplificá-lo para nossas pretensões aqui, descrevendo, nos termos de Marcilese *et al.* (2015), dois tipos de concordância: a redundante e a não redundante. Chamamos concordância verbal redundante aquela em que todos os elementos do sintagma recebem marcação morfofonológica explícita de plural. Chamamos não redundante aquelas concordâncias verbais que são feitas com pelo menos um dos elementos do sintagma verbal não recebendo essa marcação explícita. Observemos os exemplos (1) e (2), extraídos de Scherre e Naro (1998), o primeiro com uma concordância redundante e o segundo com uma não redundante:

(1) Eles ganham demais da conta.

(2) Eles ganha demais.

No entanto, esse fenômeno é mais complexo do que o contraste entre as sentenças em (1) e (2) sugere. Há fatores linguísticos, sociais, e até psicológicos que favorecem ou desfavorecem determinadas realizações,

se pensarmos em apontamentos como os de Tamminga, MacKenzie e Embick (2016) para o estudo da variação intra-individual. Por exemplo, no âmbito da concordância verbal, foco deste trabalho, temos visto, no campo dos fatores linguísticos, a saliência fônica, a posição do sujeito e a quantidade de material interveniente entre sujeito e verbo como importantes fatores no enquadramento do fenômeno.

Alguns desses fatores condicionantes para a variação, no entanto, podem ser entendidos como parte de uma realidade mais psicológica e intra-individual, ponto defendido por Tamminga, MacKenzie e Embick (2016), e não como um todo de condicionadores puramente linguísticos. Não nos parece que seja a melhor opção, de fato, colocar em um mesmo balaio todos esses fatores, sobretudo levando em consideração suas naturezas distintas. Por isso, neste trabalho, abordaremos os fatores geralmente entendidos pela sociolinguística como internos em duas categorias: fatores internos do tipo linguístico e fatores internos do tipo psicológico (condicionares *i* e *p*, na nomenclatura de Tamminga, MacKenzie e Embick (2016)); ambos, na sociolinguística, são geralmente entendidos como tão somente condicionadores internos.

A literatura aponta que, quanto maior a saliência fônica entre formas plurais e singulares, maiores são as chances de que a palavra seja marcada explicitamente como plural (em outras palavras, que receba uma concordância redundante) (BORTONI-RICARDO, 2008; LEMLE; NARO, 1977; NARO, 1981), ou seja, há mais chances de que a concordância não redundante seja feita em pares como *falam x fala*, em que a oposição é feita a partir de relativamente pouco material fônico entre uma e outra<sup>1</sup>, do que em pares como *falaram x falou*, em que há mais massa fônica na diferença entre as formas.

No que diz respeito à posição do sujeito, a literatura tem apontado que sujeitos prepostos (como em *eles fala*), por exemplo, favorecem a marcação redundante, ao passo que sujeitos pospostos (como em *falou os menino*) favorecem a marcação não redundante (SCHERRE; NARO, 2007).

---

<sup>1</sup> Agradecemos à sugestão de um dos pareceristas, que nos atentou para o fato de que a oposição em um par como *falam x fala* tem não só o acréscimo da nasalização, mas também a possibilidade de alterações vocálicas diversas, como já aponta Rocha (2021).

Já no que se refere à quantidade de material interveniente entre sujeito e verbo, a literatura indica que, quanto mais material entre os dois, maiores são as chances de que se realize uma marcação não redundante (SCHERRE; NARO, 2007). Em Tamminga, MacKenzie e Embick (2016), encontramos argumentos que nos fazem pensar se esse tipo de condicionamento não seria mais psicológico do que linguístico, uma vez que poderia estar ligado à capacidade individual de *span* de memória que cada indivíduo tem. Indivíduos com maiores capacidades de armazenamento na memória de trabalho poderiam ser capazes de manter a informação de que o sujeito está, por exemplo, no plural, mesmo quando há mais material interveniente, ao passo que indivíduos com espaços menores no *span* podem acabar perdendo essa informação mais rapidamente quando há material interveniente, produzindo instâncias não redundantes. Em alguns casos, especialmente quando o sujeito é singular e o verbo é conjugado no plural por conta de um material interveniente plural (como em *O tecido das cortinas rasgaram*), a literatura costuma chamar esse fenômeno de *erro de atração* (cf. RODRIGUES, 2006; LEE; PHILLIPS, 2023).

Ainda, há o chamado paralelismo linguístico, formal ou discursivo, fator que influencia a produção linguística a partir do postulado de que marcas levam a marcas e zeros levam a zeros, favorecendo ou desfavorecendo uma marcação ou outra a partir de elementos precedentes. Ou seja, se foi preferida uma forma (redundante ou não redundante, seja em um VP ou em um NP, por exemplo), há maiores chances de que o indivíduo mantenha a aplicação dessas marcas (POPLACK, 1980; SCHERRE, 1994; 1998; VIEIRA, 2013). Para nós, baseados em Tamminga, MacKenzie e Embick (2016), esse princípio pode ser encarado como um efeito atuante de *priming*. Dessa forma, consideramos que haja, nesses casos, um efeito psicológico mais geral que atua na produção linguística (condicionar do tipo *p*, nas palavras dos autores), mais do que um efeito puramente linguístico (condicionador do tipo *i*, também nas palavras dos autores), como parece defender a literatura sociolinguística.

No campo social, fatores como sexo, idade, classe social e escolarização geralmente são levados em consideração quando se tentam entender quais características podem estar associadas a um tipo ou outro de produção. Nesse sentido, a literatura aponta que homens menos escolarizados, de estratos sociais menos favorecidos, e mais jovens tendem a produzir mais a concordância chamada não redundante (LEMLE; NARO, 1977; SCHERRE; NARO, 1998; VIEIRA; BAZENGA, 2013).

No escopo da escolaridade, e olhando especificamente para universitários, grupo experimental do estudo aqui reportado, vale ressaltar que as políticas de ações afirmativas, desde 2003 na Universidade do Estado do Rio de Janeiro<sup>2</sup>, e a partir de 2012, nacionalmente (BRASIL, 2012), ampliaram o acesso ao ensino superior, provavelmente formando, nas Universidades, um ambiente social ainda mais rico para o estudo da concordância verbal variável: indivíduos de regiões menos favorecidas socialmente, em que a concordância não redundante tende a ser mais produtiva, chegam cada vez mais a níveis altos de escolarização<sup>3</sup>, complexificando um cenário que já não era tão simples assim – levando em consideração que dados como os dos *corpora* do Projeto NURC/UFRJ, cujos informantes são considerados indivíduos cultos/letrados do Rio de Janeiro, já demonstravam algum grau de variabilidade na concordância verbal entre esse grupo (cf. GRACIOSA, 1991), certamente relativamente mais homogêneo do que é hoje em dia. Vale ressaltar que dados mais atuais no escopo da concordância variável de P6 têm demonstrado que o maior acesso à educação contribui para o aumento dos índices de marcação redundante (cf. ROCHA, 2021; VIEIRA; BAZENGA, 2013, 2015), embora ainda haja algum grau de variação atestado nos grupos das amostras. Aqueles indivíduos oriundos de contextos economicamente menos favorecidos receberiam influência da escolarização, passando a realizar mais concordâncias redundantes do que seus pares de origem social, pelo menos em contextos de maior monitoramento. Como essa variabilidade se organiza em termos de gramática internalizada ainda é ponto de debate, com discussões, sem dúvida, bastante produtivas. Siqueira e Freitag (2022), inclusive, trazem a questão da mobilidade para o escopo da variação dentro do contexto universitário, em Sergipe; em uma cidade de grandes dimensões, como é o Rio de Janeiro, estudos assim seriam bastante elucidativos para entendermos melhor possíveis efeitos advindos da mobilidade e como ela influencia os padrões de variação linguística em universitários, principalmente aqueles advindos de contextos menos favorecidos socioeconomicamente.

---

<sup>2</sup> “Uerj é a 1ª a adotar sistema de cotas para negros”. Disponível em: <https://www1.folha.uol.com.br/folha/educacao/ult305u12494.shtml>. Acesso em 1 set. 2022.

<sup>3</sup> Não necessariamente via universidade pública, haja vista que programas como PROUNI e FIES, a despeito de seus reveses (ARRUDA, 2011) também aumentaram o acesso ao ensino superior.

Ainda, esses apontamentos não conseguem dar conta da natureza do fenômeno, que por vezes precisa ter nuances simplificadas para o contexto experimental, como fizemos aqui. Por exemplo, marcações não redundantes recaem em um *continuum* mais complexo do que tão somente “marcar apenas o sujeito para plural”, ou “marcar apenas o determinante para plural”. Em uma construção como *Os meninos comem doce*, é possível ter *Os menino come*, mas também *Os meninos come*, e também *Os menino comem*. Dados de literatura também apontam ocorrências de variação no determinante “os”, embora em quantidade pouco expressiva, e em contextos específicos (cf. SCHERRE, 2001)<sup>4</sup>.

Pensando em uma interface entre a sócio e a psicolinguística, alguns estudos já fizeram primeiros apontamentos no âmbito da concordância verbal variável<sup>5</sup>. Por exemplo, os trabalhos de Marcilese *et al.* (2015, 2017) apontam custos maiores de processamento, refletidos em tempos de resposta (TR, daqui em diante) maiores, para construções com concordância não redundante, seja em testes de leitura (técnica *maze task*) ou escuta automonitorada, ambos com indivíduos que tinham pelo menos iniciado o ensino superior.

A partir do trabalho de 2015, de escuta automonitorada, Marcilese *et al.* concluíram que, apesar dos tempos de escuta maiores associados à concordância não redundante, os participantes compreendiam esse tipo de construção como veiculadora de ideias de plural, i.e., os participantes conseguiam depreender plural de construções como *as garota emagreceu*. Esse tipo de resultado encontra consonância na literatura pelo menos em dois outros estudos (CORRÊA; AUGUSTO; FERRARI-NETO, 2005; FERRARI-NETO, 2008), que atestaram a mesma possibilidade de interpretação de construções não redundantes como plural por crianças por volta dos dois anos de idade, no escopo da concordância nominal com pseudopalavras, como em *mostra os dabo pro Dedé*. Já em 2017, com a

---

<sup>4</sup> Embora mais rarefeitos, há registros na literatura de concordância nominal variável dados em que o determinante ou elemento da primeira posição não recebe a marca explícita, como em “na casa **do meus pais**”, “gosto de tê **a minhas coisas**”, “**na minhas coisa**” (SCHERRE, 2001). De fato, são construções mais complexas, não do âmbito especificamente verbal, e com a presença de pronomes possessivos, mas é interessante notar a existência, também, desses padrões.

<sup>5</sup> Alguns dos estudos que apontaremos aqui trabalharam com concordância nominal, mas nós entendemos que esses achados também contribuem para o estudo da concordância verbal.



técnica de *maze task*, Marcilese *et al.* investigaram efeitos de distância linear para a concordância verbal variável (ex. *Os alunos/O aluno (#) escutaram/escutou a professora*, com “#” sendo preenchido ou não com uma quantidade específica de palavras, variando de acordo com a quantidade de material interveniente entre sujeito e verbo), e obtiveram resultados que apontavam custos maiores para o processamento de concordância não redundante quando comparada à concordância redundante. O estudo apontou ainda efeitos específicos da quantidade de material interveniente entre sujeito e verbo, que diminuía os custos de processamento da concordância não redundante à medida que havia mais material interveniente entre eles. Esse dado é facilmente pareado com os achados na literatura sociolinguística que colocam mais chances de produção de concordância não redundante quando há mais material interveniente entre sujeito e verbo (SCHERRE; NARO, 2007). Estudos futuros podem, por exemplo, pensar em dados semelhantes de maneira intra-individual, verificando questões associadas ao *span* de memória e sua influência na compreensão linguística, partindo de pressupostos apontados por Tamminga, MacKenzie e Embick (2016).

