

A frequência e o processamento dos clíticos de 2SG: uma análise experimental com rastreador ocular

Frequency and processing of 2SG clitics:
an experimental analysis with eye tracker

Thiago Laurentino de Oliveira¹

Resumo: Neste estudo, apresento os resultados de um experimento de leitura de frases com rastreador ocular, através do qual analisei o processamento dos clíticos acusativos de 2ª pessoa do singular. Os objetivos desta investigação eram observar como falantes do português brasileiro processam a referência ao interlocutor a partir dos clíticos *te*, *lhe* e *o/a* e verificar se a percepção desses pronomes é influenciada pelas diferenças existentes na frequência de uso. Como referencial teórico, adoto os pressupostos da Linguística Centrada no Uso (BYBEE, 2007; DIVJAK; CALDWELL-HARRIS, 2015). Os resultados sustentam a hipótese de correlação entre uso e cognição: o clítico *te*, que é o pronome mais frequentemente encontrado nos dados de *corpora*, foi a forma que demandou menor custo de processamento aos participantes durante a tarefa experimental.

Palavras-chave: Clíticos acusativos. 2ª pessoa do singular. Rastreamento ocular. Frequência de uso. Pronomes pessoais.

Abstract: In this paper, I present the results of a sentence reading experiment with an eye tracker, through which I analyzed the processing of 2nd person singular accusative clitics. The aims of this investigation were to observe how Brazilian Portuguese speakers process the reference to listener from pronouns *te*, *lhe* and *o/a* (you-accusative 2SG) and to verify if the perception of these pronouns is influenced by existing differences in frequency of use. As a theoretical framework, I adopt the assumptions of Usage-Based Linguistics (BYBEE, 2007; DIVJAK; CALDWELL-HARRIS, 2015). The results support the hypothesis of correlation between use and cognition: clitic *te*, which is the pronoun most frequently found in corpora data, was the form that demanded the lowest processing cost from participants during the experimental task.

Keywords: Accusative clitics. 2nd person singular. Eye tracking. Frequency of use. Personal pronouns.

¹ Universidade Federal do Rio de Janeiro, Faculdade de Letras, Departamento de Letras Vernáculas, Rio de Janeiro, RJ, Brasil. Endereço eletrônico: thiagolaurentino@letras.ufrj.br.

Introdução

Neste trabalho, apresento os resultados de um experimento de leitura de frases com rastreador ocular², através do qual busquei analisar o processamento dos clíticos pronominais de referência à 2ª pessoa do singular (doravante, 2SG) na função de objeto direto. Refiro-me, especificamente, às formas *te*, *lhe* e *o/a*, exemplificados nas frases em (01):

(01)

- a. A Maria disse que *te* encontrou no mercado.
- b. A Maria disse que *lhe* encontrou no mercado.
- c. A Maria disse que *o/a* encontrou no mercado.

A possibilidade de alternância entre esses clíticos no PB atual é resultante de um processo de convergência de paradigmas pronominais diferentes ao longo da história do português (CAMARA Jr., 1985; WILLIAMS, 1994). O clítico *te* constitui um resquício do sistema de flexão casual do latim, língua na qual correspondia ao pronome *tu* no caso acusativo e estabelecia referência à 2SG. Já os clíticos *o/a* e *lhe* derivam das formas acusativa e dativa do demonstrativo latino *ille* e se tornam, em português, pronomes de referência à 3ª pessoa do singular (3SG), relacionados ao pronome pessoal *ele/ela*. Com a gramaticalização e difusão de *você* (< *Vossa Mercê*) no sistema pronominal do português, esses clíticos passam a poder estabelecer também referência à 2SG (cf. LOPES *et al.*, 2018).

Os estudos sobre o tema realizados nos últimos anos apontam que as origens distintas desses clíticos influenciam os seus usos e, principalmente, a sua distribuição nos *corpora* analisados. Apesar da difusão de *você* como principal estratégia de 2SG na posição de sujeito, verifica-se a sobrevivência do clítico *te* nas posições de complemento verbal, registrando expressiva frequência de uso (cf. MACHADO, 2011; OLIVEIRA SILVA, 2011; SOUZA, 2014). Já os clíticos *lhe* e *o/a* apresentam frequências oscilantes nas amostras e seus usos estão geralmente correlacionados com a variedade dialetal dos informantes, a modalidade da língua e com o grau de formalidade da situação comunicativa (cf. CAMARGO Jr., 2007; ALMEIDA, 2009).

A partir desses resultados, majoritariamente baseados em análise de *corpora*, levanto dois problemas: como falantes do PB devem processar enunciados em que os clíticos *te*, *lhe* e *o/a* estejam presentes, atuando na referência à 2SG? As diferentes frequências de uso desses

² Os resultados do experimento apresentado neste artigo fazem parte da tese de Oliveira (2018).

pronomes influenciam o acesso à informação de 2SG? Diante desses questionamentos, proponho como objetivos deste trabalho (i) aprofundar as análises acerca dos clíticos de 2SG no PB, utilizando uma abordagem experimental, ainda pouco explorada nas investigações sobre o tema, e (ii) discutir sobre a relação entre gramática, uso e percepção linguística, destacando o modo como tal relação pode ser verificada no fenômeno em pauta.

Para a viabilização dos objetivos delineados, adotarei como aparato teórico a Linguística Centrada no Uso, com destaque para os trabalhos que tratam da frequência de uso e do processo de enraizamento (BYBEE, 2007; DIVJAK; CALDWELL-HARRIS, 2015). Desses pressupostos gerais, elaboro as seguintes hipóteses, a serem testadas no experimento: (i) os pronomes *te*, *lhe* e *o/a* não são processados da mesma maneira, sendo o clítico *te* a estratégia mais eficiente no acesso à informação de 2SG; (ii) a diferença no processamento é reflexo das frequências de uso distintas, pois as formas linguísticas mais frequentes são acessadas mais facilmente pelos falantes, com menor custo cognitivo envolvido. Sendo assim, assumo, nesta análise, o pressuposto segundo o qual as construções gramaticais armazenadas na mente dos usuários da língua são direta e constantemente influenciadas pela experiência linguística concreta dos falantes.

Estruturo este artigo em seis seções, incluindo a introdução. Na segunda seção, revisito os resultados de algumas pesquisas que analisaram os clíticos de 2SG em dados de *corpora* do PB, evidenciando as diferenças atestadas nas frequências de uso. Na terceira seção, apresento brevemente os postulados teóricos que advogam o impacto da frequência de uso nas representações linguísticas dos falantes. Descrevo o desenho do experimento de leitura de frases com rastreamento ocular elaborado para analisar o processamento dos clíticos de 2SG na quarta seção. Reporto, na penúltima seção, os principais resultados obtidos, analisando-os à luz dos pressupostos e hipóteses da pesquisa. Por fim, sublinho as contribuições mais relevantes da investigação, tecendo as palavras finais do artigo.

