

CAPÍTULO 8

METODOLOGIAS E ANÁLISES EM ESTUDOS DE PREDICAÇÃO TRANSITIVA DIRETA COM PRONOME *SE*: UM ESTUDO PUXA OUTRO

Eneile Santos Saraiva de Pontes
Marcia dos Santos Machado Vieira
(Universidade Federal do Rio de Janeiro)

8.1 INTRODUÇÃO

Este capítulo tem como principal objetivo apresentar o passo a passo e as metodologias empregadas em estudos desenvolvidos no Projeto Predicar, desde 2008, os quais se têm dedicado a analisar os usos de construções transitivas diretas com pronome *SE*, em textos escritos e orais do Português Brasileiro (PB). A seguir, há exemplos das construções:

(1) O ganho por produtividade somado às baixas remunerações, **paga-se**, em média, [3 reais] por tonelada extraída, compelem esses homens, muitas vezes, a ultrapassarem os limites físicos. [Artigo de opinião, texto jornalístico]

(2) Enquanto isso, **deve-se repor** os líquidos perdidos, comer (com moderação) para repor a necessidade energética do corpo e esperar até poder sair de casa para pular o Carnaval novamente. [Entrevista, DID, Projeto NURC-RJ].

(3) **Tem-se**, tão-somente, uma relação semântica: o locutor anuncia o tópico sobre o qual vai falar para depois fazer um comentário por meio de uma sentença completa. [Artigo científico, Revista Diadorim, n 2, 2007].

(4) Com um *corpus* de palavras escritas por 81 alunos, **elaborou-se** um instrumento de coleta de dados ancorado nos contextos fonológicos apontados por Paiva (1996) como favoráveis ao processo de monotongação. [Dissertação, Letras, 2020]

(5) **Pretende-se reduzir** a carga de impostos sobre o lucro das empresas de 34 para cerca de 20. A perda de arrecadação seria compensada com a volta do Imposto de Renda sobre os dividendos distribuídos aos acionistas. [Editorial, Folha de São Paulo, 2019]

Os dois primeiros dados apresentados em (1) e (2) compuseram o *corpus* da primeira fase da pesquisa desenvolvida entre 2008 e 2011.¹ Na ocasião, nossa atenção centrava-se em ocorrências de predicação transitiva direta com o pronome SE, com predicador simples (1) ou complexo (2), que projetasse um participante não agentivo (tema, afetado ou efetuado) manifesto por SN plural. Na amostra que reuniu ocorrências com essa configuração, analisamos a regra variável de concordância: flexão verbal em acordo ou não com o SN plural (variantes singular e plural). Isso foi feito a partir dos pressupostos da Teoria da Variação e Mudança (WEINREICH; LABOV; HERZOG, 1968; LABOV, 1994), em uma abordagem sociofuncionalista (TAVARES, 2003; GÖRSKI et al., 2003; PAIVA, 1998).

O terceiro exemplo (3) é um dado da amostra examinada numa segunda fase de análise (2011-2013)² em que o nosso principal objetivo era avaliar o uso da construção *tem-se X* (ou *ter-se X*), que entraria em competição, principalmente na modalidade escrita, com a forma verbal *há X* (ou *haver X*). Mais uma vez, seguimos os pressupostos da Teoria da Variação e Mudança e, mais especificamente, da Sociolinguística Quantitativa articulados à Linguística Funcional, o que redundou num olhar sociofuncionalista que considerou a questão da gramaticalidade/gramaticalização (HOPPER, 1991) de *tem-se/têm-se* na esfera discursiva de falantes mais escolarizados em práticas escritas, como recurso a conter, em certa medida, uma inclinação ao acionamento de *ter* existencial e a se alinhar a construções com SE mais presentes naquela esfera.

Já os dois últimos exemplos (4) e (5), com dados de predicadores que projetam um participante não agentivo que se manifesta via SN no singular, estão associados a amostras do estudo desenvolvido na terceira fase (2018-2022).³ Interessa-nos, nas amostras de dados com esse contorno avaliar a variação de usos entre três construções saber:

1 Período de pesquisa em Iniciação Científica.

2 Período de pesquisa de Mestrado em Língua Portuguesa (SARAIVA, 2013).

3 Período da pesquisa de Doutorado em Língua Portuguesa (SARAIVA DE PONTES, 2022).

(i) [Predicador_{TD} + SE (participante 2 [SN/SO]) (*participante 1 suspenso*)]_{predicação}; (ii) [Verbo_{(SEMI-)AUXILIAR} + SE + Verbo Principal_{TD} (VP_{TD})]_{predicador complexo} (participante 2 [SN/SO]) (*participante 1 suspenso*)_{predicação}; e (iii) [Predicador_{TER} + SE (participante 2 [SN/SO]) (*participante 1 suspenso ou inexistente*)]_{predicação}.

Dessa forma, todas as ocorrências de (1) a (5), com predicadores simples ou complexos que projetam participante manifesto por SN no singular ou plural e SO, passam a ser contempladas a partir de uma análise (socio)construcionista (BYBEE, 2010, 2013; FILLMORE, 1988; GOLDBERG, 1995, 2006; LANGACKER, 1987, 1991, 2008; MACHADO VIEIRA, 2017; TRAUOGOTT; TROUSDALE, 2013; WIEDEMER; MACHADO VIEIRA, 2018). E as construções de predicação transitiva com pronome SE começam a ser observadas considerando-se os domínios discursivos em que ocorrem e a tipologia textual neles implicados: jornalístico e acadêmico; argumentativo ou expositivo. E, assim, centramos nossa atenção na relação entre construções de predicação transitiva direta com pronome SE e construções textuais-discursivas.

A sequência de três fases da pesquisa de dados de construção de predicação transitiva direta com pronome SE revela a constante preocupação em mapear diferentes fenômenos que afetam sua configuração (concordância, gramaticalidade/gramaticalização de verbo nela envolvido, variação construcional) e ocorrências da construção em contextos reais de uso para, dessa maneira, propor a descrição gramatical de suas características morfossintáticas, semântico-pragmáticas, sociodiscursivas e cognitivas. Também revela o poder que temas correlacionados têm na dinamização de um fazer científico: somos movidos por perguntas que nos levam a algumas respostas e também a mais perguntas.

Aqui, procuramos discutir: (i) o uso de estatística e a constituição e análise de *corpus* nos estudos linguísticos; e (ii) metodologias e alguns resultados desses três momentos da pesquisa já citados. Assim, pretendemos, de forma panorâmica, traçar breve apresentação das fundamentações teóricas dos estudos de predicação transitiva direta com pronome SE, da relação entre problemas e hipóteses, motivações e expectativas de pesquisa, de questões relativas à etapa de levantamento de dados e composição do *corpus*, do processo de análise qualitativa e quantitativa e do *design* de conclusões e generalizações a partir das análises empreendidas.

8.2 A LINGUÍSTICA DE *CORPUS* E O USO DE ESTATÍSTICA

Levshina (2015, p. 01) afirma que a estatística tem papel fundamental no desenvolvimento dos estudos linguísticos. A autora destaca que as análises pautadas em percentuais estatísticos tiveram um papel importante nas chamadas disciplinas linguísticas híbridas: psicolinguística, sociolinguística, linguística aplicada, computacional e de *corpus*.

Levshina também ressalta que a estatística não era acionada nos estudos embasados nas teorias estruturalista e gerativista, pois tais correntes partem do pressuposto de que a gramática é autônoma e, dessa forma, “um observador linguístico, portanto,

pode descrever os hábitos de fala de uma comunidade sem recorrer a estatísticas”⁴ (BLOOMFIELD, 1935, p. 37). Entendemos que a autora põe, então, em proeminência a ótica em que a descrição gramatical pode estar centrada em virtualidades, reunidas a partir da vivência do analista, que também é usuário da língua e, como tal, se envolve em práticas sociocomunicativas, além de observar, numa comunidade, outras práticas de que não se envolve como enunciador.

Todavia, a partir da mudança de perspectiva, em que a língua passa a ser analisada a partir do uso, o tratamento estatístico de dados de línguas ganha um espaço importante nas análises linguísticas e a autora destaca o trabalho de Langacker (1987) como um importante marco, pois defende que o conhecimento linguístico é moldado pelo uso da linguagem.

Dito isso, é importante salientar também o papel da Linguística de *Corpus* que se ocupa do levantamento e da exploração de *corpora* textuais selecionados com o objetivo de atenderem a diversas pesquisas que examinem fenômenos de uma língua ou de uma variedade linguística. São grandes as discussões em torno de a Linguística de *Corpus* ser considerada como uma teoria ou metodologia (cf. ASSUNÇÃO; ARAÚJO, 2019), mas destaca-se aqui a sua pertinência em trazer para o debate a temática de elaboração de *corpus* para análises linguísticas e:

esta área representa uma nova abordagem filosófica para os estudos da linguagem. Svartvik (1996) concorda com Leech, que afirma que “a linguística de corpus não define somente uma metodologia emergente para o estudo da linguagem, mas uma nova maneira de fazer pesquisa, e de fato uma nova abordagem filosófica para este assunto. O computador, como uma ferramenta tecnológica de poder indiscutível, tornou este novo tipo de linguística possível” (LEECH, 1992, p. 106 citado em SVARTVIK, 1996, p. 12). Entretanto, cabe aos linguistas, com suas próprias intuições sobre a língua, instruir estes programas para extrair as evidências linguísticas com as quais irão trabalhar (OLIVEIRA, 2009, p. 48).

Notamos como o advento do acesso ao computador pessoal encaminhou os linguistas a elaborarem grandes *corpora* para a investigação de fenômenos linguísticos. E isso se intensifica a cada dia, com o rápido desenvolvimento da tecnologia de informação e de computação.