Levando em consideração também os achados na seara da aquisição, dados de Molina, Marcilese e Name (2017), com crianças na faixa dos cinco a seis anos de idade, em uma tarefa de identificação de figuras, vão nessa mesma direção dos trabalhos de Marcilese *et al.* 2015, Corrêa, Augusto e Ferrari-Neto (2005) e Ferrari Neto (2008). Para essas crianças, instâncias como *Comeu doce* podem veicular tanto a ideia de singular, como em *Ele comeu doce* ou plural, como em *Eles comeu doce*. A partir dos 10 anos essa possibilidade passa a ser mais restrita, com um pareamento de *Comeu doce* mais ligado à ideia de singular, provavelmente porque, no processo de escolarização, passa-se a esperar, de maneira mais consciente, que um verbo com marcação morfofonológica de singular (ou sem marcação morfofonológica explícita de plural) seja pareado com um sujeito também no singular.

Molina, Marcilese e Name (2018), inclusive, em um estudo que coletou medidas que refletiam mais o processamento *online* da tarefa de identificação de figuras reportada pelas autoras em 2017, verificando o tempo entre o término da escuta da sentença e a escolha por uma das figuras apresentadas na tarefa, observaram que sentenças apresentadas no plural eram pareadas mais rapidamente com figuras com mais de um agente, enquanto as sentenças no singular eram mais custosas nesse

mesmo cenário – o que as fez inferir que, no PB, sentenças no plural apresentam informações mais robustas para quem está ouvindo, pois não deixam brechas para mais de uma possibilidade de interpretação, como as do singular. Esses dados reforçam a ideia de um *input* variável no PB, em que o traço de número pode ser pouco confiável nos primeiros momentos; afinal, um verbo sem marca explícita de plural pode ser pareado para um sujeito no plural.

No âmbito do *input* variável, Jakubów (2018) e Jakubów e Corrêa (2021) fizeram contribuições relevantes a partir de experimentos de produção eliciada, mostrando a presença de concordância verbal variável em crianças em fase pré-escolar e no 6º ano do ensino fundamental, tanto na rede pública quanto na privada da cidade do Rio de Janeiro, com taxas de produtividade significativas, a despeito dos estratos sociais a que pertenciam. Embora houvesse diferenças entre os grupos experimentais no que dizia respeito à produção de concordância redundante e não redundante, com as crianças do 6º ano do ensino privado preferindo a concordância redundante, por exemplo, foi possível verificar produções variáveis em todos os grupos, o que mais uma vez aponta para algum tipo de acomodação de variabilidade nas gramáticas dos indivíduos durante a aquisição de linguagem.

Embora fora do escopo da interface sócio-psicolinguística, Maia e Nascimento (2020) realizaram um experimento com *eye-tracking* que trouxe resultados interessantes para este trabalho. Eles expuseram universitários à leitura de sentenças com construções como (...) *tivesse todos os livros*, (...) *tivesse todos os livro*, e (...) *comesse todos os livros*, para a investigação de custos de processamento sintático e semântico. Olhando especificamente para o par (...) *tivesse todos os livros* e (...) *tivesse todos os livro*, os resultados sugeriram tempos maiores de fixação, tanto progressiva quanto regressiva, nas concordâncias que temos chamado aqui de não redundantes (*os livro*) quando comparadas àquelas que temos chamado concordância redundante (*os livros*). Assim como em testes de leitura automonitorada, em estudos com *eye-tracking*, maiores tempos de fixação podem estar associados a maiores custos de processamento. Esse resultado faz sentido, assim como os de Marcilese *et al.* (2015, 2017), ao apontar custos maiores para a não redundância, sobretudo se pensarmos em contextos universitários, em que há mais expectativa, de modo geral, de um uso normativo da língua, que preconiza a concordância redundante.

Até agora, listamos alguns trabalhos que trazem à baila a noção de *custo*, geralmente refletido em um tempo maior comparativamente entre condições, seja de leitura, de escuta, ou de fixação de olhar. Mas o que exatamente esse custo está refletindo, haja vista que é uma medida numérica, coletada em contextos experimentais? Via de regra, geralmente esses custos são interpretados como *engajamento cognitivo*, de modo que, se há um custo maior, é sinal de que houve mais engajamento cognitivo, comparativamente. Nesse sentido, o que exatamente subjaz esse custo? Seria um grau de maior ou menor familiaridade com uma das formas? Seria um efeito de monitoramento sociolinguístico em decorrência de contexto? Em que medida é possível supor efeitos associados à modalidade dos estímulos (i.e., escritos ou falados)?

Essas são perguntas que, sem dúvida, interessam à agenda de pesquisa a que o presente estudo se vincula. Aqui, daremos atenção especial a uma outra questão, para nós tão importante quanto essas primeiras, relacionada com a dimensão desses custos de processamento. Em que medida, por exemplo, variáveis dicotômicas em um experimento podem acabar sendo pareadas com ideias como “certo” e “errado”, por exemplo, no sentido em que um custo de processamento maior poderia ser ligado a uma forma “incorreta”?

Embora na linguística essa nem seja mais uma discussão e lidemos relativamente bem com ideias de custo de processamento como refletindo uma realidade subjacente da gramática, essas associações poderiam ser feitas por leitores menos familiarizados com a área. Uma das maneiras de entender a proporcionalidade desse custo, ou pelo menos o quão alto um custo de processamento pode ser, ou ainda o que é de fato o custo de processamento associado a um erro, pode ser incluir, nos experimentos, construções agramaticais, que permitiriam ter uma compreensão mais robusta dos custos associados a um erro de fato – o que não é o caso de uma variante linguística estigmatizada. Além disso, entender essa proporcionalidade também é necessário para compreender melhor o que subjaz cognitivamente esse custo associado ao processamento da variante, o que condiz com algumas das perguntas que fizemos há pouco. Se não temos essa condição de controle (nesse caso, agramatical), como saber a que processos cognitivos subjacentes atribuir esse custo? Ao lançar mão dessa proporcionalidade, podemos conseguir observar diferenças envolvidas no processamento daquilo que é variante e daquilo que é agramatical, contribuindo no entendimento

do aumento no custo de processamento como quebra de expectativa ou violação de eventuais regras gramaticais, ou ainda, no campo da concordância verbal, como violação de computação sintático-semântica entre sujeito e verbo, por exemplo.

Vale ressaltar, ainda, que nem sempre uma construção necessariamente agramatical precisa ser incorporada a um experimento, uma vez que é difícil às vezes definir o limite da agramaticalidade (SOTO; ALMEIDA, 2021), mas construções não atestadas ou pouco atestadas podem fazer parte dos experimentos com essa mesma funcionalidade, como fez Squires (2014), por exemplo, que manipulou em experimentos construções como  $SUJ_{SING} + DOESN'T$  e  $SUJ_{SING} + DON'T$ , ambas atestadas na literatura sociolinguística do inglês estadunidense e a construção  $SUJ_{PL} + DOESN'T$ , não atestada por essa mesma literatura. No experimento reportado neste trabalho, especificamente, construímos instâncias agramaticais (ou, pelo menos, não atestadas na literatura) a partir da combinação de um sujeito de terceira pessoa do plural com verbos de primeira pessoa do singular (Cf. Seção 4).

Ainda, no que diz respeito a como lidamos com o problema da falta de invariabilidade (HEALD; KLOS; NUSBAUM, 2006), haja vista que conseguimos entender instâncias variáveis costumeiramente sem maiores problemas, precisamos entender quais são os fatores cognitivos que estão associados a uma possível adaptabilidade à medida que interagimos com essas instâncias. O efeito de *priming*, por exemplo, pode ser um aliado nesse sentido, uma vez que traços linguísticos já processados tendem a ser reprocessados com mais facilidade (HASTING *et al.*, 2007; LEDOUX *et al.*, 2007; TRAXLER *et al.*, 2000); para a produção, inclusive, essa já é uma aposta feita por Tamminga, MacKenzie e Embick (2016).

### 3 Processamento como pareamento de traços

Neste trabalho, assumimos uma visão mais formalista de linguagem, e no que se refere à variação linguística, precisamos reconhecer que modelos mais formais geralmente encontram alguma dificuldade em articular uma realidade variável e intra-individual da língua – realidade atestada por trabalhos da sociolinguística variacionista e já pré-anunciada pelo menos desde 1968 com o trabalho de Wenreich, Labov e Herzog (2006). A opção por um modelo formal de linguagem advém de evidências da neurofisiologia da linguagem para um nível de especialização linguístico

que parece ter algum *status* diferenciado na cognição humana – afasias e transtornos especificamente linguísticos apontam nessa direção (ROSA, 2020; GARRAFFA; FYNDANIS, 2020, para uma revisão).

Contudo, a especificidade do conhecimento linguístico não deve ser um impedimento para que se assuma a interação desse conhecimento com outros domínios ou aparatos cognitivos gerais, como o da memória, o da capacidade de imitação e efeitos de *priming*, por exemplo (TAMMINGA; MACKENZIE; EMBICK, 2016) – especialmente quando se consideram condições como quantidade de material interveniente e paralelismo como influentes na produção linguística de concordância de número; ambos, para nós, mais conectados a uma realidade psicológico-cognitiva do que propriamente linguística. Também não deve estar desarticulada do conhecimento social que acumulamos desde o início da aquisição da linguagem (NARDY; CHEVROT; BARBU, 2014; SMITH; DUHRAM; RICHARDS, 2013) – a organização de, por exemplo, uma espécie de monitor sociolinguístico (LABOV *et al.*, 2011), que não necessariamente acumula um conhecimento especificamente linguístico, mas que capta, reage a e processa informações de cunho social advindas de estímulos linguísticos, deve dar conta, em algum grau, dessa integração entre língua e o conhecimento social acumulado. Em suma, estamos assumindo um conhecimento especificamente linguístico, relativamente organizado de maneira modular na mente, mas permeado por interfaces não especificamente linguísticas que se articulam com as noções de variabilidade do sistema.