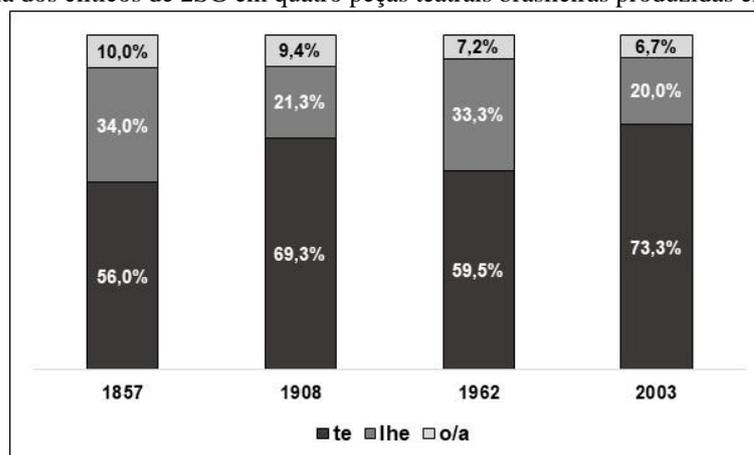
A frequência dos clíticos de 2SG em dados de *corpora*

As pesquisas que vêm sendo realizadas nos últimos anos acerca dos clíticos pronominais de 2SG no PB se inserem majoritariamente no quadro teórico da sociolinguística variacionista. Tais investigações exploram os clíticos *te*, *lhe* e *o/a* tanto de uma perspectiva diacrônica quanto de uma perspectiva sincrônica. As análises diacrônicas mapeiam as frequências de uso atestadas em *corpora* ao longo do tempo, observando como a difusão de *você* no sujeito repercute na complementação verbal de 2SG. Já os estudos sincrônicos, além de registrar a distribuição das variantes nas amostras, relacionam os usos dos pronomes a fatores linguísticos e sociais.

Menciono, nas próximas linhas, algumas dessas pesquisas, destacando os dados de frequência que elas registram acerca dos clíticos em discussão.

No plano diacrônico, destaco a tese de Machado (2011), que investigou o comportamento das formas de tratamento ao interlocutor em peças teatrais brasileiras e portuguesas escritas entre os séculos XIX e XX. O foco principal da autora era as mudanças pronominais ocorridas na posição de sujeito, mas ela também registra os usos verificados em outros contextos morfossintáticos. Das 29 peças analisadas, 14 foram escritas por brasileiros e eram ambientadas no Rio de Janeiro. Nas 14 obras brasileiras, Machado (2011) encontrou 1052 dados de formas pronominais de 2SG em posição de complemento verbal, sendo 752 ocorrências de clíticos de 2SG. Desse conjunto de dados, 43,6% (328 oco.) correspondiam ao clítico *te*, 39,5% (297 oco.), ao clítico *lhe* e 16,9% (127 oco.), ao clítico *o/a*.

Figura 1 - Frequência dos clíticos de 2SG em quatro peças teatrais brasileiras produzidas em sincronias distintas



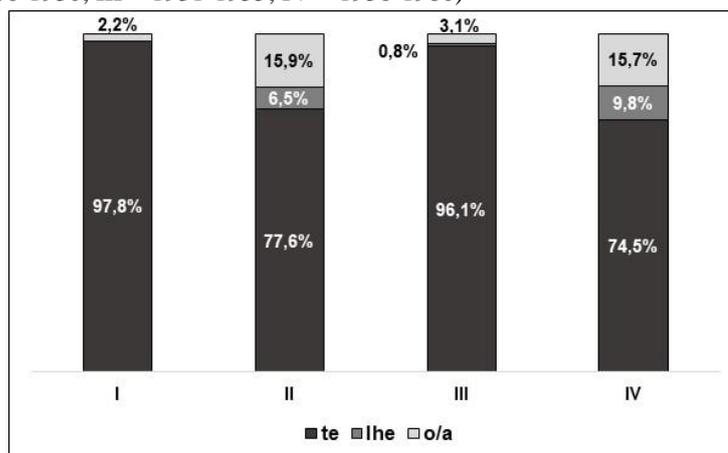
Fonte: elaborado pelo autor, com base em Machado (2011, p. 152)

A partir dos dados apresentados por Machado (2011, p. 152), elaborei a figura 1, na qual destaco as frequências dos clíticos analisados em peças teatrais escritas em momentos distintos da diacronia estudada pela autora: meados do século XIX (*Demônio familiar*, de José de Alencar), início do século XX (*Quebranto*, de Coelho Neto), meados do século XX (*Toda donzela tem um pai que é uma fera*, Gláucio Gill) e início do século XXI (*Síndromes*, de Maria Carmen Barbosa e Miguel Falabella). Como se pode notar, o clítico *te* prevalece, em todas as peças, sobre os demais, correspondendo a 56% dos dados de 1857 (107/191 oco.), 69,3% de 1908 (52/75 oco.), 59,5% de 1962 (25/42 oco.) e 73,3% das ocorrências de 2003 (22/30 oco.). Outro fato a ressaltar é a ordenação das frequências das variantes, visto que *te* sempre aparece em primeiro lugar, seguido por *lhe*; *o/a* é a forma clítica menos frequente nas quatro obras destacadas. Ao comentar os dados de complemento verbal, Machado observa que

O alçamento de *você* à posição de principal [...] estratégia pronominal de referência à segunda pessoa do discurso gera um aumento nos índices de usos dos pronomes oblíquos e possessivos de 3ª pessoa. Todavia, mesmo com o “desaparecimento” em algumas décadas do século XX, de *tu* na função de sujeito, observa-se a permanência, principalmente, do pronome oblíquo átono *te* [...]. Tal fato ilustra a conservação das formas de P2 e a possível associação destas a *você* para a designação de um mesmo interlocutor por um determinado falante. (MACHADO, 2011, p. 224)

Ainda na perspectiva diacrônica, cito a dissertação de Souza (2014), que mapeou as formas variantes de 2SG na posição de acusativo em cartas pessoais escritas por residentes na cidade do Rio de Janeiro, entre 1880 e 1980. A autora coletou 433 ocorrências de pronomes acusativos de 2SG, sendo 337 dados de *te*, 40 de *o/a* e 17 de *lhe* (4%). Essas ocorrências evidenciam a predominância de *te* também na escrita epistolar. Na figura 2, apresento a distribuição dos dados de clíticos na diacronia investigada por Souza (2014):

Figura 2 - Frequência dos clíticos de 2SG em cartas pessoais brasileiras de quatro recortes temporais diferentes (I = 1880-1905; II = 1906-1930; III = 1931-1955; IV = 1956-1980)