Em 1964, foi lançado o primeiro *corpus* linguístico eletrônico, o *Brown University Standard Corpus of Present-Day American English* e os desafios eram imensos, pois o acervo:

4 Tradução nossa para: A linguistic observer therefore can describe the speech-habits of a community without resorting to statistics (BLOOMFIELD 1935: 37).

continha uma quantidade invejável de dados para a época: um milhão de palavras. Há 35 anos as dificuldades de se informatizar um conjunto de textos eram tremendas. Vale lembrar, por exemplo, que os textos tiveram de ser transferidos para o computador por meio de cartões, perfurados um a um, tal era a tecnologia da época. Este feito, por si só, já traria respeito e admiração à empreitada (SARDINHA, 2000, p. 323).

O avanço tecnológico leva-nos a um novo aparato descritivo para estudar língua escrita e língua falada. Não é à toa que, segundo Machado Vieira et al. (2022), a Associação Brasileira de Linguística já tem em mira uma Plataforma da Diversidade Linguística Brasileira, que se preste a (re)uso amplo, para além de fins específicos relativos à pesquisa linguística de caráter técnico-científico. E também há linguistas brasileiros a investirem em caminhos de ciência aberta (cf. SARDINHA et al., 2021).

Refletir sobre um *corpus* é, portanto, importante se queremos lidar com investigação de dados oriundos da realidade da língua em uso. E, nas pesquisas que apresentaremos aqui, vamos mostrar como selecionamos os *corpora* que julgamos pertinentes para examinar estruturas mais prototípicas em domínios discursivos, tipos e gêneros textuais específicos, como veremos adiante. Vamos tratar também de programas para análise estatística. Há diversos programas estatísticos que nos auxiliam na investigação dos fenômenos linguísticos. Utilizamos, nas nossas pesquisas, o Programa estatístico Goldvarb X desenvolvido por Sankoff, Tagliamonte e Smith (2005), que possibilita uma gama de possibilidades para a realização de análise (quali-)quantitativa e:

a partir de seu uso, podem ser feitas análises univariadas (ou unidimensionais), análises multivariadas (ou multidimensionais) e tabulações cruzadas. As análises univariadas são casos em que se testam o efeito de uma variável independente sobre uma variável dependente. Tais resultados aparecem sob a forma de frequências absolutas e relativas. As multivariadas permitem investigar situações em que a variável linguística em estudo é influenciada por vários elementos do contexto, ou seja, múltiplas variáveis independentes. Essa investigação mede os efeitos, bem como a significância dos efeitos, dessas variáveis independentes sobre a ocorrência das realizações da variável dependente. Os resultados obtidos se apresentam como pesos relativos. A tabulação cruzada, por sua vez, mostra as relações – ou a falta delas – entre as variáveis independentes (GUY; ZILLES, 2007 apud BERLINCK; BIAZZOLI, 2018, p. 260).

Utilizamos, ainda, na pesquisa de doutorado de Saraiva de Pontes (2022), a linguagem R, mais especificamente, o programa RStudio para proceder à análise colostru-

cional. A partir dela, verificamos a força de atração dos verbos na atuação da predicação transitiva direta com pronome SE. Considerando essas possibilidades apresentadas, passaremos, então, nas próximas seções, a abordar as pesquisas que foram desenvolvidas.

8.3 AS CONSTRUÇÕES DE PREDICAÇÃO COM PRONOME SE E SN PLURAL: A (NÃO) ATIVAÇÃO DA RELAÇÃO DE CONCORDÂNCIA.

Empreendemos uma pesquisa de cunho variacionista sobre o fenômeno de flexão verbal variante no singular ou no plural em contexto de construção de predicação transitiva direta com pronome SE e com SN plural (MACHADO VIEIRA; SARAIVA, 2011). Isso foi feito a partir de estudos (cf. BAGNO, 2001; DUARTE, 2008; SCHERRE, 2005) que alertavam para o considerado percentual de não relação de concordância em passivas sintéticas e debates sobre a função sintática exercida pelo SN projetado pelo verbo predicador ser sujeito – como prevê a perspectiva tradicional – ou objeto direto – como propõem estudos linguísticos da literatura linguística (por exemplo, SCHERRE, 2005).

Primeiramente, avaliamos como poderíamos proceder para constituir o *corpus* para análise com base na literatura. E, guiados por estudos (cf. DUARTE, 2007, entre outros) que apontavam para o uso do pronome SE principalmente na escrita, mais monitorada, resolvemos priorizar o levantamento de dados em textos que refletissem usos da norma culta nessa modalidade. Espaços de norma culta como fonte de dados foram priorizados também em razão de outro intento: lograr relacionar a variação na flexão do primeiro verbo presente em dados de locuções verbais ao reconhecimento destas como unidades funcionais e, por conseguinte, do SN plural como potencial termo sujeito a afetar a flexão do primeiro verbo daquelas.

Para tanto, foram consultados materiais do domínio acadêmico (resumos de eventos acadêmico-científicos, monografias finais de disciplinas de Pós-Graduação em Letras Vernáculas, textos de provas e trabalhos de graduandos de disciplinas de Português do curso de Letras) e jornalístico (artigos de opinião, notícias, entrevistas transcritas, materiais coletados nas revistas *Veja* e *Caros Amigos*, bem como em jornais e/ou boletins da UFRJ e em *O Globo*). Também consideramos as elocuições formais (gravações de aulas, conferências, palestras etc.) do acervo NURC-RJ,⁵ a fim de avaliar a frequência de uso das variantes em predicções com pronome SE em contextos de fala espontânea comparativamente a contextos de escrita, mesmo com a expectativa inicial de que as construções com pronome SE não seriam produtivas nessa modalidade.

Seguimos os pressupostos teóricos do Sociofuncionalismo e, dessa maneira, a partir da meta de proceder a uma análise multivariada dos dados, buscamos analisar os

5 O acervo do Projeto NURC-RJ (Projeto da Norma Urbana Oral Culta do Rio de Janeiro) é referência nacional para estudos da variante culta da língua portuguesa. Para maiores informações, eis o endereço eletrônico: <https://nurcrj.letras.ufrj.br/>.

grupos de fatores (variáveis independentes) que poderiam coatuar na (não) marcação de flexão verbal no plural, ou na concordância verbal com o SN plural se perspectívamos os dados de predicação transitiva direta com pronome SE como construção de voz passiva sintética.

De acordo com Tavares (2003), na articulação entre o Funcionalismo e a Sociolinguística:

i) o objeto de estudo é a língua em uso, cuja natureza heterogênea abriga a variação e a mudança; (ii) os fenômenos linguísticos que constituem o alvo das investigações são analisados em situações de comunicação real em que falantes reais interagem; (iii) a língua está continuamente se movendo, mudando e interagindo; (iv) a mudança espalha-se de forma gradual ao longo do espectro social, considerando-se fatores como região, geração, classe social etc., sendo o aumento de frequência de uso compreendido como índice de difusão sociolinguística; (v) fenômenos de mudança podem ser atestados através de tratamento empírico com quantificação estatística (TAVARES, 2003, p. 114).

Trabalhamos com dois dos 5 questionamentos propostos por Weinreich; Labov; Herzog (1968):⁶ a *questão dos fatores condicionantes* (em que as mudanças são observadas considerando fatores linguísticos e extralinguísticos que podem favorecer ou não a utilização de uma variante inovadora) e a *questão da implementação* (que busca pensar nos motivos que levam à mudança linguística em uma determinada época).

Nesse momento preliminar da pesquisa, estávamos preocupados em analisar o maior número de dados em *corpora* múltiplos para entender as potencialidades de manifestação do fenômeno da variação. Líamos todos os materiais a que tínhamos acesso com o objetivo de capturar manifestações da construção de predicação a descrever. E, como tínhamos interesse somente em dados que contivessem SNs plurais como manifestação do participante não agente do predador transitivo direto da predicação, o rol de dados era mais específico. Assim, lidamos com uma pequena amostra de dados, apesar de muitos textos terem sido consultados.

Conseguimos somar um total de 286 dados e trabalhamos com as seguintes expectativas iniciais: (i) por lidarmos com textos produzidos em esferas de uso da norma culta, esperávamos, no cômputo geral, índices expressivos de marcação de flexão verbal no plural, maior inclinação à concordância do predador com o SN plural; (ii) no caso de dados de predicação via locuções verbais, a expectativa para as construções com verbo (semi-)auxiliar era a de diferença balizada por conta da natureza desse verbo (o primeiro da locução) – obtendo-se maior percentual de ocorrência de dados com flexão verbal no plural, se o primeiro verbo de uma locução apresentar mais atri-

6 **Completam:** a *questão da transição* (busca avaliar os caminhos da mudança na estrutura social, do encaixamento (que observa as mudanças linguísticas articuladas, entrelaçadas entre si e na estrutura social) e da avaliação (reflete-se sobre os efeitos da mudança sobre a estrutura e o uso da língua).

butos de auxiliaridade, e menor percentual de flexão do verbo no plural, se o verbo apresentar menor perfil de auxiliaridade –; e (iii) quanto à modalidade de expressão, a expectativa de que partimos foi a de que a estruturação de predicções com pronome SE não fosse tão frequente na fala e, se estivesse presente, apresentaria índices maiores de não flexão verbal no plural, de não concordância, pelo fato de ser associada a uma maior espontaneidade e a maior grau de planejamento textual local, não passando geralmente, então, por processos de planejamento prévio e revisão a que podem ser submetidos os textos escritos em análise. Com relação ao que afeta a variação flexão verbal singular e flexão verbal plural, nossa hipótese de partida foi a de que fatores de ordem morfossintática e de ordem sociodiscursiva pudessem afetar os índices de acionamento das variantes. Essas expectativas e hipótese nortearam o curso da análise multivariada. Uma análise multivariada permite examinar como a variável linguística sob exame (concordância ou não) é influenciada por fatores de variáveis independentes (atributos de ordem formal e funcional), quais são os fatores condicionantes relevantes e se há restrições (contextos de (não) ativação da concordância em 100% dos casos, por exemplo).