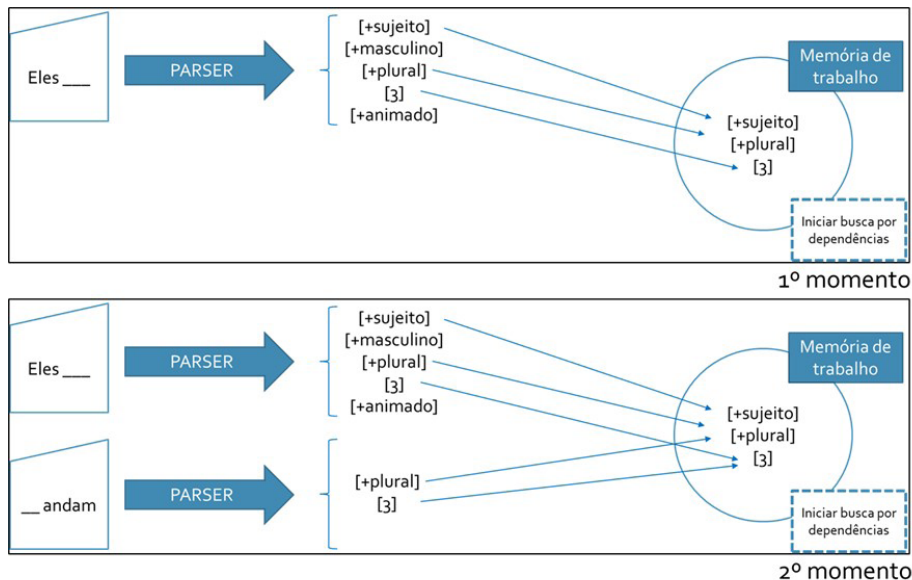
Dentre os modelos formais que estabelecem conosco alguma correlação, o da Morfologia Distribuída (MD) parece contribuir de maneira significativa para a questão da falta de invariabilidade da língua, como cunharam Heald, Klos e Nusbaum (2006). Não entraremos aqui em uma explanação teórica detalhada do modelo, que pode ser obtida em trabalhos mais direcionados (BASSANI; LUNGUINHO, 2011; EMBICK; NOYER, 2006; HALLE; MARANTZ, 1993;), principalmente porque, para prosseguirmos com as questões que nosso trabalho pretende responder, basta que compreendamos uma visão de gramática como um conjunto de traços, como veremos a seguir.

Uma assunção basilar do modelo da MD é a de que a gramática é dividida em três listas de informações: a lista 1, com traços gramaticais/semânticos, por exemplo [1], [2], [3], para se referir às pessoas do discurso, [singular] e [plural] para número, [presente], [passado], [futuro]

para tempo, etc; a lista 2, com os itens de vocabulário, que são as formas disponíveis na língua para a representação dos traços armazenados na lista 1 (nesse caso, traços como [2, singular (ou [-plural]), +dativo] na lista 1 corresponderiam à forma *lhe*, no português); e a lista 3, com o conhecimento enciclopédico, ou o *significado* especial dessas formas (BASSANI; LUNGUINHO, 2011).

Esse tipo de visão para nós é essencial porque assumimos, junto com Lee e Phillips (2023), que processar enunciados linguísticos, e especialmente *compreender* enunciados linguísticos, requer solucionar dependências. Isso significa que qualquer enunciado será decomposto em traços (como esses que a MD aponta), que serão armazenados na memória de trabalho para serem comparados a outros traços na solução de dependências e na extração de significado. Por exemplo, no âmbito da concordância verbal, há algum tipo de correlação (estrutural e/ou sintático-semântica) entre sujeito e verbo, e essa correlação pode ser feita, por exemplo, via checagem de traços, como os de número e pessoa. Para que esse mecanismo funcione, Lee e Phillips propõem um princípio de recuperação baseada em pistas (*cue-based retrieval*), que guardaria na memória traços referentes a uma palavra (como um pronome) e faria comparações com traços de outras palavras (como um verbo), estabelecendo relações de dependência. Se esses traços não são pareados de maneira satisfatória, o custo de processamento pode aumentar, seja pela busca de novos traços para que o *matching* seja satisfeito (pressuposto de modelos de ativação), seja pelo desencadeamento de uma reanálise para checar, ou acomodar, ou ainda consertar, o que acabou de ser processado (modelos de acesso direto) (LEE; PHILLIPS, 2023; VASISHTH, 2019). A Figura 1 tenta ilustrar, em dois momentos, a checagem de traços armazenados na memória de trabalho.

FIGURA 1 - Esquema ilustrativo do parseamento e armazenagem de traços na memória de trabalho e busca por instâncias que solucionem a dependência, em dois momentos, baseado na proposta de Lee e Phillips (2023)



Fonte: Almeida (2022).

Mas e quando temos casos como o da concordância verbal variável, em que uma forma verbal como *anda* não tem, pelo menos não explicitamente representado, o traço [+plural] (ou tem o traço [-plural]) para ser pareado com o de um pronome como *Eles*? Para nós, as soluções podem variar entre os falantes.

Por exemplo, imaginemos um falante que tenha como forma menos marcada, ou mais comum, a concordância verbal não redundante. Para esse falante, um pronome como *eles* seria decomposto em traços como [3], [+plural], e uma forma verbal não redundante subsequente, como *gosta*, seria decomposta em traços como [3] e, supomos, [presente], e ficaria subespecificada, nos termos da MD (cf. ALMEIDA, 2022; BASSANI; LUNGUINHO, 2011), para número, uma vez que essa forma, para esse falante, pode ser interpretada tanto em [+plural] quanto em [-plural]. A nossa aposta é que, para esses falantes, a forma de número não seja levada em consideração para a satisfação das dependências, haja

vista que é variável (ou subespecificada). Lembremo-nos das crianças dos experimentos de Molina, Marcilese e Name (2017, 2018), que entre 5 e 6 anos mapearam formas como *Comeu doce* como podendo fazer menção a mais de um agente realizando a ação. No âmbito da concordância nominal, Augusto e Ferrari-Neto (2005) e Ferrari-Neto (2008) mostraram que crianças por volta de 2 anos se apoiavam no determinante para extrair informação de número (*Mostra os dabo pro Dedé*). Isso parece indicar que, no PB, se levarmos em conta um modelo de processamento como o de Lee e Phillips (2023), o traço de número pode ser pouco robusto ou pouco confiável no início da aquisição, o que deve mudar a partir do processo de escolarização e pressão/expectativa social de indivíduos mais escolarizados, o que nos faz chegar a um segundo tipo de falante, o mais escolarizado – que pra nós é especialmente interessante haja vista nosso experimento ter sido conduzido somente com universitários.

Esses falantes aprenderam explicitamente sobre concordância verbal (e nominal) e sabem que sujeito e verbo “devem” concordar em número. É possível pensar que esse tipo de conhecimento metalinguístico, associado à aquisição de padrões sociais e expectativas de um “falar corretamente”, aguce, ainda que inconscientemente, o tipo de traço que é checado pelo *parser* no momento da decomposição. Para uma mesma sentença como *Comeu doce*, que vimos acima, crianças de 10 anos e adultos já não têm as mesmas tendências que crianças mais novas – passam a parear essa forma mais com imagens em que apenas um sujeito realize a ação (MOLINA; MARCILESE; NAME, 2017, 2018). O que isso pode nos indicar é que, à medida que somos escolarizados e adquirimos mais conhecimento sociolinguístico, passamos a dar relativa mais importância para o traço de número, mesmo que inconscientemente. Daí, quando os *matches* não são satisfeitos como esperamos, podem acontecer aumentos de custo, seja por busca por outros itens linguísticos que satisfaçam o *match*, seja por reanálise, como vimos há pouco em Lee e Phillips (2023) e em Vasisht *et al.* (2019).

Dependendo do ambiente linguístico a que estejamos expostos na infância, esse processo de valoração pode ocorrer mais ou menos rapidamente. Por exemplo, crianças que crescem em um ambiente muito variável, ou com altas taxas de concordância não redundante, como os socioeconomicamente menos favorecidos, com maior presença masculina, podem demorar mais para começar a valorar o traço de número como robusto e confiável, ao passo que crianças que cresçam em ambientes com altas taxas de concordância redundante podem fazer essa valoração mais rapidamente.



Esse tipo de efeito esperado, o de aumento de custo de processamento de uma instância que não faça um *match* esperado (por conta do processo de escolarização, das pressões sociais e do conhecimento sociolinguístico, que todos os falantes, em maior ou menor grau, mais ou menos conscientemente, têm) pode ser pareado com o modelo do monitor sociolinguístico cunhado por Labov e colegas em 2011, inicialmente pensado para verificar como falantes reagiriam frente à quantidade de vezes em que são expostos a um estímulo, levando em consideração também a avaliação social do traço em questão (se é mais ou menos estereotipado socialmente, por exemplo).

Em uma bateria de experimentos de julgamento de adequação para um vaga de trabalho de jornalista baseada em estímulos linguísticos e na quantidade de apresentação desses estímulos, Labov *et al.* (2011) demonstraram alguns aspectos que teriam atuação no que chamaram monitor sociolinguístico, dos quais destacamos sensibilidade e assimetria, para nós aspectos de especial interesse. No experimento de Labov *et al.* (2011), um dos aspectos controlados foi a proporção entre variantes (estigmatizada e de prestígio) a que os sujeitos experimentais ouviam e como isso impactava o julgamento que eles faziam de uma aspirante à jornalista a partir de instâncias da forma não padrão /in/ (como em *workin* ') comparadas a instâncias da forma padrão /iŋ/ (como em *working*). O que os resultados indicaram é que, mesmo em proporções baixas de estímulos estigmatizados (ex. 10% ou 20%), em comparação com os prestigiados, os sujeitos experimentais já reagiram de maneira sensível àquela forma, fazendo julgamentos negativos. A essa capacidade de reagir mesmo quando a quantidade de estímulos marcados é baixa, os autores chamaram *sensibilidade*. Esse componente tem relação, também, com o da assimetria, que indica que os sujeitos experimentais podem reagir ainda mais negativamente se o contexto não favorece aquela instância (e aqui, havemos de convir que não há contexto experimental ecológico que possa despistar a assimetria totalmente – e nem é exatamente isso o que queremos, diga-se de passagem, por entendermos que o contexto experimental é, por definição, diferente de um contexto natural). Uma replicação do estudo do monitor sociolinguístico em língua portuguesa pode ser verificada em Freitag (2020).

Uma outra articulação que consideramos importante de ser feita é com elementos que Tamminga, MacKenzie e Embick (2016) consideraram importantes para a variação linguística intra-individual. O

efeito de *priming* para eles é bastante importante na produção linguística, assim como memória, imitação e planejamento [de fala]. Nós derivamos esse modelo para a compreensão por entendermos que as forças que atuam na produção são semelhantes àquelas da compreensão, dentro do pacote que chamamos de processamento. Nossa atenção especial está nos efeitos de *priming*. Em relação a eles, Tamminga, MacKenzie e Embick (2016) defendem que, uma vez que uma opção de produção de uma forma linguística tenha sido feita, há chances maiores de que essa forma seja mantida, dependendo do indivíduo – o que a sociolinguística, especialmente na concordância de número, pensa como paralelismo linguístico, com o princípio de que marcas levam a marcas e zeros levam a zeros (POPLACK, 1980; SCHERRE, 1994; 1998; VIEIRA, 2013). De maneira análoga, na compreensão, uma vez que uma forma tenha sido processada, ela deve abrir caminho para que formas análogas a ela sejam processadas de maneira mais facilitada também.

Ainda sobre *priming*, é importante ressaltarmos que a literatura tem sugerido que esse tipo de efeito ocorre apenas como um reflexo de algo que já estava registrado na mente anteriormente (HASTING *et al.*, 2007; LEDOUX *et al.*, 2007; TRAXLER *et al.*, 2000), ou seja, especificamente o tipo de *priming* a que estamos fazendo referência aqui reflete conhecimento linguístico. Supomos a partir disso que a compreensão de uma variante linguística, faça ela parte ou não do repertório do indivíduo, é facilitada à medida que esse indivíduo seja mais exposto a essa variação. No caso da concordância verbal variável, levando em conta os dados de literatura, especialmente aqueles que sugerem *input* variável no Rio de Janeiro (JAKUBÓW, 2018, por exemplo), supomos que todos os indivíduos tenham tido algum contato com possibilidades redundantes e não redundantes de concordância.