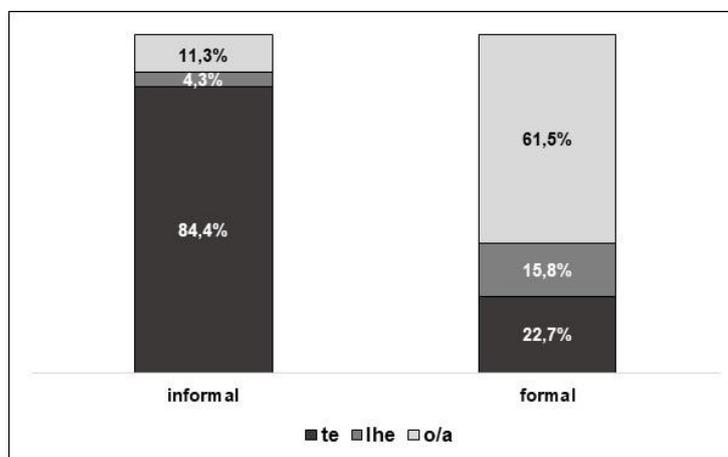


Fonte: elaborado pelo autor, a partir dos dados de Souza (2014)

Evidencia-se, mais uma vez, que a alta frequência de ocorrência da forma clítica *te* não se restringe a uma sincronia específica, visto que essa variante predomina em todos os recortes cronológicos analisados por Souza (2014). Cabe ressaltar, entretanto, que, nos dados das cartas pessoais, a frequência do clítico *lhe* foi muito baixa em comparação com *te* e mesmo com *o/a*: enquanto nas peças teatrais (figura 1) essa variante oscila entre 20% e 30%, nas cartas seus percentuais não ultrapassam 10%. De modo geral, vemos que as pesquisas diacrônicas revelam uma expressiva utilização de *te* em textos dos séculos XIX e XX frente aos outros clíticos, mesmo em um contexto de mudança, em que a forma nominativa *tu* torna-se de uso bastante restrito, dando lugar à forma *você*, principalmente na posição de sujeito.

Em relação às pesquisas sobre o tema no plano sincrônico, aponto a dissertação de Camargo Jr. (2007). O autor analisou as formas pronominais acusativas de 2SG em produções textuais de alunos do segundo segmento do ensino fundamental de um colégio da rede privada da cidade de São Paulo. Para estimular a utilização de pronomes de 2SG, Camargo Jr. (2007) pediu aos alunos que escrevessem uma carta em duas modalidades discursivas – formal e informal. O pesquisador obteve 1.524 ocorrências de objeto direto de 2SG, sendo 595 dados da modalidade formal (39%) e 929 da modalidade informal (61%). Os clíticos foram majoritários nas produções informais (83,2%).

Figura 3 - Frequência dos clíticos de 2SG em redações de estudantes paulistanos do ensino fundamental durante tarefa escolar



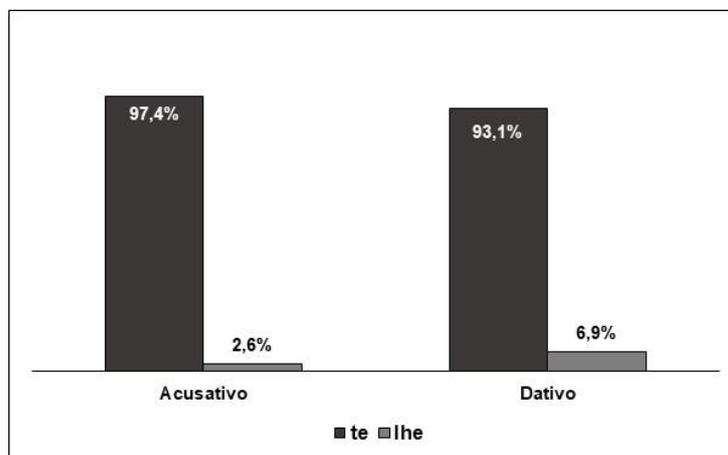
Fonte: elaborado pelo autor, com base em Camargo Jr. (2007, p. 50; 65)

Reuni na figura 3 as frequências dos dados de formas clíticas de 2SG que Camargo Jr. (2007) encontrou nas redações dos alunos, dividindo-as segundo a variável modalidade discursiva, controlada pelo autor. Percebe-se que, embora os três clíticos tenham ocorrido nos textos das duas modalidades, suas frequências se diferenciam consideravelmente. Nos textos informais, prevaleceu o uso de *te*, enquanto nos textos formais *o/a* foi majoritário. A forma *lhe* foi a variante menos utilizada nas duas modalidades. Os resultados sugerem uma influência significativa do tipo de registro sobre o uso das formas pronominais, sobretudo no que se refere ao clítico *o/a*. É digno de nota, contudo, o registro do clítico *te* na modalidade formal: correspondendo a 22,7% dos dados (56/247 oco.), essa frequência ilustra a “força” de uso do pronome, presente em uma tarefa de produção escrita escolar.

Em sua dissertação, Oliveira Silva (2011) também analisou sincronicamente as formas clíticas de 2SG. Baseada em roteiros de cinema cujas histórias se passavam nas cidades do Rio de Janeiro, São Paulo e Porto Alegre, a autora encontrou um total de 445 dados de formas pronominais com referência ao interlocutor, sendo 186 ocorrências na posição de acusativo e

259 na posição de dativo. Os pronomes clíticos correspondem a 386 dados, representando, pois, a maior parcela do conjunto.

Figura 4 - Frequência dos clíticos de 2SG em roteiros de cinema segundo o contexto sintático de complementação



Fonte: elaborado pelo autor, com base em Oliveira Silva (2011, p. 27)

Na figura 4, insiro apenas as ocorrências de clíticos de 2SG levantadas por Oliveira Silva (2011). A forma *te* também foi a estratégia predominante nessa amostra: como ilustram as frequências, a variante registrou índices acima dos 90% em ambas as posições sintáticas. O clítico *lhe* foi bem menos frequente, correspondendo a 2,6% na posição acusativa e 6,9% na posição dativa. A ausência do clítico *o/a* na figura 4 deve-se ao fato de Oliveira Silva (2011) não ter contabilizado nenhum dado dessa variante nos roteiros analisados.

Em suma, procurei demonstrar brevemente, nesta seção, os principais resultados de algumas pesquisas sociolinguísticas que investigaram os clíticos *te*, *lhe* e *o/a* em dados do PB. Todas essas pesquisas evidenciam, de algum modo, um mesmo fato: a constante e expressiva frequência do clítico *te*, seja nas amostras diacrônicas, seja nas amostras sincrônicas. Diante disso, levanto a seguinte questão: quais são as consequências da alta frequência de uso do clítico *te* na posição de complemento verbal para a representação pronominal de 2SG?

