

Marcia dos Santos Machado Vieira
organizadora

PREDICAR

Uma rede de perspectivas metodológicas



Blucher Open Access



PREDICAR: UMA REDE DE PERSPECTIVAS
METODOLÓGICAS

CONSELHO EDITORIAL

André Costa e Silva

Cecilia Consolo

Dijon de Moraes

Jarbas Vargas Nascimento

Luis Barbosa Cortez

Marco Aurélio Cremasco

Rogério Lerner

Marcia dos Santos Machado Vieira

PREDICAR: UMA REDE DE PERSPECTIVAS METODOLÓGICAS



2022

Predicar: uma rede de perspectivas metodológicas

© 2022 Marcia dos Santos Machado Vieira

Editora Edgard Blücher Ltda.

Publisher Edgard Blücher

Editor Eduardo Blücher

Coordenação editorial Jonas Eliakim

Produção editorial Aline Fernandes

Diagramação Joyce Rosa

Revisão de texto Samira Panini

Capa Laércio Flenic

Imagem da capa iStockphoto

Blucher

Rua Pedroso Alvarenga, 1245, 4º andar

04531-934 – São Paulo – SP – Brasil

Tel.: 55 11 3078-5366

contato@blucher.com.br

www.blucher.com.br

Segundo o Novo Acordo Ortográfico, conforme 5. ed. do *Vocabulário Ortográfico da Língua Portuguesa*, Academia Brasileira de Letras, março de 2009.

É proibida a reprodução total ou parcial por quaisquer meios sem autorização escrita da editora.

Todos os direitos reservados pela Editora Edgard Blücher Ltda.

Dados Internacionais de Catalogação na Publicação (CIP) Angélica Ilacqua CRB-8/7057

Predicar : uma rede de perspectivas metodológicas / organizado por Marcia dos Santos Machado Vieira. - São Paulo : Blucher, 2022.
296 p.

Bibliografia

ISBN 978-65-5550-265-7 (impresso)

ISBN 978-65-5550-262-6 (eletrônico)

1. Línguas clássicas e vernáculas 2. Língua portuguesa I. Machado Vieira, Marcia dos Santos

22-1710

CDD 469.07

Índices para catálogo sistemático:

1. Línguas clássicas e vernáculas

O presente livro foi produzido com apoio financeiro do Programa de Pós-Graduação em Letras Vernáculas – UFRJ e do Programa de Excelência Acadêmica (PROEX) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES).



APOIO NA DIVULGAÇÃO CIENTÍFICA

Programa de Pós-Graduação em Letras Vernáculas, Universidade Federal do Rio de Janeiro (UFRJ – Brasil)

<http://www.posvernaculas.letras.ufrj.br/pt/>

Projeto Predicar – Formação e expressão de predicados complexos e predicções: estabilidade, variação e mudança construcional

<https://projeto-predicar.wixsite.com/predicar>

(UFRJ – Brasil)

Projeto VariaR – Variação em línguas românicas

<https://variatar.wixsite.com/variatar>

(UPVM e UFRJ – França e Brasil)

Design e Diagramação

Jeane Nunes da Penha

Marcia dos Santos Machado Vieira

(Universidade Federal do Rio de Janeiro – UFRJ, Brasil)

Equipe de revisão

Marcia dos Santos Machado Vieira

Eneile Santos Saraiva de Pontes

Fábio Rodrigo Gomes da Costa

Gabriel Lucas Martins

Jeane Nunes da Penha

Laís Lima de Souza

Mariana Gonçalves da Costa

Natércia Almeida Lacerda

Pâmela Fagundes Travassos

Ravena Beatriz de Sousa Teixeira

(Universidade Federal do Rio de Janeiro – UFRJ, Brasil)



COMITÊ CIENTÍFICO AVALIADOR

Alexandre Ferreira Martins (Leitor brasileiro (MRE/CAPES) na *Hankuk University of Foreign Studies* (Coreia do Sul), Mestre pela *Université Paul-Valéry/Montpellier 3* (France))

Carla Valeria de Souza Faria (Docente da *Università Ca' Foscari Venezia* (Itália), Doutora pela Universidade Federal do Rio de Janeiro)

Habiba Naciri (Docente da Faculdade de Letras e Ciências Humanas, Universidade Mohammed V Rabat – Marrocos, Doutora pela Faculdade de Letras e Ciências Humanas, Universidade Mohammed V Rabat – Marrocos)

Liliane Santos (Docente da *Université de Lille* (France), Doutora pela *Université de Nancy 2* (France))

Marcos Luiz Wiedemer (Docente da Universidade do Estado do Rio de Janeiro – Faculdade de Formação de Professores de São Gonçalo, Doutor pela Universidade Estadual Paulista, Procientista/UERJ)

Maria João Marçalo (Docente da Universidade de Évora (Portugal), Doutora pela Universidade de Évora)

Mariana Meireles de Oliveira Silva (Docente (Substituta) de Português língua estrangeira do *Collège Romain Rolland - Sartrouville* (France), Mestre pela *Université Sorbonne Nouvelle/Paris 3*, Mestranda pela *Université Paris Nanterre*)

Vanessa Meireles (Docente da *Université Paul-Valéry – Montpellier 3* (France), Doutora pela *Université Paris 8 Vincennes – Saint-Denis* (France))

Vinicius Maciel de Oliveira (Docente da Universidade do Estado do Rio de Janeiro – Faculdade de Formação de Professores de São Gonçalo, Doutor em Letras Vernáculas pela Universidade Federal do Rio de Janeiro)

CONTEÚDO

APRESENTAÇÃO	15
Aprender e fazer em rede: estabilidade, variação e mudança	15
PRESENTATION	23
Learning and doing on a network: stability, variation and change	23
1. CODAZO EN PLENA CARA Y PUÑETAZO EN TODA LA BOCA: LA FAMILIA DE CONSTRUCCIONES SOMÁTICAS DE GOLPE	31
Abstract	31
1.1 Introducción	32
1.2 Descripción de la construcción [V _[impacto] /V (con) N en PLENO N _[parte del cuerpo]]	33
1.3 Más allá de [en PLENO N _[parte del cuerpo]]: la construcción [V _[impacto] /V (con) N en TODO DET N _[parte del cuerpo]]	43
1.4 Conclusiones	50
Bibliografía	51
2. TAKING A LOOK AT THE SUPPORT VERB CONSTRUCTION V A LOOK: A DEMONSTRATION OF METHODOLOGY	57

2.1 Introduction: how we express visual perception	57
2.2 <i>Take a look</i> and <i>have a look</i> in some dictionaries of English	59
2.3 Getting a grip of real data for <i>look</i> with English-Corpora.org	63
2.4 Implications for the organization of lexical and grammatical knowledge	85
2.5 Ideas for further discussion and practice activities	88
References	89

3. COLETA E TRATAMENTO DE DADOS NO CORPUS DO PORTUGUÊS 93

3.1 Introdução	93
3.2 Quais os aspectos positivos e negativos dos bancos de dados online?	94
3.3 <i>Corpus</i> do Português: conhecendo os <i>corpora</i>	95
3.4 A plataforma de busca e o processo de coleta	97
3.5 Considerações finais	122
Referências	122

4. OS PRIMEIROS PASSOS EM ANÁLISE COLOSTRUCIONAL: PESQUISAS DE PREDICAÇÃO PARA ILUSTRAR UM FAZER 123

4.1 Introdução	124
4.2 Pilares e tipos de análise colostrucional	125
4.3 Análise colostrucional de colexemas simples	139
4.4 Análise colostrucional de colexemas distintivos	142
4.5 Análise colostrucional de colexemas covariantes	145
4.6 Trabalho com softwares de funções de planilha para uma análise baseada no sistema	147
4.7 Notas sobre interpretação e descrição de resultados	150
4.8 Análise comparativa por subamostras organizadas textual-discursivamente	151
4.9 Linha executiva de investigação de colocações	156
4.10 Considerações finais	157
Referências	158

5. A PESQUISA EXPERIMENTAL NA ANÁLISE DE CONSTRUÇÕES COM VERBO SUPORTE	161
5.1 Introdução	161
5.2 Elaboração de testes de percepção e avaliação subjetiva: etapas e cuidados a serem tomados	163
5.3 Análise crítica dos resultados	167
5.4 Pesquisa experimental em plataformas digitais	171
5.5 Considerações finais	172
Referências	173
6. PROPOSTA DE TESTES EXPERIMENTAIS PARA ANÁLISE DA VARIAÇÃO CONSTRUCIONAL DE PREDICADOR COMPLEXO	175
Referências	190
7. DISCURSO DE PARLAMENTARES: UM PERFIL METODOLÓGICO DE ANÁLISE DE <i>CORPUS</i> DE DOMÍNIO POLÍTICO	193
7.1 Introdução	193
7.2 Discurso político: algumas reflexões	194
7.3 A constituição do <i>corpus</i> no discurso político	195
7.4 Amostra de pesquisa: análise de <i>corpus político</i>	197
7.5 Considerações finais	205
Referências	205
8. METODOLOGIAS E ANÁLISES EM ESTUDOS DE PREDICAÇÃO TRANSITIVA DIRETA COM PRONOME <i>SE</i>: UM ESTUDO PUXA OUTRO	207
8.1 Introdução	207
8.2 A linguística de <i>corpus</i> e o uso de estatística	209
8.3 As construções de predicação com pronome <i>SE</i> e <i>SN</i> plural: a (não) ativação da relação de concordância.	212
8.4 A construção <i>tem-se</i> no português brasileiro: estaria em variação com a forma verbal <i>há</i> ?	217
8.5 A variação de usos de predicação transitiva direta com pronome <i>SE</i>	225

8.6 Considerações finais	230
Referências	231

9. ANÁLISE COMPARATIVA PARA DETECTAR AUXILIARIDADE VERBAL E PREDICADORES COMPLEXOS **235**

9.1 Introdução	235
9.2 O tratamento da auxiliaridade	239
9.3 Para o tratamento da auxiliaridade	245
9.4 Para a análise comparativa dos (sub)tipos de verbos gramaticais e de construções de predicadores complexos (verbais ou verbo-nominais)	254
9.5 Considerações finais	281
Referências	283

POSFÁCIO **285**

Il progetto Predicar e lo studio delle costruzioni	285
--	-----

POSFÁCIO **287**

La nécessaire articulation entre une certaine manière de faire de la recherche sur les prédicats et les prédications et l'enseignement de la langue	287
---	-----

AS AUTORAS E OS AUTORES **291**

APRESENTAÇÃO

APRENDER E FAZER EM REDE: ESTABILIDADE, VARIAÇÃO E MUDANÇA

Marcia dos Santos Machado Vieira (Universidade Federal do Rio de Janeiro, CNPq/Faperj)

Este livro é uma homenagem aos vinte anos de desenvolvimento de pesquisas no âmbito do Projeto Predicar, coordenado, desde sua inauguração em 2002, por mim, Marcia dos Santos Machado Vieira, no Programa de Pós-Graduação em Letras Vernáculas na Universidade Federal do Rio de Janeiro. É fruto de ricas (inter)ações mobilizadas dentro do Projeto Predicar, a partir dele ou além dele.

Em 20 anos de atuação, o Projeto Predicar tem trilhado variados caminhos, sempre contando com a participação e a interlocução de estudantes (de iniciação científica, pesquisa de conclusão de curso de graduação, mestrado, doutorado). Tem contribuído para a formação de muitos, como docentes-pesquisadores, no que diz respeito a predicação, predicator complexo, multifuncionalidade verbal e temas correlatos ligados a processos de estabilização, variação e mudança. Tem colaborado para a descrição dos fenômenos de auxiliaridade, impersonalização, intensificação, atenuação, representação, passividade, idiomatismo, temporalidade, aspectualidade, modalidade. E, para tanto, tem articulado diferentes perfis de pesquisa empírica: observacional, de dados em *corpora* (orais e escritos, sincrônicos ou diacrônicos, de variedades do Português dentro e fora do Brasil); ou experimental, de dados coletados na experi-

ência de uso ou manipulados, mas sempre tomando por base essa experiência capturada em *corpora*. Geralmente, desenvolve-se com base em análise comparativa (inclusive, via recorte longitudinal) e análise multivariada e, mais recentemente, também com base em análise colostrucional. E lida com tratamento qualitativo e quantitativo de dados ou de registros de experimentos do tipo *offline*.

Atualmente, o estudo de coocorrências, de colocações de unidades construcionais, de análise colostrucional, como se verá num dos capítulos desta obra, tem mobilizado a equipe. Sob o título *Formação e expressão de predicados complexos e predicções: estabilidade, variação e mudança construcional*, explora-se, sob uma ótica socioconstrucionista (MACHADO VIEIRA; WIEDEMER, 2019, 2020),¹ a potencialidade dos processos de variação construcional, de mudança (mudança construcional ou construcionalização gramatical e lexical, TRAUGOTT; TROUSDALE, 2013)² e de (relativa) estabilização que capturamos ao representar a rede de construções com base nas inclinações detectadas por meio de análise de frequências de tipos construcionais e de ocorrências e/ou por meio de análise colostrucional. Entram em jogo (i) a tensão entre processos e forças de unidade e variabilidade, (ii) a relação entre repetição e convencionalização de um lado e inovação, *mismatch* e criatividade de outro, (iii) a triangulação entre iconicidade, indexicalidade e arbitrariedade, e (iv) a ligação entre identidade normalizada ou padronizada, identidade multifacetada (de indivíduo e de comunidade) em meio à diversificada e densa natureza de redes sociais que os falantes de uma língua integram ou constroem. O Projeto Predicar iniciou-se sob o título *Formação e expressão de predicados complexos: polifuncionalidade verbal*. Nessa ocasião, os esforços voltaram-se para o estudo de unidades verbais em contexto, para a compreensão de sua polifuncionalidade na experiência do uso, das propriedades envolvidas na categorização de verbos, considerando-se um continuum, com interseções, de membros mais ou menos exemplares das categorias percebidas na análise linguístico-textual e membros periféricos, bem como para a detecção dos predicadores complexos formados a partir de verbos. E, em alguma medida, diferentes tônicas relativas a fases ou ondas de investigações do Projeto estão contempladas neste livro.

Este livro reúne textos sobre metodologia para investigação de idiomatismos e fra-seologia, estudo de variação construcional, pesquisa experimental, pesquisa empírica de *corpora*, análise colostrucional, análise multivariada e análise comparativa. A questão central que vai mobilizar os autores dos capítulos aqui reunidos é a metodologia empregada para pesquisar língua(s), concebida numa perspectiva em que im-

1 MACHADO VIEIRA, Marcia dos Santos; WIEDEMER, Marcos Luiz. Sociolinguística Variacionista e Gramática de Construções: Variationist Sociolinguistics and Construction Grammar: os desafios e as perspectivas de compatibilização. In: *Dimensões e Experiências em Sociolinguística*. São Paulo: Blucher, 2019, p. 85-120.

MACHADO VIEIRA, Marcia dos Santos; WIEDEMER, Marcos Luiz. A variação no modelo construcionista da Linguística Funcional-Cognitiva. In: BRESCANCINI, Cláudia Regina; MONARETTO, Valéria. Neto de Oliveira (orgs.). *Sociolinguística no Brasil: textos selecionados*. Porto Alegre: Editora da PUCRS, 2020, p. 265-304.

2 TRAUGOTT, E. C.; TROUSDALE, G. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press, 2013.

porta relacionar atributos de forma e função/significado e interessa pensar no potencial da ciência linguística a que se chega com diferentes perspectivas metodológicas para o espaço de sensibilização e formação de pesquisadores, de ensino sobre processos de pesquisa e de tratamento gramatical no ensino de língua (materna ou não materna). O processo de estudo científico de gramática de uma língua com base em variadas vias de análise que possam ser postas em prática precisa ganhar relevo também em sala de aula, para além dos laboratórios da universidade. E assim, pode-se colaborar para fazer emergir o espírito investigativo, bem como o interesse por diferentes abordagens e processos para detectar e mapear a dinamicidade e a heterogeneidade linguísticas.

O primeiro capítulo deste livro foi escrito por Pedro Ivorra Ordines (Universidade de Santiago de Compostela, *Universitat Pompeu Fabra*). Em Codazo em plena cara y puñetazo en toda la boca: *La familia de construcciones SOMÁTICAS DE GOLPE*, o autor trata de construções morfossintática e semântico-pragmaticamente semelhantes que podem ser descritas como uma família de construções no âmbito da Gramática de Construções. Lida com expressões idiomáticas de construção em que uma preposição e um adjetivo são especificados lexicalmente, enquanto três *slots* são lexicalmente abertos, com restrições semânticas. O objetivo é descrever essas construções de comportamento semelhante em uma rede bidimensional. Utilizando o *corpus* esTenTen18 (*Sketch Engine*), o autor realiza uma análise das restrições semânticas desses *slots*, atentando para produtividade por meio de frequência de tipo e hápax legomena. E assim, o autor põe em evidência outro assunto importante em Gramática de Construções, que, em alguma medida, constitui a tônica a mobilizar linguistas a essa concepção de língua em que gramática e léxico estão integrados, mas que, também, nem sempre tem tido a atenção que seria esperada por conta disso.

O segundo capítulo, escrito por Bert Cappelle (*Universidade de Lille*) e Pâmela Fagundes Travassos (Universidade Federal do Rio de Janeiro) ilustra uma tônica no Projeto Predicar: a interação entre docentes (neste caso, coordenador) e discentes (neste caso, doutoranda), a atenção ao que é *mainstream* na área de Letras e Linguística (sem perder de vista o que antecede e leva a isso), o potencial interrelacional de diferentes instituições em prol da descrição de línguas. Sob o título *Taking a look at the support verb construction V a look: a demonstration of methodology*, os autores apresentam o passo a passo sobre como conduzir uma pesquisa de variação construcional, ilustrada aqui com base numa que desenvolveram a partir do diálogo e da orientação que vêm mantendo desde que Pâmela Fagundes Travassos fez seu doutorado sanduíche na Universidade de Lille.

O terceiro capítulo, escrito por Ravena Beatriz de Sousa Teixeira e Jeane Nunes da Penha (ambas da Universidade Federal do Rio de Janeiro), perspectiva a pesquisa linguística centrada em *corpora* anotados, focalizando os do *Corpus* do Português, de Mark Davies. Tal perspectiva é muito cara a áreas, como Sociolinguística, Linguística de *Corpus*, Gramática de Construções Baseada no Uso, enfim, ao que se refere como Linguística Baseada no Uso. No texto *Construções com verbo suporte: coleta e trata-*

mento de dados no *Corpus do Português*, as autoras mostram formas e recursos para explorar essa plataforma, tratam de aspectos positivos e negativos de bancos de dados online e de características da constituição de amostras e mostram um passo a passo de trabalho com *corpora* anotados.

Os primeiros passos em análise colostrucional: pesquisas de predicação para ilustrar um fazer é um texto escrito a muitas mãos porque, no âmbito do Projeto Predicar, revela um fazer que se vem edificando a partir de leituras, vozes e vivências diversas. Foi escrito por Marcia dos Santos Machado Vieira, Mariana Gonçalves da Costa, Ravena Beatriz de Sousa Teixeira, Pâmela Fagundes Travassos, Pedro Poppolino e Lais Lima de Souza (todos da Universidade Federal do Rio de Janeiro). É, antes de tudo, uma homenagem à equipe que, guiada pela meta de refletir e estudar em parceria questões de metodologia de análise de coocorrências linguísticas, se aprimora e se fortalece cooperativamente para desenvolver a chamada análise colostrucional, sem perder de vista os limites e as implicações desta, sem perder de vista sua articulação a outras metodologias que permanecem a orientar caminhos do grupo de pesquisa. Nesse capítulo, os autores exploram três tipos de procedimentos de análise colostrucional via linguagem R e aplicativo RStudio, bem como passam brevemente pela modelagem de análise multivariada e pela modelagem de análise da relação entre unidades construcionais a construções textuais-discursivas, a padrões discursivos, via Excel.

Os dois capítulos seguintes revelam justamente outra metodologia de trabalho implementada no Projeto Predicar no início do século XXI, quando nem de perto tinha a visibilidade que hoje tem em estudos brasileiros: a metodologia experimental. Desde Machado Vieira e Esteves (2008), muitas foram as pesquisas centradas em avaliação subjetiva de usos de predicações e predicadores no Português (avaliação inspirada no problema de avaliação subjetiva em WEINREICH; LABOV; HERZOG, 1968). Então, uma amostra dessa perspectiva metodológica também está nesta obra, via os dois capítulos que passam a ser resumidos.

Jeane Nunes da Penha, Ravena Beatriz de Sousa Teixeira e Pâmela Fagundes Travassos (todas da Universidade Federal do Rio de Janeiro) mostram, em *Pesquisa experimental na análise de construções com verbo suporte*, as principais orientações para o tratamento experimental de dados do uso com base na experiência de suas pesquisas sobre construções com verbos suportes.

Pâmela Fagundes Travassos (Universidade Federal do Rio de Janeiro, professora da rede municipal do Rio de Janeiro), em *Proposta de testes experimentais para análise da variação construcional de predicador complexo*, reúne e ilustra uma parcela dos materiais de sua proposta de testes para pesquisa experimental apresentada por ocasião de sua candidatura a doutorado sanduíche na Universidade de Lille. Nesse capítulo, a autora mostra modelos de testes experimentais para análise da variação entre predicadores complexos com verbo suporte DAR: *dar um empurrãozinho*, *dar uma forcinha*, por exemplo. E, assim, lida com o problema da avaliação subjetiva da possi-

bilidade de equivalência e intercambialidade de predicadores como esses por formas verbais simples, com atitudes de usuários quanto a esse tipo de pregação.

Fábio Rodrigo Gomes da Costa (Universidade Federal do Rio de Janeiro, professor da SEEDUC), em *Discurso de parlamentares: um perfil metodológico de análise de corpus de domínio político*, trata do percurso de uma análise multivariada de dados de pregação com verbo *trazer*, em que este pode figurar ou como verbo predicator/pleno ou como verbo suporte e pode, neste caso, integrar um predicator complexo de perfil não composicional mais ou menos evidente. Ao fazer isso, ele aborda grupos de fatores e hipóteses que são interligados para ilustrar um encaminhamento desse tipo de análise. Também aborda aspectos da constituição do discurso de domínio político em que a pregação de atenuação discursiva tem proeminência.