## 4 Experimento

A partir de agora, reportaremos o experimento que, lembramos, investigou custos de processamento de concordância verbal variável, e suas contrapartes no processamento de agramaticalidade, assim como a atuação de efeitos de adaptação via *priming* durante esse processamento. A pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (CEP) do Instituto de Estudos e Saúde Coletiva da Universidade Federal do Rio de Janeiro (IESC – UFRJ); CAAE 44809221.9.0000.5286; parecer

4.693.846. Todos os *scripts* e dados coletados nesse experimento, em atendimento aos preceitos da Ciência Aberta, podem ser encontrados em <<https://osf.io/5nphu/>>. Os dados são livres para retestagem, que é por nós, inclusive, incentivada.

O experimento foi conduzido levando em consideração três hipóteses: a primeira de que o processamento associado a variantes linguísticas (concordâncias redundante e não redundante) desencadearia custos diferentes daqueles associados à condição agramatical, devido ao *status* cognitivo subjacente diferenciado entre as formas; a segunda de que existiriam efeitos de sensibilidade e assimetria que os participantes exibiriam no experimento ao terem contato com a concordância não redundante e com a condição agramatical; e a terceira de que haveria um efeito de adaptação via *priming* para a leitura de concordância não redundante em comparação com a redundante que não seria constatado em relação à condição agramatical quando comparada à condição redundante.

Levando isso em consideração, três expectativas foram consideradas, portanto: (i) a de que as médias dos tempos de leitura dos segmentos críticos para cada uma das condições seriam significativamente diferentes; (ii) a de que haveria uma distância, nos momentos iniciais do experimento, entre a concordância não redundante e a condição agramatical, quando comparadas à condição redundante, causadas por efeitos de assimetria e sensibilidade nos participantes; e (iii) a de que a diferença entre as médias dos tempos de leitura das condições redundante e não redundante tendesse a diminuir depois desse primeiro momento, o que não aconteceria com a condição agramatical.

#### 4.1 Metodologia

Para esse experimento, foi utilizado o paradigma experimental de leitura automonitorada, em que os participantes fazem a leitura dos estímulos segmento a segmento, a partir de seu próprio ritmo de leitura, com os respectivos tempos de leitura de cada segmento registrados pela plataforma que hospeda o experimento. O experimento foi aplicado remotamente, a partir da plataforma PC IBEX (SCHWARZ; ZERH, 2021), que é uma plataforma gratuita desenvolvida para experimentos comportamentais remotos, utilizando linguagem JavaScript. Os experimentos desenvolvidos na plataforma podem ser rodados em qualquer computador com acesso à internet, sem custos extras para o experimentador ou para o participante, por meio de um *link*.

Durante o teste de leitura automonitorada, os sujeitos experimentais leram estímulos que simulavam trechos de entrevista em três condições: concordância redundante (como em *eles falam*), concordância não redundante (como em *eles fala*), e agramatical, do tipo sujeito<sub>p6</sub> + verbo<sub>p1</sub> (como em *\*eles falo*). Em cada estímulo, havia três ocorrências (OC1, OC2, e OC3) de sintagmas formados por *eles + verbo*, sempre nas mesmas posições (palavras 11-12, 23-24, e 32-33). A escolha por entrevistas foi intencional, uma vez que queríamos estímulos que emulassem, de maneira relativamente mais ecológica, a espontaneidade da fala, embora isso seja uma tarefa sempre desafiadora em contextos experimentais, e em que pudéssemos manipular mais de uma instância de segmento crítico no mesmo estímulo.

Na subseção a seguir, veremos especificamente como os estímulos foram controlados na elaboração da atividade.

#### 4.1.1 Materiais

Simulando trechos de entrevista, nossos estímulos foram pensados para compreender temas do cotidiano das pessoas, como rotina dos filhos, animais, lazer, faxina, festas de fim de ano e trabalho, por exemplo. Os estímulos eram sempre precedidos por uma pergunta geradora (apresentada como *Pergunta da entrevista*) e eram formados por uma sentença introdutória, em que o “entrevistado” introduziria a sua resposta e duas sentenças na sequência, sempre falando sobre um conjunto de seres que pudesse ser retomado pelo pronome *eles*, conforme podemos ver no exemplo (3) a seguir. Note que os verbos críticos, em negrito, estão conjugados nas três condições experimentais: concordância redundante, concordância não redundante e concordância agramatical. As condições eram dadas para o estímulo todo, de modo que todas as ocorrências dentro de um mesmo estímulo seguiam a mesma regra de concordância.

(3) Pergunta da entrevista: “Qual a rotina dos seus filhos, em casa?”

Eu sou pai de dois: o Pedro e o João. **Eles gost-am/a/o** bastante de estudar e **eles cheg-am/a/o** da escola com muito dever. Ao chegar, **eles tom-am/a/o** banho pra almoçar e brincar a tarde toda.

Além das posições em que sujeito e verbo ocorriam, também foram controlados o tempo verbal (presente do indicativo), o tipo de verbo (todos regulares), a conjugação (sempre a primeira), o tamanho

dos verbos (não mais do que três sílabas) e a quantidade de aparições dos verbos no conjunto experimental (somente duas vezes por condição, nos estímulos experimentais). Os estímulos eram todos menores do que 40 palavras. Também foram evitadas construções que pudessem ativar um efeito de *priming* de plural, de modo que instâncias como *dois filhos*, por exemplo, foram substituídas por *um menino e uma menina*.

Ao todo, nosso conjunto de estímulos continha 110 itens (90 experimentais, 30 em cada uma das três condições, e 20 distratores), que foram pseudorandomizados em 12 listas/subconjuntos de acordo com um quadrado latino. Cada participante teve contato com um subconjunto formado por 35 itens, a saber, 15 experimentais (5 em cada uma das três condições) e os 20 distratores, a ele atribuído aleatoriamente pela plataforma PC IBEX. Nenhum participante viu um dado estímulo em mais de uma condição, nem estímulos em uma mesma condição sequencialmente.

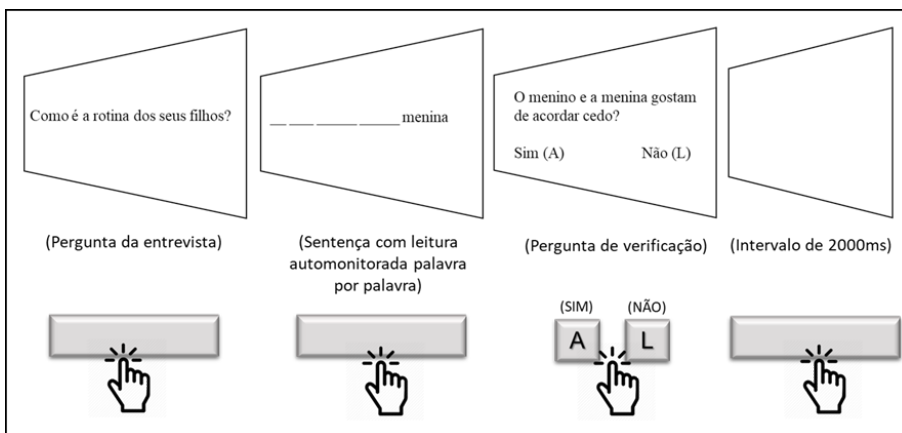
#### 4.1.2 Procedimentos

Tendo em vista que o experimento foi aplicado totalmente de maneira remota, os participantes foram recrutados por meio de postagens nas mídias sociais, que os direcionavam para um formulário na plataforma Google Forms, onde o Termo de Consentimento Livre e Esclarecido (TCLE) era apresentado e, caso o participante aceitasse participar do experimento, concordado. Além disso, alguns dados iniciais, de caráter socioeducacional, eram coletados. A exploração desses dados será feita em um momento mais oportuno, em outro trabalho.

Ao finalizar o preenchimento desse formulário, o participante automaticamente recebia um *link* para prosseguir para a plataforma PC IBEX, onde seria exposto novamente ao TCLE, e à explicação e ao treinamento do experimento, com três distratores-treino. O participante era instruído a apertar a barra de espaço toda vez que quisesse mudar de tela ou ler uma nova palavra de um estímulo. Todos os estímulos foram apresentados palavra por palavra, com tempos de resposta sendo coletados a cada vez que o participante pressionasse a barra de espaço. Como tarefa solicitada ao participante, ao final de cada um dos estímulos, ele deveria responder a uma pergunta de compreensão do tipo sim/não, apertando a tecla A para *sim* e L para *não*. Após cada pergunta de compreensão, o próximo estímulo era apresentado somente pelo menos após 2000 ms (esse tempo poderia variar para mais, caso o participante

demorasse para apertar o botão para prosseguir, mas nunca seria menor do que 2000 ms). Na metade do experimento, havia uma tela de pausa, em que o participante era instruído a descansar um pouco, se quisesse, mas sem fechar a tela. Na Figura 2, é possível ver uma ilustração da apresentação dos estímulos e da pergunta de verificação.

FIGURA 2 – Esquema ilustrativo dos procedimentos do teste de leitura automonitorada



Fonte: Almeida (2022).

#### 4.1.3 Participantes

Utilizamos como critério de recrutamento de participantes aqueles que fossem cariocas, nascidos e criados (pelo menos por maior parte do tempo durante a infância) na cidade do Rio de Janeiro, com pelo menos ensino superior em curso, que não fossem estudantes de cursos de Letras, dos quais se espera um olhar mais atento para questões de cunho linguístico, sobretudo em termos de variação linguística, e que tivessem entre 18 e 30 anos de idade, para que se evitasse ruído nos dados por meio de diferenças geracionais. Os participantes também deveriam ter computador disponível para a execução do experimento. Foram oferecidos certificados de atividades complementares para aqueles que optassem por recebê-lo. Utilizamos como critério de exclusão de participantes, após a realização do experimento, um ponto de corte a partir das médias das acurácias nas perguntas de verificação do grupo subtraída de três vezes o

desvio padrão. Dos 73 participantes que fizeram o experimento, apenas 1 foi excluído por essa razão (acurácia de 80%, em um grupo em que o limite mínimo foi de 83,40%;  $= \bar{x}$  95,06%; DP = 4,31%).

Desse modo, portanto, tivemos 72 participantes (48 mulheres, 23 homens e 1 pessoa que marcou a opção de gênero *outros*). Em média, os participantes tinham 22 anos e 6 meses (DP = 2;10). Também foram coletados, como informações complementares dos participantes, os bairros em que moravam, o nível de instrução dos cuidadores na infância, área de conhecimento na universidade, hábitos de leitura, e tipos de escola nos ensinos fundamental e médio. Foram rodados modelos mistos (cf. Materiais suplementares, disponíveis em <<https://osf.io/5nphu/>>) que incluíram todos esses dados coletados como possíveis variáveis preditoras dos tempos de resposta, mas nenhum resultado se mostrou significativo naquele momento, exceto *bairro*, que, com significâncias aleatórias entre grupos de bairros, não puderam contribuir com a discussão neste momento.

## 5 Resultados

Antes da análise estatística, os dados foram filtrados para a exclusão de *outliers*. Foram considerados *outliers* todos os tempos de resposta (TR) inferiores a 100 ms e superiores a 3000 ms, o que correspondeu a cerca de 1% do quantitativo de dados efetivamente analisados. Entendemos que antes de 100 ms seria improvável que alguém tivesse de fato lido a palavra e que, após 3000 ms, a leitura seria mais monitorada do que o esperado para um teste dessa natureza.