Pressupostos teóricos: a frequência de uso na cognição linguística

Como arcabouço teórico deste trabalho, adoto os pressupostos da Linguística Centrada no Uso. Segundo Martelotta (2011), a Linguística Centrada no Uso (Trad. ing. ‘Usage-Based Linguistics’) é “um tipo de abordagem que [...] considera haver uma relação estreita entre a estrutura das línguas e o uso que os falantes fazem delas nos contextos reais de comunicação” (p. 55-56). Na análise da relação entre estrutura e uso, os estudiosos dessa abordagem examinam, também, o papel da cognição humana nesse processo. Segundo o mesmo autor,

Nessa perspectiva, são levados em conta, na análise das línguas, aspectos relacionados a restrições cognitivas que incluem a captação de dados da experiência, sua compreensão e seu armazenamento na memória, assim como aspectos associados à capacidade de organização, acesso, conexão, utilização e transmissão adequada desses dados. (MARTELOTTA, 2011, p. 56)

Dentro da referida abordagem, uma peça-chave nas pesquisas centradas no uso é a aferição da frequência das construções linguísticas. Há décadas, diferentes linguistas (sobretudo aqueles filiados à perspectiva funcionalista e cognitivista) têm se dedicado a descrever e analisar a relação entre frequência de uso e estruturas linguísticas, tentando compreender como as diferenças de frequência moldam/afetam o conhecimento gramatical dos usuários da língua. De acordo com Divjak e Caldwell-Harris,

Na psicolinguística e na linguística cognitiva, frequência geralmente se refere ao número de vezes que uma unidade específica da língua (como um fonema, palavra ou sintagma) ocorre em um ambiente especificado. A frequência é normalmente usada em um sentido relativo, para categorizar alguns estímulos como sendo mais ou menos predominantes no ambiente do que outros estímulos. (DIVJAK; CALDWELL-HARRIS, 2015, p. 54, *tradução minha*)³

No que se refere à razão pela qual as formas linguísticas apresentam distribuições diferenciadas nos textos, ou seja, por que algumas palavras/construções são mais frequentemente encontradas do que outras, cito Bybee (2007). A linguista, que é uma das principais referências sobre o tema, afirma que os itens linguísticos se tornam frequentes por diferentes motivos, sendo alguns deles os seguintes: (i) a questão sobre o que as pessoas querem falar, ou seja, o conteúdo da mensagem; (ii) o modo como os falantes estruturam o discurso deles, já que os padrões discursivos são convencionalizados na gramática; (iii) a generalidade/flexibilidade do significado veiculado, pois itens com significados mais gerais e abstratos podem ocorrer em mais contextos discursivos diferentes. Nesse sentido, a autora sintetiza que

[...] a resposta para a questão de se a frequência é uma causa ou um efeito é complexa. Por um lado, a frequência é apenas uma contagem, um padrão observável nos textos, o que é certamente um efeito. Por outro lado, a frequência ou repetição de experiências tem um impacto nas representações

³ Na versão original, em inglês: “Within psycholinguistics and cognitive linguistics, frequency most often refers to the number of times a particular chunk of language (such as a phoneme, word, or phrase) occurs in a specified environment. Frequency is typically used in a relative sense, to categorize some stimuli as being more or less prevalent in the environment than other stimuli”.

cognitivas e, dessa maneira, torna-se uma causa [...]”⁴. (BYBEE, 2007, p. 18, *tradução minha*)

Sobre esse tópico, Divjak e Caldwell-Harris (2015) também tecem algumas breves considerações, destacando que a frequência de ocorrência das palavras é condicionada por um conjunto de características específicas. Segundo elas, palavras de alta frequência possuem um comprimento mais curto, exprimem significação concreta em vez de abstrata, são mais facilmente imagináveis e seriam adquiridas em uma fase mais precoce da aquisição. As autoras mencionam ainda evidências experimentais de que parte do efeito de frequência é causado pela experiência real dos falantes.

Ao examinar os efeitos da frequência sobre a língua, Bybee e outros pesquisadores costumam distinguir dois grandes modos de contabilizar a frequência de uso dos itens linguísticos. Esses dois modos são rotulados pela autora como frequência de ocorrência (*token frequency*) e frequência de tipo (*type frequency*), descritos por ela da seguinte maneira:

A frequência de ocorrência conta o número de vezes que uma unidade aparece no texto em questão. Qualquer unidade específica, como uma consoante particular [s], uma sílaba [ba], uma palavra *cachorro* ou *o*, uma locução *dar um tempo* ou até mesmo uma frase como *Sua torrada apareceu* pode ter uma frequência de ocorrência. A frequência de tipo é uma espécie de contagem muito diferente. Somente padrões da língua têm frequência de tipo, porque isso se refere a quantos itens distintos são representados pelo padrão. (BYBEE, 2007, p. 9, *tradução minha*)⁵

Mais do que uma mera divisão metodológica, esses dois modos distintos de contagem se revelam, em vários estudos, como importantes ferramentas analíticas, pois evidenciam os diferentes efeitos da frequência na cognição. Tendo em vista os propósitos deste artigo, focalizarei apenas os efeitos relacionados à frequência de ocorrência, descrita e analisada nos estudos sociolinguísticos sobre os clíticos de 2SG. Bybee (2007) destaca três principais efeitos decorrentes da alta frequência de ocorrência: o Efeito de Conservação (*Conserving Effect*), o Efeito de Redução (*Reducing Effect*) e a Autonomia (*Autonomy*). Assumindo que o Efeito de Conservação e a Autonomia são mais salientes na expressão pronominal de 2SG, comento, a seguir, esses efeitos em termos conceituais.

⁴ Em inglês: “the answer to the question of whether frequency is a cause or an effect is complex. On the one hand, frequency is just a tally, a pattern observable in texts, which is of course an effect. On the other hand, frequency or repetition of experiences has an impact on cognitive representations and in this way becomes a cause”.

⁵ Em inglês: “Token frequency counts the number of times a unit appears in running text. Any specific unit, such as particular consonant [s], a syllable [ba], a word *dog* or *the*, a phrase *take a break*, or even a sentence such as *Your toast popped up* can have a token frequency. Type frequency is a very different sort of count. Only patterns of language have type frequency, because this refers to how many distinct items are represented by the pattern”.

O Efeito de Conservação está intimamente relacionado à dimensão cognitiva das formas linguísticas. Segundo Bybee (2007), o uso continuamente repetido de itens e construções fortalece as representações mnemônicas deles, tornando-os mais acessíveis. Como consequência dessa representação constantemente reforçada, verifica-se que as formas de alta frequência de uso são mais resistentes a processos de mudança linguística, sendo conservadas na gramática. Além disso, o Efeito de Conservação interage com o processo cognitivo que Langacker (1987) denominou de enraizamento (*entrenchment*):

Todo uso de uma estrutura tem um impacto positivo sobre o seu grau de enraizamento, enquanto períodos prolongados de desuso têm um impacto negativo. Com o uso repetido, uma nova estrutura se torna progressivamente enraizada, a ponto de se tornar uma unidade; além disso, as unidades são enraizadas de maneira variável, dependendo da frequência de sua ocorrência.” (LANGACKER, 1987, p. 59 *apud* DIVJAK; CALDWELL-HARRIS, 2015, p. 60, *tradução minha*)⁶

Analisando a caracterização de enraizamento apresentada por Langacker, Divjak e Caldwell-Harris (2015) ressaltam que a visão do autor sublinha o papel que tal processo cognitivo desempenha na representação, relacionando-o ao armazenamento e organização de estruturas no inventário mental dos falantes. Em outras palavras, afirma-se, de um lado, que o aumento da frequência de ocorrência aprofunda o enraizamento das formas linguísticas; de outro, que o aumento do enraizamento pode gerar diferenças qualitativas na representação, a ponto de gerar *unicização*, isto é, a criação de unidades na memória. Por essa razão, tanto o Efeito de Conservação quanto o enraizamento têm implicações diretas para a percepção dos falantes: devido à representação fortalecida na memória, desencadeada pela repetição de uso, o acesso a formas de alta frequência demanda um esforço cognitivo menor, que se reflete no processamento mais rápido, fácil e preciso das estruturas mais frequentes.