Em *Metodologias e análises em estudos de pregação transitiva direta com pronome SE: um estudo puxa outro*, Eneile Santos Saraiva de Pontes e Marcia dos Santos Machado Vieira (ambas da Universidade Federal do Rio de Janeiro) mostram o passo a passo envolvido em frentes de pesquisa relativa a dados de pregação transitiva direta com pronome SE (indicativo de passiva sintética, de supressão de participante indutor e, por consequência, de indeterminação do responsável ou do causador do estado de coisas). Na verdade, mostram como elas se interligam, dado o fato de que um estudo acabou por desdobrar-se em outros. E, então, põem em evidência o desenrolar muito frequente de (inter)ações e repercussões em pesquisas científicas que são feitas no âmbito de um projeto maior.

O décimo capítulo, escrito por Marcia dos Santos Machado Vieira (Universidade Federal do Rio de Janeiro, CNPq/Faperj), centra-se em critérios para a identificação de predicadores complexos e de subtipos verbais no que diz respeito especialmente ao caráter de auxiliaridade e é construído perspectivando espaços de tratamento dessa temática no ensino de Português, língua materna e não materna. Sob o título *Análise comparativa para detectar auxiliaridade verbal e predicadores complexos*, são apresentados critérios relativos à associação de usos verbais às categorias de verbo (semi-)auxiliar, (semi-)suporte, verbo relacional, e, por conseguinte, à detecção de predicadores complexos. A partir de diferentes critérios e alguns usos verbais, bem como de pesquisas desenvolvidas no Projeto Predicar, a associação de verbos a categorias e graus de auxiliaridade e a certos *slots* em predicadores complexos é a temática que recebe atenção.

Este livro conta, ainda, com textos, posfácios, que traduzem olhares, fora do Brasil, sobre o valor de contribuições descritivas respaldadas em ciência linguística (aquí, algumas das que são frutos e têm lugar nas (inter)ações do Projeto Predicar ou a partir deste) ao ensino de Português.

O posfácio em italiano foi escrito por Carla Valeria de Souza Faria (Universidade Ca'Foscari Venezia), docente-pesquisadora com quem a interação é viabilizada especialmente pela articulação do Projeto Predicar ao Projeto VariaR, sendo ela colega-membro deste.

O posfácio em francês foi escrito por Liliane Santos (Universidade de Lille), com quem a equipe do Projeto Predicar pôde dialogar durante o I Congresso Internacional do Programa de Pós-Graduação em Letras Vernáculas, *Vozes e escritas nos diferentes espaços da língua portuguesa*, em razão do seminário em que Liliane Santos foi uma das coordenadoras. Um resultado desse espaço de interação é a entrevista Santos et al. (2021).³ No posfácio desta obra, a docente-pesquisadora trata da interligação entre caminhos de pesquisa linguística e caminhos de ensino de língua materna e não materna, inclusive para a percepção da natureza pluricêntrica do Português.

As valiosas contribuições quanto a metodologias da pesquisa linguística e quanto a olhares sobre essa rede de perspectivas que aqui estão reunidas dão uma mostra do potencial de rumos na área e de interlocuções além dela. A todos os que colaboraram para o resultado que se materializa na forma deste livro, a manifestação de muitos agradecimentos não traduz o conforto e a alegria de poder contar com os pontos de vista e as expertises que este livro mobilizou. Certamente essa forma não capta a riqueza dos diálogos e das experiências de pesquisa e interlocução que nos (re)elaboram e nos instigam a aprender e caminhar sempre, mas mostram certamente um pouco do perfil multimetodológico e interrelacional que caracteriza fazeres no Predicar desde sempre.

A versão final de cada capítulo é de responsabilidade intelectual de seu(s) autor(es), que teve/tiveram total liberdade para definir o perfil de abordagem temática e textual. Naturalmente, a contribuição de colegas especialistas da área de Letras e Linguística tornou mais rico o processo de (inter)ação que um livro requer. Sou grata a cada um do comitê científico que colaborou, com seu tempo e seu know-how, para a melhor consecução deste projeto de partilha.

Sou especialmente grata ao Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro e à Capes/Proex (Coordenação de Pessoal de Pós-Graduação e Pesquisa/Programa de Excelência Acadêmica) pelo apoio financeiro, que permitiu e incentivou a concretização de um trabalho de interlocução nacional e internacional, como o envolvido aqui. Agradeço também a cada um dos envolvidos na revisão, diagramação e divulgação do livro. Agradeço a Gabriel Lucas Martins (Universidade Federal do Rio de Janeiro) a colaboração na versão desta apresentação em inglês.

Torço para que essa rede de perspectivas metodológicas possa ajudar a comunidade de discentes e docentes pesquisadores e que, em alguma medida, possa ser bem aproveitada em espaços que deem atenção à formação de pesquisadores em cursos de graduação de Bacharelado e Licenciatura e ainda noutras práticas de investigação além das da universidade. Afinal, métodos de observação e mapeamento de usos linguísticos podem ter consequências em diversas áreas do saber: política, publicidade,

3 SANTOS, L. et al. O português como língua não materna: reflexões sobre ensino e tradução – Entrevista com a Professora Doutora Liliane Santos. Diadorim: *revista de estudos linguísticos e literários*, v. 23, n. 1, 2021. <https://revistas.ufrj.br/index.php/diadorim/article/view/41164>. Acesso: 01 abr. 2022.

inteligência artificial (com linguagem cada vez menos “artificial”, com atenção a diversas interações, diversos espaços de construção do binômio linguagem e poder, inclusive poder de influenciar e mover (inter)ações).

Bom proveito!

PRESENTATION

LEARNING AND DOING ON A NETWORK: STABILITY, VARIATION AND CHANGE

Marcia dos Santos Machado Vieira (Universidade Federal do Rio de Janeiro,
CNPq/Faperj)

This book is a tribute to twenty years of research development within the scope of the Predicar Project, coordinated, since its inauguration in 2002, by me, Marcia dos Santos Machado Vieira, in the Postgraduate Program in Vernacular Letters at the Federal University of Rio de Janeiro. It is the result of rich (inter)actions mobilized within the Predicar Project, from or beyond it.

In 20 years of vitality, the Predicar Project has followed different paths, always counting on the participation and dialogue of students (from scientific initiation, research for the conclusion of undergraduate, master's, doctoral courses). It has contributed to the training of many as professors/teachers-researchers, with regard to predication, complex predicates, verbal multifunctionality and related topics linked to processes of stabilization, variation and change. It has contributed to the description of the phenomena of auxiliary, impersonalization, intensification, attenuation, representation, passivity, idiom, temporality, aspectuality, modality. And, for that, it has articulated different profiles of empirical research: observational profile, of corpora data (oral and written ones, synchronic or diachronic ones, of varieties of Portuguese inside and outside Brazil); or experimental profile, from data collected in

the experience of use or manipulated data but always based on this experience captured in corpora. Generally, it is developed on the basis of comparative analysis (including longitudinal analysis) and multivariate analysis and, more recently, also on the basis of collostructional analysis. And it deals with qualitative and quantitative treatment of data or records of experiments of the offline type.

Currently, the study of co-occurrences, of placements of construction units, of collostructional analysis, as will be seen in one of the chapters of this book, has mobilized the team. Under the title *Formação e expressão de predicados complexos e predicacões: estabilidade, variação e mudança construcional / Formation and expression of complex predicates and predications: stability, variation and constructional change*, from a socioconstructionist perspective (MACHADO VIEIRA; WIEDEMER, 2019, 2020),¹ the potential of the processes of constructional variation, of change (constructional change or grammatical and lexical constructionalization, TRAUOGOTT; TROUSDALE, 2013)² and of (relative) stabilization that we capture by representing the network of constructions based on the tendencies detected through constructional type and occurrence frequency analysis and/or through collostructional analysis. What come into play are (i) the tension between processes and forces of unity and variability, (ii) the relationship between repetition and conventionalization on the one hand and innovation, mismatch and creativity on the other, (iii) the triangulation between iconicity, indexicality and arbitrariness/conventionality, and (iv) the link between normalized or standardized identity, multifaceted (individual and community) identity in the midst of the diverse and dense nature of social networks that speakers of a language integrate or build. The Predicar Project began under the title *Formação e expressão de predicados complexos: polifuncionalidade verbal / Formation and expression of complex predicates: verbal polyfunctionality*. On that occasion, efforts turned to the study of verbal units in context, to the understanding of their multifunctionality in the experience of use, of the properties involved in the categorization of verbs, considering a continuum, with intersections, of more or less exemplary members of the verbs categories perceived in the linguistic-textual analysis and of the peripheral members, as well as for the detection of complex predicates formed from verbs. And, to some extent, different emphasis related to phases or waves of investigations of the Project are contemplated in this book.

This book brings together texts on methodology for investigating idioms and phraseology, for studying constructional variation, for experimental research, for empirical corpora research, for collostructional analysis, multivariate analysis and

1 MACHADO VIEIRA, M. dos S.; WIEDEMER, M. L. *Sociolinguística Variacionista e Gramática de Construções: Variationist Sociolinguistics and Construction Grammar: os desafios e as perspectivas de compatibilização*. In: *Dimensões e Experiências em Sociolinguística*. São Paulo: Blucher, 2019, p. 85-120.

MACHADO VIEIRA, M. dos S.; WIEDEMER, M. L. *A variação no modelo construcionista da Linguística Funcional-Cognitiva*. In: BRESCANCINI, C. R.; MONARETTO, Valéria. V. N. de O. (orgs.). *Sociolinguística no Brasil: textos selecionados*. Porto Alegre: Editora da PUCRS, 2020, p. 265-304.

2 TRAUOGOTT, E. C.; TROUSDALE, G. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press, 2013.

comparative analysis. The central question that will mobilize the authors of the chapters gathered here is the methodology used to research language(s), conceived in a perspective in which it is important to relate attributes of form and function/meaning and it is interesting to think about the potential of linguistic science that is reached with different methodological perspectives for the space of sensitization and training of researchers, for teaching about research processes and for dealing with the grammatical treatment in (mother or non-mother) language teaching. The process of scientifically studying the grammar of a language, based on various avenues of analysis that can be put into practice, needs to gain prominence also in the classroom, in addition to university labs. And, thus, we can collaborate to bring out the investigative spirit, as well as the interest in different approaches and processes to detect and map linguistic dynamics and heterogeneity.

The first chapter of this book was written by Pedro Ivorra Ordines (Universidade de Santiago de Compostela, Universitat Pompeu Fabra). In *Codazo en plena y puñetazo en toda la boca: La familia de construcciones SOMÁTICAS DE GOLPE*, the author deals with morpho-syntactically and semantic-pragmatically similar constructions that can be described as a family of constructions within the Construction Grammar framework. He deals with idioms where a preposition and an adjective are lexically specified, while three slots are lexically open, with semantic restrictions. The objective is to describe these constructions of similar behavior in a two-dimensional network. Using the esTenTen18 corpus (Sketch Engine), the author develops an in-depth analysis of the semantic restrictions of these slots, paying attention to productivity through type frequency and hapax legomena. And, thus, the author highlights another important issue in Construction Grammar, which, to some extent, constitutes the tonic to mobilize linguists to this conception of language in which grammar and lexicon are integrated, but which, also, does not always have received the attention that would be expected.

The second chapter, written by Bert Cappelle (University of Lille) and Pâmela Fagundes Travassos (Federal University of Rio de Janeiro) illustrates a tendency in the Predicar Project: the interaction between professors (in this case, a co-supervisor) and students (in this case, a doctoral student), attention to which is mainstream in the area of Letters and Linguistics (without losing sight of what precedes and leads to it), the interrelational potential of different institutions in favor of the description of languages. Under the title *Taking a look at the support verb construction V a look: a demonstration of methodology*, the authors present step-by-step instructions on how to conduct a study of constructional variation, illustrated here with one that they developed from the dialogue and guidance they have maintained since Pâmela Fagundes Travassos made her sandwich doctorate at the University of Lille.

The third chapter, written by Ravena Beatriz de Sousa Teixeira and Jeane Nunes da Penha (both from the Federal University of Rio de Janeiro), focuses on linguistic research centered on annotated corpora, focusing on those from the Corpus of Portuguese, by Mark Davies. Such a perspective is very important to areas such as Sociolinguistics, Corpus Linguistics, Usage-Based Construction Grammar, in brief,

what is referred to as Usage-Based Linguistics. In the text *Constructions with support verbs: data collection and treatment in the Corpus of Portuguese / Construções com verbo suporte: coleta e tratamento de dados no Corpus do Português*, the authors show ways and resources to explore this platform, deal with positive and negative aspects of online databases and characteristics of the constitution of samples and show a step-by-step work with annotated corpora.

Os primeiros passos em análise colostrucional: pesquisas de predicação para ilustrar um fazer / The first steps in collostructional analysis: predication research to illustrate a doing is a text written by many hands because, within the scope of the Predicar Project, it reveals a doing that has been built from readings, voices and different experiences. It was written by Marcia dos Santos Machado Vieira, Mariana Gonçalves da Costa, Ravena Beatriz de Sousa Teixeira, Pâmela Fagundes Travassos, Pedro Poppolino and Lais Lima de Souza (all of them from the Federal University of Rio de Janeiro). It is, above all, a tribute to the team that, guided by the goal of reflecting and studying in partnership questions of methodology for the analysis of linguistic co-occurrences, improves and strengthens itself cooperatively to develop the so-called collostructional analysis, without losing sight of the limits and its implications, without losing sight of its articulation with other methodologies that continue to guide the research group's paths. In this chapter, the authors explore three types of collostructional analysis procedures via R language and RStudio application, as well as briefly pass through the multivariate analysis modeling and the analysis modeling of the relationship between constructional units to textual-discursive constructions, to discursive patterns, via Excel.

The two following chapters reveal precisely another methodology of work implemented in the Predicar Project at the beginning of the 21st century when it was not nearly as visible as it is today in Brazilian studies: the experimental methodology. From Machado Vieira; Esteves (2008), many studies have focused on the subjective evaluation of uses of predications and predicates in Portuguese (evaluation inspired by the evaluation problem in WEINREICH; LABOV; HERZOG, 1968). So, a sample of this methodological perspective is also in this work, via the two chapters that are now summarized.

Jeane Nunes da Penha, Ravena Beatriz de Sousa Teixeira and Pâmela Fagundes Travassos (all from the Federal University of Rio de Janeiro) show, in *Pesquisa experimental na análise de construções com verbo suporte / Experimental research in the analysis of constructions with support verbs*, they present the main guidelines for the experimental treatment of usage data based on the experience of their research on constructions with support verbs.

Pâmela Fagundes Travassos (Federal University of Rio de Janeiro, teacher at the Municipal Secretary of Rio de Janeiro), in *Proposta de testes experimentais para análise da variação construcional de predicador complexo / Proposal for experimental tests for the analysis of the constructional variation of a complex predicate*, gathers and illustrates a portion of the materials of her proposal of tests for experimental research presented on the occasion of her candidacy for a sandwich doctorate at the University

of Lille. In this chapter, the author shows models of experimental tests to analyze the variation between complex predicates with support verb DAR: *dar um empurrãozinho*, *dar uma forcinha* (something like give a little help), for example. And, thus, it deals with the problem of subjective evaluation of the possibility of equivalence and interchangeability of predicates like these in simple verbal forms, with users' attitudes towards this type of predication.

Fábio Rodrigo Gomes da Costa (Federal University of Rio de Janeiro, teacher at SEEDUC), in *Discurso de parlamentares: um perfil metodológico de análise de corpus de domínio político / Discourse of parliamentarians: a methodological profile of corpus analysis in the political domain*, deals with the path of a multivariate analysis of predication data with the verb *trazer* (to bring), in that this can appear either as a predicate/full verb or as a support verb and can, in this case, integrate a complex predicate with a more or less evident non-compositional profile. In doing so, he addresses clusters of factors and hypotheses that are interlinked to illustrate a route in order to deal with this type of analysis. It also addresses aspects of political domain discourse constitution in which the predication of discursive attenuation has prominence.

In *Metodologias e análises em estudos de predicação transitiva direta com pronome SE: um estudo puxa outro / Methodologies and analyzes in direct transitive predication studies with SE pronoun: one study leads to another*, Eneile Santos Saraiva de Pontes and Marcia dos Santos Machado Vieira (both from the Federal University of Rio de Janeiro) show the step by step involved in research fronts related to data of direct transitive predication with SE pronoun (indicative of synthetic passive, of suppression of the inducing participant and, consequently, of indeterminacy of the person responsible for or the cause of the state of affairs). In fact, they show how the fronts are interconnected, given the fact that one study ended up unfolding in others. And, then, they highlight the very frequent unfolding of (inter)actions and repercussions in scientific research that are carried out within the scope of a larger project.

The tenth chapter, written by Marcia dos Santos Machado Vieira (Federal University of Rio de Janeiro), focuses on criteria for the identification of complex predicates and verbal subtypes, especially with regard to the auxiliary character. And it is built in view of spaces for the treatment of this theme in the teaching of Portuguese, mother and non-mother language. Under the heading *Análise comparativa para detectar auxiliabilidade verbal e predicadores complexos / Comparative analysis to detect verbal auxiliary and complex predicates*, criteria related to the association of verbal uses to the categories of (semi-)auxiliary verb, (semi-)support and relational verbs and, therefore, the detection of complex predicates are presented. Based on different criteria and some verbal uses, as well as research developed in the Predicar Project, the association of verbs to categories and degrees of auxiliary and to certain slots in complex predicates is the theme that receives attention.

This book also has texts, afterwords, which translate views, inside and outside Brazil, on the value of descriptive contributions supported by linguistic science (here,

some of those that are fruits and take place in the (inter)actions of the Predicar Project (or from there) to the teaching of Portuguese.

The postscript in French was written by Liliane Santos (University of Lille), with whom the Predicar Project team could dialogue during the I International Congress of the Postgraduate Program in Vernacular Letters, *Voices e escritas nos diferentes espaços da língua portuguesa / Voices and Writings in the Different Spaces of the Portuguese Language*, in reason for the seminar of which Liliane Santos was one of the coordinators. One result of this space of interaction is the interview Santos et al. (2021).³ In the afterword of this work, the professor-researcher deals with the interconnection between paths of linguistic research and paths of teaching mother and non mother language, including for the perception of the pluricentricity of Portuguese.

The valuable contributions regarding the methodologies of linguistic research and the perspectives on this network of perspectives that are gathered here show the potential for directions in the area and for dialogues beyond it. To all those who contributed to the result that is materialized in the form of this book, the expression of many thanks does not translate the comfort and joy of being able to count on the points of view and expertise that this book mobilized. This form certainly does not capture the richness of the dialogues and research and dialogue experiences that (re) elaborate and instigate us to always learn and walk, but they certainly show a little of the multi-methodological and interrelational profile that characterizes doings in Predicar since always.

The final version of each chapter is the intellectual responsibility of its author(s), who had complete freedom to define the profile of the thematic and textual approach. Naturally, the contribution of fellow specialists in the area of Letters and Linguistics enriched the process of (inter)action that a book requires. I am grateful to each one of the scientific committee who collaborated, with their time and know-how, for the best achievement of this sharing project.

I am especially grateful to the Postgraduate Program in Vernacular Letters at the Federal University of Rio de Janeiro and to Capes/Proex (Coordination of Postgraduate and Research Personnel/Academic Excellence Program) for the financial support, which allowed and encouraged the realization of a work of national and international dialogue like the one involved here. I also thank each one of those involved in the review, layout and dissemination of the book. I thank Gabriel Lucas Martins (Federal University of Rio de Janeiro) for collaborating on the English version of this presentation.

I hope that this network of methodological perspectives can help the community of research students and professors and that, to some extent, it can be put to good use in spaces that pay attention to the training of researchers in Bachelor and Licentiate

3 SANTOS, L. et al. O português como língua não materna: reflexões sobre ensino e tradução – Entrevista com a Professora Doutora Liliane Santos. *Diadorim: estudos linguísticos e literários*, v. 23, n. 1, 2021. <https://revistas.uff.br/index.php/diadorim/article/view/41164> Acesso: 01 abr. 2022.

undergraduate courses and in other practices of research beyond that of the university. After all, methods of observation and mapping of linguistic uses can have consequences in several areas of knowledge: politics, advertising, artificial intelligence (with an increasingly less “artificial” language, with attention to different interactions, different spaces for the construction of the binomial language and power, including power to influence and move (inter)actions).

Enjoy it!

CAPÍTULO 1

CODAZO EN PLENA CARA Y PUÑETAZO EN TODA LA BOCA: LA FAMILIA DE CONS- TRUCCIONES SOMÁTICAS DE GOLPE

Pedro Ivorra Ordines (Universidade de Santiago de Compostela, Universitat
Pompeu Fabra)

ABSTRACT

This paper focuses on morpho-syntactically and semantically-pragmatically similar constructions that can be described as a family of constructions within the framework of Construction Grammar. These are described as constructional idioms in which a preposition and an adjective are lexically specified, while three slots are lexically open, with semantic restrictions. The aim is to describe these similarly behaving constructions in a two-dimensional network. Using the esTenTen18 corpus (Sketch Engine), this study intends on carrying out an in-depth analysis of the semantics restrictions of these slots paying close attention to their productivity by means of their type frequency and hapax legomena.

1.1 INTRODUCCIÓN

El presente trabajo¹ se enmarca en el proyecto de investigación *CONSTRIDIOMS (Las construcciones fraseológicas del alemán y el español en contraste a través de los corpus, FFI2019-108783RB-100)*, llevado a cabo por el grupo de investigación FRA-SESPAL. A través del marco teórico de la Gramática de Construcciones, con especial atención a su vertiente cognitiva basada en el uso (Goldberg 1995, 2006, 2019), el principal objetivo es el estudio de las construcciones fraseológicas en alemán y español, un tipo de construcción caracterizada por su naturaleza semiesquemática con ítems léxicos saturados léxicamente y *slots* que se actualizan en el discurso (véase Mellado Blanco 2021), por lo que se sitúan a medio camino dentro del continuo léxico-gramática.

Llegar a las construcciones fraseológicas que servirán de objeto de estudio, no obstante, no es una tarea fácil, pues el último objetivo del lingüista es buscar e identificar construcciones para posteriormente analizarlas. A modo de metáfora, Hilferty (2003: 49) compara la búsqueda e identificación de construcciones con la actividad de coleccionar mariposas. En definitiva, encontrar generalizaciones a partir de la lectura de muchos constructos, tarea que requiere de mucha práctica, aunque hay lingüistas que parecen tener un talento natural para captar las rarezas gramaticales de la lengua.