Foram incluídos na análise estatística os TR dos verbos críticos e dos dois segmentos subsequentes, com a intenção de encapsular possíveis efeitos de *spill over*<sup>6</sup>. Para análise, utilizamos o *software* RStudio (versão

---

<sup>6</sup> Conforme lemos em Almeida (2022, p. 71):

esse efeito consiste em um ‘atraso’ na detecção de maior engajamento cognitivo no segmento que está sendo manipulado pelo pesquisador. Isso acontece porque leitura e processamento não são processos que acontecem simultaneamente. O engajamento no processo linguístico de um trecho e seus eventuais custos relacionados podem ocorrer quando o participante já está lendo a palavra ou trecho seguinte. Dessa forma, o efeito esperado sobre o segmento crítico (um aumento ou diminuição do TR) é ‘derramado sobre’ (daí *spill over*) a próxima palavra ou segmento, possivelmente indicando ainda, segundo Augusto *et al.* 2020 (p. 126), ‘um efeito de reparo ou reanálise’ (cf. Rayner

4.1.0; R CORE TEAM, 2011). Como os dados brutos não tinham uma distribuição normal ( $p < 0,0001$ , no teste Jarque Bera (JARQUE; BERA, 1980)), como geralmente acontece com dados de leitura automonitorada, que costumam ter distribuição assimétrica à direita (ou *skewed to the right*), nós fizemos uma transformação do tipo *BoxCox* (BOX; COX, 1964; GUERRERO, 1993), que ajusta um valor *lambda* ótimo como parâmetro transformacional dos dados. A partir dessa transformação, nossos dados passaram a ter uma distribuição normal ( $p = 0,959$ , também no teste Jarque Bera (JARQUE; BERA, 1980)). A transformação logarítmica não atingiu os mesmos resultados ( $p < 0,0001$ ).

Primeiro, para fins de comparações gerais entre as condições, rodamos um modelo linear misto<sup>7</sup> que teve TR como variável dependente e condição como efeito fixo, considerando participantes e itens como efeitos aleatórios. Esse modelo foi considerado melhor do que o nulo, em uma análise aninhada ( $X^2(2) = 219,52$ ;  $p < 0,0001$ ). Nesse primeiro momento, então, propositalmente não consideramos as ocorrências, porque queríamos entender o comportamento de cada uma das condições, cujos tempos do *output* do modelo estão no gráfico a seguir (Figura 3). Para que a inteligibilidade do *output* do modelo não seja comprometida, reportamos os valores de média e erro padrão (SE) a partir de modelos que consideraram os tempos brutos e valores *t* e *p* a partir dos modelos que levaram em conta os TR transformados pelo parâmetro *lambda*. Desse modo, inteligibilidade e robustez podem ser preservadas.

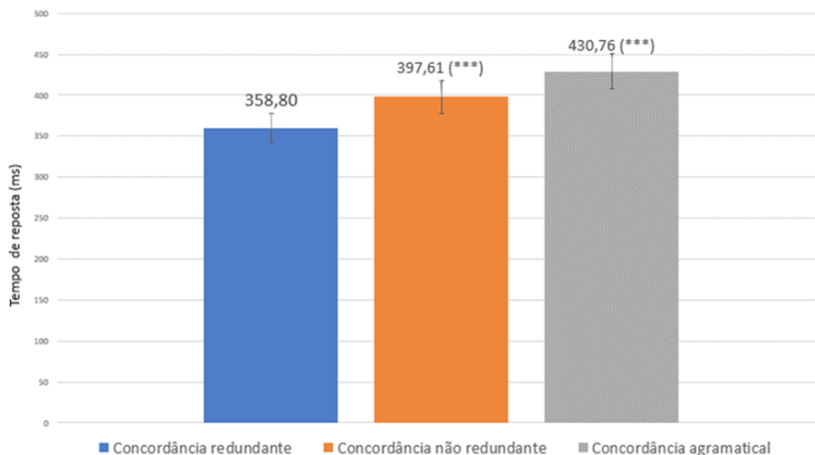
---

(1998), para mais detalhes sobre o efeito, e Augusto *et al.* 2020, para um exemplo de estudo no PB com leitura automonitorada).

<sup>7</sup> Pacote *lme4* (BATES *et al.*, 2015) no RStudio.



FIGURA 3 – Médias dos TR por condição. No gráfico, a coluna azul representa a média dos TR para a condição *redundante*, a laranja, o mesmo para a condição *não redundante*, e a cinza, o mesmo para a condição *agramatical*. As barras de erro representam 5% para mais e para menos em relação aos valores das médias



Fonte: Almeida (2022).

A análise indicou aumentos significativos no TR das condições *não redundante* ( $= \bar{x}$  397,61 ms; SE = 14,276;  $t = 7,887$ ;  $p = 3,43e-15$ ) e *agramatical* ( $= \bar{x}$  430,76; SE = 4,975;  $t = 14,894$ ;  $p < 0,0001$ ), em comparação com a condição *redundante* ( $= \bar{x}$  358,80; SE = 14,276,  $t = 1289,351$ ,  $p < 0,0001$ ). Para verificar a diferença entre as condições *não redundante* e *agramatical* rodamos uma análise *post-hoc* do tipo *Tukey*, que faz comparações de média par a par, cujo resultado mostrou uma diferença significativa entre elas (SE = 4,98;  $z = -7,009$ ;  $p < 0,0001$ ). Esses resultados demonstram que há um efeito geral de condição, que faz com que cada uma das condições seja lida em um ritmo diferente, levando-nos a inferir que cada uma das concordâncias é processada de modo distinto pelos participantes.

Para uma investigação mais profunda dos dados, pensando nos efeitos de cada uma das três ocorrências manipuladas em cada um dos estímulos, os TR obtidos foram analisados em função da variáveis condição e ocorrência, em um modelo linear misto<sup>8</sup> que também

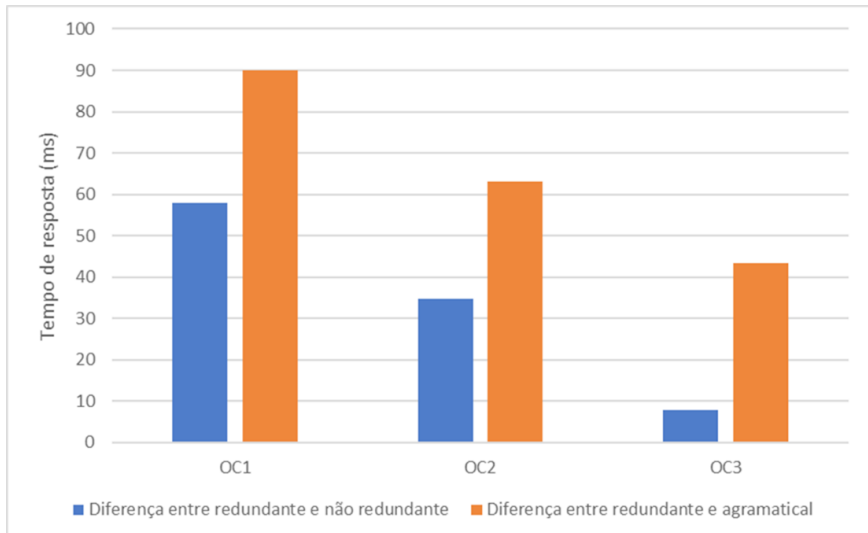
<sup>8</sup> Pacote lme4 (BATES *et al.*, 2015) no RStudio.

considerou os efeitos aleatórios de participante e item (uma análise de modelos aninhada mostrou que esse era o melhor modelo, i.e., o modelo com maior poder explicativo para os dados ( $X^2(6) = 23,832$ ;  $p = 0,0005$ )). Assim como na análise anterior, as médias dos tempos e os valores de erro padrão (SE) foram reportados com base em um *output* com os tempos brutos e os valores de significância estatística com base no modelo ajustado a partir do valor ótimo *lambda*.

Tabela 1 - Resumo dos valores médios e significância dos tempos de leitura, em ms, por condição por ocorrência. Significância comparada com o intercepto do modelo (Condição redundante / OC1). Códigos de significância: 0 ‘\*\*\*\*’ 0.001 ‘\*\*\*’ 0.01 ‘\*\*’ 0.05 ‘.’ 0.1 ‘.’ 1.

Condição / Ocorrência	OC1	OC2	OC3
<i>Redundante</i>	361,25 (SE = 15,10)	358,53 (SE = 8,54)	353,89 (SE = 8,54)
<i>Não redundante</i>	419,25 (***) (SE = 8,56)	393,22 (SE = 12,09)	361,72 (.) (SE = 12,09)
<i>Agramatical</i>	451,23 (***) (SE = 8,57)	421,65 (SE = 12,09)	397,29 (SE = 12,10)

O que esses dados mostram é que os TR aumentam de maneira significativa da condição *redundante* para as condições *não redundante* e *agramatical* na OC1 e que vão diminuindo ao longo das ocorrências. Em termos de valores absolutos do *output* do modelo, conseguimos ver que as diferenças entre os TR das condições *redundante* e *não redundante* diminuem mais rapidamente ao longo das ocorrências do que a diferença entre as condições *redundante* e *agramatical*, o que podemos ver no gráfico a seguir (Figura 4). Note que, na OC3, a diferença entre *redundante* e *não redundante* é de menos de 10 ms, enquanto entre *redundante* e *agramatical* é de mais de 40 ms.

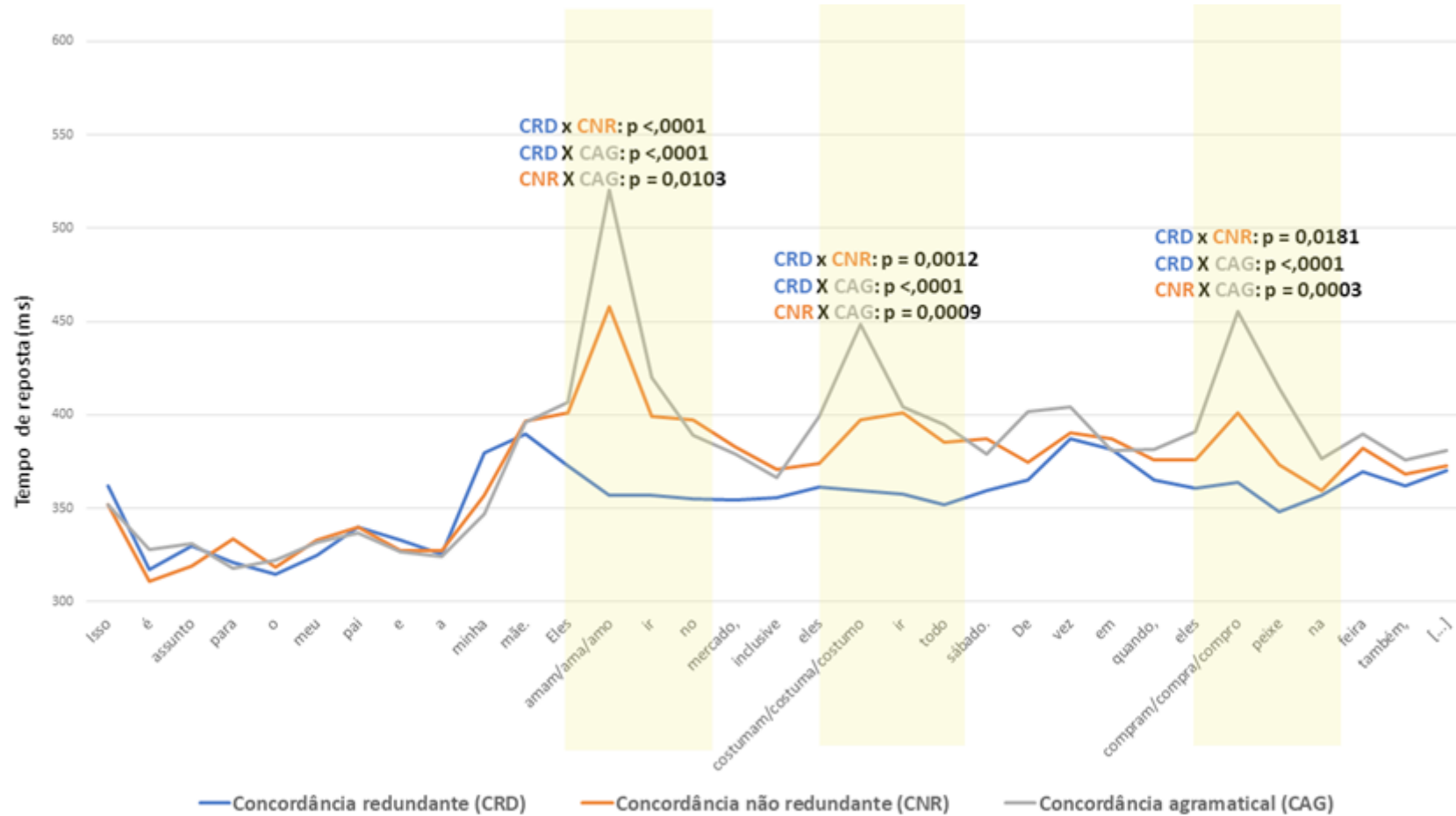
FIGURA 4 – Diferença de TR das condições *não redundante* e *agramatical* em comparação com a condição *redundante*