Passamos, então, a empreender uma análise multivariada, a fim de analisar os fatores condicionantes em coatuação no fenômeno. Os dados foram submetidos aos aplicativos *makecell* e *ivarb*, do programa estatístico *GoldvabX*, que nos oferecem valores absolutos e percentuais dos dados por variáveis independentes, bem como pesos⁷ relativos quanto à relação de coatuação destas e o acionamento das duas variantes em análise.

Trabalhamos com 2 variantes: a padrão, com marcação de flexão verbal plural/concordância com o SN plural; e não padrão, sem a marcação de flexão verbal plural/não concordância com o SN plural. Analisamos as seguintes variáveis independentes, cada uma delas com seus fatores: fonte do texto, gênero textual, modalidade expressiva, ordem e distância do argumento sujeito em relação ao predicador, caráter animado ou inanimado do argumento interno, tipo de sentença (finita ou não finita), grau de saliência fônica entre a forma singular e plural do predicador verbal, configuração semântica do SN (composto ou não por numeral), natureza semântica do argumento externo indeterminado e a configuração semântica do verbo que se flexiona.

Preparamos, então, uma planilha de dados para a análise multivariada, tendo em vista essas variáveis (seus fatores). Categorizamos cada ocorrência existente na amostra, licenciada pela construção de predicação transitiva direta com pronome SE.

Esse arquivo (planilha, também gravada em formato texto delimitado por tabulação, extensão .txt) foi submetido aos procedimentos relativos aos aplicativos do programa *Goldvarb X*,⁸ que, ao final do processamento computacional dos dados e das

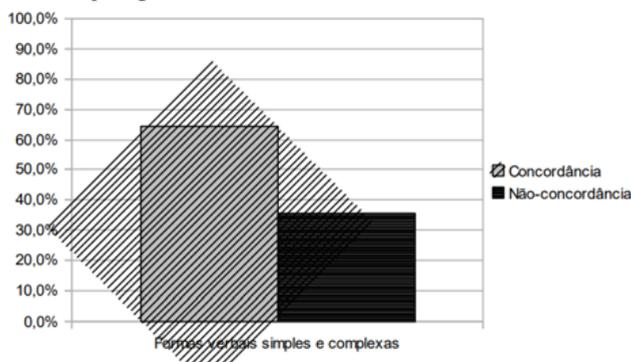
7 Para mais informações sobre como proceder no *Goldvarb X*, sugerimos a leitura de: BERLINCK, R. A.; BIAZOLLI, C. C. Ferramentas metodológicas para análises (sócio)linguísticas. In.: *Estudos Linguísticos*, 2011, p. 260-273. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/2047/1363>.

8 Em linhas gerais, o processo estrutura-se assim: checagem da codificação feita (*generate factor specifications, check tokens*) com a geração de um arquivo com extensão .res; criação de um arquivo de condições (*tokens, no recorde*) com extensão .cnd; processamento da distribuição dos dados pelos

influências investigadas, nos oferece a distribuição de dados/ocorrências em função das variáveis (dependente e independentes) em termos de valores absolutos e percentuais, bem como o resultado da rodada estatística, por regressão logística multivariada, a respeito das variáveis e fatores que coatuam nas tendências a uma ou outra variante. Destacamos aqui, desses resultados, a configuração semântica do argumento, pois notamos uma forte tendência à (não) marcação de plural na forma verbal quando o SN era composto por um numeral (como no exemplo 1) e esse foi um grupo de fatores apontado pelo programa como desencadeador da não realização de concordância. A seguir, serão apresentados dois gráficos e uma tabela (MACHADO VIEIRA; SARAIVA, 2011) que mostram alguns dos principais resultados obtidos:

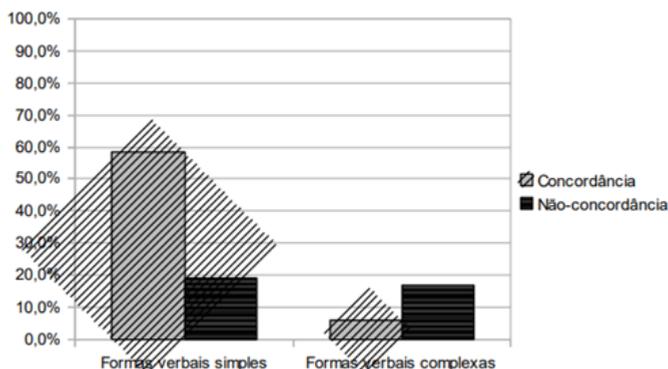
Gráfico 1: Distribuição percentual das variantes concordância vs. não concordância em estruturas passivas pronominais formadas de predicadores simples ou sequências verbais seguidos de SN plural.

3. A regra variável de flexão de número em construções passivas pronominais
3.1 No cômputo geral dos 286 dados



Fonte: MACHADO VIEIRA; SARAIVA, 2011.

Gráfico 2: Distribuição percentual das variantes concordância vs. não concordância em estruturas passivas pronominais formadas de predicadores simples e em estruturas formadas de sequências verbais.



Fonte: MACHADO VIEIRA; SARAIVA, 2011.

grupos de fatores (makecell); análise multivariada (ivarb, cells > binomial up and down); possivelmente, tabulação cruzada (cross tabulation).

Tabela 1: Distribuição percentual das variantes concordância vs. não concordância em estruturas passivas pronominais por cada domínio discursivo considerado

Domínio discursivo	Concordância		Não concordância	
	Dados / total	%	Dados / total	%
Jornalístico	102 / 140	72,9%	38 / 140	27,1%
Acadêmico	71 / 116	61,2%	45 / 116	38,8%
Conversacional	01/11/30	36,7%	19 / 30	63,3%
Total	184 / 286	64,3%	102 / 286	35,7%

Fonte: MACHADO VIEIRA; SARAIVA, 2011.

O primeiro gráfico revela que o percentual de marcação de concordância é significativo; entretanto, o índice de não flexão verbal no plural/concordância (35%) é maior do que esperávamos encontrar inicialmente. E, quando analisamos, no Gráfico 2, separadamente predicadores simples e complexos, notamos que estes apresentam maior índice de não concordância. Ao examinarmos a natureza dos verbos (semi-) auxiliares, comprovamos a hipótese de que os predicadores complexos que contêm verbos semiauxiliares com menor grau de auxiliaridade (poder, dever, pretender, por exemplo) são os que menos revelam flexão plural. Na Tabela 1, observamos apenas 30 dados no domínio conversacional, que apontaram, como prevíamos para maior índice de não concordância e, mais uma vez, tivemos uma hipótese confirmada.

Sobre a questão da implementação, traçamos algumas reflexões que poderiam motivar a não realização da concordância: por exemplo, o fato de, nas construções de predicação, o termo projetado nem sempre ser interpretado como sujeito e, sim, como um objeto direto (que não ativaria a relação de concordância).

Essa etapa inicial da pesquisa⁹ foi muito importante, pois nos auxiliou a pensar melhor nas construções de predicação com pronome SE e a formular novas perguntas, por exemplo: por que a construção é mais acionada na modalidade escrita? Quais seriam as suas funções pragmáticas? Como o usuário interpreta essas construções? E, nas etapas seguintes, voltamo-nos a dar conta desses questionamentos.

9 Outro estudo que revela resultados dessa fase de pesquisa é MACHADO VIEIRA (2015), https://www.mundoalfal.org/es/pt_Caderno7. Acesso em: 11 fev. 2022.

8.4 A CONSTRUÇÃO *TEM-SE* NO PORTUGUÊS BRASILEIRO: ESTARIA EM VARIAÇÃO COM A FORMA VERBAL *HÁ*?

Ao longo do levantamento de dados relativo à constituição de uma amostra da língua portuguesa para estudo variacionista da concordância verbal, uma construção em especial começou a chamar a atenção: *tem-se*. Passamos a configurar um arquivo somente para as suas ocorrências que apareciam entre os dados que buscávamos. Com isso, reunimos uma amostra de dados que não foram utilizados nessa primeira etapa da pesquisa. Não é raro no processo de pesquisa de um objeto linguístico esbararmos em outro(s) que não estavam no nosso campo de atenção. A pesquisa exige foco e recorte para tanto; é possível, entretanto, separar dados de outro objeto correlacionado ao que buscamos para investigação futura.

Por se tratar de uma construção que consubstancia a associação do pronome SE a um verbo estativo (não controlado e não dinâmico), “ter”, em lugar de um verbo dinâmico, a construção *tem-se* intrigava. E por quê? Porque, a princípio, o verbo estativo não se combinaria a uma estruturação na voz passiva, contraparte de estruturação na voz ativa. Também porque percebemos um uso parecido com o da forma verbal *há*, em contextos de existencialidade (6) e (7) e impersonalização discursiva (8) e (9):

(6) Assim, **há** pouca deformação plástica e a elevada dureza persiste mesmo em altas temperaturas, ao contrário dos metais. [Artigo científico, Revista Matéria, v. 16, n. 1, 2011].

(7) As formas simples contempladas configuram-se como verbos pronominais; e as formas complexas, como perífrases com verbo suporte em que o elemento não-verbal ligado a ele é cognato àquela forma simples e como construções com verbo auxiliar de passiva ligado a um verbo em sua forma nominal também cognato à forma simples. **Têm-se**, então, três tipos de construções contempladas nesse estudo. [Dissertação de mestrado, Letras, 2011].

(8) Em 1, por exemplo, **há** a valorização e ênfase do constituinte “os médicos”, modificador do SN complexo, e obscurecimento da importância do núcleo “o laudo”, como consequência do recurso formal do CV. [Artigo científico, Revista Diadorim, n 4, 2008].