Quanto à Autonomia, Bybee (2007, p. 13) comenta que ela seria o terceiro efeito da frequência de ocorrência, embora também possa ser considerada um caso extremo do Efeito de Conservação. A lógica subjacente à questão da Autonomia é que “quando palavras (ou locuções) são altamente frequentes, elas podem ser acessadas independentemente dos itens relacionados e, portanto, não estão tão interconectadas na rede” (BYBEE, 2007, p. 13-14). Tomando por base exemplos da morfologia flexional, a autora resalta que tal efeito, além de levar à resistência à mudança, permite ainda que formas de alta frequência sirvam de base para

⁶ “Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence”.

reformulações dentro de um paradigma, levando, em casos extremos, à supleção. A título de ilustração, ela menciona o caso da forma de passado *went*, que se desvinculou do verbo *wend* e se tornou o passado do verbo *go*.

Relacionando os pressupostos brevemente pontuados nesta seção com os resultados das pesquisas empíricas resenhadas na seção anterior, podemos afirmar, por hipótese, que a alta frequência de uso do clítico *te*, atestada nos dados de *corpora*, desencadeia o Efeito de Conservação (e mesmo a Autonomia) desse pronome na gramática do PB. Sendo assim, defendo que a representação dessa forma linguística estaria mais arraigada na mente dos falantes, fato que a torna mais acessível na memória e, conseqüentemente, menos custosa de ser processada, em comparação com os clíticos *lhe* e *o/a*.

A fim de pôr à prova essas hipóteses, elaborei e apliquei um experimento de leitura de frases com rastreamento ocular, cuja metodologia descrevo na próxima seção.

Metodologia experimental: rastreando evidências

As pesquisas experimentais que adotam a técnica do rastreamento ocular (*eye-tracking*) partem do pressuposto de que “os movimentos dos olhos reflectem eventuais dificuldades sentidas pelo leitor durante o processamento da informação linguística” (LUEGI; COSTA; FARIA, 2009, p. 63). Dentro dessa lógica, acredita-se que “ao direcionar o olhar para um determinado ponto, o sujeito direciona também seus recursos cognitivos a ele com vistas a processar informações nele contidas ou a ele relacionadas. (FORSTER, 2017, p. 623). Sendo assim, optei por construir um experimento com rastreador ocular, com o intuito de verificar se o tipo de clítico de 2SG afeta significativamente o processamento de frases durante a leitura.

Design experimental

O experimento possui um *design* do tipo 3, em que foi manipulado como variável independente o *tipo de clítico*, com três níveis: *te*, *lhe* ou *o/a*. Cada nível da variável independente correspondia, portanto, a uma condição experimental. As variáveis dependentes foram os tempos e o número de fixações oculares (dos clíticos, do conjunto ‘clítico-verbo’ e das frases inteiras) e os índices de respostas às perguntas interpretativas. No quadro 1, reproduzo um conjunto de frases experimentais que integravam o experimento, junto com a questão interpretativa a elas associada.

Quadro 1 – Exemplo de frases experimentais utilizadas na tarefa

Frases experimentais:

Clítico te

Paulo garantiu para Aline na frente de Ana: eu já *te* indiquei ao cargo.

Questão interpretativa: Paulo já indicou Aline ao cargo? SIM NÃO

Clítico lhe

Paulo garantiu para Aline na frente de Ana: eu já *lhe* indiquei ao cargo.

Questão interpretativa: Paulo já indicou Aline ao cargo? SIM NÃO

Clítico o/a

Paulo garantiu para Aline na frente de Ana: eu *a* indiquei para o cargo⁷.

Questão interpretativa: Paulo indicou Aline para o cargo? SIM NÃO

Fonte: elaborado pelo autor.

Materiais

Elaborei 30 enunciados experimentais, sendo 10 estímulos para cada condição. Organizei esses estímulos em três conjuntos, formados por 10 itens experimentais (5 enunciados x 2 clíticos diferentes) e 20 itens distratores. Os participantes foram divididos igualmente entre os conjuntos. Cada conjunto apresentava um par diferente de pronomes: conjunto 1 – *te* e *lhe*; conjunto 2 – *lhe* e *o/a*; conjunto 3 – *te* e *o/a*. A configuração *within-subjects* possibilitou que todos os participantes fossem expostos às mesmas frases, em que a principal diferença era o tipo de clítico.

As frases experimentais eram compostas por duas orações (cf. Quadro 1). A primeira, uma oração matriz, continha três nomes próprios e um verbo comunicativo – “Paulo”, “Aline” e “Ana”. Além de introduzir referentes com traço [+humano], essa sentença criava uma situação de interação dialógica, necessária para o estudo dos clíticos de 2SG. A segunda sentença era uma oração completiva, em que apareciam dois pronomes pessoais: “eu”, que, na situação, se referia ao sintagma nominal sujeito da primeira sentença (no Quadro 1, “Paulo”), e um dos clíticos analisados (*te*, *lhe* ou *o/a*), relacionado ao sintagma preposicionado da primeira sentença (“para Aline”), destinatário da mensagem.

Os itens experimentais e distratores foram inseridos em *slides* de apresentação do aplicativo *PowerPoint*, com fundo branco e tamanho 25,4cm x 19,05cm. Em seguida, eles foram salvos um a um na extensão pdf e importados pelo programa *TOBII Studio 2.3.2*. Os estímulos eram exibidos de modo semialeatório (nenhum participante iniciava a tarefa visualizando um enunciado experimental e nem via, consecutivamente, dois enunciados desse tipo).

⁷ As frases experimentais com o clítico *o/a* apresentam alguns ajustes para compensar a diferença gráfica quanto ao número de caracteres que existe entre os pronomes analisados.

Participantes

Participaram do experimento 30 alunos da UFRJ, 5 deles do sexo masculino, nativos do estado do Rio de Janeiro e com média de idade de 23 anos. Todos possuíam ensino superior (completo ou em andamento) e eram, majoritariamente (28), dos cursos de Letras, à exceção de um participante de Odontologia e outro de Artes Visuais. Todos os alunos participaram voluntariamente, tendo lido e assinado previamente o termo de consentimento livre e esclarecido. Posteriormente, eles receberam certificado de participação por e-mail.