Nuestro estudio se centra, concretamente, en un tipo de construcción fraseológica: las construcciones SOMÁTICAS DE GOLPE, caracterizadas por tener componentes léxicamente saturados y *slots* o casillas vacías que se actualizan en el discurso. Desde este punto de partida, analizamos los rasgos generales del patrón discontinuo [$V_{[impac-to]}/V (con) N en PLENO N_{[parte del cuerpo]}$] con la finalidad de determinar por qué nos encontramos ante una construcción según los presupuestos constructoristas. Seguidamente, presentamos el procedimiento metodológico inductivo basado en corpus que hemos seguido para seleccionar aquellos constructos que están licenciados por la construcción objeto de estudio. En un tercer paso, analizamos las restricciones semánticas a las cuales están sometidas los diferentes *slots* de la construcción parcialmente saturada (Sección 2). A partir de la premisa que las construcciones no son entidades aisladas, presentamos los tipos de relaciones que se puedan dar entre las construcciones para, posteriormente, presentar el análisis de la construcción [$V_{[impac-to]}/V (con) N en TODO DET N_{[parte del cuerpo]}$], siguiendo el mismo procedimiento que la anterior. Terminamos la Sección 3 con la discusión de cómo las construcciones interactúan dentro de la familia de construcciones SOMÁTICAS DE GOLPE. La Sección 4 concluye el estudio con una revisión crítica de los resultados.

1 Actividad financiada por la Unión Europea-NextGenerationEU, Ministerio de Universidades y Plan de Recuperación, Transformación y Resiliencia, mediante convocatoria de la Universidad Pompeu Fabra (Barcelona) (2021UPF-MS-24).

1.2 DESCRIPCIÓN DE LA CONSTRUCCIÓN [V_{[IMPACTO] / V (CON) N EN PLENO N_[PARTE DEL CUERPO]]}

1.2.1 RASGOS GENERALES DE LA CONSTRUCCIÓN

En el caso del patrón objeto de estudio de la presente investigación [V_{[impacto] / V (con) N en PLENO N_[parte del cuerpo]]², nos encontramos, siguiendo los presupuestos de la Gramática de Construcciones, ante una construcción³ según la definición que propone Goldberg (1995: 4), es decir,}

C is a CONSTRUCTION iff_{def} C is a form–meaning pair <F_p S_i> such that some aspect of F_i or some aspect of S_i is not strictly predictable from C's components parts or from other previously established constructions.

A partir de la definición que ofrece Goldberg observamos que se trata de un emparejamiento de forma (un verbo –principalmente de impacto–, la preposición *con* puede ser opcional dependiendo de si el N que le sigue funciona como complemento circunstancial o complemento directo y un sintagma preposicional encabezado por la preposición *en*, seguido del adjetivo *pleno* que puede variar entre masculino o femenino según el género del sustantivo que le sigue y el sustantivo somático) y significado ('alcanzar plenamente N_[parte del cuerpo]' con gran intensidad'). En la medida que hablamos de emparejamiento indisoluble de forma y significado⁴, se trata pues de un elemento unitario, en vez de una mera forma que se podría describir sin hacer alusión al conocimiento del hablante de una lengua. Un argumento importante a la hora de determinar que se trata de una construcción gramatical es el criterio de la no predictibilidad⁵: cualquier aspecto o bien de la forma o del significado no se puede inferir

- 2 Desde la Semántica de Marcos, se defiende una descripción formal detallada de la construcción: «[s]ince the information we record is supposed to be relevant to the syntactic description of a given lexical unit, we require that the frame elements we attend to are in *grammatical construction* with the lexical unit being described» (Fillmore 2008: 51).
- 3 Para más definiciones de la noción de construcción: desde la Gramática de Construcciones de Fillmore, Kay y colaboradores, véase Fillmore 1988: 36, Fillmore 1999: 113, Fillmore 2013: 126; desde la Gramática de Construcciones Cognitiva, véase Goldberg 2006: 5, Goldberg 2019: 7; y desde la Gramática Cognitiva, véase Langacker 1991: 5, 548.
- 4 De la misma manera que otras teorías cognitivo-funcionales de la lengua, en la Gramática de Construcciones el significado tiene una especial relevancia, dado que «a focus on form to the neglect of function is like investigating a human organ such as the liver, without attending to what the liver does: while this is not impossible, it is certain to fail to be explanatory» (Goldberg 2006: 168; véase Fillmore 1989: 21, Goldberg 1995: 229).
- 5 Para hablar del criterio de la no predictibilidad, los construccionistas remiten al principio de composicionalidad, es decir, «if you know the meanings of the words and you know *all* the rules that combine words and phrases into larger formal units, while simultaneously combining the meanings of the smaller units into the meanings of the larger ones, then you know the forms and meanings of all the larger units, including all the sentences» (Kay / Michaelis 2013: 2274; véase Michaelis 2019).

de la suma de sus constituyentes o derivar de otras construcciones existentes ya establecidas y, por lo tanto, no hay otra alternativa que nos lleva a la conclusión de que la expresión ha sido aprendida como emparejamiento de forma y significado, como construcción. En este caso concreto, consecuentemente, hablamos de construcción gramatical porque aquello que resulta impredecible es que el adjetivo PLENO seguido de un N somático esté vinculado pragmáticamente con el impacto y los golpes⁶ (véanse ejemplos 1 y 2). Es precisamente la frecuente aparición de un determinado patrón en un contexto lo que puede considerarse como un argumento a favor de la postulación de una construcción, dado que «linguistic units can be identified as constructions not only on the basis of their morphosyntactic or semantic conventional makeup, but also on the basis of their discursal and/or sociocultural currency» (Nikiforidou 2009: 25; véase Fillmore 2008, Croft / Cruse 2004, Finkbeiner 2019).

(1) La daga le alcanzó **en pleno pecho**, tomando por sorpresa al ladrón, y un Orión cabreado le tomó por el cuello. (18678239130)

(2) Ambos resultan fatalmente heridos: el bravo capitán Pedro de Baeza recibe una estocada **en plena boca** que la atraviesa la quijada y le quita varias muelas. (13769276051)

La generalización de la noción de construcción gramatical⁷ para explicar la totalidad de la lengua del hablante implica que la gramática se entienda más bien como un continuum de unidades simbólicas que se sitúan a lo largo del continuum léxico-gramática; en otras palabras, desde las construcciones sustantivas a las esquemáticas, desde las concretas a las abstractas y desde las atómicas a las complejas. En el caso de la construcción que nos concierne en el presente estudio, nos encontramos ante un «partially filled idiom» (Goldberg 2006: 5) en el que uno o varios de sus componentes están saturados léxicamente (sería el caso de la preposición *en* y el adjetivo *pleno* que varía entre la forma en masculino o femenino según el género del sustantivo que le sigue), mientras que los *slots* se actualizan en el discurso con ciertas restricciones semánticas.

Por su naturaleza no saturada en su totalidad, sería un *partially lexically filled idiom*, construcción caracterizada por la idea de *continuum* y gradualidad. Otras denominaciones comunes que se han utilizado desde enfoques construccionistas son «formal or lexically open idioms» (Fillmore / Kay / O'Connor 1988), «schematic idioms» (Croft / Cruse 2004) o «constructional idioms» (Booij 2002: 301-302), quien las define como «syntactic constructions with a (partially or fully) non-compositional

6 Resulta curioso ver que en el DUE (2007: 2329), para la acepción de «pleno», nos encontramos una nota que sugiere el significado con el que hemos descrito la construcción, especialmente en el primero de los ejemplos: «[a] veces, se emplea como expresión enfática para acentuar el hecho de que se trata precisamente de la cosa representada por el nombre: 'Recibió el golpe en plena cara. Iba con abrigo en pleno mes de julio'». Siguiendo a Albelda Marco (2007: 143), observamos que la intensificación se expresa de manera paradigmática, es decir, el valor intensificador está léxicamente presente en la construcción.

7 Conviene recordar, en este punto, la diferencia entre construcción y constructo: mientras que las construcciones son generalizaciones a partir del conocimiento lingüístico convencional, los constructos son realizaciones de construcciones en el discurso real (véase Figura 1).

meaning contributed by the construction, in which – unlike idioms in the traditional sense – only a subset (possibly empty) of the terminal elements is fixed».

1.2.2 TRABAJAR CON SKETCH ENGINE: LA BÚSQUEDA DEL PATRÓN DISCONTINUO V N EN PLENO X

Para el análisis de la construcción objeto de estudio utilizamos el corpus⁸ *esTenTen18* del gestor de corpus Sketch Engine. Se trata de un macrocorpus con un total de 19 600 041 678 *tokens*, el cual, a su vez, se divide en dos subcorpus de tamaño similar: el *European Spanish Corpus* (49,32%) y el *American Spanish Corpus* (46,46%). El 4,22% restante es de procedencia indeterminada. Además, está compilado de forma automática a partir de internet y constituido por contenido que, en gran parte, puede estar generado por el usuario, por lo que es común encontrar erratas.

El primer paso en la búsqueda del patrón es la selección del tipo de búsqueda que mejor se adecúe a nuestro objeto de estudio. Teniendo en cuenta que se trata de un patrón discontinuo en el que tenemos una parte fija y constituyentes que se van actualizando en el discurso, nos decantamos por la herramienta *Corpus Query Language (CQL)* de Sketch Engine. Se trata de un lenguaje de búsqueda de corpus para la detección de patrones con un nivel de precisión y exhaustividad mucho mayor que las otras opciones de búsqueda (véase Hidalgo-Ternero / Corpas Pastor 2021 para una revisión crítica de los tipos de búsqueda que ofrece Sketch Engine). De este modo, para realizar la búsqueda de la construcción $[V_{[impacto]}/V (con) N en PLENO N_{[parte\ del\ cuerpo]}]$, utilizamos el siguiente esquema CQL: `[word="en"] [lemma="pleno"] [tag="N.*S.*"]`⁹. Dicho de otro modo: utilizamos el código *word* para los elementos invariables, mientras que *lemma* indica que puede variar en género y número. Finalmente, el código *tag* nos permite elegir la categoría gramatical “N.*” (para sustantivo) y el número “S.*” (para singular), por lo que la delimitación del número incidiría, al mismo tiempo, en el número del adjetivo que aparece en posición anterior, salvo errores tipográficos que puedan ofrecer los resultados. Esta primera búsqueda arroja un total de 535 160 ocurrencias, pero no todas forman parte del objeto de estudio aún estando el *slot* actualizado con un ítem somático (véase captura de pantalla 1).

8 Desde las aproximaciones basadas en construcciones, los corpus de grandes dimensiones se han convertido en una herramienta esencial para el estudio y la descripción de las construcciones en la medida que “reveal how constructions vary, often in ways that the researcher did not anticipate” (Wiedemer / Machado Vieira / Cezario 2019: 31; véase Gries 2013, Yoon / Gries 2016, Hoffmann / Horsch / Brunner 2018).

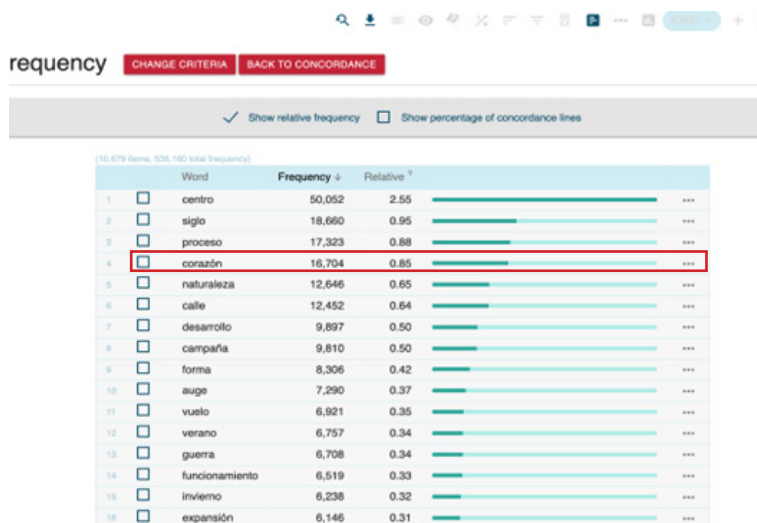
9 Tanto el slot del verbo como el del sustantivo que puede ir precedido o no de la preposición no están especificados en la búsqueda CQL, puesto que la sintaxis del español es bastante flexible, por lo que resultaría muy complicado predecir cómo se comporta la construcción a priori, tal y como podemos observar en los ejemplos (1) y (2).

Captura de pantalla 1: Resultados sin filtrar a partir del esquema CQL [word="en"][lemma="pleno"] [tag="N.*S.*"]

Details	Left context	KWIC	Right context	
1	algq.es	i del sol, pero que no le achicharra como	en pleno verano	(aunque todo depende del cambio climáti
2	etbiogedealfonso...	gos (León), tuvo la fortuna de comprobar	en pleno corazón	de los Picos de Europa la eficacia y coord
3	fioreasalud.es	la fórmula de abonado. </> Un árbol	en plena producción	responde muy bien a la siguiente aplicack
4	fioreasalud.es	la misma cantidad cuando el árbol esté	en plena floración	, esto ayuda a promover el amame de la tr
5	mejor.es	r Azca fundado en el año 2009 y ubicado	en pleno centro	financiero de la ciudad de Madrid. </>
6	mejor.es	ta de expansión de la ciudad de Málaga.	en pleno corazón	económico y financiero de la ciudad. </>
7	sociedadartia...	in del Reina Sofía, un museo que estaba	en plena decadencia	y que lo han vuelto a convertir en lo que ti
8	joldom.es	ta, en la Plaza Nueva. </> Otra joya	en pleno centro	de la ciudad. </> Suscripción a nuest
9	cuadernosedemedi...	nente inédito suceso de la muerte súbita	en plena competición	ante una multitud de testigos. </> Al i
10	menorcaairport.es	mbre a noviembre, aunque es cierto que	en pleno mes	de enero hay días de sol radiante que invi
11	trnd.es	nos. Una pena que el proyecto emplee	en pleno otoño	. </> Con lo bien que nos habría veni
12	alimentacion.es	i celebra el primer domingo de cada mes	en pleno casco	antiguo de Bullas e incluye en cada edicó
13	oicisismoafondo...	ici si aprobaba. </> Siete suspensos	en plena adolescencia	pudieron privarle de un futuro que ya asor
14	oicisismoafondo...	trabajaría para él. </> Sin embargo,	en plena escapada	, tras varios kilómetros, sus compañeros l
15	etperyon.es	ta desde la ciudad amurallada y sentirse	en pleno centro	histórico de la ciudad en el mar, a poco
16	hotelpiramide.e...	</> Los hoteles Piramide están situados	en pleno corazón	de las Pías Bajas, a menos de 1,5 km de
17	votzon.com	fa de desprestigio que los nacionalistas,	en plena deriva	independentista del catalanismo político, i
18	etdeportedejaen...	rajada, para completar sus 75 kilómetros	en pleno corazón	de los Pirineos y siendo el tiempo previst

Una vez obtenemos todos los resultados, observamos que no todas las instancias están licenciadas por la construcción que hemos descrito en el apartado anterior, teniendo en cuenta que el *slot* tiene la restricción semántica de que tiene que ser una parte del cuerpo humano y, además, debe cumplir con el significado que hemos descrito más arriba (véanse los ejemplos destacados de la Captura de pantalla 1). Para poder ir refinando la búsqueda para encontrar y seleccionar las ocurrencias que nos interesan, utilizamos la herramienta *frecuencia* para ver los diferentes *types* que actualizan el *slot* de la búsqueda ajustando la ordenación a la última palabra del KWIC (véase Captura de pantalla 2). Así pues, en un siguiente paso vamos seleccionando todos aquellos *types* que no nos interesan del listado para que, de este modo, el gestor de corpus nos enseñe aquellas ocurrencias que no hemos descartado. En este proceso, obtenemos un total de 26 *types* (21752 ocurrencias, es decir, un 4,06% del total de las instancias) que ordenamos por frecuencia *token*: *corazón*, *cara*, *rostro*, *pecho*, *pulmón*, *boca*, *frente*, *cabeza*, *ojo*, *estómago*, *jeta*, *vena*, *nariz*, *cuello*, *nuca*, *vientre*, *mandíbula*, *oreja*, *hocico*, *abdomen*, *mentón*, *sien*, *cocorota*, *boca del estómago*, *careto* y *morro*.

Captura de pantalla 2: Listado de los types por orden de frecuencia token.



En un tercer paso, llevamos a cabo el vaciado del corpus seleccionando las instancias que estén licenciadas por la construcción fraseológica somática [V_[impacto]/V (con) N en PLENO N_[parte del cuerpo]]: ‘alcanzar plenamente N[parte del cuerpo], con gran intensidad’. De las 21752 ocurrencias que obtenemos del paso anterior, contabilizamos un total de 1671 instancias licenciadas por dicha construcción (un 7,68% del vaciado somático se corresponde con la construcción). Encontramos, de este modo, otros emparejamientos de forma y significado como las construcciones con un valor eminentemente locativo (véanse ejemplos 3 y 4).

(3) El pasado 14 de diciembre abrió las puertas al público un nuevo restaurante en Valencia, Viva Tapas Bar, ubicado en un edificio del siglo XIX en la calle Santo Tomás 18, **en pleno corazón**¹⁰. (288516960)

(4) Cuando se le ve sin su casco tipo samurái el personaje, tal como se muestra en la imagen, tiene una cresta mohicana y un tatuaje **en plena frente** que da un aspecto siniestro a su enjuto y alargado rostro de nariz pequeña y achatada y prominentes cejas. (2097266808)

1.2.3 RESTRICCIONES SEMÁNTICAS, PRODUCTIVIDAD Y PREFERENCIAS COLOCACIONALES

Siguiendo la idea del continuum léxico-gramática que se defiende desde los postulados construccionistas, en este trabajo analizamos un tipo de construcción bajo la denominación de «construcción fraseológica». Uno de los rasgos que nos interesa de este tipo de construcción es el hecho de que está formada por constituyentes saturados

¹⁰ Aparte del valor locativo, el lema «corazón» no interviene con el sentido somático, sino como «centro» de un lugar (véase también ejemplo 13).

léxicamente y *slots* o casillas vacías que se actualizan en el discurso. Así pues, la restricción semántica que observamos en el patrón con el que hemos realizado el vaciado es que el *slot* esté actualizado por una parte del cuerpo humano, por lo que resulta útil el estudio pormenorizado por medio de corpus para poder llevar a cabo una descripción completa desde un punto de vista semántico-discursivo. Atendiendo a la semántica de la construcción, García-Page (2008: 363) apunta que

la presencia de un sustantivo que designa una parte del cuerpo favorece la transparencia [de la construcción] y, por tanto, su analizabilidad semántica, gracias, fundamentalmente, a la posibilidad de establecer una asociación entre la parte del cuerpo y una de sus propiedades o funciones (biológica, fisiológica, psicomotriz, física...), la actividad externa para la que está capacitada, el lugar donde está ubicada o el gesto o ademán que pueda acompañar su enunciación (somatismo cinésico).

Estrechamente relacionado con el carácter parcialmente esquemático de la construcción objeto de estudio es su naturaleza productiva, derivada de la posibilidad que ofrece el patrón con *slots*, teniendo en cuenta que pueden ser actualizados con diferentes ítems léxicos. En el contexto constructorista, para hablar de la productividad del *slot* de una construcción se recurre a la noción de frecuencia *type*, es decir, «how many different items occur in the schematic slot of constructions» (Bybee 2013: 61). Así pues, contabilizamos 26 ítems léxicos somáticos diferentes de un total de 1671 ocurrencias licenciadas por la construcción¹¹, por lo que tenemos 0,015 *types* por cada instancia licenciada por la construcción (véase tabla 1).

Tabla 1: Types por orden de frecuencia de la construcción [V_[impacto]/V (con) N en PLENO N_[parte del cuerpo]]

Parte del cuerpo	Frecuencia token	Porcentaje de la construcción
cara	621	37,16%
rostro	497	29,74%
pecho	134	8,01%
cabeza	66	3,95%
corazón	61	3,65%
boca	60	3,59%
frente	49	2,93%
estómago	31	1,85%

11 Una idea muy interesante desde los postulados basados en construcciones es la de herencia, de manera que una construcción saturada léxicamente hereda las propiedades de una construcción más abstracta situada en un nivel superior jerárquico, tratándose pues de herencia en el eje vertical («mother construction» según la terminología de Audring 2019). Desde esta perspectiva, los constructos heredarían las características de la construcción más esquemática.

jeta	21	1,25%
mandíbula	15	0,89%
nariz ¹²	15	0,89%
nuca	15	0,89%
cuello	14	0,83%
hocico	11	0,66%
vientre	10	0,60%
mentón	9	0,54%
sien	9	0,54%
abdomen	6	0,36%
boca del estómago	6	0,36%
ojo	6	0,36%
oreja	5	0,30%
cocorota	4	0,24%
careto	2	0,12%
morro	2	0,12%
vena	2	0,12%
pulmón	1	0,06%

Al principio de la sección 2.1. describíamos la construcción como un emparejamiento indisoluble de forma con el significado ‘alcanzar plenamente $N_{[parte\ del\ cuerpo]}$ con gran intensidad’. De las instancias de la construcción somática que hemos ido analizando a lo largo del trabajo, hemos ejemplificado con algunas ocurrencias que tienen una mayor frecuencia *token* como *en pleno pecho* (ejemplo 1) y *en plena boca* (ejemplo 2). Por sí solas, fuera de contexto, tienen un valor eminentemente locativo para indicar ‘en el medio de $N_{[parte\ del\ cuerpo]}$ ’. No obstante, el apoyo contextual nos permite refinar la definición de la construcción con la que estamos trabajando, prestando especial atención a los enlaces asociativos que se establecen entre los ítems léxicos que suelen coaparecer con el patrón en cuestión. De hecho, Goldberg (1995: 229) ya habla de estudiar las construcciones como «ordered triplets of form, meaning and context», por lo que se le concede al contexto un rol fundamental en el ajuste y la matización de la interpretación de una construcción. Para ello, también nos disponemos a estudiar los otros dos *slots* que hemos descrito de la construcción para analizar las preferencias colocaciones del sintagma preposicional *en pleno* $N_{[parte\ del\ cuerpo]}$. En primer lugar, en cuanto al *slot* verbal, contabilizamos un total de 99 verbos diferentes de las 1671 ocurrencias licenciadas por la construcción (véase Tabla 2), por lo que constatamos que no

12 Desde la Gramática de Construcciones, las contaminaciones son una consecuencia lógica de las interrelaciones que mantienen las construcciones entre sí. Así pues, a partir de la locución en las narices, observamos una cierta recurrencia del patrón en plena nariz, tal y como podemos ver en el ejemplo: «Con un movimiento repentino, rápido como un relámpago, una manzana salió de la mano de Sam y golpeó a Bill en plena nariz» (13944705230) (véase Ivorra Ordines / Mellado Blanco 2021, Ivorra Ordines 2021).

es muy productivo teniendo en cuenta que los diez primeros *types* conforman casi el 50% de las ocurrencias¹³. Otro indicador de la productividad de un *slot* es el índice de ocurrencia *hápx legómena*. A diferencia de la aproximación anterior que está basada en la frecuencia *type* de una construcción, esta postura está basada en la frecuencia *token*. Concretamente, la idea que se defiende es la siguiente: si aquellas instancias con una alta frecuencia *token* se almacenan en la memoria como unidades, un indicador de la productividad de un *slot* es el número de instancias que aparece raras veces en los corpus, lo que implica que no puedan llegar a ser formas convencionalizadas de decir las cosas (Zeschel 2012: 174). En la medida que puedan interpretarse correctamente por los hablantes, son una evidencia del uso productivo de la construcción. En este caso concreto contabilizamos un total de 46 ítems verbales que aparecen una sola vez en el corpus de los 99 *types* (46,46% de *hápx*), lo que incide no solo en la productividad de la construcción, sino también en el *entrenchment* del *slot*.