(9) Em linhas gerais, **tem-se** o intuito de verificar a distribuição estatística das formas verbais em alternância pelos contextos linguísticos e extralinguísticos estipulados para o exame do fenômeno variável, identificar suas motivações funcionais, estudar a avaliação das formas simples e complexas e a influência desta sobre a seleção e o emprego dessas formas na composição de textos, averiguar a configuração e a funcionalidade de perífrases verbo-nominais na construção discursiva. [Dissertação de mestrado, Letras, 2011].

Em (6) e (7), notamos o acionamento de *há* e *têm-se* em contextos considerados em Saraiva (2013) como mais existenciais, em que a permuta entre as duas variantes e a forma verbal existe(m) é mais proeminente (*têm-se/há/existem* então três tipos de construções contempladas... / *há/tem-se/existe* a valorização e ênfase do constituinte

te...); Só que a predicação com *existir* é pessoal intransitiva. Na de verbo *haver*, a predicação é impessoal transitiva. Detectamos a potencialidade de conceptualização da predicação quer sob o ponto de vista do estado-efeito/resultado, cuja expressão conta ainda com unidades verbais como ocorrer, acontecer, dar-se, além de ter(-se) e haver (por exemplo, Assim, *há/ocorre/dá-se/resulta* pouca deformação plástica...), quer sob o ponto de vista da ação/causação (por exemplo, *há a valorização e ênfase do constituinte «médicos»/valoriza-se e enfatiza-se o constituinte «médicos»*). Já em (8) e (9), mapeamos a recuperação do participante 1 suspenso que se relaciona, referencialmente, à 1ª pessoa discursiva (Em 1, por exemplo, *observa-se/detecta-se a valorização e ênfase do constituinte... / objetiva-se verificar a distribuição estatística das formas verbais...*). O cientista/linguista observa, objetiva.

Então, nosso interesse voltou-se a entender como emergiu, na língua portuguesa, a construção *ter-se*, como se gramaticalizou como uma forma alternante a *haver* e *ter* existenciais.

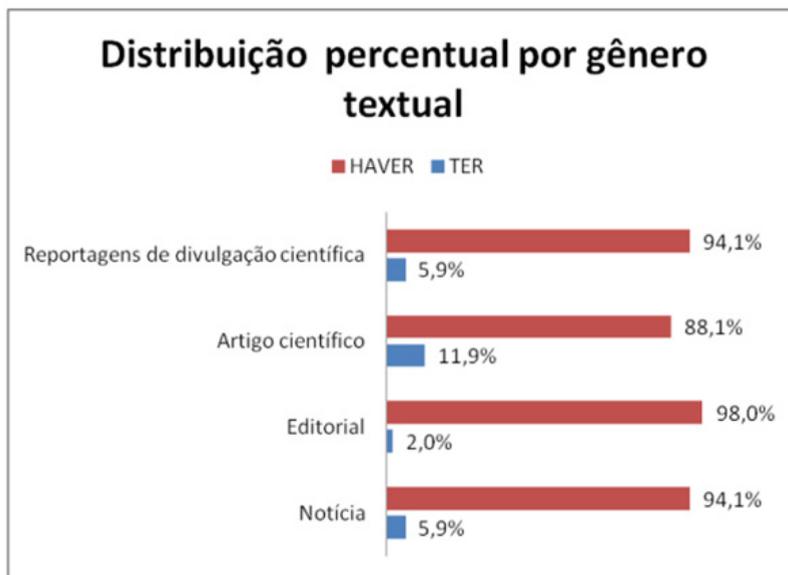
A constituição do *corpus* de análise aconteceu em 3 fases. Na primeira,

foram levantados dados do Jornal O Globo, da Revista Superinteressante, da Revista Diadorim e da Revista Matéria. Inicialmente, foram pesquisados 68 textos (34 notícias e 34 editoriais) do Jornal O Globo, escolhidos aleatoriamente em edições dos anos de 2010 e 2011, 68 artigos científicos (34 da Revista Diadorim – volumes 2, 4 e 6 – e 34 da Revista Matéria – volume 16, do terceiro trimestre de 2011, volume 15, do primeiro e do terceiro trimestre de 2010, e volume 14, do quarto trimestre de 2009) e, por fim, 34 reportagens de divulgação científica da Revista Superinteressante (edições 289, 290, 294 e 298 do ano de 2011) (SARAIVA, 2013, p. 48).

Os periódicos *Diadorim: revista de estudos linguísticos e literários* e *Matéria*, respectivamente, ligam-se ao Programa de Pós-Graduação em Letras Vernáculas e ao Laboratório de Hidrogênio, Coppe – Universidade Federal do Rio de Janeiro, com publicações de pesquisadores, em sua maioria, alunos de pós-graduação, a primeira da área de Letras e a segunda da área de Ciência e Engenharia de Materiais e de Materiais para Energia. Assim, lidamos com textos da modalidade escrita, já que os dados da primeira fase de investigação nos revelaram baixos índices de construções com SE na oralidade. Dessa forma, tínhamos uma amostragem aleatória e representativa (cf. LEVSHINA, 2015), já que os textos eram selecionados casualmente, mas com a característica de serem reveladores da norma culta.

Assim, contou-se com um total de 194 dados e em 93,3% (181 dados), “constatou-se o uso do verbo *haver*”. Com relação ao verbo *ter*, em que estavam inseridos os dados com *tem-se/(têm-se)*, o percentual de uso foi de 6,7% (13 dados)”. (SARAIVA, 2013, p. 49, 50). Os resultados distribuídos entre os gêneros textuais, são observados no gráfico a seguir:

Gráfico 3: Distribuição percentual de *ter* e *haver* por gênero textual numa amostra-teste de 194 dados.



Fonte: SARAIVA, 2013.

Notamos que os usos de *tem-se* se concentravam na modalidade acadêmica, principalmente nos artigos científicos. Por essa razão, resolvemos ampliar o *corpus* e trabalhar com a variável binária *tem-se/têm-se-há*, já que os dados com *tem* eram pouquíssimos, o que nos encaminhou para a hipótese de que a gramaticalidade de *tem-se* ocorreria para continuar a competição entre as formas verbais *ter* e *haver* na escrita, já que, na fala, se computam muitos registros da forma *tem* com valor impessoal, o que não ocorre na escrita (cf. CALLOU; AVELAR, 2007), em que a forma *é*, de certa forma rechaçada. Assim, foram levantados dados nas seguintes fontes: *Jornal O Globo*, *Jornal Extra*, *Revista Caros Amigos*, *Revista Diadorim*, *Revista História Comparada*, *Revista Matéria* e dissertações das áreas de Letras, História e Engenharia Elétrica dos Programas de Pós-graduação da Universidade Federal do Rio de Janeiro.

Dessa forma, como representantes da modalidade acadêmica, além das áreas de Letras e Engenharias, incluímos a *Revista História Comparada*, uma publicação do Programa de Pós-graduação História Comparada da Universidade Federal do Rio de Janeiro, e dissertações das três áreas. E, em relação aos textos da modalidade jornalística, consultamos o *Jornal Extra*, pois é uma publicação da mesma editora do *Jornal O Globo*, que possui como público-alvo as classes C e D. E, pelo fato de os jornais serem veiculados a camadas sociais diferentes, pretendeu-se verificar se, no *Jornal Extra*, se encontravam ocorrências ou não de *tem-se*. Foram selecionados, mais uma vez de maneira aleatória:

- (1) 400 notícias e reportagens do Jornal O Globo e do Jornal Extra (200 de cada veículo);
- (2) 50 editoriais do Jornal O Globo;
- (3) 100 artigos de opinião do Jornal O Globo (as edições do Jornal O Globo e do Extra consultadas distribuíram-se entre os anos de 2011 e 2012);
- (4) 102 artigos científicos, sendo 34 da Revista Diadorim (volumes 2, 4 e 6), 34 da Revista Matéria (volume 16, do terceiro trimestre de 2011, volume 15, do primeiro e do terceiro trimestre de 2010 e volume 14, do quarto trimestre de 2009) e 34 da Revista História Comparada (volumes: 1, nº2 – dez., 2007; 2, nº1 – jun., 2007; 2, nº2 – dez., 2008; 3, nº1 – jun., 2009; 3, nº2 – dez., 2009; 4, nº1 – jul., 2010; 4, nº2 – dez., 2010; 5, nº1 – ago., 2011 e 5, nº2 – dez., 2011; e
- (5) 60 dissertações (sendo 20 da área de Letras – Mestrado em Língua Portuguesa, 20 da área de História – Mestrado em História Comparada e 20 da área de Mestrado em Engenharia Elétrica) entre os textos publicados no período compreendido entre 2009 e 2011 (SARAIWA, 2013, p. 54,55).

A fim de buscar um equilíbrio, consideramos somente as 20 páginas iniciais, a contar da introdução, das dissertações. Como aconteceu na primeira fase de pesquisa em nível de iniciação científica, também lidamos com textos escritos: publicações impressas e em formato PDF. E, nesses, contávamos com o auxílio da busca das estruturas pela ferramenta lupa. Essa etapa de levantamento de dados era muito demorada e trabalhosa, mas feita sempre com muito cuidado para que não deixássemos nenhum dado de fora. Como se tratava de uma pesquisa sociovariacionista, tínhamos como aliado, em todas as etapas, o programa Goldvarb X e a análise estatística que ele propicia. De acordo com Levshina (2015, p. 1),

A estatística está no cerne da investigação científica. Ela auxilia o pesquisador a formular e testar hipóteses teóricas, fazendo generalizações sobre uma população de interesse com base em uma amostra limitada. É indispensável em todas as etapas do ciclo empírico da pesquisa, desde a formulação de uma hipótese e coleta de dados até a análise de dados e falsificação de hipóteses.¹⁰
(LEVSHINA, 2015, p. 1).

10 Tradução nossa para: Statistics lies at the heart of scientific investigation. It helps the researcher to formulate and test theoretical hypotheses, making generalizations about a population of interest based on a limited sample. It is indispensable at all stages of the empirical cycle of research, from formulation of a hypothesis and data collection to data analysis and hypothesis falsification (LEVSHINA, 2015, p. 1).