Procedimento

Apliquei o experimento no Laboratório de Psicolinguística Experimental (LAPEX)⁸ da Faculdade de Letras da UFRJ. Antes de iniciar o experimento de fato, os participantes calibravam o olhar ao rastreador, recebiam as instruções de realização da tarefa e cumpriam um pequeno treino. Os voluntários liam as sentenças que apareciam automaticamente na tela do computador, gastando o tempo necessário para a compreensão das frases. Em seguida, eles deviam responder à pergunta interpretativa, relacionada com a frase lida. A pergunta permanecia em tela por 5 segundos, desaparecendo em seguida. Para responder, os participantes fixavam o olhar em uma das opções de resposta, “sim e “não”, situadas abaixo da pergunta. Cada sujeito levou, em média, 7 minutos para concluir a tarefa.

As frases eram exibidas em um monitor de 23”, acoplado ao aparelho de rastreamento ocular *TOBII Studio 2.3.2*. Os estímulos apareciam na tela em cor preta, com fundo branco escritos com fonte *Courier New*, tamanho 14, centralizado, ocupando uma única linha. Todas as frases (experimentais e distratores) tinham a mesma extensão: 24 sílabas na primeira frase e 11 sílabas na pergunta interpretativa.

Hipóteses e previsões

Assumo, neste trabalho, a hipótese geral da Linguística Centrada no Uso, segundo a qual as estruturas linguísticas armazenadas na mente dos falantes são continuamente influenciadas, modificadas ou reforçadas pela experiência concreta destes com a língua. Como hipótese específica da pesquisa, proponho que as diferentes frequências de uso dos clíticos de 2SG afetam distintamente o processamento desses pronomes, haja vista os diferentes níveis de enraizamento dessas formas na memória dos falantes.

⁸ Agradeço imensamente ao Professor Marcus Maia, coordenador do LAPEX, por viabilizar e supervisionar a aplicação do experimento.

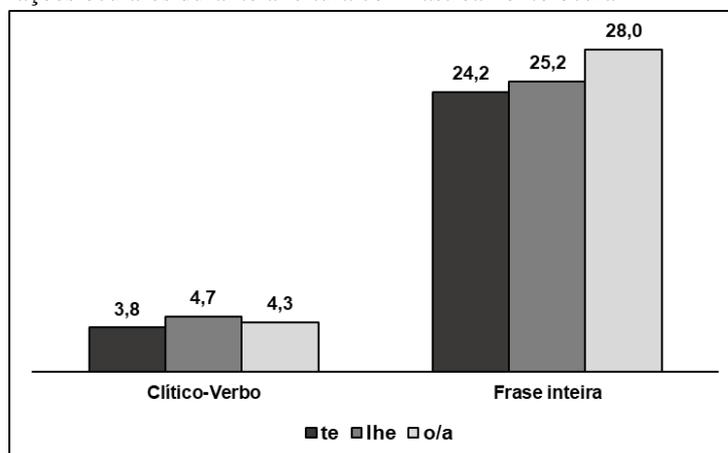
Dessas hipóteses, derivam as seguintes previsões acerca do comportamento dos participantes no experimento: (i) as frases com *te* serão fixadas em menor número e por menos tempo, além de ativarem com mais eficácia a informação de 2SG; (ii) as frases com *lhe* serão fixadas em maior número e por mais tempo (em comparação com as frases contendo *te*), embora também ativem a informação de 2SG eficientemente; (iii) as frases com o clítico *o/a* irão registrar o maior número e os maiores tempos de fixação do olhar, além de dificultarem a ativação da informação de 2SG.

Resultados

Reporto, a seguir, os seguintes resultados: (i) o número de fixações oculares no conjunto ‘clítico-verbo’ e na frase inteira, apresentado em média de ocorrências; (ii) os tempos de leitura do conjunto ‘clítico-verbo’ e da frase inteira, mensurados em milésimos de segundo; (iii) os índices de interpretação da informação de 2SG, extraídos das respostas às questões interpretativas.

A figura 5 exhibe os resultados da média de fixações do olhar capturadas na área em que aparecia o conjunto ‘clítico-verbo’ e na frase inteira. Nas barras à esquerda, vemos que o conjunto ‘te+verbo’ foi o que computou o menor número de fixações, com média de 3,8. Já o conjunto ‘o/a+verbo’ registrou uma média de 4,3 fixações. O conjunto ‘lhe+verbo’ apresenta o número mais alto de fixações, com média de 4,7. Ao considerar as médias de leitura da frase inteira, nas barras à direita, percebemos que as frases com o clítico *o/a* registraram o maior valor médio de fixações (28,0); por outro lado, as frases com *te* contabilizaram a menor média de fixações (24,2), seguidas das frases com *lhe* (média de 25,2 fixações).

Figura 5 - Média de fixações oculares durante a leitura com rastreamento ocular

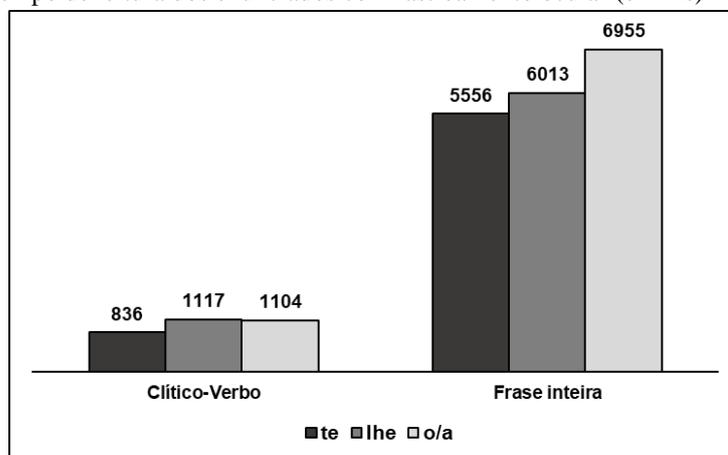


Fonte: elaborado pelo autor.

Para verificar a significância estatística das diferenças observadas entre as médias, realizou-se uma análise de variância (ANOVA) por sujeito, com base no *tipo de clítico*. A ANOVA apontou efeito principal do tipo de clítico, tanto para as diferenças nas médias do conjunto ‘clítico-verbo’ ($F(2,198) = 4,01$ $p < 0,05$), quanto nas médias da frase inteira ($F(2,198) = 3,11$ $p < 0,05$). Em relação ao conjunto ‘clítico-verbo’, houve diferença significativa entre as médias *te+verbo* e *lhe+verbo* ($t(99) = 3,04$ $p < 0,01$); nas comparações entre *o/a+verbo* e *te+verbo* ($t(99) = 1,53$ $p > 0,05$) e *o/a+verbo* e *lhe+verbo* ($t(99) = 1,25$ $p > 0,05$), as diferenças não foram significativas. Quanto às médias da frase inteira, houve diferença significativa entre ‘frases com *te*’ e ‘frases com *o/a*’ ($t(99) = 2,50$ $p < 0,05$), ao passo que as diferenças nas médias dos pares ‘frases com *te*’ e ‘frases com *lhe*’ ($t(99) = 0,59$ $p > 0,05$) e ‘frases com *lhe*’ e ‘frases com *o/a*’ ($t(99) = 1,73$ $p > 0,05$) não foram estatisticamente significativas.