Fonte: Almeida (2022).

Contudo, apenas com esse modelo, ainda não é possível saber se essas diminuições são significativas por condição, pois o modelo está tomando como intercepto a condição *redundante* e a primeira ocorrência. Por isso, nós rodamos uma análise *post-hoc* do tipo *Tukey* para comparar as médias par a par. Na Figura 5, é possível ver esses resultados e um gráfico que mostra as regiões de interesse consideradas para a análise estatística.

A partir dessa análise, foi possível notar comportamentos diferentes entre as ocorrências de cada uma das condições. Por exemplo, os p valores das diferenças entre as condições *redundante* e *agramatical* permanece inalterado ( $<0,0001$ ), independente da ocorrência. Já para as diferenças entre *redundante* e *não redundante*, podemos notar uma progressiva perda de significância dos p valores ( $<0,0001$  na OC1; 0,0012 para OC2 e 0,018 na OC3). Finalmente, quando comparamos as condições *não redundante* e *agramatical*, é possível vermos que a diferença entre elas cresce continuamente, em termos estatísticos (0,0103 na OC1; 0,0009 na OC2; e 0,0003 na OC3).

FIGURA 5 – Gráfico de palavras por condição, com comparações de médias *post-hoc* por ocorrência. No gráfico, a linha azul representa a média dos TR de cada palavra na condição *redundante*; a laranja representa o mesmo para a condição *não redundante* e a cinza representa as mesmas informações para a condição *agramatical*. As áreas sombreadas representam os segmentos críticos (verbo mais dois segmentos subsequentes) de cada uma das três ocorrências. Os p valores representados dizem respeito às comparações par a par da análise *post-hoc* do tipo *Tukey*



Fonte: Almeida (2022).

## 6 Discussão

Neste trabalho, reportamos um experimento de leitura automonitorada desenvolvido no escopo da concordância verbal variável no PB, especificamente o falado na cidade do Rio de Janeiro, para o qual foram traçadas três hipóteses: (i) o processamento associado a variantes linguísticas (concordâncias redundante e não redundante) desencadeariam custos diferentes daqueles associados à condição agramatical, devido ao *status* cognitivo subjacente diferenciado entre as formas; (ii) existiriam efeitos de sensibilidade e assimetria que seriam detectados no experimento ao terem contato com a concordância *não redundante*, com efeitos similares para a condição *agramatical*; e (iii) haveria um efeito de adaptação em função de um efeito de *priming* para a leitura da condição *não redundante* quando comparada à redundante que não seria detectado para a condição *agramatical* no mesmo tipo de comparabilidade.

A partir dos dados coletados, é possível ver que os resultados obtidos se mostraram compatíveis com as previsões experimentais estabelecidas, e permitem que as hipóteses nulas sejam rejeitadas. Em primeiro lugar, como é possível ver na Tabela 1 e na Figura 3, as médias dos tempos de leitura para cada uma das condições, independentemente das ocorrências, se mostraram diferentes, de modo que a concordância redundante foi lida mais rapidamente do que a não redundante, que, por sua vez, foi lida mais rapidamente do que a condição agramatical. Esse achado é coerente com a literatura, e pode ser articulado com a segunda hipótese, a que previa efeitos de assimetria e sensibilidade para a condição não redundante – e efeitos semelhantes para a condição agramatical.

Nesse sentido, o que ocasionaria o acréscimo nas médias dos TR para a condição não redundante seriam os efeitos de assimetria e sensibilidade causados por uma variante estigmatizada. Nossos participantes, como um grupo de universitários, certamente não esperariam essas formas em um experimento linguístico feito no âmbito de uma universidade, ainda que remotamente, embora eventualmente possam fazer uso delas. Em outras palavras, a forma não redundante é assimétrica para o contexto do experimento e as expectativas criadas, ainda que inconscientemente, em torno dele, e os participantes foram sensíveis a isso logo na primeira ocorrência (como podemos ver na Tabela 1 e na Figura 5).

Para a forma agramatical, também há uma espécie de sensibilidade para uma forma diferente, mas o esforço cognitivo refletido nos TR foi maior, e acreditamos que isso se deva ao fato de a forma agramatical, como

a construímos, não ser uma opção acessível dadas as regras combinatórias dos morfemas presentes na Lista 1 de qualquer falante. Também por esse motivo, não podemos cravar um efeito de sensibilidade *stricto sensu*, como previsto no monitor sociolinguístico laboviano, haja vista que a forma agramatical não é exatamente uma variante linguística – e esse efeito deve mesmo estar relacionado com formas linguísticas atestadas. Assim, é provável que tenha havido para a condição agramatical um esforço maior como uma tentativa de mapear a forma com algum tipo de significado, possivelmente com processos de reintegração e retestagem de traços armazenados, como previsto por Lee e Phillips (2023).

Há ainda que se levar em consideração<sup>9</sup> que há a possibilidade de os sujeitos do experimento terem feito, para a condição agramatical, uma leitura diferente do que esperávamos, o que pode, também, ter influenciado os resultados. Fato é que há realizações fonéticas no português falado no Rio de Janeiro que possibilitam que *eles tomo*, por exemplo, seja uma pronúncia possível para *eles tomam*, como apontado no trabalho de Rocha (2021). Ou ainda que seja mapeado como passado, com *eles tomo* sendo lido como *eles tomô* (próximo a *eles tomou*). No primeiro caso, parece-nos improvável que esse tipo de leitura seja a mais prototípica para os nossos sujeitos experimentais – o que, inclusive, tornaria a construção menos distante daquelas que consideramos gramaticais neste estudo. Sendo esse o caso, haveria uma diminuição mais expressiva dos TRs, o que não foi observado. Já no segundo caso, o da possível leitura como passado, teríamos, ainda assim, uma quebra de expectativa influenciada pelo contexto, que é todo construído no presente – o que justificaria o aumento dos TRs. Para ambos os casos, apesar de nossas especulações, não há como garantir inequivocamente que leitura está sendo de fato feita, sendo necessários experimentos paralelos que tenham como foco específico essa questão.

Apesar dos possíveis impasses no que diz respeito à condição agramatical, a terceira previsão é a que contribui de maneira mais consistente para o entendimento da compreensão linguística diante do problema da falta de invariabilidade, se confirmada. Levando em consideração que na análise *post-hoc*, que comparou médias par a par (Figura 3), obtivemos uma aproximação consistente das condições

---

<sup>9</sup> Agradecemos especialmente a um dos pareceristas, que nos atentou para essa importante observação.

*redundante e não redundante*, é plausível supor a atuação de um efeito de adaptação via *priming* (TAMMINGA; MACKENZIE; EMBICK, 2016). Uma vez que os participantes foram capazes de fazer pareamentos da forma não redundante com o disponível em seu conhecimento gramatical (uma vez que essa é uma forma produtiva no PB), o esforço cognitivo refletido em TR foi paulatinamente diminuído, em comparação à concordância redundante, mais esperada para esse tipo de experimento. A partir do aumento dos p valores, o que reflete uma perda de significância estatística entre as diferenças de TR das duas condições, vemos que elas passam a ser lidas, progressivamente, no mesmo ritmo. A nossa sugestão é que isso é reflexo do fato de ambas as formas serem atestadas e internalizadas na gramática do falante carioca, talvez com níveis de familiaridade variados entre os participantes, a depender do grau de contato com a variante.

Nesse sentido, o efeito de *priming* também tem sua contribuição, pois é cada vez mais fácil processar uma forma já anteriormente processada, via ativação na memória. Nos estudos sociolinguísticos, esse efeito geralmente é referido por meio da ideia que marcas levam a marcas e zeros levam a zeros (especialmente para concordância, cf. POPLACK, 1980; SCHERRE, 1994; 1998; VIEIRA, 2013). Curiosamente, esse efeito não acontece com a condição redundante, cujas diferenças de TR não são significativas ao longo das três ocorrências (ver Tabela 1). É possível supor que a forma redundante seja tão esperada que não haja efeitos significativos em processá-la anteriormente, mas o mesmo não acontece com a concordância não redundante, para qual o *priming* facilita o acesso a uma forma inesperada e socialmente estigmatizada, e faz com que o custo de processamento na compreensão seja atenuado, tendendo a ser tão automatizado quanto à concordância não redundante no experimento, passadas as questões iniciais de assimetria.

A nosso ver, não é provável que esses mesmos efeitos de *priming* tenham colaborado para a leitura mais rápida da concordância agramatical ao longo das ocorrências. Isolada, essa condição só foi lida mais rapidamente em relação a si mesma, e não em relação às outras condições – se distanciando, inclusive, da leitura não redundante, como aferem os valores de significância da análise *post-hoc*. Mas fato é que houve um decréscimo no tempo de leitura, e acreditamos que isso possa estar associado a um tipo de adaptação viso-perceptual, que armazenaria a forma na memória e a deixaria mais fácil de ser acessada, ou a um processamento estratégico que modularia a expectativa para a violação a ponto de amenizar o efeito dessa violação em termos de carga cognitiva.

Não deve se tratar, portanto, de um tipo de *priming*, porque a literatura tem defendido a ideia de *primings* para formas já existentes na língua, o que não é o caso desse tipo de concordância (HASTING *et al.*, 2007; LEDOUX *et al.*, 2007; TRAXLER *et al.*, 2000), pelo menos enquanto levamos em consideração, para elas, a leitura da forma verbal de primeira pessoa do singular com um sujeito de terceira pessoa do plural<sup>10</sup>. Fica claro quando olhamos para as diferenças de TR entre as condições agramatical e não redundante que há fenômenos distintos em operação em cada uma delas. Ou seja, podemos afirmar que variação linguística é processada de maneira diferente de agramaticalidade (que estaria mais próximo daquilo que puristas apontam como erro ou que psicolinguistas às vezes denominam violação morfossintática).