Assim, a análise quantitativa e qualitativa forneceu-nos bases para trabalharmos com algumas hipóteses, dentre as quais, destacaremos as seguintes: (i) por ser a variante inovadora, *tem-se/têm-se* seria utilizada em menores percentuais, como já verificamos na primeira etapa de levantamento de dados; (ii) a referida construção seria mais acionada nos textos acadêmicos, pela sua semântica de promover a opacificação do autor do texto; e (iii) o usuário da língua interpretaria tal estrutura como se fosse um único bloco, ou seja, gramaticalizada tal qual uma unidade verbal que entra em variação com outras unidades da língua, *haver* e *ter*. A seguir, um gráfico com importantes informações sobre o estudo sociovariacionista:

Tabela 2: Distribuição dos dados

Formas	Dissertações Letras	Dissertações História	Dissertações Engenharias	Artigos Diadorim	Artigos História	Artigos Matéria	Not. e Rep. O Globo	Not. e Rep. O Globo	Artigos O Globo	Artigos Caros Amigos	Editoriais O Globo	Total
Há	130	29	31	122	90	38	52	25	44	89	46	696
Tem-se/ Têm-se	27	4	25	18	8	5	0	0	0	2	1	90
Tem	0	0	0	0	0	0	4	4	2	0	0	10

Fonte: SARAIVA (2013).

Foram contabilizados um total absoluto de 796 dados e ainda constam, na tabela apresentada, as ocorrências com *tem* para demonstrar que, pouco utilizadas na escrita, não aparecem nos textos acadêmicos, o que corrobora a hipótese de que *tem-se* seria mais acionado na escrita por conta da forma *tem* ser rechaçada e por conta de o texto acadêmico ser um lugar em que unidades construcionais com pronome SE são mais acionadas. Após essa etapa, passamos, então, para a terceira e última, que considerou somente os dados levantados em textos acadêmicos (527 ocorrências),

já que o enfoque principal da pesquisa é averiguar se a construção tem-se(/têm-se) estaria perpassando por um processo de gramaticalidade e, como ela demonstra-se mais produtiva somente na modalidade acadêmica, decidiu-se, assim, trabalhar somente com textos presentes nesta classificação (SARAIVA, 2013, p. 56).

Passamos, então, a analisar 11 grupos de fatores para entendermos melhor os condicionamentos linguísticos e extralinguísticos que podem influenciar na ativação de uma das variantes disponíveis no Português: (1) fonte, (2) estrutura da predicação, (3)

presença e natureza de elemento antecedente proclisador, (4) posição das variantes no período, (5) tipo de sentença, (6) localização do contêiner no período, (7) localização da construção especificadora no período, (8) grau de envolvimento do autor na expressão do enunciado, (9) equivalência entre as formas *tem-se*/(*têm-se*) e *há*, (10) estruturação do elemento nominal (SN) e (11) natureza semântica do SN selecionado.

Como o enfoque deste capítulo está em expor e ilustrar as metodologias utilizadas, abordamos aqui somente o resultado referente ao grupo (3) presença e natureza do elemento antecedente proclisador, pois, mediante a presença de um vocábulo atrator do pronome (advérbio de negação, palavras do grupo *qu-*) houve uma resistência ao acionamento de “se tem”, fato que demonstra a cristalização e gramaticalidade da construção *tem-se*. Eis os resultados obtidos:

Tabela 3: Presença e natureza de elemento antecedente proclisador

Presença e natureza de elemento antecedente proclisador	Tem-se/ Têm-se		Há		PR
	Dados/Total	%	Dados/Total	%	
Período sem atrator	75/321	23.4%	246/321	76.6%	.77
Presença de vocábulo atrator no período	1/13	7.7%	12/13	92.3%	.73
Presença de vocábulo atrator imediatamente antes da variante	1/149	0.7%	148/149	99.3%	.06

Fonte: SARAIVA, 2013.

Esse grupo de fatores foi selecionado pelo GoldVarbX como relevante para a análise e para o acionamento de *tem-se/têm-se*. E, de acordo com a análise da tabela, percebemos que seu uso é prioritário (75 das 77 ocorrências) em períodos sem elemento atrator. E essa conclusão nos orientou a defender a hipótese de que a construção era acionada e processada pelo usuário da língua como um bloco único, pois é evidente a inclinação a evitar a variante em contextos em que a colocação do -SE ocorre em posição de ênclise.

Sobre o estudo da gramaticalização a partir da perspectiva funcionalista, há a preocupação em investigar as fases do processo, desde o surgimento de uma forma linguística até uma *substituição* (possível de acontecer, mas não obrigatória) de uma forma mais antiga por uma mais atual ou *especificação*, cuja forma gramaticalizada estaria em uso restrito em determinado domínio funcional. Em Bybee (2003), notamos que é feito um destaque à frequência de uso como um dos principais fatores que contribuiriam para a gramaticalização, pois favorece o processo de mudança. Dessa forma, a partir da repetição, sequências, de palavras ou de morfemas, começam a ser acionadas e interpretadas como uma unidade única do processamento linguístico.

Com o objetivo de averiguar como o usuário da língua percebe a construção *tem-se*, elaboramos um teste de atitudes em que, dentre outras questões levantadas, o informante diria como ele interpretaria sintaticamente o termo projetado pela construção, como sujeito ou objeto direto e se ela é considerada como mais típica da oralidade ou da escrita. A partir dos resultados,

acredita-se que têm-se(/têm-se) será apontada prioritariamente como uma estrutura inerente à língua escrita e haverá uma associação da construção como estrutura passiva sintética, em virtude da transitividade direta do verbo ter (SARAIVA, 2013, p. 46).

O teste foi aplicado a um conjunto de estudantes de graduação de quatro áreas: letras (30 alunos da UFRJ), história, pedagogia e matemática (18 alunos da UFRJ). Os formulários preenchidos pelos participantes da área de Letras foram analisados separadamente, pois os estudantes, que cursaram a disciplina Morfossintaxe do português, possuíam conhecimento sobre construções passivas sintéticas, mais especificamente, sobre a prescrição de marcação de plural na forma verbal em caso de projeção de SN plural e esse, inclusive, foi tema discutido em sala de aula. Foram selecionadas 32 sentenças que continham as formas *tem-se/têm-se* e *há*, predicções transitivas diretas com SE e outras estruturas, como estruturas com o uso de *a gente*, por exemplo, que funcionaram como sentenças distratoras. Os testes *offline*, feitos em formulário em papel, foram respondidos por estudantes voluntários, como atividades de língua portuguesa em que eles registraram o que percebiam como mais pertinente a cada questão de análise de enunciados-estímulos da língua. Nenhum dado de ordem pessoal e/ou que pudesse identificar o participante foi coletado/anotado via formulário. Os formulários continham apenas enunciados relativos à língua portuguesa para que os participantes indicassem a alternativa que percebiam mais ligada ao enunciado-estímulo.

O teste foi respondido pelos estudantes que gentilmente se disponibilizaram a participar. Após a análise das respostas, notamos que não poderíamos afirmar com precisão se as construções com *tem-se* eram associadas à voz passiva, pois, por mais que as predicções transitivas diretas sejam bem perceptíveis nas produções acadêmicas, a maioria dos estudantes das áreas de história, pedagogia e matemática informou que não reconhecia características das construções passivas sintéticas/pronominais:

Ficou evidenciado que estruturas passivas sintéticas não são tão naturais para os estudantes, mesmo ocorrendo (como construções de indeterminação) em seus textos e, por este motivo, não é possível afirmar, com precisão, se realmente a estrutura temse(/têm-se) não foi apreendida como passiva, uma vez que, na aplicação do teste, muitos informantes evidenciaram que só lembravam dos termos “passiva sintética/pronominal/se apassivador”, mas não recordavam quais eram as características da referida construção (SARAIVA, 2013, p. 134).

Quando os resultados dos registros feitos por estudantes de letras foram analisados, percebemos que o fator concordância foi fundamental, pois, quando a sentença apresentava a forma *têm-se* (com \wedge indicando marcação de plural), os alunos marcaram que o SN projetado exercia a função de sujeito; e, nos demais casos, com SN no plural e a forma *tem-se*, os participantes apontavam que estávamos diante de um SN com função sintática de objeto direto:

os informantes de Letras, quando observaram a – relação de concordância com têm-se – plural selecionado SN plural – (construção que contém um VTD), assinalaram a sentença como uma construção passiva. Estes não levaram em conta o fato de o verbo ser “estativo” e teoricamente não ser utilizado nessa construção, salvo os casos já descritos nesta dissertação (SARAIVA, 2013, p. 136).

Em relação ao fato de as construções com SE serem mais associadas ao texto escrito ou não, os participantes do teste precisavam encaixar as sentenças em uma escala de 1 a 5:

sendo que 1 indica que a forma está fortemente relacionada ao contexto da oralidade, 2 indica que a estrutura é mais encontrada na oralidade do que na escrita, 3 indica que a estrutura pode aparecer tanto na oralidade quanto na escrita, mais ou menos com a mesma frequência, 4 indica que a forma é mais encontrada na escrita do que na oralidade e 5 indica que a forma está fortemente relacionada ao contexto da escrita (SARAIVA, 2013, p. 137).

A partir da análise desta segunda etapa do teste, notamos que os dados que continham *tem-se/têm-se* foram mais associados ao contexto da escrita e classificados com 4 (23 de 90 análises) ou 5 (34 de 90 análises). A forma *há* recebeu mais avaliações com 3 (38 de 90), e esses números corroboram a nossa hipótese, pois, a partir da escala proposta, os participantes do teste associaram os dados com *tem-se/têm-se* como sendo mais característicos da língua escrita.