A média de tempo de leitura, outra variável dependente controlada com o auxílio do rastreador ocular, aparece na figura 6. Analisando as barras à esquerda da figura, nota-se que as fixações no conjunto ‘*te+verbo*’ registraram o menor tempo médio de leitura (836ms). Quanto aos conjuntos ‘*lhe+verbo*’ e ‘*o/a+verbo*’, percebem-se tempos médios de leitura mais elevados (1117ms e 1104ms, respectivamente). Nas barras à direita da figura 6, vemos que o tempo médio de leitura das frases com *te* foi de 5556ms, o menor tempo em relação às frases com *lhe* (6013ms) e com *o/a* (6955 ms). Essas diferenças demonstram que os participantes fixaram o olhar por menos tempo quando encontravam no conjunto ‘clítico-verbo’ o pronome *te*, sugerindo que esse pronome era acessado mais rapidamente durante o processamento da leitura. Além disso, a presença de *te* nas frases parece ter facilitado também o processamento integral da frase, demandando menor esforço cognitivo.

Figura 6 – Média de tempo de leitura dos enunciados com rastreamento ocular (em ms)

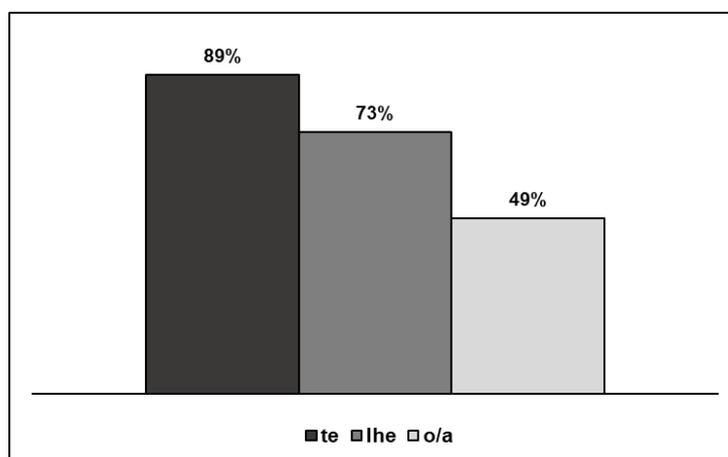


Fonte: elaborado pelo autor.

Realizando a ANOVA, verificou-se que houve efeito principal do tipo de clítico sobre os tempos de leitura do conjunto ‘clítico+verbo’ ($F(2,198) = 7,68 p < 0,001$) e da frase inteira ($F(2,198) = 6,77 p < 0,01$). A diferença entre as médias de tempo de leitura de ‘te+verbo’ e os demais conjuntos foi significativa (te-lhe: $t(99) = 3,90 p < 0,001$; te-o/a: $t(99) = 3,41 p < 0,001$); porém, o mesmo não ocorreu entre as médias de ‘lhe+verbo’ e ‘o/a+verbo’ ($t(99) = 0,14 p > 0,05$). Na análise estatística das médias de leitura da frase inteira, foram significativas as diferenças envolvendo as frases com o/a (te-o/a: $t(99) = 3,80 p < 0,001$; lhe-o/a: $t(99) = 2,25 p < 0,05$), mas não houve significância na diferença entre os tempos de leitura das frases com *te* e com *lhe* ($t(99) = 1,22 p > 0,05$).

A figura 7 apresenta o resultado do índice de interpretação da informação de 2SG, obtido a partir da resposta dada pelos participantes às questões interpretativas que apareciam após as frases com os clíticos. Verificamos que a condição envolvendo o clítico *te* foi a que mais conduziu os participantes à interpretação da informação de 2SG, (89% das respostas). Em segundo lugar, aparece a condição com o clítico *lhe*, cujo percentual de interpretação da informação de 2SG foi de 73%. A condição com o clítico *o/a* foi a que menos propiciou a interpretação de 2SG, com índice de 49%.

Figura 7 - Índice de interpretação da informação de 2SG



Fonte: elaborado pelo autor.

Para avaliar estatisticamente se as diferenças entre as condições eram significativas, utilizou-se o teste de qui-quadrado. Os resultados do teste sinalizaram que todas as diferenças entre os índices são significativas (te-lhe: $\chi^2 = 8,31, p < 0,01$; te-o/a: $\chi^2 = 37,40, p < 0,001$; lhe-o/a: $\chi^2 = 12,10, p < 0,001$).

Discussão

Os resultados encontrados a partir do estudo com rastreador ocular vão na direção das hipóteses levantadas para este trabalho. De modo geral, o comportamento distinto dos participantes durante a tarefa evidencia a existência de diferenças na percepção dos clíticos *te*, *lhe* e *o/a* em termos do processamento da informação de 2SG. Tais diferenças parecem estar diretamente relacionadas com a frequência de uso desses pronomes, como pretendo argumentar nesta breve discussão dos dados experimentais.

As frases do teste que continham *te*, o clítico apontado como o mais frequente por vários estudos baseados em *corpora*, foram as que exigiram dos participantes menor esforço cognitivo, verificado tanto pelas médias de fixação do olhar (3,8 na região do clítico e 24,2 na frase inteira) quanto de tempo de leitura (836ms na região do clítico e 5556ms na frase inteira). Interessantemente, o maior índice de interpretação da informação da 2SG também foi registrado nas frases com *te* (89%). Esses dados salientam a eficiência do clítico *te* para acessar a informação de 2SG na memória dos falantes.

Por outro lado, nota-se que os participantes apresentaram uma performance bastante diferente diante dos clíticos *lhe* e *o/a*. Em relação às frases com o clítico *lhe*, as médias de fixação e de tempo de leitura da frase inteira não foram significativamente mais altas do que nas frases com *te*. No entanto, houve diferenças estatisticamente relevantes na análise do conjunto ‘clítico-verbo’ (média de 4,7 fixações oculares e de 1117ms de tempo de leitura), o que significa que a presença do clítico *lhe* nas frases exigiram maior esforço cognitivo dos participantes para acessar esse pronome. Além disso, o índice de interpretação da informação da 2SG (73%) nas frases com *lhe*, embora tenha sido alto, também foi significativamente inferior ao mesmo índice obtido nas frases com *te*.

As frases com o pronome *o/a* registraram, em média, 28 fixações oculares (significativamente maior do que nas frases com *te*) e 6955ms de tempo de leitura (significativamente maior do que nas frases com *te* e *lhe*). A área onde aparecia esse clítico também demandou maior tempo médio de leitura (1104ms), comparativamente com o clítico *te*. Soma-se a isso o fato de que o índice de interpretação da informação de 2SG nas frases com *o/a* foi de 49%, o mais baixo dentre as três condições controladas no estudo.