Levando em consideração que as previsões foram confirmadas, a partir de agora apontaremos como esses dados poderão ser mais bem explorados no futuro, e que outras ferramentas e metodologias poderão nos ajudar na compreensão do fenômeno.

## 7 Apontamentos futuros

Embora os resultados desse estudo tenham lançado luz sobre a compreensão do processamento de instâncias variáveis gramaticais e de instâncias agramaticais, contribuindo para a interface entre a psicolinguística e a sociolinguística, principalmente no que diz respeito aos efeitos de adaptabilidade via *priming* atuante, há desdobramentos que podem muito enriquecer e aprimorar o entendimento que temos até o momento.

Primeiro, julgamos que, em estudos futuros, é necessário que se façam retestes, incluindo mais ocorrências em cada estímulo com o objetivo de verificar se finalmente concordâncias redundante e não redundante podem ser lidas com o mesmo ritmo, sem diferença significativa entre elas. Além disso, é importante pensar em outros grupos para testagens futuras, levando o teste, por exemplo, para estudantes do ensino médio, especialmente aqueles da rede pública de regiões menos favorecidas na cidade do Rio.

---

<sup>10</sup> Se, por acaso, os sujeitos experimentais tiverem tido outras leituras para essa construção, possíveis a partir do estudo de Rocha (2021), o não efeito de *priming* na construção aqui, até o momento, entendida como agramatical precisará ser repensado. Como não é possível, ainda, garantir que não foi feita de fato a leitura que esperávamos, essa é uma discussão que deverá ter lugar em futuros trabalhos.



Além disso, é necessário que se façam testagens com outras técnicas experimentais. A eletroencefalografia (EEG), por exemplo, pode nos dar resultados temporais mais precisos e efeitos de comparabilidade mais robustos, refletindo processos mais automatizados e inconscientes. Há pouquíssimos experimentos que tenham levado em consideração essa técnica no escopo da variação linguística (cf. Soto e Almeida (2021), para uma reflexão crítica da literatura; Loudermilk (2013), para um estudo de EEG e variável morfofonológica; e Zaharchuk, Shevlin e Van Hell (2021), para uma variável sintática).

Finalmente, novas investigações podem explorar ainda mais a ecologia dos dados, levando em consideração até mesmo aspectos de covariação não redundante dentro dos sintagmas (um estudo relativamente mais inflado, mas, sem dúvida, bastante interessante).

## 8 Conclusão

No escopo da interface entre a psico e a sociolinguística, especialmente no que diz respeito à concordância verbal variável, alguns trabalhos já vinham apontando custos maiores para o processamento de concordância não redundante, quando comparada à concordância redundante (MARCILESE *et al.*, 2015, 2017). Com isso em mente, o presente trabalho pôde contribuir para a literatura da área, agora incorporando efeitos de acomodação/adaptação via *priming* como atuantes no processo de compreensão (socio)linguística, além de uma aposta em estímulos mais ecológicos. Sem dúvidas, ainda há bastante a ser descoberto, mas esperamos que esses dados contribuam, empiricamente, para atestar o que há muito temos falado na linguística, especialmente na sociolinguística: não chamar de erro o que a linguística chama variação.

## Agradecimentos

Agradeço à minha orientadora e primeira leitora crítica desse texto, Profa. Marije Soto, assumindo quaisquer erros ou má formulações persistentes como meus; à professora Juliana Novo Gomes pela disponibilização do *script* do PC IBEX que serviu como base para esse experimento; e à professora Christina Gomes por algumas das referências muito específicas de concordância variável. Agradeço também a cada voluntário que dedicou seu tempo em contribuir com o experimento aqui

reportado, e aos pareceristas que, com leitura atenta, apontaram melhorias importantes para o texto. Este trabalho foi realizado com financiamento da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES), no âmbito do PROEX.

## Referências

ADGER, D.; SMITH, J. Variation in agreement: a lexical feature-based approach. *Lingua*, cidade, v. 120, n. 5, p. 1109-1134, 2010. DOI: <https://doi.org/10.1016/j.lingua.2008.05.007>.

ALMEIDA, W. C. *Processamento e percepção da concordância verbal variável de P6 entre universitários da cidade do Rio de Janeiro*. 2022. 165f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2022.

ARAÚJO, S. S. F. A concordância verbal nos *continua* sociolinguísticos do português brasileiro e luandense. *Interdisciplinar*, Itabaiana, v. 24, 2016. Disponível em: <<https://seer.ufs.br/index.php/interdisciplinar/article/view/5399/4424>>. Acesso em: 14 mar. 2023.

ARRUDA, A. L. Políticas da Educação Superior no Brasil: expansão e democratização: um debate contemporâneo. *Espaço do Currículo*, João Pessoa, v. 3, n. 2, p. 501-510, 2011. DOI: <https://doi.org/10.15687/rec.v3i2.9661>.

AUGUSTO, M. R. A.; SOTO, M.; SENA, N.; BERNARDES, J. Resumptivos em relativas de objeto direto: resultados de leitura automonitorada. *Revista Letras*, Curitiba, v. 101, p. 114-143, 2020. DOI: <http://dx.doi.org/10.5380/rel.v101i0.72650>.

BASSANI, I. S.; LUNGUINHO, M. V. Revisitando a flexão verbal do português à luz da Morfologia Distribuída: um estudo do presente, pretérito imperfeito e pretérito perfeito do indicativo. *ReVEL*, edição especial n. 5, p. 199-227, 2011. Disponível em: <[http://www.revel.inf.br/files/artigos/revel\\_esp\\_5\\_revisitando\\_a\\_flexao\\_verbal\\_do\\_portugues.pdf](http://www.revel.inf.br/files/artigos/revel_esp_5_revisitando_a_flexao_verbal_do_portugues.pdf)>. Acesso em: 14 mar. 2023.

BATES, D.; MAECHLER, M.; BOLKER, B.; WALKER, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*,

Innsbruck, v. 67, n. 1, p. 1-48, 2015. DOI: <https://doi.org/10.18637/jss.v067.i01>.

BORTONI-RICARDO, S. M. A concordância verbal em português: um estudo de sua significação social. In: VOTRE, S.; RONCARATI, C. (orgs.). *Anthony Julius Naro e a Linguística no Brasil: uma homenagem acadêmica*. Rio de Janeiro: 7Letras, 2008. p. 362-380.

BOX, G. E. P.; COX, D. R. An analysis of transformations. *JRSS B*, cidade, v. 26, n. 2, p. 211–246, 1964. Disponível em: <<https://www.jstor.org/stable/2984418>>. Acesso em: 14 mar. 2023.

BRASIL. Lei nº 12.711 de 29 de agosto de 2012. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. *Diário Oficial da União*, seção 1, página 1, publicado em 30 ago. 2012.

BYBEE, J. Morphology: A Study of the Relation between Meaning and Form, *Typological Studies in Language*. Amsterdam: John Benjamins, 1985.

CORRÊA, L. M. S.; AUGUSTO, M. R. A.; FERRARI-NETO, J. The early processing of number agreement in the DP: evidence from the acquisition of Brazilian Portuguese. *Boston University Conference on Language Development* (Apresentação de trabalho). 2005.

EMBICK, D.; NOYER, R. Distributed Morphology and the Syntax/Morphology Interface. In: RAMCHAND, G.; REISS, C. (orgs.). *Oxford Handbook of Linguistics Interfaces*. New York: Oxford University Press, 2006. p. 298-324.

FERRARI-NETO, J. *Aquisição de Número Gramatical no Português Brasileiro: Processamento de Informação de Interface e Concordância*. 2008, 154f. Tese (Doutorado em Letras) – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, 2008.

FREITAG, R. M. K. Effects of the Linguistics Processing: Palatals in Brazilian Portuguese and the Sociolinguistic Monitor. *University of Pennsylvania Working Papers in Linguistics*, Pennsylvania, v. 25, n. 2, p. 21-30, 2020. Disponível em: <<https://repository.upenn.edu/pwpl/vol25/iss2/4>>. Acesso em: 14 mar. 2023.

GARRAFFA, M.; FYNDANIS, V. Linguistic theory and aphasia: an overview. *Aphasiology*, cidade, v.34, n.8, p.905-926, 2020. DOI: <https://doi.org/10.1080/02687038.2020.1770196>.

GRACIOSA, D. *Concordância verbal na fala culta carioca*. 1991. 125f. Dissertação (Programa de Pós-Graduação em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, 1991.

GUERRERO, V. M. Time-series analysis supported by power transformations. *Journal of Forecasting*, v. 12, n. 1, p. 37–48, 1993. DOI: <https://doi.org/10.1002/for.3980120104>.

HALLE, M.; MARANTZ, A. Distributed Morphology and pieces of inflection. In: HALE, K.; KEYSER, S. J. (eds.) *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. Cambridge: MIT Press, 1993. p. 111-176.

HASTING, A. S.; KOTZ, S. A.; FRIEDERICI, A. D. Setting the Stage for Automatic Syntax Processing: The Mismatch Negativity as an Indicator of Syntactic Priming. *Journal of Cognitive Neuroscience*, v. 19, n. 3, p. 386-400, 2007. DOI: <https://doi.org/10.1162/jocn.2007.19.3.386>.

HEALD, S.; KLOS, S.; NUSBAUM, H. Understanding Speech in the Context of Variability. In: HICKOK, G.; SMALL, S. (eds.). *Neurobiology of Language*. Chicago: Elsevier, 2016. p. 195-208. DOI: <https://doi.org/10.1016/B978-0-12-407794-2.00017-1>.

JAKUBÓW, A. P. S. P. *Language acquisition based on variable input: the case of number agreement in Brazilian Portuguese*. 2018. 191f. Tese (Doutorado em Letras/Estudos da Linguagem) – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, 2018.

JAKUBÓW, A. P. S. P.; CORRÊA, L. S. The production of variable number agreement in Brazilian Portuguese: a procedural and developmental account. In: AVRAM, L.; SEVCENCO, A.; TOMESCU, V. (eds.). *L1 Acquisition and L2 Learning. The view from Romance*. John Benjamins Publishing Company, 2021. p. 109-132. DOI: <https://doi.org/10.1075/lald.65.05jak>.

JARQUE, C. M.; BERA, A. K. Efficient test for normality, homoscedasticity, and serial independence of residuals. *Economic Letters*, Holanda do Norte, v. 6, n. 3, p. 255-259, 1980. DOI: [https://doi.org/10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5).

LABOV, W. Some sociolinguistic principles. In: PAULSTON, C. B.; TUCKER, G. R. (eds.). *Sociolinguistics: the essential readings*. Massachusetts: Blackwell Publishing, 2003. p. 235-250.

LABOV, W.; ASH, S.; RAVINDRANATH, M.; WELDON, T.; BARANOWSKI, M.; NAGY, N. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, v. 15, n. 4, 431-463, 2011. DOI: <https://doi.org/10.1111/j.1467-9841.2011.00504.x>.

LEDOUX, K.; TRAXLER, M. J.; SWAAB, T. Y. Syntactic priming in compression: evidence from event-related potentials. *Psychol Sci*, v. 18, n. 2, 135-143, 2007. DOI: 10.1111/j.1467-9280.2007.01863.x.