Tabla 2: Los quince colocados verbales más frecuentes de la construcción $[V_{[\text{impacto}]} / V (\text{con}) N \text{ en PLENO } N_{[\text{parte del cuerpo}]}]$

Forma verbal	Frecuencia <i>token</i>	Porcentaje respecto al total de formas verbales
dar	231	19,61%
recibir	211	17,91%
golpear	125	10,61%
propinar	44	3,73%
explotar	37	3,14%
estallar	36	3,05%
pegar	33	2,80%
impactar	33	2,80%
disparar	27	2,29%
alcanzar	25	2,12%
acertar	25	2,12%
lanzar	22	1,87%
soltar	22	1,87%
clavar	21	1,78%
caer	19	1,61%

Según la Semántica de Marcos (Fillmore 2008), los marcos son estructuras conceptuales a través de las cuales se entienden y se definen las construcciones. En este caso concreto, nos encontramos ante un marco de TRANSACCIÓN, la naturaleza del cual nos exige entender un marco o guion en el que se desarrollan una serie de eventos. En ese guion, una persona (a la que llamamos INFRACTOR) provoca daño a una parte del cuerpo de otra persona (a la que llamamos PARTE AFECTADA) con un

¹³ Contabilizamos un total de 1178 formas verbales, por lo que el *slot* verbal tiene un alto grado de *entrenchment* dentro de la construcción, con un 70,50% de saturación léxica.

objeto o parte del cuerpo (INSTRUMENTO DE DAÑO). Por lo tanto, observando la naturaleza de los verbos, vemos que se trata de verbos transitivos que implican golpe¹⁴ (*golpear, propinar, pegar, impactar*, etc.). En aquellas instancias con verbos que no inciden directamente en el golpe, la semántica del verbo se diluye a favor del significado de la construcción por un proceso de coerción (véase Michaelis 2003). En otras palabras, esta observación es un argumento más a favor de la idea de que «hearers try not simply to match the constructional semantics onto the lexeme, but also contextually work out the speaker’s intended interpretation» (Leclercq 2019: 282), tal y como podemos ver con los verbos *besar* (ejemplo 5), *plantar* (ejemplo 6) o *imprimir* (ejemplo 7).

(5) Lentamente inclinó la cabeza y la besó **en plena boca**. Amy había abierto los labios al oír el término cariñoso en boca de Nicky, y los abrió aún más cuando él los cubrió con los suyos. (18869735468)

(6) La dueña de la casa no perdió la cabeza: plantó al santo una bofetada **en plena cara**, como si se tratase de un criado tonto y perezoso, le ordenó que cediera su puesto a los gendarmes y fuese inmediatamente a cuidar el ganado. (12963504987)

(7) –¡No está mal, Steve! –le felicito mientras lanzo sendos rayos repulsores en dirección a mi oponente, buscando desestabilizarlo para imprimirle un puñetazo **en pleno rostro**. (8332398514)

En segundo lugar, en cuanto al primer *slot* nominal, contabilizamos un total de 260 ítems léxicos diferentes¹⁵, lo que tiene una notable repercusión en el carácter productivo del *slot* (véase Tabla 3). Otro factor que incide en la productividad del *slot* son las ocurrencias que tienen una única aparición en los corpus, teniendo en cuenta que de los 260 *types* un 60,31% responden a *hápax legómena* (un total de 175 ocurrencias). Dado que no pueden llegar a convencionalizarse, se convierten en un indicador importante a la hora de determinar la productividad de un *slot*.

Tabla 3: Los quince colocados nominales más frecuentes de la construcción $[V_{[impacto]} / V (con) N en PLENO N_{[parte\ del\ cuerpo]}]$

Sustantivo	Frecuencia token	Porcentaje respecto al total de formas nominales
puñetazo	116	12,38%
bofetada	91	9,71%
golpe	83	8,86%
patada	36	3,84%
tiro	23	2,45%
disparo	22	2,35%

14 Desde una teoría basada en las representaciones ejemplares (véase Bybee 2013), los tres verbos más frecuentes *dar, recibir y golpear* se considerarían los verbos prototípicos (representan casi el 50% del total de los ítems léxicos verbales), los cuales presentan efectos prototípicos.

15 Contabilizamos un total de 937 formas nominales, por lo que el *slot* nominal tiene un alto grado de *entrenchment* dentro de la construcción, con un 56,07% de saturación léxica.

escupitajo	20	2,13%
bofetón	18	1,92%
cachetada	17	1,81%
impacto	16	1,71%
balazo	15	1,60%
beso	15	1,60%
bala	13	1,39%
codazo	13	1,39%
cabezazo	10	1,07%

Siguiendo tanto los porcentajes de la Tabla 3 que inciden en los ítems léxicos nominales más frecuentes como los indicadores de productividad de la frecuencia *type* y *hápax legómena*, observamos que se trata de un *slot* muy productivo, el cual, según la Semántica de Marcos, responde al INSTRUMENTO DE DAÑO. Si estudiamos detenidamente los ítems léxicos que actualizan este *slot*, podemos constatar una cierta regularidad. Nos estamos refiriendo a la notable presencia de sustantivos con sufijos para la designación de golpe¹⁶ o acciones bruscas, repentinas o sorpresivas: *-azo*, *-ada* y *-ón*¹⁷ (véase NGLÉ 2009: 111-113). Se trata de sufijos de derivación nominal que se utilizan para expresar el instrumento (ejemplos 8 y 9) o la parte del cuerpo (ejemplos 10 y 11) con la que se da un golpe.

(8) Fue hacia ellos a la carrera, pero antes de alcanzar su posición, uno fue atravesado por una espada para después recibir un **hachazo en plena sien**. (3778287322)

(9) Lemunao fue agredido a trompadas y patadas, pero la lesión más grave le fue provocada por una **pedrada** que le impactó **en pleno rostro** y a corta distancia arroja por uno de los individuos con una gomera. (16250890765)

(10) Para demostrar lo que decía, Nephausto trabó la cuchilla con su báculo, la apartó para deshacer la defensa de su rival y le asestó una **patada en pleno vientre**. Rómulo retrocedió y antes de recuperarse el báculo se estampó en su cara haciéndole caer rotando en el aire como una peonza. (308169696)

(11) Le propiné una patada a su espada, alejándolo de sus dedos tendidos. Se revolvió furioso, propinándome un **rodillazo en pleno estómago** que me dejó sin aliento. Rápido como una centella se puso en pie. Aun en medio del aturdimiento, atiné a ponerle la zancadilla, que Ethan esquivó. (12571659430)

El recuento manual de las instancias que tienen sustantivos con sufijos de golpe arroja un total de 528 ocurrencias (56,35% respecto del total de ítems léxicos nomina-

16 Para un estudio sobre la productividad de la construcción fraseológica [*a S_{plural/acción violenta}*]: *a pedradas, a cuchilladas, a guantazos, a codazos*, etc., véase López Meirama (2016).

17 Dentro de la construcción que estamos analizando para este trabajo, se podría plantear la construcción semiesquemática [*N_{sufijo de golpe}*], en la que tendríamos el sufijo saturado, mientras que el sustantivo se tendría que actualizar en el discurso (véase Booij 2010 sobre el estudio de sufijos desde una perspectiva construccionalista).

les), lo que incide en el alto grado de *entrenchment* del *slot* con sustantivos terminados con un sufijo de golpe, tratándose de *-azo* del sufijo más frecuente con 310 ocurrencias, seguido por *-ada* con 189 ocurrencias y, finalmente, *-ón* con 29 ocurrencias (véase Monge 1972). Asimismo, constatamos una cierta productividad de los sufijos de golpe¹⁸: contabilizamos un total de 115 *types* (21,78%), mientras que vemos un total de 76 *hápax legómena* (66,08%).

1.3 MÁS ALLÁ DE [EN PLENO N_[PARTE DEL CUERPO]]: LA CONSTRUCCIÓN [V_[IMPACTO]/V (CON) N EN TODO DET N_[PARTE DEL CUERPO]]

1.3.1 RELACIONES ENTRE CONSTRUCCIONES: MADRES Y HERMANAS

Uno de los pilares fundamentales desde los postulados construccionistas es que el conocimiento lingüístico se almacena en la memoria en forma de construcciones, es decir, como emparejamientos de forma y significado à la Saussure. El repertorio de construcciones no conforma un conjunto desestructurado, sino más bien las construcciones forman parte de redes o familias de construcciones, relacionadas a través de enlaces de herencia que motivan las propiedades de las construcciones más particulares (Goldberg 1995). De este modo, la red de herencia nos permite llegar a generalizaciones entre construcciones, lo que explica que las taxonomías basadas en construcciones «emerge in a bottom-up fashion» (Hoffmann 2017: 315), desde lo más sustantivo, concreto y atómico a lo más esquemático, abstracto y complejo.

Hasta ahora, hemos descrito relaciones verticales entre construcciones relacionadas entre sí en el eje vertical, de manera que la construcción más específica hereda los rasgos de la construcción más esquemática ('construcción madre'). En la bibliografía, asimismo, muchos construccionistas defienden la existencia de enlaces horizontales ('construcciones hermanas') para expresar diferentes tipos de relaciones sintagmáticas y paradigmáticas entre construcciones en un mismo nivel de abstracción (véase Van de Velde 2014, Traugott 2018, Diessel 2019, entre otros).

Antes de abordar qué relaciones se establecen entre las construcciones somáticas [V_[impacto]/V (con) N en PLENO N_[parte del cuerpo]] y [V_[impacto]/V (con) N en TODO DET N_[parte del cuerpo]], conviene describir las características principales de ésta última con el mismo procedimiento inductivo que hemos utilizado en la construcción anterior a partir del gestor de corpus Sketch Engine.

18 Desgranamos aquí los datos de la productividad de los sufijos:

- azo: 80 *types* (52,80%) y 49 *hápax legómena* (61,25%);
- ada: 25 *types* (13,23%) y 17 *hápax legómena* (68%);
- ón: 9 *types* (31,03%) y 6 *hápax legómena* (66,7%).

1.3.2 RASGOS GENERALES DE LA CONSTRUCCIÓN

Siguiendo la misma metodología que con la construcción anterior, nos disponemos a estudiar el emparejamiento de forma $[V_{[\text{impacto}]} / V (\text{con}) N \text{ en TODO DET } N_{[\text{parte del cuerpo}]}]$ y significado ‘alcanzar plenamente $N_{[\text{parte del cuerpo}]}$, con gran intensidad’. Desde un punto de vista formal, en este caso, nos encontramos con el adjetivo intensificador *todo*, el cual se caracteriza por conservar el artículo determinativo cuando precede al nombre (DUE 2007: 2892).

A partir del esquema CQL [word=“en”] [lemma=“todo”] [lemma=“el”] [tag=“N.*S.*”], obtenemos un total de 1 250 443 ocurrencias. Utilizando la herramienta *frecuencia*, ordenamos los *types* a partir de la última palabra del KWIC y nos centramos en aquellos ítems léxicos que hacen referencia a una parte del cuerpo. De este proceso de vaciado, nos quedamos con un total de 34 *types* (8244 ocurrencias, o un 0,66% del total de las instancias) que enumeramos a continuación por orden de frecuencia *token*: *cara, boca, rostro, cabeza, espalda, pecho, pierna, corazón, frente, ojo, brazo, jeta, mano, abdomen, uña, cuello, morro, estómago, torso, careto, lomo, culo, nuca, vientre, hocico, mandíbula, oreja, cráneo, barriga, retina, rodilla, cocorota, dedo y boca del estómago*. No obstante, no todas las ocurrencias de los *types* somáticos están licenciadas por la construcción objeto de estudio, por lo que identificamos otras construcciones con un valor eminentemente locativo (ejemplos 12 y 13).

(12) Generalmente se aplican **en toda la cara** como medida preventiva antienviejamiento. (117086049)

(13) Por eso, con ayuda internacional, se construyó hace más de un decenio un hotel, **en todo el corazón** del pueblo, que cuenta con 12 habitaciones. (151205756)

De las 8244 ocurrencias que hemos obtenido del vaciado anterior, seleccionamos aquellas que están licenciadas por la construcción de la que nos estamos ocupando y tenemos un total de 3271 ocurrencias (39,68%), cifras que inciden en un mayor grado de *entrenchment* de la construcción, teniendo en cuenta que «[e]ach token of use of a word or sequence of words strengthens its representation and makes it more easily accessed» (Bybee 2007: 324) (véase Tabla 4).

Tabla 4: Types por orden de frecuencia de la construcción $[V_{[\text{impacto}]} / V N \text{ en TODO DET } N_{[\text{parte del cuerpo}]}]$

Parte del cuerpo	Frecuencia token	Porcentaje de la construcción
boca	1653	50,53%
cara	920	28,12%
cabeza	148	4,52%
jeta	74	2,26%
frente	67	2,05%
pecho	59	1,80%
rostro	54	1,65%
morro	44	1,34%

espalda	34	1,04%
ojo	27	0,82%
estómago	26	0,79%
nuca	25	0,76%
careto	21	0,64%
mandíbula	19	0,58%
hocico	16	0,49%
cocorota	16	0,49%
corazón	13	0,40%
culo	13	0,40%
cuello	12	0,38%
lomo	10	0,30%
pierna	10	0,30%
mano	7	0,21%
barriga	7	0,21%
cráneo	6	0,18%
abdomen	4	0,12%
torso	4	0,12%
rodilla	4	0,12%
boca del estómago	3	0,09%
brazo	3	0,09%
oreja	3	0,09%
dedo	2	0,06%
uña	1	0,03%
vientre	1	0,03%
retina	1	0,03%

Un dato curioso de la productividad del *slot* somático es que tan solo dos *types* licencian casi el 80% de las instancias. Esto se traduce en una alta concentración de *tokens* en un número reducido de *types*, lo que implica una escasa variación y productividad en la construcción. Por una parte, en cuanto al *slot* verbal, contabilizamos un total de 90 formas verbales diferentes de las 3271 ocurrencias¹⁹, por lo que observamos una escasa repercusión sobre la productividad del *slot* verbal (es decir, 0,027 *types* por cada ocurrencia), hecho estrechamente relacionado con la presencia de *types* con una frecuencia *token* muy elevada (véase Ivorra Ordines 2021). De la productividad por medio de las ocurrencias únicas, contabilizamos 41 *hápax legómena* de los 90 ítems léxicos diferentes (45,5%).

¹⁹ Contabilizamos un total de 1651 formas verbales, por lo que el *slot* nominal tiene un alto grado de *entrenchment* dentro de la construcción, con un 50,47% de saturación léxica.

Tabla 5: Los quince colocados verbales más frecuentes de la construcción $[V_{[\text{impacto}]} / V (\text{con}) N \text{ en TODO DET } N_{[\text{parte del cuerpo}]}]$

Forma verbal	Frecuencia token	Porcentaje respecto al total de formas verbales
dar	736	44,58%
pegar	102	6,18%
recibir	86	5,21%
llevarse	83	3,99%
golpear	66	3,45%
hacer	57	3,45%
meter	44	2,66%
soltar	38	2,30%
propinar	33	1,99%
caer	28	1,69%
lanzar	27	1,63%
estampar	27	1,63%
explotar	26	1,57%
arrear	21	1,27%
impactar	18	1,09%

Por otra parte, en cuanto al *slot* nominal, contabilizamos un total de 255 ítems léxicos diferentes²⁰ (véase Tabla 6). En este punto, queremos subrayar la notable presencia de *zas* en la actualización del *slot*, dado que ocupa más del 50% de las instancias de la construcción y, además, tiene una frecuencia diez veces mayor que el segundo sustantivo más frecuente. Otro factor que incide en la productividad de la construcción son los *hápax legómena*, teniendo en cuenta que aquellas instancias que aparecen raras veces en los corpus no pueden llegar a convencionalizarse: de los 255 *types*, un 59% son instancias que aparecen una vez en el corpus (153 ocurrencias).

Tabla 6: Los quince colocados nominales más frecuentes de la construcción $[V_{[\text{impacto}]} / V (\text{con}) N \text{ en TODO DET } N_{[\text{parte del cuerpo}]}]$

Sustantivo	Frecuencia token	Porcentaje respecto al número de instancias totales
zas	1392	54,03%
puñetazo	130	5,04%
patada	103	3,99%
zasca	96	3,72%
golpe	75	2,91%

²⁰ Contabilizamos un total de 2576 formas nominales, por lo que el *slot* nominal tiene un alto grado de entrenchment dentro de la construcción, con un 78,75% de saturación léxica.

bofetada	57	2,21%
hostia	54	2,09%
bofetón	33	1,28%
beso	25	0,97%
tiro	24	0,93%
corrida	18	0,69%
codazo	18	0,69%
guantazo	16	0,62%
tortazo	15	0,58%
rodillazo	14	0,54%

De la misma manera que hemos procedido con la construcción anterior, contabilizamos las instancias que contienen sustantivos con sufijos de golpe y del recuento obtenemos un total de 676 ocurrencias (26,24% respecto del total de ítems léxicos nominales), lo que indica que hay un menor grado de *entrenchment* con sustantivos de golpe respecto a la construcción anterior. En este caso, el orden de frecuencia también se repite: siendo *-azo* el sufijo más frecuente con 417 ocurrencias, seguido por *-ada* con 204 ocurrencias y, finalmente, *-ón* con 55 ocurrencias. Asimismo, observamos una cierta productividad con los sustantivos con sufijos de golpe²¹: 138 ítems léxicos diferentes (20,41%) y 79 instancias únicas (57,24%).

Atendiendo a la alta frecuencia *token* de ‘zas’ dentro del *slot* nominal de la construcción, realizamos una búsqueda del esquema CQL [*zas en todo*] [tag=“D.*”] [1...1] para analizar con qué parte del cuerpo se suele asociar el patrón [*zas en TODO DET*] (véase Stefanowitsch / Gries 2003 sobre el estudio de la interacción entre las palabras y las construcciones). De esta manera, contabilizamos un total de 739 ocurrencias del patrón *zas en toda la boca*²². Desde una aproximación basada en el uso, «the frequencies of repeated linguistic processing events translate into different strengths of associations in the network representing linguistic knowledge» (Schmid / Küchenhoff 2013: 566). Por lo tanto, siguiendo la definición de Goldberg (2006) de construcción, una alta frecuencia *token* y un alto grado de *entrenchment* justifican la postulación de un nodo separado²³. Aunque no siempre hay una conexión directa entre la informaci-

21 Desgranamos aquí los datos de la productividad de los sufijos:

-azo: 105 types (25,18%) y 62 hápax legómena (59,04%);

-ada: 21 types (10,29%) y 9 hápax legómena (42,85%);

-ón: 12 types (21,82%) y 8 hápax legómena (66,7%).

22 Frase pegadiza popularizada por el personaje Peter Griffin en la sitcom estadounidense *Family Guy* y posteriormente por Sheldon Cooper en la sitcom *The Big Bang Theory* en el doblaje al español en España, como traducción de «Pow! Right in the kisser!» y «bazinga», respectivamente (https://es.wiktionary.org/wiki/zas,_en_toda_la_boca#:~:text=Expresi%C3%B3n%20que%20indica%20que%20se,palabras%20u%20opiniones%20de%20alguien).

23 En esta misma línea de razonamiento se expresa Filatkina (2018: 33), quien pone de manifiesto que «in the process of lexicon expansion, for example, a sporadic innovation only has a chance to enter into the lexicon if it is supported by a sufficient number of speakers, i.e. if the item is frequently used by them in a new form and / or meaning and function?».

ón de los corpus y el grado de *entrenchment*, se considera que una construcción está más *entrenched*, si se encuentra con frecuencia. Así pues, podríamos considerar *zas en toda la boca* como una microconstrucción y, por lo tanto, con estatus de nodo independiente dentro de la familia de construcciones SOMÁTICAS DE GOLPE (véase Sommerer / Baumann 2021, Sommerer 2022). Además, dentro del continuum léxico-gramática, nos encontraríamos ante una instancia cognitivamente fijada de una construcción fraseológica que se sitúa más cerca del polo léxico por estar más lexicalizada, convencionalizada y ser más idiomática.

1.3.3 RELACIONES ENTRE LAS CONSTRUCCIONES SOMÁTICAS: HERMANAS Y MADRES

En la sección 3.1. partíamos de la premisa que las construcciones no son entidades aisladas las unas de las otras, sino que más bien se organizan en redes estructuradas de construcciones. Para entender las relaciones que se establecen entre las construcciones nos tenemos que remitir a principios psicológicos relevantes que nos permitan explicar la organización de la lengua. Aducimos pues, al Principio de No Sinonimia de Goldberg (1995: 67): si dos construcciones son sintácticamente diferentes, tienen que ser semántica o pragmáticamente diferentes. Más concretamente, nos centramos en el corolario A, que dice así: si dos construcciones son sintácticamente diferentes y semánticamente sinónimas, no tienen que ser pragmáticamente sinónimas (véase Ungerer 2021: 394).

Traducido a nuestro objeto de estudio, hemos visto que desde un punto de vista sintáctico ambas construcciones difieren en el tipo de adjetivo intensificador que utilizan, conservando el adjetivo *todo* el determinante que precede al sustantivo somático que acompañan. De la misma manera, para hablar de la semántica de la construcción, gracias a la Semántica de Marcos, aducimos que nos encontramos ante el marco de la TRANSACCIÓN en el que hemos distinguido tres argumentos principales: el INFRACTOR que es quien provoca el daño a una parte del cuerpo (PARTE AFECTADA) con un objeto o parte del cuerpo (INSTRUMENTO DE DAÑO) (véanse ejemplos 14 y 15).