Desse modo, as evidências experimentais parecem sustentar a hipótese de que a alta frequência de uso do clítico *te* desencadeou o Efeito de Conservação desse pronome, nos termos de Bybee (2007), no sistema pronominal do PB. Logo, a representação dessa forma pronominal está mais enraizada na mente dos falantes e, conseqüentemente, é mais facilmente acessada na

memória e envolve menos custo de processamento, como ilustram os dados obtidos durante a leitura de frases com rastreamento ocular.

Palavras finais

Este trabalho traz contribuições relevantes para a área de estudos sobre o sistema pronominal do português brasileiro, visto que focaliza os clíticos de 2SG (pouco estudados, em comparação com os clíticos de 3SG) utilizando a metodologia experimental (pouco aplicada nas análises dos pronomes com referência ao interlocutor). Do ponto de vista teórico, os resultados oferecem evidências favoráveis ao pressuposto da Linguística Centrada no Uso, segundo o qual a experiência concreta dos falantes com a linguagem influencia direta e continuamente a representação das unidades linguísticas armazenadas na memória. Desse modo, as análises dos clíticos sugerem uma estreita correlação entre uso e percepção, sobretudo no que tange à forma *te*: além de ser a estratégia amplamente utilizada pelos falantes do PB, o clítico *te* revela-se como o mais eficiente no processamento da informação de 2SG.

A fim de enfatizar a importância da correlação entre uso e cognição, encerro citando Bybee (2007). Ao destacar o papel da frequência para o Efeito de Conservação, a autora afirma que “palavras que são fortes na memória e fáceis de acessar provavelmente não serão substituídas por novas formas criadas com o padrão regular”⁹ (BYBEE, 2007, p. 271). Dito de outro modo, a linguista assinala que “exemplares de construções altamente enraizados com itens lexicais particulares podem continuar a ser usados, mesmo que novos padrões produtivos tenham se tornado correntes na língua”¹⁰ (BYBEE, 2007, p. 277).

Aplicando essa explicação ao sistema pronominal de 2SG, em que a forma *você* parece se impor sobre *tu*, pode-se propor que, apesar da difusão de *você* no PB, demonstrada por inúmeras pesquisas, *te* continuará a ser a principal forma de 2SG como complemento. A razão disso seria a elevada frequência do clítico nas construções verbo-pronominais, que gera o seu enraizamento na representação mental, conservando-o mesmo após a emergência de *você*.

Referências

ALMEIDA, G. de S. **Quem te viu quem lhe vê**: a expressão do objeto acusativo de referência à segunda pessoa na fala de Salvador. 2009. 193 f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Letras e Linguística, Universidade Federal da Bahia, Salvador, 2009. Disponível em: <http://repositorio.ufba.br/ri/handle/ri/10995>.

⁹ “Words that are strong in memory and easy to access are not likely to be replaced by new forms created with the regular pattern”.

¹⁰ “highly entrenched examples of constructions with particular lexical items can continue to be used even though new productive patterns have become current in the language”.

BYBEE, J. **Frequency of use and the organization of language**. Oxford: Oxford University Press, 2007.

CAMARA Jr., J. M. **História e Estrutura da Língua Portuguesa**. Rio de Janeiro: Padrão, 1985.

CAMARGO Jr., A. R. **A realização do objeto direto em referência ao interlocutor**. 2007. 98 f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Filologia e Língua Portuguesa, Universidade de São Paulo, São Paulo, 2007. Disponível em: <https://teses.usp.br/teses/disponiveis/8/8142/tde-03032008-114747/pt-br.php>.

DIVJAK, D.; CALDWELL-HARRIS, C. L. Frequency and entrenchment. *In*: DABROWSKA, E; DIVJAK, D. (Eds.). **Handbook of Cognitive Linguistics**. Berlim/Boston: De Gruyter Mouton, 2015. p. 53-75.

FORSTER, R. Aspectos da utilização do rastreamento ocular na pesquisa psicolinguística. **DELTA**, São Paulo, v. 33, n. 2, p. 609-644, jun. 2017. Disponível em: <https://doi.org/10.1590/0102-445095461720767529>.

LOPES, C. R. dos S. *et al.* A reorganização no sistema pronominal de 2a. pessoa na história do português brasileiro: outras relações gramaticais. *In*: LOPES, C. R. dos S. (Coord.). **Mudança sintática das classes de palavra: perspectiva funcionalista; história do português brasileiro**. São Paulo: Contexto, 2018. p. 142-185.

LUEGI, P.; COSTA, A.; FARIA, I. H. Analisando os comportamentos oculares durante a leitura. **Linguística**, v. 5, n. 1, p. 62-80, 2009. Disponível em: <https://revistas.ufrj.br/index.php/rl/article/view/4424/3196>.

MACHADO, A. C. M. **As formas de tratamento no teatro brasileiro e português dos séculos XIX e XX**. 2011. 217 f. Tese (Doutorado em Letras Vernáculas) – Programa de Pós-Graduação em Letras Vernáculas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

MARTELOTTA, M. E. **Mudança Linguística**. Uma abordagem baseada no uso. São Paulo: Cortez, 2011.

OLIVEIRA, T. L. **Língua e percepção: O Processamento dos clíticos com referência ao interlocutor no Português Brasileiro**. 2018. 258 f. Tese (Doutorado em Letras Vernáculas) – Programa de Pós-Graduação em Letras Vernáculas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018. Disponível em: <https://drive.google.com/file/d/1qamg9vOzB0LAINZ1ecuowHNR7Calq8w0/view?usp=sharing>

OLIVEIRA SILVA, D. **A expressão pronominal do acusativo e do dativo na segunda pessoa no português brasileiro: análise de roteiros cinematográficos**. 2011. Dissertação (Mestrado em Linguística Românica) – Eberhard-Karls Universität Tübingen, Tübingen, 2011.

SOUZA, C. D. **Eu te amo, eu lhe adoro, eu quero você**: a variação das formas de acusativo de 2ª pessoa em cartas pessoais (1880-1980). 2014. 156 f. Dissertação (Mestrado em Letras Vernáculas) – Programa de Pós-Graduação em Letras Vernáculas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em:

<http://www.posvernaculas.letras.ufrj.br/images/Posvernaculas/3-mestrado/dissertacoes/2014/3-SouzaCD.pdf>.

WILLIAMS, E. B. **Do latim ao português**: fonologia e morfologia históricas da língua portuguesa. Rio de Janeiro: Tempo Brasileiro, 1994.

Sobre o autor

Thiago Laurentino de Oliveira (Orcid iD: <https://orcid.org/0000-0002-9537-5264>)

Doutor e mestre em Letras Vernáculas (Língua Portuguesa) pela Universidade Federal do Rio de Janeiro (UFRJ); bacharel e licenciado em Letras - Português/Literaturas pela mesma instituição. É professor adjunto do Departamento de Letras Vernáculas (Setor de Língua Portuguesa) da Faculdade de Letras da UFRJ.

Recebido em agosto de 2020.

Aprovado em outubro de 2020.