A partir dessa segunda etapa do estudo, desenvolvida no âmbito do mestrado, percebemos que as construções com SE poderiam promover a demissão do participante 1 suspenso igual ao eu-enunciador, o que nos levaria a investigar um *continuum* de impersonalização discursiva na terceira fase de pesquisa, a de doutorado. Essa fase do estudo também revelou que a construção *tem-se/têm-se*, como já havíamos levantado a hipótese, aparece predominantemente no discurso científico e com espaço expressivo de uso na língua escrita.

Vale destacar, por fim, o caráter inédito dessa descrição (SARAIVA, 2013) que teve curso a partir da observação de certos dados de *tem-se* entremeados noutros dados de

construções de predicação transitiva direta com pronome SE para os quais se voltava a atenção na primeira fase, a de iniciação científica. Afinal, até onde sabemos, até a pesquisa de mestrado,¹¹ na literatura tal fenômeno variável não tinha sido objeto de exame empírico: a unidade construcional *tem-se* como uma variante a entrar em relação com *haver* e *ter* existenciais/impessoais no Português do Brasil escrito (acadêmico e jornalístico). E, assim, vemos como uma pesquisa pode surpreender-nos ao nos levar a outros dados que não eram esperados, a outro caminho de estudo.

8.5 A VARIAÇÃO DE USOS DE PREDICAÇÃO TRANSITIVA DIRETA COM PRONOME SE

Nessa terceira etapa da pesquisa, passamos a analisar a variação de usos das construções transitivas diretas com o pronome SE. Assim, avaliamos o contexto da impersonalização (quando o participante desfocalizado é a primeira pessoa discursiva) ou da indeterminação (quando o participante desfocalizado é a terceira pessoa discursiva) e da existencialidade, sendo esse último acionado somente pela construção *tem-se*. A seguir, exemplos com esses usos:

(10) Na sequência, **deve-se analisar** o Livro IV do Código Civil, sobre o Direito das Sucessões, para, comparando seu conteúdo ao correspondente no Projeto de FELICIO DOS SANTOS, verificar continuidades entre um texto e outro, que configurem possíveis contribuições deste para aquele. [Tese, Direito, UFMG, 2017]

(11) No fórum suíço, **divulgaram-se** indicadores desta concentração de riquezas nos últimos anos, enquanto crescia a aplicação de novas tecnologias, a Revolução 4.0 em 40 anos, a diferença entre a remuneração de um executivo e de um trabalhador médio aumentou 970. [Editorial, Jornal O Globo, 2019]

(12) Dessa maneira, **tem-se** que certas atribuições acabam por recair aos órgãos competentes para a representação internacional da Argentina. [Dissertação, Direito, 2017]

(13) Paralelamente, **tem-se** no Rio uma Polícia Militar altamente letal, que também contribui para aumentar o número de mortes que não serão elucidadas. [Editorial, Jornal O Globo, 2020]

Em (10), notamos que se recupera o participante suspenso como o próprio autor da tese, que informa o que ele deve analisar na sequência do seu trabalho (Na sequência, eu *analiso/devo analisar* o livro V...); em (11), observamos a indeterminação, pois há referência à terceira pessoa discursiva (*Participantes do Fórum suíço/Eles divulgaram indicadores...*); em (12), verificamos que a construção *tem-se* também pode ser associada à impersonalização, pois, de forma um pouco mais distanciada do que em (10), ainda é possível recuperar a primeira pessoa discursiva (Eu concluo que certas atri-

11 O texto da dissertação pode ser encontrado no site do Projeto Predicar, disponível em: <https://projeto-predicar.wixsite.com/predicar>.

buições..., *Eu tenho comigo* que certas atribuições...) e, em (13), percebemos que *tem-se* também veicula a existencialidade (*tem-se/ há/ existe* no Rio de Janeiro...).

A hipótese inicial dessa terceira etapa dos estudos das construções transitivas diretas com pronome SE é a de que as construções (i) Predicador_{TD} + SE e (ii) Verbo_{(SEMI-)AUXILIAR} + SE +VP_{TD} seriam mais acionadas nos textos científicos com o intuito de promover demissão da primeira pessoa discursiva, já que os pesquisadores descrevem etapas e metodologias de suas pesquisas e também se posicionam sobre os resultados obtidos, e seriam mais acionadas nos textos jornalísticos para promover a demissão da terceira pessoa discursiva, já que há referências a ações da terceira pessoa discursiva, principalmente personalidades do universo político. A construção PredicadorTER + SE, como demonstrou a etapa anterior, seria mais acionada no discurso acadêmico, com o objetivo de opacificar a figura do eu-enunciador.

Dessa forma, nessa terceira etapa, lidamos com um total de 1.029 dados (782 ocorrências de Predicador_{TD} + SE, 72 ocorrências de PredicadorTER + SE e 175 ocorrências de Verbo_{(SEMI-)AUXILIAR} + SE +VP_{TD}) que foram levantados em textos científicos das seguintes áreas: letras, direito e engenharias. O acervo reuniu um total de 36 textos científicos, sendo 18 teses e 18 dissertações, e 600 textos jornalísticos, sendo 300 artigos de opinião e 300 editoriais dos jornais *O Globo* e *Folha de São Paulo*.¹² Os materiais analisados configuram-se como textos escritos representantes da norma de referência do PB.

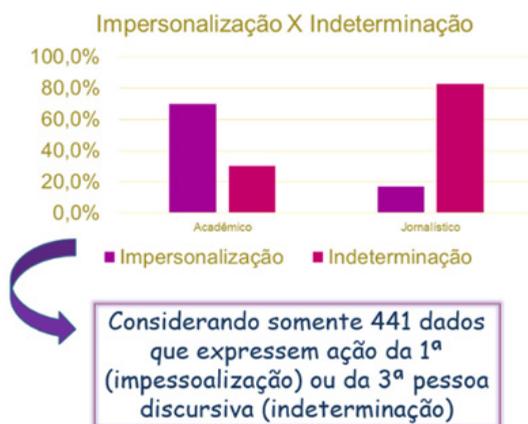
Mais uma vez, foram consideradas as 20 páginas iniciais das teses e dissertações, a partir da introdução e buscou-se um equilíbrio entre o número de palavras dos textos científicos e jornalísticos, pois importa realizar a análise colostrucional (cf. STEFANOWITSCH; GRIES, 2003) para verificar a força de atração de determinados verbos para o preenchimento dos *slots* das construções 1 e 2. Utilizamos a ferramenta lupa para identificar todas as ocorrências de pronome SE e, dessa forma, agilizar o processo de levantamento de dados, já que não lidamos com textos impressos, pois utilizamos as versões digitais dos textos jornalísticos.

Para analisarmos a variação de usos, lidamos, mais uma vez, com o programa estatístico Goldvarb X. Investigamos os seguintes grupos de fatores: (1) grau de desfocalização do P1 suspenso,¹³ considerando o envolvimento das pessoas discursivas; (2) grau de ameaça à face do enunciador; (3) influência dos gêneros textuais nos usos das construções; e (4) ativação da relação de concordância no caso da projeção de SN plural (como P1). Destacamos, a seguir, o resultado obtido em relação ao acionamento da impersonalização discursiva nos textos científicos e jornalísticos:

12 Também foram observadas notícias e reportagens, mas os dados das construções em estudo eram escassos nesses gêneros textuais.

13 Mapeamos um total de 7 graus de desfocalização. Eles são articulados às pessoas discursivas da seguinte forma: grau 1 (envolvimento da 1ª pessoa discursiva), grau 2 (envolvimento da 1ª e da 2ª pessoa discursiva), grau 3 (envolvimento da 1ª pessoa discursiva, de forma mais distanciada), grau 4 (indeterminação genérica, envolvimento da 1ª pessoa do plural), grau 5 (envolvimento da 3ª pessoa discursiva), grau 6 (esvaziamento da referenciação, não há claramente o controle de um P1) e grau 7 (predicação sem participante).

Gráfico 4: Impersonalização X Indeterminação em textos acadêmicos e jornalísticos



Fonte: SARAIVA, 2021.

Como prevíamos, a desfocalização do autor do texto é mais recorrente nas teses e dissertações, enquanto a desfocalização da terceira pessoa é mais acionada nos artigos de opinião e editoriais. Investigamos, dessa forma, a perspectiva do ajuste focal (cf. LANGACKER, 1987, 1991, 2008). A estruturação sintática dos lexemas nas sentenças, ocasionadas pelo acionamento das predicções transitivas diretas com o pronome SE, pode fornecer pistas a respeito da forma como o autor do texto quer que as informações sejam evidenciadas: em quem pratica uma determinada ação ou no que é feito. Assim, em relação à conceituação de evento (construção de uma cena) o sentido dado pelo interlocutor dependerá da maneira como os itens foram dispostos pelo enunciadador a partir dos recursos disponíveis no sistema.

Destacamos, dessa forma, que as construções em estudo são ativadas para assegurar a demissão do participante responsável pela predicação, geralmente com os traços + humano, + controlador e com o objetivo de focalizar no que é feito, com inúmeras funções pragmáticas em jogo, como: atenuar uma opinião da primeira pessoa discursiva ou promover menor comprometimento ao não citar claramente a ação de uma terceira pessoa discursiva que pode ser inferida pelo contexto.

Para mapearmos os usos das construções, recorreremos ao tratamento da variação no âmbito da Gramática de Construções (WIEDEMER; MACHADO VIEIRA, 2018). Esses autores apontam que, no Brasil, ainda contamos com poucos estudos que abordam a variação na perspectiva dos estudos funcional-construcionistas:

No Brasil, a percepção que temos é a de que, às vezes, tal fenômeno é até cogitado e admitido entre observações/generalizações referentes ao que se vem esboçando sobre o rótulo de competição de formas linguísticas (padrões construcionais ou construtos). Isso se dá com base na ideia de que, nesse caso, o

resultado é algum tipo de tendência à substituição de uma forma por outra e, então, ao fortalecimento cognitivo de uma forma em detrimento da outra na mente do falante e, com a convencionalização desse processo, na memória linguística de uma comunidade de fala (WIEDEMER; MACHADO VIEIRA, 2018, p. 42).