(14) La daga le alcanzó **en pleno pecho**, tomando por sorpresa al Ladrón, y un Orión cabreado le tomó por el cuello. (18678239130)

(15) Voy a terminar con una contusión cerebral, de la palmada tan fuerte que acabo de darme **en toda la frente**. (4622148116)

De la misma manera, desde un punto de vista pragmático, argumentábamos que era la función ilocutiva de la intensificación la que caracteriza ambas construcciones, entendida como una estrategia evaluativa del contenido proposicional. Esta función intensificadora, por lo tanto, refuerza la implicación del hablante en la comunicación e «imprime un grado mayor de compromiso con lo dicho que, en consecuencia, produce efectos a nivel comunicativo y social» (Albelda Marco 2007: 114). Desde esta perspectiva, ambas construcciones son sinónimas desde un punto de vista pragmático.

co al compartir la intensificación como función ilocutiva. Sin embargo, otros rasgos acerca del uso de las construcciones, «including register²⁴ (e.g. formal or informal), dialect variation and so on, are stated as part of the construction as well» (Goldberg 2006: 10). No es tanto en el *slot* nominal (véanse Tabla 1 y Tabla 4) donde encontramos en ambas construcciones sustantivos coloquiales tales como *jeta*, *cocorota*, *morro* o *cocorota* (véanse ejemplos 16 y 17) con frecuencias relativas similares ni en el *slot* verbal donde aparecen verbos coloquiales de manera residual tales como *zurrar* o *zumar* en la construcción con *pleno* (ejemplos 18 y 19),

(16) Yo estaba tan cabreado por la injusticia sufrida que me defendí asestándole una galleta **en plena jeta**, y nos liamos a forcejear. (4729238699)

(17) Que no te detenga el asco que producen las cucarachas pues aquí tienes la oportunidad de darles un buen martillazo **en toda la cocorota**. (13689415648)

(18) No sólo Maldonado me zumbó un zas **en toda la boca** ayer. Senna también. (7706393471)

(19) Es que la primera ola del primer plano me zurró bien **en toda la cara**. (4550398255)

sino más bien la notable presencia de la onomatopeya *zas* (1391 ocurrencias, o un 54,03%), así como su variante *zasca* (94 ocurrencias: 2,94%), las que sirven a modo de catalizador a la hora de encontrar la construcción con el adjetivo *todo* en contextos más coloquiales (véanse ejemplos 20 y 21). Desde una perspectiva construccionista se considera que determinados aspectos pragmáticos forman parte del conocimiento que aprenden los hablantes junto con las formas lingüísticas concretas (véase Kay / Fillmore 1999). Por lo tanto, este tipo de información se considera que es integrante del conocimiento convencional de la construcción y debe ser incorporado en las especificaciones construccionales, por ejemplo, como [regist col] para captar que se tratan de usos convencionales del empleo de un registro determinado de la construcción (Finkbeiner 2019: 172).

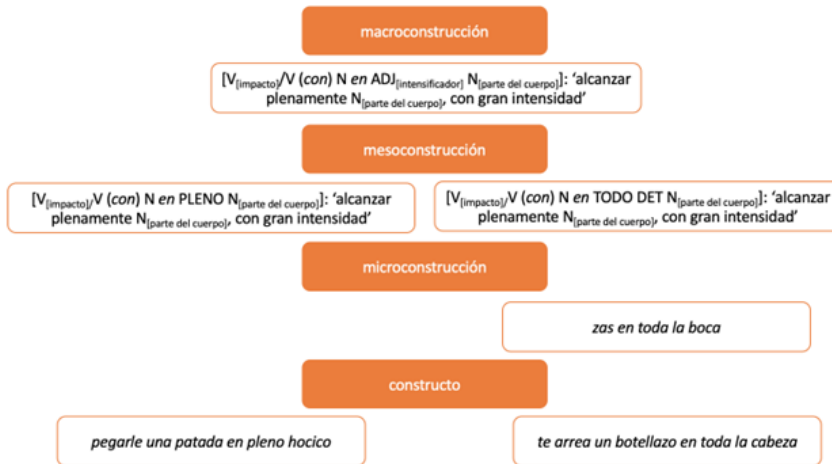
(20) [...] creo debían haber sido mas prudentes, porque pueden llevarse un buen zas **en toda la cara**. (4331544177)

(21) La respuesta del secretario general de Podemos ha sido lo que se dice un zasca **en toda la boca** con mucho, mucho éxito entre los tuiteros. (3837780312)

Partiendo de la premisa que «any quirk of a construction is sufficient to represent that construction as an independent node» (Croft / Cruse 2004: 263), postulamos dos construccionistas distintas [V_[impacto]/V (con) N en PLENO N_[parte del cuerpo]] y [V_[impacto]/V (con) N en TODO DET N_[parte del cuerpo]], que se sitúan en un mismo nivel de abstracción (relaciones horizontales, o construcciones hermanas), ya que son formalmente similares y presentan el mismo nivel de especificidad. Según la terminología de Traugott (2008: 236), nos encontraríamos ante mesoconstrucciones, es decir, «sets of similarly behaving specific constructions». Asimismo, en un nivel mayor de abstracción, estaríamos ante la macroconstrucción [V_[impacto]/V (con) N en ADJ_[intensificador] N_[parte del cuerpo]],

con el significado ‘alcanzar plenamente $N_{[parte\ del\ cuerpo]}$ con gran intensidad’. En niveles más específicos, podríamos postular la microconstrucción *zas en toda la boca*, a partir de la alta frecuencia *token*, así como el alto grado de *entrenchment* de la construcción. Finalmente, los constructos son instancias que encontramos en el discurso (véase Figura 1).

Figura 1: La familia de las construcciones SOMÁTICAS DE GOLPE.



1.4 CONCLUSIONES

El presente trabajo ha abogado por el estudio de la familia de construcciones SOMÁTICAS DE GOLPE desde un enfoque basado en construcciones, es decir, en tanto que emparejamientos indisolubles de forma y un significado. Partiendo de enfoque holístico basado en el uso, se adopta un tratamiento de igual entre los diferentes patrones gramaticales (es decir, no existen determinadas formas más centrales que otras) y, además, todas las dimensiones de la lengua (sintaxis, semántica, pragmática, discurso, morfología, fonología, prosodia) contribuyen de la misma manera a la hora de describir las expresiones lingüísticas (Fried 2015: 1000). Desde esta perspectiva, la noción de construcción se concibe como pieza gramatical idónea para captar las dimensiones de la lengua en una concepción holística, dado que las construcciones especifican una forma y una función.

A partir de una aproximación basada en corpus, hemos podido llevar a cabo un estudio minucioso de las construcciones $[V_{[impacto]}/V (con) N en PLENO N_{[parte\ del\ cuerpo]}$ y $[V_{[impacto]}/V (con) N en TODO DET N_{[parte\ del\ cuerpo]}$. Desde esta perspectiva, hemos analizado una serie de parámetros que nos han permitido vislumbrar las similitudes y las diferencias entre ambas construcciones somáticas, observando, en un primer paso, un mayor *entrenchment* de la construcción con el adjetivo *todo* (39,68%, respecto al 7,68% de la construcción con *pleno*). Con el vaciado de las instancias, hemos estudiado la variabilidad de los tres *slots* para determinar cuán productivos son en cada una de las construcciones. Respecto al *slot* somático, observamos que el patrón con

todo es más productivo, teniendo en cuenta que tenemos 0,027 *types* por cada ocurrencia licenciada por la construcción semiesquemática. No de la misma manera ocurre con el *slot* verbal: hay una menor disparidad de formas verbales (90 *types*) con el patrón *todo* en relación con los 99 *types* del patrón con *pleno*, aún siendo el primero la forma con más ocurrencias en los corpus. En esta misma línea, observamos una mayor diversidad de formas sustantivas de golpe: 260 *types* en el patrón con el adjetivo *pleno*, mientras que 255 ítems léxicos diferentes en el patrón con *todo*. El análisis pormenorizado de dicho *slot* nos permite constatar la restricción semántica predominante: nos referimos a los sustantivos que utilizan sufijos de golpe (-*azo*, -*ada* y -*ón*, por orden de frecuencia en ambas construcciones), hecho que incide en una mayor fijación cognitiva del golpe en el *slot*, es decir, un 56,35% en el patrón con el adjetivo *pleno* y un 26,24% en el patrón con *todo*.

En el caso de estas construcciones objeto de estudio con varios *slots*, un aspecto interesante que nos permite ver una metodología inductiva basada en corpus es el grado de saturación léxica de dichos *slots*. Así pues, podemos constatar que en la construcción con *pleno* hay una mayor especificidad léxica de la forma verbal con un 70,50%, en comparación con el 50,47% de la construcción con *todo*. Al revés ocurre con el *slot* nominal de golpe: en la construcción con *todo* hay una actualización del *slot* del 78,75%, mientras que en la construcción con *pleno* es del 56,07%. Estos porcentajes nos indican cómo se distribuye la actualización de los *slots* dependiendo de cada construcción (véase Herbst 2020), hecho que podría considerarse otro factor determinante a la hora de postular dos construcciones independientes.

Concluimos el trabajo con un boceto de la familia de construcciones SOMÁTICAS DE GOLPE a partir de la premisa que las construcciones no son entidades aisladas dentro del constructo mental, sino que interactúan entre ellas tanto en el eje vertical (relaciones de herencia o construcciones madres) como en el eje horizontal (construcciones hermanas).

BIBLIOGRAFÍA

- Albelda Marco, Marta. 2007. *La intensificación como categoría pragmática: revisión y propuesta*. Peter Lang.
- Audring, Jenny. 2019. Mothers or sisters? The encoding of morphological knowledge. *Word Structure* 12 (3), 274-296.
- Booij, Geert. 2002. Constructional Idioms, Morphology, and the Dutch Lexicon. *Journal of Germanic Linguistics* 14 (4), 301-327.
- Booij, Geert. 2010. Construction Morphology. *Language and Linguistics Compass* 4 (7), 543-555.
- Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford University Press.

- Bybee, Joan. 2013. Usage-Based Theory and Exemplar Representations of Constructions. En: Thomas Hoffmann / Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford University Press, 49-69.
- Croft, William / Cruse, Allan. 2004. *Cognitive Linguistics*. Cambridge University Press.
- Diessel, Holger. 2019. *The Grammar Network. How Linguistic Structure is Shaped by Language Use*. Cambridge University Press.
- DUE. Moliner, María. 2007. *Diccionario de uso del español*. Gredos.
- Filatkina, Natalia. 2018. Expanding the lexicon through formulaic patterns. The emergence of formulaicity in language history and modern language use. En: Sabine Arndt-Lappe / Angelika Braun / Claudine Moulin / Esme Winter-Froemel (eds.), *Expanding the Lexicon. Linguistic Innovation, Morphological Productivity and Lucidity*. De Gruyter, 15-42.
- Fillmore, Charles J. 1988. The Mechanisms of "Construction Grammar". *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35-55.
- Fillmore, Charles J. 1989. Grammatical Construction Theory and the Familiar Dichotomies. En: Rainer Dietrich / Carl Graumann (eds.), *Language Processing in Social Context. An Interdisciplinary Context*. Elsevier Science Publishers, 17-38.
- Fillmore, Charles J. 1999. Inversion and Constructional Inheritance. En: Gert Webehluth / Jean-Pierre Koenig / Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*. University Chicago Press, 113-128.
- Fillmore, Charles J. 2008. Border Conflict: FrameNet Meets Construction Grammar. En: Janet Ann DeCesaris, Elisenda Bernal (eds.), *Proceedings of the XIII EU-RALEX International Congress (Barcelona, 15-19 July 2008)*. Institut Universitari de Lingüística Aplicada, 49-68.
- Fillmore, Charles J. 2013. Berkeley Construction Grammar. En Thomas Hoffmann / Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford University Press, 111-132.
- Fillmore, Charles J. / Kay, Paul / O'Connor, Mary Catherine. 1988. Regularity and Idiomaticity in Grammatical Constructions. The Case of *Let Alone*. *Language* 64 (3), 501-538.
- Finkbeiner, Rita. 2019. Reflections on the role of pragmatics in Construction Grammar. *Constructions and Frames* 11 (2), 171-192.
- Fried, Mirjam. 2015. Construction Grammar. En: Tibor Kiss / Artemis Alexiadou (eds.), *Syntax – Theory and Analysis. An International Handbook. Volume 2*. Walter de Gruyter, 974-1003.
- García-Page, Mario. 2008. *Introducción a la fraseología española. Estudio de las locuciones*. Anthropos.

- Goldberg, Adele. 1995. *A Construction Grammar Approach to Argument Structure*. The University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at Work. The Nature of Generalization in Language*. Oxford University Press.
- Goldberg, Adele. 2019. *Explain me this: Creativity, Competition and the Partial Productivity of Constructions*. Princeton University Press.
- Gries, Stefan Th. 2013. Data in Construction Grammar. En: Thomas Hoffmann / Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford University Press, 93-110.
- Kay, Paul / Michaelis, Laura A. 2013. Constructional Meaning and Compositionality. En: Claudia Maienborn / Klaus von Stechow / Paul Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*. De Gruyter, 2271-2296.
- Herbst, Thomas. 2020. Constructions, generalizations, and the unpredictability of language. Moving towards collostruction grammar. *Constructions and Frames* 12 (1), 56-95.
- Hidalgo-Tenero, Carlos Manuel / Corpas Pastor, Gloria. 2021. La variación fraseológica: análisis del rendimiento de los corpus monolingües como recursos de traducción. *Études Romanes de Brno* 42(1), 359-379. [https://digilib.phil.muni.cz/bitstream/handle/11222.digilib/144095/1_EtudesRomanesDeBrno_51-2021-1_21.pdf?sequence=1]
- Hilferty, Joseph. 2004. *In defence of grammatical constructions*. Tesis Doctoral. Universidad de Barcelona. [<http://www.ub.edu/grelc/eng/?doctoral=in-defence-of-grammatical-constructions>]
- Hoffmann, Thomas. 2017. Construction Grammars. En: Barbara Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*. Cambridge University Press, 310-329.
- Hoffmann, Thomas / Horsch, Jakob / Brunner, Thomas. 2018. The more data, the better: A usage-based account of the English comparative correlative construction. *Cognitive Linguistics* 30 (1), 1-36.
- Ivorra Ordines, Pedro. 2021. *Les construccions comparatives intensificadores de la lletjor en català, espanyol, anglés i francès. Un estudi basat en corpus*. Tesis Doctoral. Universitat Pompeu Fabra. [<https://www.tdx.cat/handle/10803/672700>]
- Ivorra Ordines, Pedro / Mellado Blanco, Carmen. 2021. *Más tontos que el novio de la Chelo*. La intensificación de la estulticia en foros y chats por medio de comparaciones creativas: una aproximación desde la Gramática de Construcciones. *Estudios Románicos. La intensidad en las lenguas románicas como estrategia comunicativa* 30, 39-58. [<https://revistas.um.es/estudiosromanicos/article/view/471241/308431>]

- Kay, Paul / Fillmore, Charles J. 1999. Grammatical constructions and linguistic generalizations. *The What's X Doing Y? construction*. *Language* 75, 1-34. [<https://www.jstor.org/stable/417472>]
- Langacker, Ronald. 1991. *Foundations of Cognitive Grammar: Volume II: Descriptive Application*. Stanford University Press.
- Leclercq, Benoît. 2019. Coercion. A case of saturation. *Constructions and Frames* 11 (2), 270-289.
- López Meirama, Belén. 2016. *A tiros y a balazos: análisis construccional*. En: Gloria Corpas Pastor (ed.), *Computerised and Corpus-Based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Editions Tradulex, 340-348.
- Mellado Blanco, Carmen. 2021. Proyecto de investigación CONSTRIDIOMS: las construcciones fraseológicas del alemán y el español en contraste a través de los corpus. *Linred: Lingüística en la Red* 18. [http://www.linred.es/informacion_pdf/LR-informacion25-22052021.pdf]
- Michaelis, Laura A. 2003. Headless constructions and coercion by construction. En: Elaine J. Francis / Laura A. Michaelis (eds.), *Mismatch: Form-Function Incongruity and the Architecture of Grammar*. CSLI Publications, 259-310.
- Michaelis, Laura A. 2019. Constructions are Patterns and so are Fixed Expressions. En: Beatrix Busse / Ruth Moehlig-Falke (eds.), *Patterns in Language and Linguistics*. De Guyter, 193-220.
- Monge, Félix. 1972. Sufijos españoles para la designación de 'golpe'. *Homenaje a Francisco Ynduráin*. Facultad de Filosofía y Letras, 229-247.
- NGLE. Real Academia Española y Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*. Espasa.
- Nikiforidou, Kiki. 2009. Constructional Analysis. En: Frank Brisard / Jan-Ola Östman / Jef Verschueren (eds.), *Grammar, Meaning and Pragmatics*. John Benjamins, 16-32.
- Schmid, Hans-Jörg / Küchenhoff, Helmut. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24 (3), 531-577.
- Sommerer, Lotte. 2022. *Day to day and night after night*. Temporal NPN constructions in Present Day English. En: Lotte Sommerer / Evelin Keizer (eds.), *English Nouns Phrases from a Functional-Cognitive Perspective. Current Issues*. John Benjamins, 363-394.
- Sommerer, Lotte / Baumann, Andreas. 2021. Of absent mothers, strong sisters and peculiar daughters: The constructional network of English NPN constructions. *Cognitive Linguistics* 32 (1), 97-131.

- Stefanowitsch, Anatol / Gries, Stefan Th. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8 (2), 209-243.
- Traugott, Elizabeth Closs. 2008. Grammaticalization, constructions and the incremental development of language: Suggestions from the development of Degree Modifiers in English. En: Regine Eckardt / Gerhard Jäger / Tonjes Veenstra (eds.), *Variation, Selection, Development: Probing the evolutionary model of language change*. De Gruyter, 219-250.
- Traugott, Elizabeth Closs. 2018. Modeling Language Change with Constructional Networks. En: Salvador Pons Bordería / Óscar Loureda Lamas (eds.), *Beyond Grammaticalization and Discourse Markers. New Issues in the Study of Language Change*. Brill, 17-50.
- Van de Velde, Freek. 2014. Degeneracy: The maintenance of constructional networks. En: Rooney Boogart / Timothy Colleman / Gijsbert Rutten (eds.), *Extending the Scope of Construction Grammar*. De Gruyter, 141-180.
- Wiedemer, Marcos Luiz / Machado Vieira, Marcia dos Santos / Cezario, Maria Maura. 2019. Uma discussão sobre a relação entre variação e mudança na Gramática de Construções: entrevista com *Martin Hilpert*. *Diadorim* 21 (2), 30-43. [<https://revistas.ufrj.br/index.php/diadorim/article/view/30979/17761>]
- Yoon, Jiyoun / Gries, Stephan (eds.). 2016. *Corpus-based Approaches to Construction Grammar*. John Benjamins.
- Zeschel, Arne. 2012. *Incipient Productivity: A Construction-Based Approach to Linguistic Creativity*. De Gruyter.

CAPÍTULO 2

TAKING A LOOK AT THE SUPPORT VERB CONSTRUCTION *V A LOOK*: A DEMONSTRATION OF METHODOLOGY

Bert Cappelle¹ (Université de Lille)

Pâmela Fagundes Travassos (Federal University of Rio de Janeiro)

2.1 INTRODUCTION: HOW WE EXPRESS VISUAL PERCEPTION

When we talk about acts of looking, we frequently adopt expressions and constructions that we also use to deal with a rather different sort of event, namely that of people or objects moving from one place to another (Cappelle 2020; Cifuentes-Férez 2014; Gruber 1967; Kawachi 2020; Matsumoto 2001, 2020; Matsumoto et al. 2021; Talmy 1996, 2000, 2018). In English, for example, someone can *gaze around the room* or *look out (of) the window*, just like they can *pace around the room* or *jump out (of) the window*. For both kinds of events – merely looking or actually moving – we can combine a verb – either a perception verb or a motion verb – and a prepositional phrase (PP) expressing the trajectory of the immaterial gaze or of the material person. Similarly, you can *throw a glance at your phone* just like you could, say, *throw a lance at a drone*, combining in both cases a verb of ballistic motion, a noun phrase, and a PP indicating the ‘target’. Vision thus finds parallels in self-agentive and caused motion.

¹ University of Lille, Research Unit UMR 8163 STL, Univ. de Lille and CNRS, France.

You can also *get a glimpse* or *catch sight of something*, just like you can *get* or *catch* a ball that is thrown at you. In this alternative way of conceptualizing visual perception, the direction of motion is reversed, from the perceived object to the perceiver. When an object moves towards a person, it can change possession, as when a pop singer throws her jacket into the crowd and the person catching it can consider themselves the new owner. As can be expected, therefore, visual perception is often also expressed by means of expressions and constructions that involve notions of possession. Thus, apart from with the verb *get*, we also find visual perception encoded with *have* and *take* or by means of the ditransitive (double-object) construction, as in *give someone an angry look*, which, admittedly, is not just a matter of perceiving but also of making a facial expression. As is well known, the ditransitive construction is closely related to the so-called prepositional dative construction (as in *give something to someone*), which itself is structurally similar to the already-mentioned caused-motion construction, whereby the subject entity moves something to a certain goal or in a particular direction, as in *toss a startled glance towards the door*. In short, visual perception is frequently encoded in language in terms of the semantically interlinked concepts of motion, transfer, and possession.

At the same time, it should be clear that we can't take these expressions too literally. Consider the following authentic utterance, from the Corpus of Contemporary American English (Davies 2008-):

(1) Rita takes a forlorn glance towards the press area, where all of the other cameras are set up. (COCA)

This sentence mixes two different conceptualizations, that of the perceived object coming into the perceiver's possession (*take*) and that of the perceiver's gaze moving in the direction of (*towards*) the perceived object. Strictly speaking, these conceptualizations are incompatible. That (1) is not felt to be incomprehensible nonsense can only be seen as evidence that there is a great deal of semantic bleaching in the expressions used for visual perception: as language users, we treat the expressions, at least some of them, as somewhat dead metaphors. Moreover, while *giving* and *taking* are opposites when it comes to actual transfer, we find that different languages can use these opposites as default options for expressing visual perception. Consider English and Portuguese: English *take a look (at someone or something)* corresponds with (mainly Brazilian) Portuguese *dar uma olhada (em algo ou alguém)*, literally 'give a look (at someone or something)'. Finally, in English itself, *take a look* and *have a look* practically mean the same, while outside of these (and other support verb) expressions, *take* generally refers to the gaining of possession while *have* is the result of this event.