LEE, E. K. Y; PHILLIPS, C. Why non-native speakers sometimes outperform native speakers in agreement processing. *Bilingualism: Language and Cognition*, v. 26, n. 1, 2023, 152-164. DOI: 10.1017/S1366728922000414.

LEMLE, M.; NARO, A. J. Competências básicas do português. *Relatório final de pesquisa apresentado às instituições patrocinadoras Fundação Movimento Brasileiro (MOBRAL) e Fundação Ford*. Rio de Janeiro, 1977.

LOUDERMILK, B. C. *Cognitive Mechanisms in the Perception of Sociolinguistic Variation*. 2013. 202f. Tese (Doutorado em Linguística). Universidade da Califórnia, 2013.

MAIA, M.; NASCIMENTO, G. Anomalias de forma e de conteúdo em português brasileiro: um estudo preliminar de rastreamento ocular da leitura e de avaliação de aceitabilidade. *Cadernos de Tradução*, Florianópolis, v. 40, n. esp. 2: Leitura, Tradução e Cognição, p. 45-73, 2020. DOI: <https://doi.org/10.5007/2175-7968.2020v40nesp2p45>.

MARCILESE, M; HENRIQUE, K. S.; AZALIM, C; NAME, C. Processamento da concordância variável no PB em uma perspectiva experimental. *Revista Linguística*, Rio de Janeiro, v. 11, n. 1, p. 118-134, 2015. DOI: 10.17074/2238-975X.2015v11n1p118.

MARCILESE, M.; RODRIGUES, E. S.; AUGUSTO, M. R. A.; HENRIQUE, K. S. Efeitos de distância linear e marcação no processamento da concordância verbal variável do PB. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 25, n. 3, p. 1291-1325, 2017. DOI: <http://dx.doi.org/10.17851/2237-2083.25.3.1291-1325>.

MOLINA, D. S. L. *Aquisição da linguagem e variação linguística: um estudo sobre a flexão verbal variável na aquisição do PB*. 2018. 277f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, 2018.

MOLINA, D. S. L.; MARCILESE, M.; NAME, C. Ora está, ora não está: *input* variável e aquisição da flexão verbal de 3ª pessoa do plural no PB. *Matraga*, Rio de Janeiro, v. 24, n. 41, p. 288-309, 2017. DOI: <http://dx.doi.org/10.12957/matraga.2017.28498>.

MOLINA, D. S. L.; MARCILESE, M.; NAME, C. Aquisição da linguagem e variação linguística em diálogo: investigando a produção e a compreensão da flexão verbal de terceira pessoa do plural no PB. *Revista Linguística*, Rio de Janeiro, v. 14, n. 3, p. 35-54, 2018. DOI: <http://dx.doi.org/10.31513/linguistica.2018.v14n3a21334>.

MOLINARO, N.; BARBER, H.; CARREIRAS, M. Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, v. 47, n. 8, p. 908-930, 2011. DOI: [10.1016/j.cortex.2011.02.019](https://doi.org/10.1016/j.cortex.2011.02.019).

NARDY, A.; CHEVROT, J.; BARBU, S. Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study. *Language Variation and Change*, Cambridge, v. 26, n. 3, p. 273–301, 2014. DOI: [doi:10.1017/S0954394514000131](https://doi.org/10.1017/S0954394514000131).

NARO, A. J. The Social and the Structural Dimensions of a Syntactic Change. *Language*, cidade, v. 57, n. 1, p. 63-98, 1981. DOI: <https://doi.org/10.2307/414287>.

POPLACK, S. The notion of the plural in Puerto Rico Spanish: competing constraints on /s/ deletion. In: LABOV, W. (ed.). *Locating language in time and space*. Philadelphia: University of Pennsylvania, 1980, p. 55-67. Disponível em: <<http://www.sociolinguistics.uottawa.ca/shanapoplack/pubs/articles/Poplack1980Plural.pdf>>. Acesso em: 15 mar. 2023.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Acesso em: 14 mar. 2023.

RAYNER, K. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, v. 124, n. 3, p. 372-422, 1998. DOI: [10.1037/0033-2909.124.3.372](https://doi.org/10.1037/0033-2909.124.3.372).

ROCHA, J. Concordância verbal na variedade urbana do Rio de Janeiro: um estudo sobre a face fonética do morfema verbal de P6. *Estudos Linguísticos*, São Paulo, v. 50, n. 3, p. 1207-1224, 2021. DOI: <https://doi.org/10.21165/el.v50i3.3035>.

RODRIGUES, E. S. *Processamento da concordância de número entre sujeito e verbo na produção de sentenças*. Tese (Doutorado em Letras) – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, 2006.

ROSA, M. C. *Introdução à (Bio)Linguística*. São Paulo: Contexto, 2020.

SCHERRE, M. M. P. Aspectos da concordância de número no português do Brasil. *Revista Internacional de Língua Portuguesa*. Lisboa, v. 11, n. 1, p. 37-49, 1994. Disponível em: <<http://www.ai.mit.edu/projects/dm/bp/scherre94-number.pdf>>. Acesso em: 14 mar. 2023.

SCHERRE, M. M. P. Paralelismo Linguístico. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 7, n. 2, p. 29-59, 1998. DOI: <http://dx.doi.org/10.17851/2237-2083.7.2.29-59>.

SCHERRE, M. M. P.; NARO, A. J. Sobre a concordância de número no português falado no Brasil. In: RUFFINO, G. (org.). *XXI Congresso Internazionale di Linguistica e Filologia Romanza*, v. 5: *dialettologia, geolinguistica, sociolinguistica*, 1998, Tübingen. *Atti...*, Centro di Studio Filologici e Linguistici Siciliani, Universidade de Palermo, 1998, p. 509-523. Disponível em: <<http://www.ai.mit.edu/projects/dm/bp/scherrenaro98.pdf>>. Acesso em: 15 mar. 2023.

SCHERRE, M. M. P. Phrase-level parallelism effect on noun phrase number agreement. *Language Variation and Change*, Cambridge, v. 13, p. 91-107, 2001. Disponível em: <<http://www.ai.mit.edu/projects/dm/bp/scherre.pdf>>. Acesso em 14 mar. 2023.

SCHERRE, M. M. P.; NARO, A. J. Sobre o deslocamento do controle da concordância verbal. *Revista Linguística*, Rio de Janeiro, v. 3, n. 1, 2007. DOI: <https://doi.org/10.31513/linguistica.2007.v3n1a4398>.

SCHERRE, M. M. P; NARO, A. J. Sociolinguistic correlates of negative evaluation: Variable concord in Rio de Janeiro. *Language Variation and Change*, Filadélfia, v. 19, n. 2, p. 181-190, 2014. Disponível em: <<https://repository.upenn.edu/pwpl/vol19/iss2/20>>. Acesso em: 14 mar. 2023.

SCHWARZ, F.; ZEHR, J. Tutorial: Introduction to PCIBex – An Open-Science Platform for Online Experiments: Design, Data-Collection and Code-Sharing. In: Annual Meeting of the Cognitive Science Society, v. 43, 2021, Viena. *Proceedings...* Viena, Universidade de Viena, 2021, p. 15-16. Disponível em: <<https://escholarship.org/uc/item/1ng1q4c6>>. Acesso em: 14 mar. 2023.

SIQUEIRA, M.; FREITAG, R. M. K. Can mobility affect grammar at the morphosyntactic level? A study in Brazilian Portuguese. *Organon*, Porto Alegre, v. 37, n. 73, p. 14–35, 2022. DOI: <https://doi.org/10.22456/2238-8915.122586>.

SMITH, J.; DURHAM, M.; RICHARDS, H. The social and linguistic in the acquisition of sociolinguistic norms: Caregivers, children, and variation. *Linguistics*, Glasgow, v. 51, n. 2, p. 285-324, 2013. DOI: <https://doi.org/10.1515/ling-2013-0012>.

SQUIRES, L. Processing, Evaluation, Knowledge: Testing the Perception of English Subject-Verb Agreement Variation. *Journal of English Linguistics*, v. 42, n. 2, p. 144-172, 2014. DOI: <https://doi.org/10.1177/0075424214526057>.

SQUIRES, L. Processing Grammatical Differences: Perceiving versus Noticing. In: BABEL, A. M. (ed.). *Awareness and Control in Sociolinguistic Research*. Cambridge: Cambridge University Press, 2016. p. 80-103. DOI: <https://doi.org/10.1017/CBO9781139680448>.

SOTO, M.; ALMEIDA, W. C. Entre a agramaticalidade e a variação: concordância verbal, sociolinguística e neurociência da linguagem. *ReVEL*, v. 19, n. 36, p. 1-29, 2021. Disponível em: <<http://www.revel.inf.br/files/beaf22e19c16fe7af902167722857468.pdf>>. Acesso em: 14 mar. 2023.

TAMMINGA, M.; MACKENZIE, L.; EMBICK, D. The dynamics of variation in individuals. *Linguistic Variation*, v. 16, n.2, p. 300-336, 2016. DOI: 10.1075/lv.16.2.06tam.

TRAXLER, M. J.; FOSS, D. J.; SEELY, R. E.; KAUP, B.; MORRIS, R. K. Priming in sentence processing: intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, v. 29, n. 6, p. 581-594, 2000. DOI: <https://doi.org/10.1023/A:1026416225168>.



THOMAS, E. R. Sociolinguistic variables and cognition. *Cognitive Science*, Nova Jersey, v. 2, n. 6, p. 701-716, 2011. DOI: <https://doi.org/10.1002/wcs.152>.

VASISHTH, S. Using Approximate Bayesian Computation for estimating parameters in the cue-based retrieval model of sentence processing. *MethodsX*, v. 7, p. 1-6, 2019. DOI: <https://doi.org/10.1016/j.mex.2020.100850>.

VIEIRA, S. R. Concordância verbal. In: VIEIRA, S. R.; BRANDÃO, S. F. *Ensino de gramática: descrição e uso*. São Paulo: Contexto, 2013. p. 85-102.

VIEIRA, S. R.; BAZENGA, A. Patterns of third person verbal agreement. *Journal of Portuguese Linguistics*, cidade, v.12, n. 2, p. 7-50, 2013. DOI: <https://doi.org/10.5334/jpl.67>.

VIEIRA, S. R.; BAZENGA, A. A concordância de terceira pessoa plural: padrões em variedades do Português. In: VIEIRA, S. R. (org.). *A concordância verbal em variedades do português: a interface fonética-morfossintaxe*. Rio de Janeiro: Vermelho Marinho, 2015. p. 29-75.

WEINREICH, U.; LABOV, W.; HERZOG, M. I. *Fundamentos empíricos para uma teoria da mudança linguística*. Tradução de Marcos Bagno. São Paulo: Parábola, 2006 [1968].

ZAHARCHUK, H. A.; SHEVLIN, A.; VAN HELL, J. G. Are our brains more prescriptive than our mouths? Experience with dialectal variation in syntax differentially impacts ERPs and behavior. *Brain and Language*, v. 218 p. 1-24, 2021. DOI: [10.1016/j.bandl.2021.104949](https://doi.org/10.1016/j.bandl.2021.104949).

ZEHR, J.; SCHWARZ, F. *PennController for Internet Based Experiments (IBEX)*, 2018. Disponível em: <https://osf.io/md832/>. DOI: <https://doi.org/10.17605/OSF.IO/MD832>. Acesso em: 11 nov. 2021.