Wiedemer; Machado Vieira (2018) também apontam para três tipos de variação: por *allostructions* (aloconstruções) e metaconstruções; por semelhança simbólica e por paradigma/padrão discursivo. Neste estudo, analisamos a variação por aloconstrução em que o falante pode dizer algo de duas ou mais formas (HILPERT, 2014; CAPPELLE, 2006). Assim, acreditamos que as três construções em estudo poderiam ser acionadas em determinados contextos:

(14) No Capítulo 3, **observa-se** o arcabouço teórico utilizado para a análise dos dados. Com o objetivo de segmentar melhor cada base conceitual, esse capítulo subdivide-se em cinco seções, sendo a primeira sobre a Teoria da Variação e Mudança, que funciona como a alavanca para as questões sociolinguísticas que aqui são abordadas. [Dissertação, Letras, UFRJ, 2017]

(15) No gotejador 5, **pode-se observar** que as maiores vazões foram obtidas nos tratamentos com a posição da linha voltada para cima. Houve diferença significativa no tratamento 7 PBG2C3, com vazão de 0,96 L h⁻¹. Não houve diferença significativa dos tratamentos para o gotejador 6 Figura 9. [Tese, Engenharias, USP, 2017]

(16) No primeiro caso, **tem-se a observação** da realidade e, no outro, tem-se a análise da realidade e a sua avaliação possível dentro dos padrões da instituição jornalística. [Tese, Letras, UFRJ, GLND, 2016]

Os exemplos (14), (15) e (16) podem demonstrar que as três construções poderiam ser acionadas: (No Capítulo 3, *pode-se observar/tem-se observado* o arcabouço teórico... / No gotejador 5, *observa-se/tem-se observado* que as maiores vazões... / No primeiro caso, *observa-se/pode-se observar* a realidade...). Entretanto, destacamos que há nuances de diferenças entre os usos das construções, por exemplo, o verbo *poder*, usado como (semi-)auxiliar, garante maior modalização do discurso. Dessa forma, com a variação sendo abordada nos estudos da Gramática de Construções, lida-se com questões discursivo-pragmáticas que podem influenciar nas reconfigurações gramaticais, já que essas se pautam em “regularidades e em estímulos novos experienciados, filtrados ou inferidos no uso situado, o uso da língua afeta o que é armazenado na memória” (MACHADO VIEIRA; WIEDEMER, 2018, p. 6).

Por fim, destacamos que essa pesquisa ainda conta com a análise colostrucional (cf. STEFANOWITSCH, 2013), colexêmica simples (sendo realizada com o apoio do programa RStudio), que permite avaliar a força de atração dos verbos para o preenchimento dos *slots* das construções. Para mais detalhes sobre esse tipo de análise, recomendamos a leitura do capítulo 4 desta obra. Acreditamos, a partir de uma análise prévia do *corpus*, que os verbos de cognição, que refletem processos mentais, como

saber, observar, concluir, sejam mais atraídos para o preenchimento do *slot* verbal na construção Predicador_{TD} + SE e o *slot* de V2 (VP) na construção Verbo_{(SEMI-)AUXILIAR} + SE + VP_{TD}. Podemos observar, nas tabelas a seguir, que essa hipótese foi comprovada:

Tabela 4: Força colostrucional para atração de predicadores verbais para o *slot* da construção Predicador_{TD} + SE

SLOT VERBAL	CORP.FREQ	OBS	EXP	ASSOC	COLL. STR.	LOGL SIGNIF
1. SABER	358	44	0.6	attr	29.476.410	*****
2. OBSERVAR	199	38	0.3	attr	29.031.044	*****
3. PERCEBER	123	23	0.2	attr	17.424.607	*****
4. NOTAR	70	17	0.1	attr	13.859.941	*****
5. ESPERAR	132	19	0.2	attr	13.309.310	*****
6. USAR	213	19	0.4	attr	11.406.042	*****
7. ESTIMAR	50	13	0.1	attr	10.795.712	*****
8. DESTACAR	108	15	0.2	attr	10.386.160	*****
9. CONSIDERAR	530	20	0.9	attr	8.565.124	*****
10. ESTABELECECER	207	14	0.4	attr	7.606.836	*****

Fonte: SARAIVA DE PONTES, 2022.

Tabela 5: Força colostrucional para atração de predicadores verbais para o *slot* 2 da construção Verbo_{TD(SEMI-)AUXILIAR} + SE + VPTD

SLOT 2	CORP.FREQ	OBS	EXP	ASSOC	COLL. STR.	LOGL SIGNIF
1. DIZER	664	20	0.2	attr	14.016.263	*****
2. OBSERVAR	199	8	0.1	attr	6.015.700	*****
3. AFIRMAR	297	8	0.1	attr	5.370.951	*****
4. CONCLUIR	58	5	0.0	attr	4.533.426	*****
5. VER	371	7	0.1	attr	4.199.203	*****
6. IDENTIFICAR	176	5	0.1	attr	3.401.647	*****
7. IGNORAR	33	3	0.0	attr	2.749.858	*****
8. ESTIMAR	50	3	0.0	attr	2.491.994	*****
9. MENOSPREGAR	7	2	0.0	attr	2.333.846	*****
10. CONSIDERAR	530	5	0.2	attr	2.314.616	*****

Fonte: SARAIVA DE PONTES, 2022.

Nas três etapas da pesquisa, o processo de codificação se deu no programa Excel. Para a execução da análise colostrucional, contamos com a ferramenta para anotação de *corpus* do Portulan Clarin¹⁴ que possibilita a identificação do lema das palavras e sua classe gramatical e, com o auxílio da ferramenta lupa, do programa Word, foi possível identificar e contabilizar cada verbo que foi acionado para preencher os *slots* das construções.

8.6 CONSIDERAÇÕES FINAIS

Percebemos, ao longo das três fases da pesquisa, como uma mesma construção de predicação verbal pode ter os dados que licencia capturados e delineados segundo objetos de pesquisa, problemas e olhares de investigação diferentes. Todas as etapas, entretanto, dialogam entre si e fornecem bases para o desenvolvimento da pesquisa sobre predicações transitivas diretas com pronome SE como um todo.

Destacamos o nosso foco de sempre estudar a língua em contextos reais de uso, pois, dessa forma, acreditamos que, com o apoio da análise estatística e de programas computacionais que nos auxiliem a lidar com ela, realizamos análise qualitativa e quantitativa de forma mais coerente. Entretanto, temos a consciência de que sempre estamos diante de tendências que podem revelar usos de determinada comunidade linguística; ou seja, não temos a pretensão de propor afirmações generalizadas sobre, no nosso caso, a língua portuguesa falada no Brasil.

Dessa forma, propomos descrições para as construções de predicações transitivas diretas com pronome SE para além dos enfoques que tradicionalmente são atribuídos a elas: (i) quando pensamos em passivas analíticas sintéticas/pronominais, o enfoque dado nos materiais didáticos está na relação de concordância em caso de projeção de um SN plural e na articulação com a sua “correspondente” voz passiva analítica (cf. SARAIVA, 2018); (ii) as construções com verbo (semi-)auxiliar carecem de descrição nos materiais didáticos e somente em Bechara (1988, p. 185) analisamos uma discussão sobre se considerar que, na construção, há duas possibilidades descritivas em jogo, uma perífrase verbal ou um verbo pleno que projeta uma oração reduzida; e (iii) as construções com verbo *ter* mais pronome SE não são contempladas nos compêndios tradicionais.

Passamos a considerar, então, as funcionalidades pragmáticas, discursivas e semânticas que podem estar envolvidas no acionamento das construções realizado pelo usuário da língua. Intuitivamente, ao perceber que possibilitariam a demissão do participante 1 suspenso, elas, então, são acionadas como ferramentas que podem promover a impersonalização discursiva e são amplamente utilizadas em textos dos domínios discursivos acadêmicos e jornalísticos do PB.

Destacamos que, hoje, contamos com conjuntos de *corpora* que nos fornecem o acesso a textos do PB de forma mais rápida e dinâmica (*Corpus Now*, levantamentos

14 5 A Portulan Clarin é uma Infraestrutura de Investigação para a Ciência e Tecnologia da Linguagem. Disponível em: <https://portulanclarin.net/>.

de dados pelo R Tweeter, entre outros) e reconhecemos que eles são importantíssimos e fundamentais para os estudos linguísticos recentes. Como vimos, a elaboração e disponibilização de banco de dados são tópicos frequentemente discutidos pelos linguistas. Esses acervos permitem o acesso a um número maior de dados e comparações, por exemplo, entre sincronias diferentes.

Para finalizar, entendemos que a seleção de *corpus* (com análises qualitativas de textos) como a que realizamos nas três etapas da pesquisa também tem o seu espaço. No caso do nosso fenômeno em especial, que se demonstrou produtivo em textos escritos acadêmicos e jornalísticos, foi muito pertinente para abrir novos horizontes de pesquisa. Enfatizamos que o olhar atento do pesquisador é indispensável para que seja realizada a análise cuidadosa dos dados e dos resultados obtidos, inclusive dos estatísticos. Esse olhar está presente desde a configuração da amostra de dados e da definição de variáveis. Esse olhar está presente, na definição de variáveis e fatores destas com base nas quais se categorizam dados para proceder a uma análise estatística/quantitativa. Podemos ter uma amostra muito interessante, com muitas rodadas em programas estatísticos, mas nada disso substitui o olhar atento do cientista para encontrar configurações de dados, problemas, respostas, interpretar os percentuais e pesos relativos e propor soluções para os problemas que possam vir a aparecer, bem como traçar generalizações e representações consistentes e ancoradas no que foi observado.