The first aim of this chapter is to get a grasp of the *have/take a look* support verb construction in English. A support verb construction is one where a general-purpose verb (a 'light verb'), which is mainly restricted to carrying tense and aspect information, combines with a noun phrase whose head is semantically richer than the verb; together, the verb and the noun phrase express a certain action (see, e.g., Danlos 1992; Langer 2004; Namer 1998). Further examples in English are *have a drink*, *have a chat*,

have a fight, take a bath, take a rest, give a scream, make a comment and do the washing-up. Our second and no less important aim is to demonstrate how we can map a relatively small area of lexico-grammar like this one by using a widely available corpus tool: the corpus interface created by Mark Davies.² We thereby hope to give students and linguists new to corpus linguistics a glimpse of how much can be achieved with relatively little import of tool-external processing of the data – although we will also give a practical introduction to some of the many things the statistical tool R has to offer. Finally, we will attempt to integrate our empirical findings with some theoretical reflection about how lexico-grammatical knowledge is organized.

2.2 TAKE A LOOK AND HAVE A LOOK IN SOME DICTIONARIES OF ENGLISH

In the introduction, we mentioned in passing that *take a look* and *have a look* have practically identical meanings. Indeed, if we consult some of the major dictionaries of English, the impression that is raised is that *have* and *take* can be used interchangeably with *a look*. Consider for instance in Figure 1 some of the information (clipped here) in the *Longman Dictionary of Contemporary English* (LDOEC Online), s.v. *look*:

Figure 1: Fragment of LDOEC Online, s.v. *look*

look² ●●● **S1** **W1** noun 🔊 🔊

1 LOOK AT [countable usually singular] an act of looking at something
have/take a look (at somebody/something)
 🔊 Let me have a look at that – I think it's mine.
 🔊 Take a **good look** at the photo and see if you recognize anyone in it.
 🔊 I **took one look** at the coat and decided it wasn't worth £50.
have/take a look around (also **have/take a look round** British English) (=look at all the things in a particular place)
 🔊 I have a special interest in old houses. Do you mind if I take a look around?

2 EXPRESSION [countable] an **expression** that you make with your eyes or face
give somebody a look
 🔊 Mike gave him such a severe look he didn't dare argue.
 🔊 Why has Jake been giving me **dirty looks** (=unfriendly looks) all morning?

3 CONSIDER [singular] an act of examining something and thinking about it
have/take a look (at somebody/something)
 🔊 Have you had a chance to take a look at my proposal yet?
 🔊 I asked the vet to have a **quick look** at the puppies as well.
 🔊 It's time to take a **fresh look** at the old problem of low pay.
 🔊 This month, take a **long hard look** (=examine very carefully) at where your money is going.

2 Another popular tool is SketchEngine, which allows you to search some very large corpora (<https://www.sketchengine.eu/>).

LDOEC Online provides several different meanings for the noun *look*. The first and the third meaning are expressed with either *have* or *take*; the second (which we also mentioned in the introduction) is combined with *give* and can combine with both the singular and the plural of *look*. The first meaning also manifests itself in the specific expression *have/take a look (a)round* (with *round* the preferred variant in British English), where again both support verbs are presented as alternatives. In the example sentences, some words are put in boldface, presumably because these are common combinations, known as ‘collocations’: *good look*, *took one look*, *dirty looks*, *quick look*, *fresh look* and *long hard look*. As for the latter, LDOEC Online actually also presents *take a (long) hard look at something/somebody* as a separate entry. This suggests that when *look* is modified by *hard* or *long hard*, the support verb will likely be *take*, not *have*. Something similar may be the case when the numeral *one* is used as a determiner: *take one look* could be more common than *have one look*.

As can be seen in Figure 2, somewhat similar information is provided by the *Oxford English Dictionary* (OED Online).

Figure 2: Fragment of OED Online, s.v. *look*, n.

d. colloquial. to have (also take) a look: to look at a person or thing, esp. (in later use) for the purpose of examination; to inspect or scrutinize someone or something. Also in extended use: to investigate or consider an idea, prospect, or other immaterial thing. Cf. *LOOK* *v.* 1a, *LOOK* *v.* 4a.

(a) With the object of one's gaze or consideration implied.

- 1673 J. DRYDEN *Assignment* III. i. 25 I must have a look before we go.
 1777 J. WHITFIELD *Conjectures on Tyndaris of Horace* 21 Stept to Lloyd's, to take a look.
 1830 J. P. COBBETT *Jrnl. Tour in Italy* 195 A young Roman lawyer..who is anxious, like myself, to have a look from the top of Mount Vesuvius.
 1894 T. W. HYDE *Following Greek Cross* xxv. 153 About this time, Generals Meade and Warren came up on the rocks to take a look.
 1920 P. GIBBS *Realities of War* iv. vii. 189 Come up and have a look, Jack.
 1980 B. JACOBY *Girl in Mirror* II. 65 *Doctor.* How is she? *Mother.* No change. *Doctor.* Let's have a look, shall we?
 2006 I. RANKIN *Naming of Dead* i. 13 I saw the sign yesterday... The one leading off the main drag. Decided I'd take a look.

(Hide quotations)

(b) With *at*. With the object of one's gaze or consideration specified. Cf. Phrases 2b(b).

- 1725 D. DEFOE *Tour Great Brit.* II. iii. 22 He would not go away without taking a look at the fine Gardens.
 1769 G. S. CAREY *Shakespeare's Jubilee* 16 Let's have a look at you before you go.
 1831 *New-Eng. Mag.* Dec. 494 A great Irishman was brought up into the operating theatre, the other morning, for the class to have a look at.
 1885 M. COLLINS *Prettiest Woman in Warsaw* I. ii. 29 In the meantime I shall have a look at Warsaw.
 1925 A. LOOS *Gentlemen prefer Blondes* iv. 109 So then Mons. Broussard stopped crying and put on his glasses and took a good look at us.
 1960 J. IGGULDEN *Storms of Summer* 169 I'll get the quack at the Bush Hospital to have a look at it in the morning.
 2012 *Action on Hearing Loss* Aug. 24 The government will soon replace Disability Living Allowance (DLA) with Personal Independence Payment (PIP) as part of its radical welfare reforms—we take a look at how this may affect you.

(Hide quotations)

(c) With adverbs or prepositional phrases, forming phrases corresponding to phrasal verbs at *LOOK* *v.*

Esp. in **to have (also take) a look around (or round).**

- 1829 *Blackwood's Edinb. Mag.* July 77/2 You'd bether just take a bit of a look round on the worlt, an' bid it good by.
 1898 H. G. WELLS *Jimmy Goggles God in Cosmopolitan* Nov. 109/1 I stopped once, and..had a look up.
 1913 *Bull. Geogr. Soc. Philadelphia* 11 11 On approaching a strong rapid we usually took a look ahead to ascertain where the best line of advance lay.
 1925 L. H. MYERS *Clio* (1945) vii. 49 I had a look through the window.
 1932 J. BUCHAN *Gap in Curtain* ii. 97 He hoped, while in the country, to have a look round.
 a1965 W. G. PURCELL *St. Croix Trail Country* (1967) 97 I stepped over to have a look down from the edge of the hill.
 1982 P. REDMOND *Brookside* (Mersey TV shooting script) (O.E.D. Archive) Episode 4. 70 Just take a look around and see how much is going on the side.
 2011 A. LEWIS *Everybody Jam* xiv. 99 I was having a quick look around.
 2011 *Independent* 4 Aug. 13/1 Punk's not dead—it's just having..a look round a stately home.

(Hide quotations)

We can observe a few differences here with what the LDOEC presents. First, the OED makes a distinction between the use of *have/take a look* on its own (where the

target of the gaze remains implicit) and *have/take a look* combined with an *at*-PP. Second, the OED does not consider the third meaning listed by LDOEC, that of “examining something and thinking about it”, as a separate meaning but does mention this as a specific purpose of looking that arose as a later and extended use. Third, the expression *have/take a look (a)round* is given special prominence. The adverb *(a)round* is treated as identical to the particle in phrasal verbs with *look* as a verb (*to look (a)round*) and as essentially a very common choice alongside other adverbs or PPs that can occur (e.g. *take a look ahead, have a look through the window, take a look round a stately home*). Fourth, the expression *give somebody a look* is not mentioned separately. However, elsewhere in the entry of *look* as a noun (though this is not shown in Figure 2), the OED lists the meaning “[a]n expression of a specified thought or feeling by looking”. The verb *give* is not seen here as one that is privileged to express that meaning: the quotations contain also the verb *cast* or *flash*, as well as other grammatical environments (e.g. *He was regarding her with a tender, worried look*, an example from 1956). In its presentation of this use, the OED also mentions two collocations: *come-hither look* and, as in LDOEC Online, *dirty look*. Fifth, the OED presents *have* as the basic option: by mentioning “*also take*” between brackets, it suggests that *have* is more common.

Finally, let’s consider one more dictionary, *Collins English Dictionary* (Collins Online). In Figure 3, we show a part of the dictionary entry of *look*.

Figure 3: Fragment of Collins online, s.v. *look*

1. **VERB**

If you **look** in a particular direction, you direct your eyes in that direction, especially so that you can see what is there or see what something is like.

I looked down the hallway to room number nine. [VERB preposition/adverb]

She turned to look at him. [VERB preposition/adverb]

He looked away, apparently enraged. [VERB preposition/adverb]

If you look, you'll see what was a lake. [VERB]

Synonyms: see, view, consider, watch [More Synonyms of look](#)

Look is also a noun.

Lucille took a last look in the mirror.

Assisi has a couple of churches that are worth a look if you have time.

2. **VERB**

If you **look at** a book, newspaper, or magazine, you read it fairly quickly or read part of it.

You've just got to look at the last bit of Act Three. [VERB + at]

Look is also a noun.

A quick look at Monday's British newspapers shows that there's plenty of interest in foreign news. [+at]

3. VERB

If someone, especially an expert, looks at something, they examine it, and then deal with it or say how it should be dealt with.

Can you look at my back? I think something's wrong. [VERB + at]

[Also VERB]

Look is also a noun.

The car has not been running very well and a mechanic had to come over to have a look at it. [+ at]

4. VERB

If you **look at** someone in a particular way, you look at them with your expression showing what you are feeling or thinking.

She looked at him earnestly. 'You don't mind?' [VERB + at]

Look is also a noun.

He gave her a blank look, as if he had no idea who she was.

Sally spun round, a feigned look of surprise on her face. [+ of]

Noticeable differences are that there are no separate senses mentioned for *look* as a noun. Instead, the meanings we have already encountered are mentioned under the verb uses, where we just read: “*Look* is also a noun”. This might appear as though the noun use is not given proper credit, but one could argue that this choice underscores the linguistic insight that *look* is really part then of a support verb construction, with the noun occurring with the same range of meanings that the related verb can express. As in LDOEC Online, the ‘examine’ meaning is treated separately (here in sense 3). In addition, there’s a special sense in which *look*, both as verb and as noun in the support verb construction, conveys the idea of viewing something quickly or partially (sense 2). The sense of looking in a particular way to reveal your thoughts or feelings is again mentioned separately (sense 4), and just like in OED Online, the verb *give* is not treated as one that is needed to trigger that interpretation, although we can see that the first example sentence does contain that verb.

We could look at more dictionaries and we would find further differences. For now, our perusal of the three dictionaries consulted allows us to distill a number of questions that should be possible to answer by means of corpus research:

Is *have a look* more frequent than *take a look*? Could there be regional preferences (e.g. British English versus American English)?

Are there specific collocations where one support verb is more common than another (e.g. *take a good look at something*)?

Related to the preceding: are there specific meanings or functions for which one support verb is more common than another?

For the meaning relating to a way of looking that is expressive of something (e.g. *a mournful look*), is the verb *give* the most typical environment?

The message we want to give here is that by just looking at information in dictionaries (the meanings as well as the examples), we can come up with a number of

empirical questions that we otherwise might not have thought of. In the next section, we will demonstrate how such questions can be tackled with the help of corpora.

2.3 GETTING A GRIP OF REAL DATA FOR *LOOK* WITH ENGLISH-CORPORA.ORG

2.3.1 GLOWBE: REGIONAL VARIATION AT A GLANCE

Can we find out whether different varieties of English – American English, British English, Australian English, etc. – have different preferences for *have a look* versus *take a look*? The answer is yes. One laborious way would be to consult several variety-specific corpora, such as the Corpus of Contemporary American English and the British National Corpus, or different corpora in the International Corpus of English (ICE) family, one after the other, and then compare the results obtained from each corpus. Fortunately, there is a much faster way, thanks to a corpus that provides data about these and many more regional varieties in one place: the Global Web-based English corpus, or GloWbE. This is housed at www.english-corpora.org. Go to that website and you will get an overview of English corpora, as displayed in Figure 4. (For ‘monitor corpora’, which keep growing with new material up to the present date, the numbers you will find for the corpus size will be larger than in this figure.)

Figure 4: Overview of corpora housed at English-corpora.org



The most widely used online corpora [guided tour](#), [overview](#), [search types](#), [variation](#), [virtual corpora \(quick overview\)](#), [BYU](#)[†].

The links below are for the online interface. But you can also [download](#) the corpora for use on your own computer.

Corpus (online access)	Download	# words	Dialect	Time period	Genre(s)
IWeb: The Intelligent Web-based Corpus	↓	14 billion	6 countries	2017	Web
News on the Web (NOW)	↓	12.9 billion+	20 countries	2010-yesterday	Web: News
Global Web-Based English (GloWbE)	↓	1.9 billion	20 countries	2012-13	Web (incl blogs)
Wikipedia Corpus	↓	1.9 billion	(Various)	2014	Wikipedia
Corpus of Contemporary American English (COCA)	↓	1.0 billion	American	1990-2019	Balanced
Coronavirus Corpus	↓	1074 million+	20 countries	Jan 2020-yesterday	Web: News
Corpus of Historical American English (COHA)	↓	475 million	American	1820-2019	Balanced
The TV Corpus	↓	325 million	6 countries	1950-2018	TV shows
The Movie Corpus	↓	200 million	6 countries	1930-2018	Movies
Corpus of American Soap Operas	↓	100 million	American	2001-2012	TV shows
<hr/>					
Hansard Corpus		1.6 billion	British	1803-2005	Parliament
Early English Books Online		755 million	British	1470s-1690s	(Various)
Corpus of US Supreme Court Opinions		130 million	American	1790s-present	Legal opinions
TIME Magazine Corpus		100 million	American	1923-2006	Magazine
British National Corpus (BNC) *		100 million	British	1980s-1993	Balanced
Strathy Corpus (Canada)		50 million	Canadian	1970s-2000s	Balanced
CORE Corpus		50 million	6 countries	2014	Web
<hr/>					
From Google Books n-grams (compare)					
American English		155 billion	American	1500s-2000s	(Various)
British English		34 billion	British	1500s-2000	(Various)

Most of these corpora have been developed and provided a uniform interface for by Mark Davies, a now-retired professor of linguistics at Illinois State University and later Brigham Young University. A few of the corpora, like the British National Corpus, pre-existed but are housed here so that they can be searched within the same interface. The website states that more than 130,000 people, in over 140 different countries, use these corpora on a monthly basis. If you are a first-time user, it is highly recommended that you take the guided tour (click on the link on top to open a pdf document). For now, we hope we can give you an idea of what the corpora and the online interface have to offer by showing you some concrete steps to be taken. Mark Davies also developed corpora for Spanish and for Portuguese (Davies 2016-a, b), for which he designed a similar interface (for documentation, see Davies 2008).

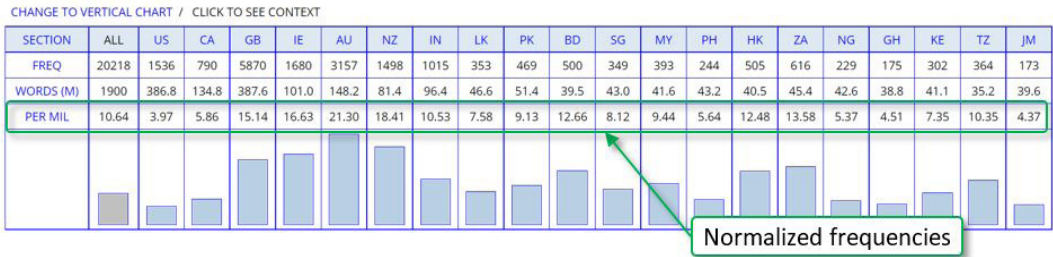
To access GloWbE, click on the name of the corpus in the list. In the search window, enter “HAVE a look” as a command (without any quotation marks), as shown in Figure 5.

Figure 5: Entering *have a look* as a string of lexical items, with *have* treated as a lemma

Capitals are used for the verb *have* to search for all inflected/contracted forms at once (*have, has, had, having, 've*). In other words, capitals force a lemma search, not a word form search. An alternative way of doing a lemma search is by enclosing the lexical item in square brackets: “[have] a look”. If you were to click “Find matching strings”, you would get to see the search results as a breakdown of the different forms of *have* combined with *a look*, with absolute frequencies in twenty different ‘World Englishes’. (You would then see that the string *'ve a look* has just one occurrence in British English and another single one in Honk Kong English.) As we’re mainly interested in discovering the relative frequencies of the expression *have a look*, whatever its inflected form, in different varieties of English, let us instead select the Chart function at the top (see Figure 6).

Figure 6: Retrieving a bar chart

If we now click “See frequency by section”, we obtain a bar chart of the ‘collapsed’ frequencies (i.e., with all the forms of *have* combined, as we did a lemma search) per variety of English (see Figure 7).

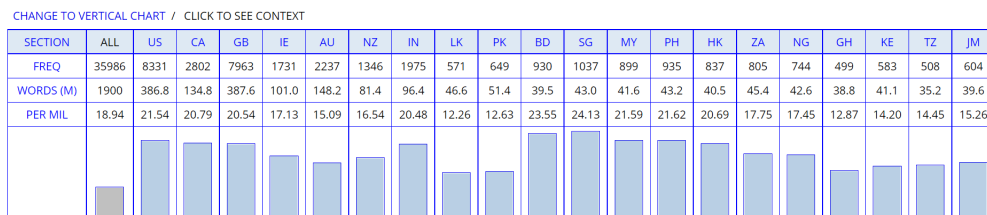
Figure 7: Bar chart of frequencies of *have a look* in twenty varieties of English worldwide

We can change the graph, if we so desire, to a so-called ‘vertical chart’, where the varieties are presented one under the other and the bars are then, in fact, horizontal. Looking at these results, we see that *have a look* is most frequent in Australian English (AU) and New Zealand English (NZ), as well as in Irish English (IE) and British English (GB). The string is also common in South African English (ZA), Bangladeshi English (BD) and Hong Kong English (HK). It will not have escaped the reader that these are all varieties spoken in countries with strong historical, cultural and economic ties, given Britain’s colonial past: countries that were part of the British Empire, which later developed into the Commonwealth of Nations, typically use the ‘British’ variant, in cases where American English and British English differ. In this case, indeed, it’s clear that we have relatively lower bars for American English (US) and Canadian English (CA), as well as for other varieties that are under ‘American’ influence, such as Philippine English (PH) and Jamaican English (JM).

There are three rows of numbers above the bars. The top row shows the absolute frequencies of the string in each regional variety. Thus, we can see that there are 5,870 tokens of this string in the British English component of GloWbE. The row underneath tells us how many words each component contains in total. For British English, this is more than 380 million words. As not all components are equally large – Tanzanian English ‘only’ has 35.2 million words – we need to know how often the string occurs as a proportion of the total size of the corpus. That is why the lowest row is most revealing: it shows the ‘normalized’ frequency of the string per million words in each corpus. It is these numbers that are represented by the bars. The first column, with the bar in grey, gives absolute and normalized frequencies of *have a look* in the entire corpus. The cell in the middle row also shows, as we could already see in Figure 1, that there are ‘1900 million’ words in the entire GloWbE, or indeed 1.9 billion words. Let’s remember that *have a look* has a frequency of 10.64 per million words in English worldwide.

We can now repeat the same procedure for *take a look*. To do this, first click on the “search” option which you’ll find above the results in the current window. This will bring you back to the screen with the search window we saw in Figures 5 and 6. Enter “TAKE a look”. The bar chart obtained looks different (see Figure 8).

Figure 8: Bar chart of frequencies of *take a look* in twenty varieties of English worldwide



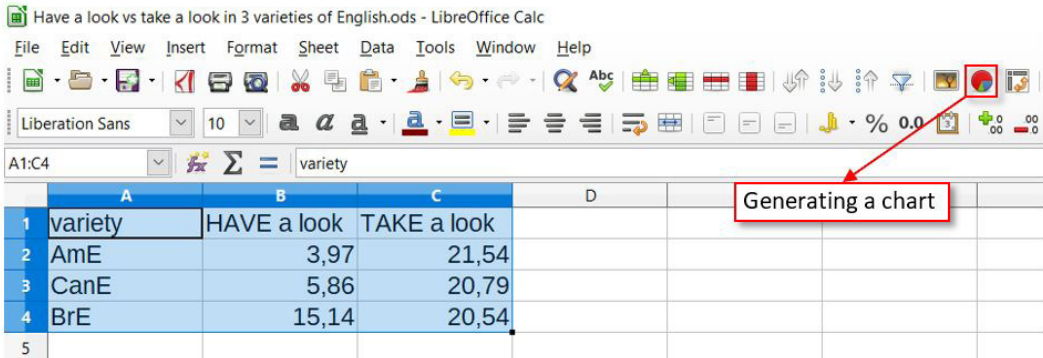
We can observe that the differences in relative frequencies across regional varieties are less pronounced for this expression. Crucially, the variant with *take* is overall more frequent than the variant with *have*: the string *take a look* occurs 18.94 times per million words, which approaches twice the frequency for the string *have a look*. This means that the OED has it backwards: it should more appropriately have written “*to take (also have) a look*”. In particular, in American English and Canadian English, where we saw that *have a look* is fairly infrequent, the frequencies are much higher for *take a look*. Even in Bangladeshi English, which we initially assumed followed British English, there is a clear preference for *take a look*.

If we want to present these results as part of a research report, we cannot just copy-paste screen grabs as we have done here. Instead, we should transfer the numbers of interest to a table in a word processor application. Numbers can also be entered in a spreadsheet application, such as Microsoft Excel or a free alternative such as LibreOffice Calc (see Figure 9).

Figure 9: Some selected corpus frequencies hand-entered in LibreOffice Calc

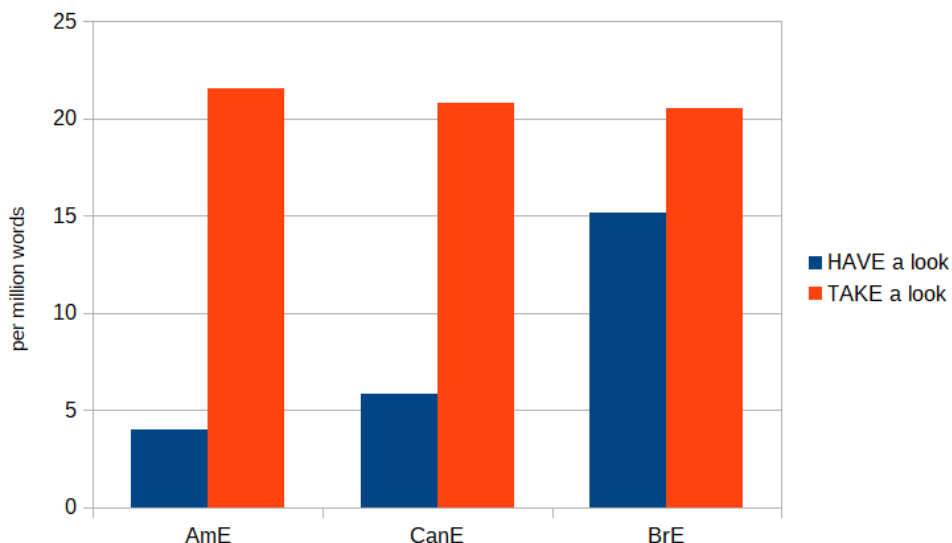
	A	B	C
1	variety	HAVE a look	TAKE a look
2	AmE	3,97	21,54
3	CanE	5,86	20,79
4	BrE	15,14	20,54

Selecting all the cells and then clicking the chart button (see Figure 10) will generate a graph.

Figure 10: Generating a bar chart and providing extra labels in LibreOffice Calc

If you are not happy with the standard color scheme in LibreOffice, you can change this by double-clicking anywhere on the chart, then right-clicking on the first series and selecting “Format Data Series...”. In the Area tab, a range of more palatable colors can be selected, including ten shades of gray. Each pair of bars can be changed this way. For our example, though, we have stuck with the standard colors.

The resulting table can then be selected and exported to your word processor application, as we have done here (see Figure 11). It might be preferable not to add any title within LibreOffice and instead add one in your word processor application, as this allows greater flexibility in terms of font size and the like. Moreover, it is not even necessary to use a separate spreadsheet application, as common word processor applications allow you to insert a chart directly into your document. Clicking the chart icon (which, depending on the application you use, looks like a pie chart icon or a bar chart icon) will then also allow you to make or adapt a data table, much like the one in Figure 9 above.

Figure 11: A simple bar chart generated via LibreOffice Calc

Now that we have this figure, two observations can be made very clearly. First of all, *take a look* is more frequent than *have a look* in all three varieties shown here. Secondly, this difference in frequency is much more pronounced in AmE and CanE than in BrE.

2.3.2 LOOKING UP COLLOCATIONAL PREFERENCES: WHICH ADJECTIVES ARE ATTRACTED TO WHICH VARIANT?

It will be remembered that *look* can be modified by one or more adjectives (*dirty, fresh, good, quick, long hard, etc.*). It would not come as a big surprise if the choice of *have* versus *make* was dependent on the occurrence of a particular adjective. How can we find out with the help of a corpus? This is possible with the “Compare” function in any of the corpora at English-corpora.org. Let’s use, for a change, the British National Corpus (BNC) (Davies 2004). In Figure 12, we show which information to enter.

Figure 12: Comparing *have a(n) ... look* and *take a(n) ... look* with respect to the adjectives that occur in them

The screenshot shows the 'Compare' function interface. At the top, there are tabs: 'List', 'Chart', 'Collocates', 'Compare' (highlighted), and 'KWIC'. Below the tabs, there are two input fields for 'Word1' and 'Word2', both containing 'HAVE a|an ADJ look' and 'TAKE a|an ADJ look' respectively. Below these is an 'ADJ' field containing 'ADJ' and a 'Collocates [POS]' field. A horizontal bar with numbers from -4 to +4 is shown, with the '1' position highlighted in green. There are 'Compare words' and 'Reset' buttons. At the bottom, there are checkboxes for 'Sections', 'Texts/Virtual', 'Sort/Limit', and 'Options'.

Mastering the Collocates and the Compare functions in English-corpora.org is an acquired skill and takes quite a bit of trial and error. Here, we use capital letters for the verbs, as before, but not for the following article, even though we would like to retrieve both *a* and *an*. However, these two words are not treated in the corpus as forms of a single lemma. We solve this by using the vertical slash symbol; this way, the interface understands that we look for “*a* or *an*”. Next up is a word that should be an adjective, so we type ADJ. If you don’t know the part-of-speech abbreviation for a given word class, click on “[POS]?” next to the search and then again click “POS” or “Insert PoS” to obtain a menu of options. We end our string with the word *look*.

Underneath the windows for the two ‘words’ – in our case, these are word combinations – we have to specify which collocates we want to retrieve. Collocates are words that often occur nearby another word or phrase. In the Collocates search window, we again type ADJ, as that is the kind of word we want to look for inside the string. In fact, though, since we already used ADJ in the string itself, we do not strictly have to fill in this window. If we leave it empty, “*” will appear automatically before we click “Compare words”. The asterisk is the ‘wildcard’ that stands for any lexical item of any word class. The bar with numbers allows us to specify how far to the left and/or to the right we want our collocates to be searched. Here, we don’t want it to be searched many words before or after the search strings. In fact, we want it to be searched *within* the search strings! Somewhat unpredictably – hence our insistence on trying things out a few times – it is the word *look* in each of the two search strings that is treated as the main context words. That is why we select the first spot to the left of our “Word 1” and “Word 2” as our position of interest. Note that for other types of searches, we can select not just a single position but a whole range, even up to nine words to the left and/or to the right of the central word. (For a span of more than four words, click once or twice on the plus sign.)

The results we obtain, after clicking “Compare words”, are displayed in Figure 13. They are quite revealing.

Figure 13: Two output tables of adjectives that are (or aren't) good collocates of *have a(n) ... look* compared to *take a(n) ... look*, and vice versa

SEE CONTEXT: CLICK ON NUMBERS (WORD 1 OR 2) [HELP...]
 SORTED BY RATIO: CHANGE TO FREQUENCY
 WORD 1 (W1): HAVE A[AN] * LOOK (0.91) WORD 2 (W2): TAKE A[AN] ADJ LOOK (1.10)

	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	LITTLE	25	0	50.0	55.1	1	FRESH	26	2	13.0	11.8
2	WEE	12	0	24.0	26.5	2	CRITICAL	12	1	12.0	10.9
3	PROPER	7	1	7.0	7.7	3	LONG	6	0	12.0	10.9
4	BETTER	9	3	3.0	3.3	4	CAREFUL	11	1	11.0	10.0
5	QUICK	55	23	2.4	2.6	5	HARD	11	1	11.0	10.0
6	GOOD	69	41	1.7	1.9	6	HONEST	5	0	10.0	9.1
7	CLOSE	11	25	0.4	0.5	7	DETAILED	5	0	10.0	9.1
8	CLOSER	21	65	0.3	0.4	8	CLOSER	65	21	3.1	2.8
						9	BRIEF	11	4	2.8	2.5
						10	CLOSE	25	11	2.3	2.1
						11	GOOD	41	69	0.6	0.5
						12	QUICK	23	55	0.4	0.4

The output of our search contains two tables. The table on the left contains a rank-ordered list of words that go together very well (towards the top, in green) or not very well (towards the bottom, in a reddish hue) with *have a(n) ADJ look*, compared to the competing environment *take a(n) ADJ look*. Remember that we set up the search in such a way that these words are adjectives appearing in the slot before the noun *look*. By the way, at the top of the table on the left, “ADJ” in Word 1 is shown as “J*”, which is just a quirk of the underlying corpus machinery – and another reason why, for research reporting purposes, it is best to produce your own table. The table on the right then shows how strongly adjectives are preferred by *take a(n) ADJ look* over the alternative with *have*.

Let’s take a more detailed (!) look at how we ought to read these tables. The number between brackets after our ‘word 1’ (W1) and ‘word 2’ (W2) – which, again, in our case are really more like lexical environment than single words – are the overall ratios of these two words. Here, there are 0.91 occurrences of *have a(n) ADJ look* in the British National Corpus for every occurrence of *take a(n) ADJ look*; conversely, there are 1.1 occurrences of the latter phrase for every occurrence of the former. Put simply, the variant with *have* is a little less frequent in British English than the one with *take*, confirming what we saw in Section 3.1 with data from GloWbE. The columns headed by W1 and W2 show the corpus frequencies for the words involved. W1/W2 is the ratio of the first environment to the second for each adjective in it. Thus, *quick* occurs 2.4 times as frequently with the *have* variant as with the *take* variant. As Mark Davies explains in the notes when you click on “[HELP...]”, when a competing variant occurs zero times, division by 0 is avoided by giving that variant a frequency of 0.5 instead of 0. That is how we get 50.0 as the W1/W2 ratio for *little*, which occurs 25 with *have* and 0 times, artificially set to 0.5 times, with *take*. The score in the right-most column of each table, finally, takes the overall relative frequency of the environment into account to either boost or attenuate the W1/W2 ratio or W2/W1 ratio. Thus, given that the *have* variant is less frequent than the *take* variant (see again its 0.91 overall ratio), if an adjective nonetheless prefers *have*, this is all the more surprising. In our case, the difference in frequency of the two competing environments is quite small, but as you

can imagine, for other cases of variation the score for a collocate might be wildly different from just its W1/W2 ratio. The score is calculated by dividing the W1/W2 ratio by the overall ratio of the host environment.

What the corpus allows us to do here is similar in spirit to a ‘distinctive collexeme analysis’, a method popular among corpus linguists to measure how strongly words are attracted to, or repulsed by, one particular construction compared to another one with a roughly equivalent meaning (Gries and Stefanowitsch 2004). However, distinctive collexeme analysis doesn’t just look for each collocate how often it occurs in one environment rather than another, competing environment, but also how often it occurs compared to all other collocates. Ultimately, it is a slightly more laborious kind of analysis than what Mark Davies’s corpus interface allows us to see.

So, what is it that we can see? The collocating adjectives reveal that the two support verb constructional alternatives are far from perfectly synonymous. Apparently, with *have* as a support verb, *look* is more often modified by ‘diminutive’ adjectives such as *little*, *wee* and *quick*, suggesting that *having a look at something* seems to be a perfunctory, fleeting action. With *take*, we find adjectives that suggest a more concentrated and considered, or a more specific way of looking: *fresh*, *critical*, *long*, *careful*, *hard*, *honest*, *detailed*, etc. Clearly, *taking a look at something* typically describes a cognitively more demanding action. It seems as though some of the active meaning of the lexical verb *take* – its dynamic and agentive semantics – is preserved in how we use it as a support verb. However, this could well be some wishful linguistic thinking. After all, there’s nothing very dynamic about *take a nap*. In any case, with *take*, the noun *look* is often synonymous to *perspective*, *view* or *approach*. This shows again that the meaning of *take a look* is not purely a matter of physically directing your eyes towards something.

The following two examples are quite representative examples of each variant of this support verb construction, in context:

(2) What are we doing? Well let’s let’s just **have a quick look** at this one. What’s this describing here? V equals large V or V nought or whatever you like, [unclear – cough] E to the minus R T over L. (BNC)

(3) The economic recession has caused European governments to **take a fresh look** at the extent to which a degree of interdependence in the past has been translated into a heavy dependence of the poorer areas on the richer north. (BNC)

Sentence (2) was taken from a tutorial lesson in the corpus. Looking at the context, the tutor *does* deal with contents for which one needs to have one’s wits together. It could be argued that the tutor tries to put the student at ease by merely *presenting* the action as something that doesn’t involve a lot of effort, as if to say: “This one will be easy for you to understand.”

2.3.3 ‘CODING’ OR ‘ANNOTATING’ A SAMPLE OF CORPUS EXAMPLES BY HAND AND THEN PROCESSING YOUR FINDINGS WITH R

Grasping the full meaning of an expression goes beyond looking at the collocating lexical items. A fuller study should involve the following three steps:

1. getting one or more samples of sentences
2. analyzing the sentences with respect to one or several variables
3. processing the results with a tool for statistical analysis

2.3.3.1 Retrieving a sample of authentic sentences

So, first, you should retrieve a sufficiently large sample of authentic utterances from the corpus. To do this, use the List function, which is selected by default, and type in “HAVE a|an ADJ look” in the search window. In the results window, you can then select 100, 200 or 500 examples (see Figure 14).

Figure 14: Asking for a random sample of one of the competing forms

HELP	ALL FORMS (SAMPLE)	FREQ	TOTAL 303 UNIQUE 100 +
1	<input type="checkbox"/> HAVE A GOOD LOOK	45	
2	<input type="checkbox"/> HAVE A QUICK LOOK	37	
3	<input type="checkbox"/> HAVE A LITTLE LOOK	22	
4	<input type="checkbox"/> HAD A GOOD LOOK	20	
5	<input type="checkbox"/> HAVE A CLOSER LOOK	18	
6	<input type="checkbox"/> HAD A QUICK LOOK	17	
7	<input type="checkbox"/> HAVE A WEE LOOK	11	
8	<input type="checkbox"/> HAVE A BETTER LOOK	8	
9	<input type="checkbox"/> HAD A CLOSE LOOK	6	
10	<input type="checkbox"/> HAVE A CLOSE LOOK	5	
11	<input type="checkbox"/> HAVE A PROPER LOOK	5	
12	<input type="checkbox"/> HAVING A LITTLE LOOK	3	

Select a random sample

The result of this action is what’s known as a set of ‘concordance lines’ (see Figure 15).

Figure 15: Part of a random sample of 100 retrieved concordance lines

CLICK FOR MORE CONTEXT		EXPLORE NEW FEATURES	SAVE	TRANSLATE	ANALYZE
1	KBC S_conv	🔍🔍🔍	🔍	🔍	🔍
need it. (SP:P51AG) Cos I just, I just want to have a quick look through it before I gave it (SP:P51AD) I'll bring it over after, as					
2	KBF S_conv	🔍🔍🔍	🔍	🔍	🔍
couldn't get over (pause) you know, I say, I had a quick look the other at Marks (pause) you know, to even do the Christmas serviettes (pause)					
3	KBW S_conv	🔍🔍🔍	🔍	🔍	🔍
(SP:P5087) (reading) The magician was fast asleep. Tim leaned over to have a good look at him, but his foot slipped and he fell in. He landed right					
4	KCY S_conv	🔍🔍🔍	🔍	🔍	🔍
I think there might be yeah (SP:P50H9) Yeah (SP:P50H9) I ain't had a proper look I just put it on and said, work a few routes out you know					
5	KD5 S_conv	🔍🔍🔍	🔍	🔍	🔍
, still (unclear) wet! (SP:P50K6) (unclear)! (SP:P50JX) (laugh) (pause) (laughing) Had a worried look on her face then! (SP:P50K6) Oh darling! (SP:P50JY) Sorry (ur					
6	KD8 S_conv	🔍🔍🔍	🔍	🔍	🔍
really nice Orion, a J reg. Anyway when they (pause) had a good look at it the front half was an Escort (pause) and the back half was an					
7	KDL S_conv	🔍🔍🔍	🔍	🔍	🔍
yes (unclear)2. (SP:P50NY) Some more over there, do you want to have a quick look? (unclear) (pause) (SP:P50NX) Trying to think what Penny would like, but					
8	KDV S_conv	🔍🔍🔍	🔍	🔍	🔍
round to the toys. (SP:P50SB) Go on then. You can have a quick look at the toys. (SP:P50SC) With you. (SP:P50SB) Can't afford to buy anything					
9	KE0 S_conv	🔍🔍🔍	🔍	🔍	🔍
the one (unclear) your dinner? (SP:P50SX) Oh ay! He was having a good look then I'd say! (SP:P50SY) (unclear) (unclear) (SP:P50SX) Didn't look at as though					
10	KE6 S_conv	🔍🔍🔍	🔍	🔍	🔍
it's my turn to buy the (pause) draw prizes. I had a quick look round yesterday but I ain't got a clue! (SP:P50X8) And what what (SP:P50XA)					
11	KPA S_conv	🔍🔍🔍	🔍	🔍	🔍
? (SP:P553C) (----) (SP:P553H) (----) no (pause) (SP:P553C) (unclear) and if we have a close look hey there's nobody in (SP:P553H) (laugh) (pause) we appear to					
12	JP3 S_meeting	🔍🔍🔍	🔍	🔍	🔍
it up on the machine to change again quickly but I've had a quick look at the situation with regard to the five per cent increase. (SP:P54GX) Yeah.					

Inspecting the sentences is important. Line 5 in the sample shows that there can be ‘noise’ in the data. This is not an example of the construction we’re interested: *a worried look* here refers to someone’s appearance. When we analyze a manageable amount of data in more detail, we have to sift out such irrelevant examples. For very large amounts of data, like those we obtained from GloWbE in Section 3.1, eliminating noise may not be practically feasible and we may just have to live with a certain degree of imprecision. Yet, we could always additionally take a small sample – say, 100 examples – and find out how precise the search is, that is, what proportion of correctly retrieved instances we have for our corpus query. We can then report on this precision rate to give the reader an idea of how reliable the graphs are. This solution, too, would present further practical problems, as the precision rate can vary across regional varieties, which would mean we have to examine samples of all varieties of interest. If there is just too much noise in the data, this may require you to think of a different search query. Ultimately, coming up with a satisfactory one is a matter of striking the right balance between precision – you want to have as little noise as possible – and so-called ‘recall’ – you want to have as many examples that really correspond to the pattern, so as not to miss any major portion. As for this latter notion, recall, you will understand that the more specific our search string is, for instance by adding *at* in our example, the higher the risk becomes that some examples that fall under the construction you are interested in will not be retrieved (e.g. a relevant example like *have a look round the new school* would not be included).

2.3.3.2 Annotating the data

Our next step, after taking a sample, is to copy the concordance lines to a spreadsheet and analyze them one by one with respect one or more ‘variables’ (also called ‘parameters’ or ‘factors’). Deciding which variables you want to focus on will depend on your specific research question(s). It is not unheard of that corpus examples are ‘coded’ or ‘annotated’ for twenty or more different variables, each of which may have several different values (or ‘factor levels’). Variables could pertain to morphological, syntactic, semantic, pragmatic, discursive and/or (when these data are available) sociolinguistic aspects of utterances and/or their producers.

In our case, we might be interested in, for instance, the pragmatic value it carries in its context, that is, whether it is used as part of a statement (most concordance lines in the fragment), a request (line 1, in a way), an invitation (lines 7 and 8), and so on. When the same exercise is done for a similarly sized sample of occurrences with “TAKE a|an ADJ look”, we might find that there are different proportions of pragmatic functions for the two competing patterns. Indeed, that is the case in our example, as we’ll see. But first, here are some details on the values we distinguished for our single variable ‘pragmatic function’:

1. Assertives

This category was assigned to the largest group of sentences, most of which are ordinary statements, e.g. *I had a furtive look round*. Note that the label was used as a shorthand for the most ordinary occurrences of the two competing patterns, including for sentences realizing speech acts that aren't assertives in a technical sense. For instance, in *Will you be all right with him while I take a better look at the other bloke?*, the proposition in which our expression occurs is not claimed by the speaker to be true (nor is it claimed to be false), as it occurs in a temporal clause that itself appears in an interrogative sentence. The *while*-clause here probably expresses a kind of intention. Yet, in order not to proliferate functional categories, we used *assertive* as a catch-all term for any sentence that wasn't of either of the next two categories.

2. Invitations and requests

These two pragmatic functions were treated as one single category, as they are closely linked: just like a request involves an action that the speaker would like to perform, an invitation refers to an action that the speaker thinks the hearer would like to perform. Note that sentences of the type in (2), with *let's...*, were also considered a kind of invitation, even if the action might not be one that the addressee is eager to perform prior to the suggestion. No further distinction was made between *let's*-utterances used to suggest a joint action by speaker and hearer and those used by the speaker to announce an examining action, as when a doctor says, *Let's have a wee look at this lump*. The latter could be argued to be a kind of request (with the permission assumed to be immediately granted) presented as an invitation, which is one more reason for why we lumped together (no pun intended) these two kinds of speech act.

3. Instructions

These come close to invitations, but this category is reserved for unmistakable orders and obligations. Instructions can also be directed at a group including the speaker, as in *It is time we take an honest look at what we do*.

We present the results of our small-scale corpus study in Table 1, which is a so-called 'contingency table':

Table 1: Pragmatic functions of the *HAVE a(n) ADJ look* and *TAKE a(n) ADJ look* support verb constructions in the BNC

	<i>Assertives</i>	<i>Invitations and requests</i>	<i>Instructions</i>	<i>Total</i>
HAVE a(n) ADJ look	50	24	6	80
TAKE a(n) ADJ look	72	9	19	100
Total	122	33	25	180

As you can see from the column with totals, twenty utterances out of a sample of a hundred retrieved tokens of *HAVE a(n) Adj look* have disappeared from the table. These were removed because they were irrelevant. All of them used *look* in the sense of appearance (e.g. *She had a seraphic look on her face*), as in line (5) in the concordance

in Figure 15. This gives us an idea of the precision rate of our corpus query: 80%. So, one in five retrieved utterances was noise. The precision rate for *TAKE a(n) ADJ look* was 100%. It also looks as though the *have* variant is more often used for invitations and requests than the *take* variant, while the latter is more often used for instructions than the former.

Before we go on, a quick word about how to organize a table. In Table 1, the two variants of the construction are presented one under the other in the column to the left. That is because they are treated as possibilities making up the ‘independent variable’. For each of these options, we are interested in finding out which pragmatic functions it can have, and these therefore make up the ‘dependent variable’. In a contingency table, the values of the dependent variable are usually presented at the top of the table, from left to right. By contrast, note that in Figure 9, ‘English variety’ (AmE, CanE, Br) was treated as the independent variable and the dependent variable consisted then of our two forms of the construction, *HAVE a look* and *TAKE a look*.

2.3.3.3 Processing the data statistically

The third step, after retrieving a sample and annotating the sentences, is to process the observed differences statistically. Are the differences in frequencies significant or are they likely to be merely coincidental variations you can get in a relatively small sample? After all, when you flip a coin just ten times, you may throw heads seven or eight times, without this having to force you to conclude that your coin is rigged and that you have a higher chance of getting heads as you do for getting tails. So, how can we then test whether or not there is any real association between the *have* vs. *take* variants of this support verb construction and the pragmatic functions realized by them?