REFERÊNCIAS

- ASSUNÇÃO, C.; ARAÚJO, C. Linguística de corpus: teoria, perspectivas metodológicas e ensino das línguas. In.: *Filologia Linguística Portuguesa*, São Paulo, v. 21, n. 2, jul./dez. 2019. Disponível em: <https://www.revistas.usp.br/flp/article/view/163496>. Acesso em: 07 mar. 2022.
- BAGNO, M. *Português ou brasileiro? Um convite à pesquisa*. São Paulo: Parábola Editorial, 2001.
- BECHARA, E. *Lições de Português pela análise sintática*. Rio de Janeiro: Padrão, 1988.
- BERLINCK, R. A.; BIAZOLLI, C. C. Ferramentas metodológicas para análises (sócio) linguísticas. In.: *Estudos linguísticos*, 2018, p. 260-273. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/2047>. Acesso em: 20 fev. 2022.
- BLOOMFIELD, L. *Language*. London: Allen & Unwin, 1935.
- BYBEE, J. Usage-based theory and exemplar representations of constructions. In.: HOFFMANN, T.; TROUSDALE, G. *The oxford handbook of construction grammar*. Oxford, UK: Oxford University Press, 2013.
- BYBEE, J. *Language, usage and cognition*. New York: Cambridge University Press, 2010.
- BYBEE, J. Mechanisms of change in grammaticization: the role of frequency. In.: BRI-

- AN, J.; RICHARD, J. D. (eds). *The Handbook of Historical Linguistics*. Blackwell, 2003.
- CALLOU, D.; AVELAR, J. Gramática e variação no Português brasileiro: considerações sobre ter~haver e de~em. In: COUTINHO, M.A. (org.); LOBO M. Textos Seleccionados- XXII Encontro da Associação Portuguesa de Linguística. 1 ed. Lisboa: Colibri, 2007, p. 183-197. Disponível em: <https://javelarnet.files.wordpress.com/2017/08/avelar-e-callou-2007a.pdf>. Acesso em: 21 fev. 2022.
- DUARTE, M. E. L. Termos da oração. In: VIEIRA, S. R.; BRANDÃO, S. F. *Ensino de gramática: descrição e uso*. São Paulo: Contexto, 2008.
- DUARTE, M. E. L. Sujeitos de referência definida e arbitrária: aspectos conservadores e inovadores na escrita padrão. In.: *Revista Linguística*, vol. 3, no 1, Rio de Janeiro: UFRJ, 2007. Disponível em: <https://revistas.ufrj.br/index.php/rl/article/view/4396>. Acesso em: 21 fev. 2022.
- FILLMORE, C. J.; KAY, P.; O'CONNOR, M. C. Regularity and idiomaticity in grammatical constructions: the case of let alone. In.: *Language*, 64, 3, 1988.
- GOLDBERG, A. E. *Constructions: a construction grammar approach to argument structure*. Chicago and London: The University Chicago Press, 1995.
- GOLDBERG, A. E. *Constructions at work*. Oxford: Oxford University Press, 2006.
- GÖRSKI, E. *et al.* Fenômenos discursivos: resultados de análises variacionistas como indícios de gramaticalização. In: RONCARATI, C.; ABRAÇADO, J. (orgs.). *Português brasileiro – contato linguístico, heterogeneidade e história*. Rio de Janeiro: 7 Letras/FAPERJ, 2003.
- HILPERT, M. Language variation and change. In: HILPERT, M. *Construction Grammar and its application to English*. Edinburgh, Edinburgh University Press, 2014.
- HOOPER, P. J. On some principles of grammaticalization. In.: TRAUGOTT, E. C.; HEINE, B. (eds). *Approaches to Grammaticalization*, Vol. I. Amsterdam: John Benjamins, 1991.
- LABOV, W. *Principles of linguistic change: internal factors*. Oxford: Wiley-Blackwell, 1994.
- LANGACKER, R. W. *Cognitive Grammar: a basic introduction*. New York: Oxford University Press, 2008.
- LANGACKER, R. W. *Foundations of cognitive grammar: Descriptive applications*. Stanford: Stanford University Press, 1991.
- LANGACKER, R. W. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press, 1987.
- LEVSHINA, N. *How to do Linguistics with R: Data exploration and statistical analysis*, Amsterdam / Philadelphia: John Benjamins Publishing Company, 2015.

- MACHADO VIEIRA, M. dos S.; BARBOSA, J. B.; FREITAG, R. M. K.; BORGES, M. M.; MEDEIROS, A. L. S. Collections of data open to society: linguistic and sociocultural memory and potential for (re)use. *Cadernos de Linguística*, v. 2, n. 1, p. e 607, 24 Jan. 2022. Disponível em: <https://cadernos.abralin.org/index.php/cadernos/article/view/607>. Acesso em: 07 mar. 2022.
- MACHADO VIEIRA, M. dos S. Expressões impessoais no discurso acadêmico brasileiro. In.: *Revista Letrônica*. Porto Alegre, v. 10, n. 1, jan-jun 2017. Disponível em: <https://revistaseletronicas.pucrs.br/ojs/index.php/letronica/article/view/25061>. Acesso em: 05 fev. 2022.
- MACHADO VIEIRA, M. dos S. Um panorama da norma de flexão verbal de número em construções com se apassivador/indeterminador. *Cuadernos de la ALFAL*, n. 7, março 2015, p. 210-230. Disponível em: <https://www.researchgate.net/publication/357063582> UM PANORAMA DA NORMA DE FLEXAO VERBAL DE NUMERO EM CONSTRUCOES COM SE APASSIVADORINDETERMINADOR AN OVERVIEW OF NUMBER VERBAL INFLECTION NORM IN CONSTRUCTIONS WITH PASSIVEINDETERMINACY SE MARKER. Acesso em: 03 jan. 2022.
- MACHADO VIEIRA, M. dos S.; SARAIVA, E. S. A concordância de número em estruturas passivas pronominais no Português Brasileiro. In: *XVI Congreso Internacional de La ALFAL* (Asociación de Linguística y Filología de América Latina), 2011, Alcalá de Henares. Documentos para el XVI Congreso Internacional de La ALFAL (Asociación de Linguística y Filología de América Latina). Alcalá de Henares: ALFAL, 2011. Disponível em: <https://www.researchgate.net/publication/357063678> A concordancia de numero em estruturas passivas pronominais no portugues brasileiro. Acesso em: 03 jan. 2022.
- OLIVEIRA, L. P. Linguística de *corpus*: teoria, interfaces e aplicação. In.: *Matraga*, Rio de Janeiro, v.16, n.24, jan./jun. 2009.
- PAIVA, M. C. Variação e especificidades funcionais no domínio da causalidade. *Revista de Estudos da Linguagem*, v.7, n.2, 1998.
- SANKOFF, D.; TAGLIAMONTE, S. A.; SMITH, E. *Goldvarb X*: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto, 2005.
- SARAIVA DE PONTES, E. S. *Predicação transitiva direta com pronome SE*: perfis de impersonalização discursiva em variação. Tese (Doutorado em Língua Portuguesa) do Programa de Pós-Graduação em Letras Vernáculas, Faculdade de Letras, UFRJ, Rio de Janeiro, 2022.
- SARAIVA, E. S. Análise-se os dados? Quem é o responsável pela análise? Apresentação de pôster no I Colóquio Internacional Variar, 2021.
- SARAIVA, E. S. Ensino de construções passivas analíticas e sintéticas em materiais didáticos e paradidáticos. In.: *Revista Vocábulo*, v. XIV, 2018. Disponí-

- vel em: https://api3.baraodemaua.br/media/16425/2019-07-22-07-28-56_788422182dbdce81dae8ca001a2d1af2098bc32f.pdf. Acesso em: 03 fev. 2022.
- SARAIVA, E. S. *A construção TEM-SE no português brasileiro escrito: uma análise sociofuncionalista*. Dissertação (Mestrado em Língua Portuguesa) do Programa de Pós-Graduação em Letras Vernáculas, Faculdade de Letras, UFRJ, Rio de Janeiro, 2013.
- SARDINHA, T. B. *et al.* Portal multimodal/multilíngue para o avanço da Ciência Aberta nas Humanidades. In.: *Cadernos de Linguística*, v. 2, p. e406, 2021. Disponível em: <https://cadernos.abralin.org/index.php/cadernos/article/view/406>. Acesso em: 03 fev. 2022,
- SARDINHA, T. B. Linguística de corpus: histórico e problemática. In.: *Delta: documentação de estudos em linguística teórica e aplicada*, São Paulo, v. 16, n. 2, 2000.
- SCHERRE, M. M. P. *DOA-SE lindos filhotes de poodle – variação linguística, mídia e preconceito*. São Paulo: Parábola Editorial, 2005.
- STEFANOWITSCH, A. Collostructional analysis. In: HOFFMAN, T.; TROUSDALE, G. *The Oxford Handbook of Construction Grammar*, 2013.
- STEFANOWITSCH, A.; GRIES, S. Collostructions: Investigating the interaction between words and constructions. In.: *International Journal of Corpus Linguistics*, v. 8, n. 2, 2003.
- TAVARES, M. A. *A gramaticalização de e, aí, daí e então: estratificação/variação e mudança no domínio funcional da sequenciação retroativo-propulsora de informações um estudo sociofuncionalista*. Tese de doutorado, Florianópolis: UFSC, 2003.
- TRAUGOTT, E. C.; TROUSDALE, G. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press. 2013.
- WEINREICH, U. W. LABOV; M. HERZOG. Empirical foundations for a theory of language change. In: W. Lehmann; Y. Malkiel (eds.). *Directions for historical linguistics*. Austin: University of Texas Press, 1968.
- WIEDEMER, M. L; MACHADO VIEIRA, M. dos S. Lexemas e construção: atração, coerção e variação. In: *Caderno Seminal Digital*. Rio de Janeiro, v. 30, 2018. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/cadernoseminal/article/view/34009>. Acesso em: 03 fev. 2022.