Many linguists use the free software environment R (R Core Team 2018) to deal with the kind of statistics question we’re facing here. R is a programming language, which means you have to type in lines of code yourself to tell the program what to do. This may be off-putting if you’re used to clicking on icons and selecting options from scroll-down menus in a fully designed interface. Fortunately, there is a vibrant community of people from all over the world offering help to one another. An excellent introduction to how to use R for linguistic research is the book by Levshina (2015). What follows is based on Chapter 9 of Levshina’s book, about “Measuring associations between two categorical variables”. Other useful textbooks on using R in quantitative linguistic approaches are those by Desagulier (2017) and Gries (2021). For beginners, there is also Jeroen Claes’s (2019) crash course for linguists (http://www.jeroenclaes.be/r_crashcourse_for_linguists/).

First you need to install the program. We suggest you also install RStudio IDE (integrated development environment), which makes your work in R a lot more user-friendly. You can download a free licence here:

<https://www.rstudio.com/products/rstudio/download/>

Once you have installed RStudio on your local computer and opened the application, you may first need to install a number of add-on ‘packages’. At the command prompt (the ‘>’ sign where the cursor is blinking in the console pane), you need to type in the following:

```
> install.packages("vcd")
```

This will automatically install a package that will allow you to visualize categorical data as well as compute the effect size of a variable.³ You then also need to load this package, as follows:

```
> library(vcd)
```

To produce something like Table 1 in R, without the marginal frequencies (i.e. the totals), use the function `rbind()`, which connects vectors (i.e., number sequences) as rows forming a table:

```
> look <- rbind(c(50, 24, 6), c(72, 9, 19))
```

To check whether this has the desired effect of producing a table, simply type the name of the contingency table (which we have just named ‘look’, but we could also have named it ‘table’ or ‘data’):

```
> look
```

This is what you then see:

```
[,1] [,2] [,3]
[1,] 50 24 6
[2,] 72 9 19
```

To add more insightful labels to the rows and columns, you can use commands like these:

```
> rownames(look) <- c("HAVE a(n) ADJ look", "TAKE a(n) ADJ look")
> colnames(look) <- c("assert", "invit/req", "instruct")
```

This produces the following result:

```
> look
      assert invit/req instruct
HAVE a(n) ADJ look   50    24    6
TAKE a(n) ADJ look  72     9   19
```

3 Another package is *Quanteda*, which you can use to conduct textual analysis, such as comparing speeches by different politicians and thereby creating word clouds that illustrate which words are typical for each of them, among many other features (<https://quanteda.io/index.html>). A package to carry out collexeme analysis, which we mentioned in passing in Section 3.2, has been developed by Susanne Flach and can be downloaded from this site: <https://sfla.ch/collostructions/>. It can then be installed and loaded in R (see Flach 2021).

It's also possible to calculate the proportions of every cell in a column, such that their sum adds up to 1.

```
> prop.table(look, 2)
```

```
assert invit/req instruct
```

```
HAVE a(n) ADJ look 0.4098361 0.7272727 0.24
```

```
TAKE a(n) ADJ look 0.5901639 0.2727273 0.76
```

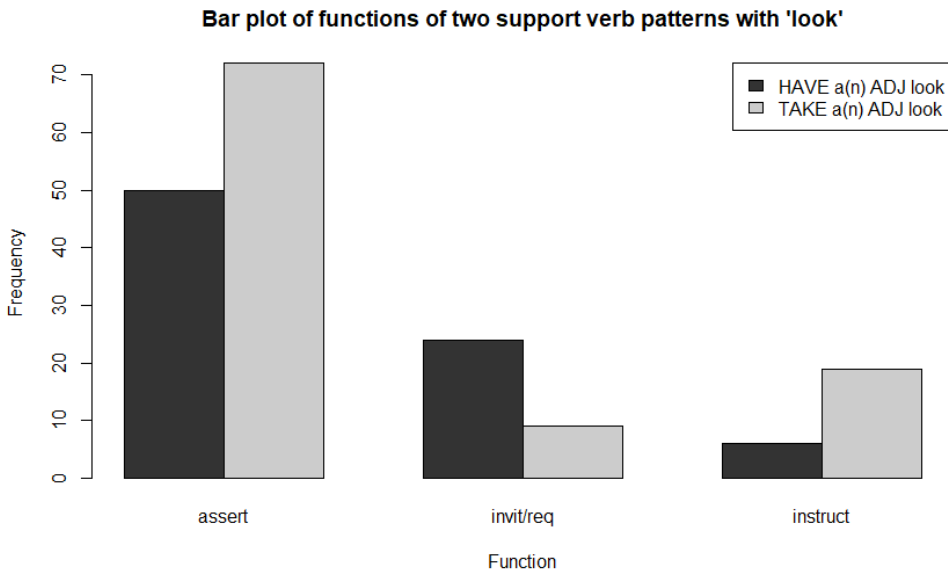
If you wanted to compute the proportions of cells making up a row, you would have to add “, 1” after the name of the table instead of adding “, 2”. (Not adding anything gives you the proportions of all cells in the table.)

You can generate a bar chart as follows:

```
> barplot(look, beside = TRUE, col = c("grey20", "grey80"), main =
"Bar plot of functions of two support verb patterns with 'look'",
xlab = "Function", ylab = "Frequency")
> legend("topright", fill=c("grey20", "grey80"), c("HAVE a(n) ADJ look",
"TAKE a(n) ADJ look"))
```

To produce a graph about something else, with another title, other labels and other variable names in the legend, simply replace what's rendered here in grey by descriptors that apply to your data. The code produces the chart shown in Figure 16:

Figure 16: Bar chart of pragmatic functions for two support verb alternatives, generated in R



So far, all of this is not anything that we couldn't have done more easily with the in-built chart function of a spreadsheet application like Microsoft Excel or LibreOffice Calc. Where R comes into its own is the possibility it offers to carry out statistical calculations. Presenting a table or a graph is just a matter of doing what is known as descriptive statistics. When we also want to test a hypothesis, we use so-called inferential statistics. This is the sort of statistics that allows us to assess whether we can generalize beyond our small sample the observation that, for instance, *have a look* appears to be more common for invitations and requests than *take a look*. Note from the proportions table above that the difference in proportions seems strongest for the value 'instructions'. But this category also represents the lowest share of the total sentences in the sample, so perhaps not that much can be inferred from these observations.

Let's start by comparing the observed frequencies in the table with the expected values, that is, the ones we would get if there were no relation whatsoever between choice of support verb and pragmatic use. We can type in the following two commands, each producing a table:

> look

```
assert invit/req instruct
HAVE a(n) ADJ look 50 24 6
TAKE a(n) ADJ look 72 9 19
```

> chisq.test(look)\$expected

```
assert invit/req instruct
HAVE a(n) ADJ look 54.22222 14.66667 11.11111
TAKE a(n) ADJ look 67.77778 18.33333 13.88889
```

The first table is the one with the real observations – it could also be produced by using the command **chisq.test(look)\$observed** – and the second one reproduces the table in such a way that all bumps and dips in the proportions are smoothed out. Note that the total observations for *have a look* is 80 and for *take a look* is 100 in our sample, which means that for each function, we could expect a four-to-five ratio under 'chance' conditions. So, per four tokens of *have a look*, there should then be five tokens for *take a look*, whatever the pragmatic function. Indeed, if you divide any expected frequency for *have a look* by the expected frequency for *take a look* just below that cell, you will get as a result 0.8 (or something very close to it, like 0.799999856). To get an idea of how far the observed frequencies are removed from the expected ones, we can ask for 'Pearson residuals' to be computed:

> chisq.test(look)\$residuals

```
assert invit/req instruct
HAVE a(n) ADJ look -0.5733931 2.437087 -1.533333
TAKE a(n) ADJ look 0.5128583 -2.179797 1.371455
```

These values are the differences between the observed and the expected values divided by the squared root of the expected value. A negative value is obtained when the observed frequency is smaller than the expected frequency; a positive value when it is greater. The higher the absolute value of a Pearson residual, the larger the difference between observed and expected frequency. Are the differences we see here sufficiently large to reject the ‘null hypothesis’, which says that there is no association between choice of verb and function? Of course, even if the null hypothesis were true, the observed frequencies are always likely to deviate somewhat from the ‘expected’ ones, which therefore, paradoxically, are not really *that* expected. The point is, though, that the null hypothesis becomes increasingly implausible the wider the gap is between the observed proportions and the ‘ideal’ ones (in our case, 4/5). To find out whether the difference is sufficiently large to be statistically significant, we can apply the ‘chi square’ (χ^2) test. This yields a p-value, which tells you how probable it is that the null hypothesis can still be true if you have a difference that is at least as big as that between the observed and the expected values. If the p-value is lower than 0.05, this means that there is less than 5% probability that the observed values can deviate as much from the expected values as they do under the null hypothesis being true. In other words, that’s rather too low a probability to still maintain that null hypothesis. Especially if the p-value dives under 0.01, let alone under 0.001, there remains hardly any reason at all to hold on to the assumption of no association between variables. This is how we can calculate chi square and the corresponding p-value in R:

```
> chisq.test(look)
```

Pearson’s Chi-squared test

```
data: look
```

```
X-squared = 15.515, df = 2, p-value = 0.0004276
```

What is meant by ‘df’ is degrees of freedom, which we always have to report, along with the chi square value and the p-value. For a two-by-three contingency table, the number of degrees of freedom is two. Why? Because there are only two cells that could be freely filled in. Once, say, two row values are randomly chosen, then we know what the third value is, as in a kind of Sudoku puzzle, based on that row’s marginal total. And then, once the entire first row is filled in, we have no choices left for the second row at all, because the marginal column totals tell us what these values should be. Importantly, here, the p-value is lower than 0.001, so we can report this result as follows: “We have found a significant association between verb choice in the English *have/take a(n) ADJ look* pattern and pragmatic function: $\chi^2(2) = 15.52$, $p < 0.001$.”

Many linguists would leave it at that, but apart from knowing now that choosing one verb rather the other has an effect on which pragmatic value is likely to be realized,

we would also like to get an idea of the effect size. For this, we can use the function `assocstats()` (which stands for ‘association statistics’):

```
> assocstats(look)
      X^2 df P(> X^2)
Likelihood Ratio 15.940 2 0.00034561
Pearson          15.515 2 0.00042759

Phi-Coefficient : NA
Contingency Coeff.: 0.282
Cramer's V      : 0.294
```

The function returns a value between 0 and 1 for three measures of effect size; however, with tables larger than two-by-two, the ϕ -coefficient is unreliable and therefore not returned (‘NA’ standing for ‘not applicable’). It is recommended that for a two-by-three table, Cramér’s *V* is reported. In our case, the effect size is moderate (between 0.21 and 0.35), which means that, while the result of the chi-square test is significant, the two categorical variables (verb and function) are only moderately associated – the effect could be weaker but it could also be stronger. (Between 0 and 0.07, the association can be considered negligible; between 0.07 and 0.21 weak; between 0.21 and 0.35, as we said, moderate; over 0.35 strong. For a two-by-two contingency table, however, the association is weak for a value between 0.1 and 0.3, moderate for a value between 0.3 and 0.5, and strong only for a value over 0.5.

To get an idea of the direction(s) of association, we can calculate the odds ratios (ORs) for pairwise comparisons between the different pragmatic uses. To do so, we divide the odds for one use by the odds for another. First, though, let’s again reproduce the table so that we have direct access at the numbers:

```
> look
      assert invit/req instruct
HAVE a(n) ADJ look  50    24    6
TAKE a(n) ADJ look  72     9   19
```

Odds are ratios already in themselves, so an odds ratio is a ratio (i.e. a division) of two ratios (i.e. two divisions). It doesn’t matter which odds you use as numerator for each division, but it may be easiest to report your results if you take the largest of the two:

```
> (24/9)/(50/72)
[1] 3.84
```

```
> (24/9)/(6/19)
```

```
[1] 8.444444
```

```
> (50/72)/(6/19)
```

```
[1] 2.199074
```

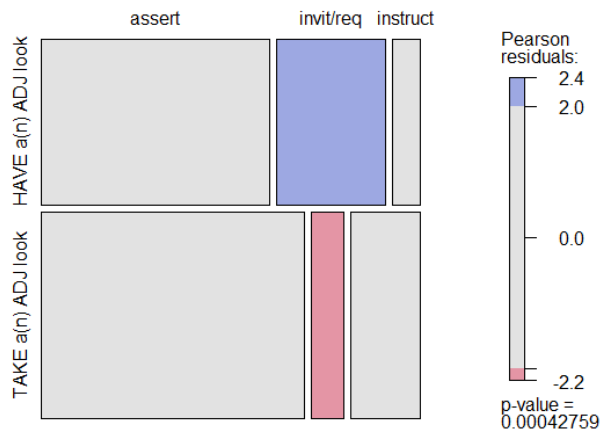
We can formulate these results as follows: “The odds that *V a(n) ADJ look* is used as an invitation or request rather than as an assertive is almost four times higher with *have* than with *take*; the odds that this construction is used as an invitation or request rather than as an instruction is more than eight times higher with *have* than with *take*; finally, the odds that it is used as an assertive rather than as an instruction is about twice as high with *have* as with *take*; Cramér’s $V = 0.294$ ”. Note that it is customary to add to such a description of the odds ratio(s) the value of an effect size measure, which we obtained above.

For a contingency table, especially one of a higher than two-by-two dimension, we can nicely capture the conspicuous overuse and underuse of certain fields in the table in a so-called ‘mosaic plot’:

```
> mosaic(look, shade = TRUE, varnames = FALSE)
```

This generates the visualisation shown in Figure 17.

Figure 17: Mosaic plot of pragmatic uses of *HAVE a(n) ADJ look* versus *TAKE a(n) ADJ look*



Blue tiles represent positive residuals (light blue: between 2 and 4). Reddish tiles represent negative residuals (pink: between -2 and -4). There are only moderate cases of over- and underuse in our data. (Values above 4 would show in dark blue and values below 4 in magenta.) The plot shows us immediately that invitations and requests are conspicuously common with *have* and conspicuously uncommon with *take*.

Finally, we may want to inspect which residuals make a significant contribution to this distribution, something which standardized residuals (i.e., residuals divided by their standard error) are used for. The following code can be entered for this:

```
> chisq.test(look)$stdres
```

This produces the table below:

```
assert invit/req instruct
HAVE a(n) ADJ look -1.355224 3.618136 -2.216885
TAKE a(n) ADJ look 1.355224 -3.618136 2.216885
```

Cells with standardized residuals greater than 1.96 or smaller than -1.96 contain values that deviate significantly from expected values at the 0.05 level of significance. With values exceeding ± 2.58 , the deviation reaches the 0.01 level of significance, and with values outside ± 3.29 , there is significance at the 0.001 level. The table we have obtained allows us to see that cells in the assertives column fail to make any significant contribution, that cells in the invitations and requests column make a statistically extremely significant contribution at the 0.001 level, and that the cells in the instructions column also contribute significantly to the obtained χ^2 -statistic value⁴, but only at the 0.05 level. This means that we can confidently conclude that instructions and especially invitations/requests are pragmatic functions for which speakers make a clearly different selection of *have a(n) ADJ look* versus *take a(n) ADJ look*.

2.3.4 GIVING SOMEONE A CERTAIN LOOK: FINDING ALTERNATIVES TO THE VERB

Remember that there is also pattern in which more than just looking is involved (cf. meaning 2 in Figure 1), namely where someone gives or sends – again metaphorically, of course – a particular look to another person (or, less likely, an

4 For a prepared spreadsheet document that processes the results in a table and produces an interpretation that you can copy and paste into your paper, go to: https://www.researchgate.net/publication/361115565_Chi_square_effect_size_and_odds_ratio_calculator_for_corpus_linguistics

object). An added aspect of meaning then is that the looker reveals their emotion while looking at someone. To find out which verbs are used in the pattern, apart from *give* or *send*, we could try to retrieve instances of the so-called ditransitive pattern, with the recipient of the look realized as a personal pronoun (*me, you, him, her, us, them*). Such a search query is shown in Figure 18. The results the interface produces are displayed in Figure 19. The search was again carried out in the British National Corpus (BNC, Davies 2004).

Figure 18: Search query designed to retrieve instances corresponding to the expression *give someone a(n) ADJ look*⁵

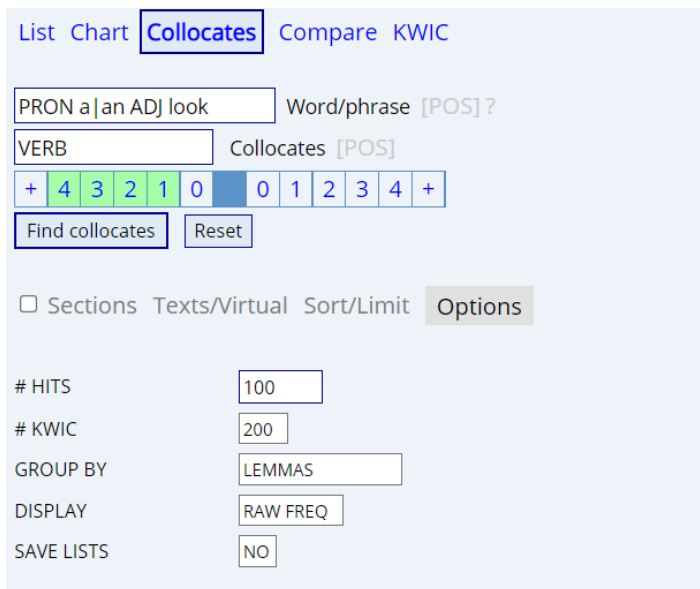


Figure 19: Results retrieved by the query: verb lemmas used in this pattern, ordered by frequency

HELP			FREQ
1	<input type="checkbox"/>	[GIVE]	224
2	<input type="checkbox"/>	[SHOOT]	47
3	<input type="checkbox"/>	[THROW]	24
4	<input type="checkbox"/>	[SEND]	13
5	<input type="checkbox"/>	[CAST]	9
6	<input type="checkbox"/>	[FLING]	5
7	<input type="checkbox"/>	[SLIDE]	4
8	<input type="checkbox"/>	[DART]	3
9	<input type="checkbox"/>	[FLICK]	2
10	<input type="checkbox"/>	[FLASH]	2
11	<input type="checkbox"/>	[LEND]	2
12	<input type="checkbox"/>	[TOSS]	1
13	<input type="checkbox"/>	[SLANT]	1
14	<input type="checkbox"/>	[OFFER]	1
15	<input type="checkbox"/>	[EARN]	1
16	<input type="checkbox"/>	[DELIVER]	1
		TOTAL	340

5 In addition to looking for this structure, with an adjective before the noun *look*, we could formulate a query that retrieves instances with *look* followed by *of ADJ NOUN*, such as *give someone a look of {utter contempt / pure disgust / mild surprise}*.

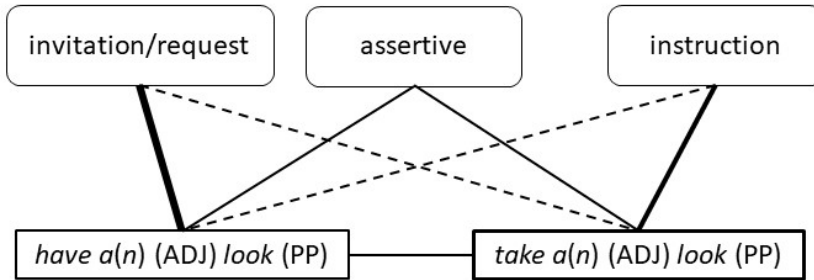
Grouping the results by lemmas, which we did in the Options tab of the search window, ensures that different inflectional forms for a single verb are not scattered in the output.

Give is clearly the most common verb, although we can't fully take the numbers at face value. If we click on [GIVE] in the result table and inspect the actual examples, we can see that a few are not of the desired type, such as *His thick moustache, neatly trimmed, gave him a handsome look*. Even when such examples are weeded out, *give* is still by far the most common verb. Note that apart from *give*, the verbs *shoot* and *throw* are very frequent, even more so than *send*. These and a few other verbs (*cast, fling, dart, flick, toss*) belong to what Levin (1993) calls 'verbs of throwing', that is, verbs describing "instantaneously causing ballistic motion" by imparting a force (Gropen et al. 1989).

2.4 IMPLICATIONS FOR THE ORGANIZATION OF LEXICAL AND GRAMMATICAL KNOWLEDGE

What we have found in the course of answering the questions raised at the end of Section 2 is that we can posit constructions of visual perception and visual expression with *look* at various levels of generality. At close to the highest level, we may find the [V *a(n)* (Adj) *look* (Adv/PP)] construction with a fairly general meaning having to do with directing one's eyes somewhere. This general construction has a few 'daughter' constructions, two of which are the by now well-known support verb constructions *have a look* and *take a look*. These are close synonyms, but each of them has nonetheless its own preferred range of uses. How can we represent their different functional profiles? In a theoretical framework such as Construction Grammar (cf., e.g., Goldberg 2003 for a quick outline of the basic tenets), idiomatic expressions are seen as constructions defined as cognitively 'entrenched' form-function units. Yet, general pragmatic functions such as asserting something, inviting someone to do something, ordering someone to do something and the like ought to be recognized as having an independent existence outside of constructions. After all, unlike the conceptual content of lexical items, such as the meaning of *cat* or *sneeze*, such a pragmatic function is not realized by a single, highly privileged linguistic form – rather, there are very many possible forms for it. We can therefore capture the relations between the forms and the pragmatic functions as in Figure 20.

Figure 20: Network of pragmatic functions associated with two roughly synonymous support verb constructions



This simple network contains the constructional idioms *have a(n) (Adj) look (PP)* and *take a(n) (Adj) look (PP)*, which are shown to be linked in a ‘horizontal’ relationship, as these are two sisters that have much the same meaning. They can be seen as ‘allostructions’ (Cappelle 2006), that is, two alternative formal manifestations of a slightly more general construction – but less general than the [V *a(n)* (Adj) look (Adv/PP)] construction mentioned above – in which the choice of verb (i.e., whether it is *have* or *take*) is left unspecified. It may be observed that [*take a(n) (Adj) look (PP)*] has been enclosed in a box with a somewhat thicker line, indicating it has higher frequency and is thus cognitively more strongly engraved or entrenched in the mind. We know from a sample of authentic data retrieved and analysed (cf. Section 3.3) that requests and invitations are more strongly associated with *have a look* than with *take a look* and this is indicated in the network with a thicker connecting line for the former and a thin dashed line for the latter. For instructions, the opposite pattern holds, but somewhat less pronouncedly so.

By representing the pragmatic values as external to each constructional alternative, the partial network in Figure 20 reflects a way of looking at the ‘construct-i-con’ (i.e., the repository of constructions in the minds of speakers) that contains more links between nodes, which are themselves less crammed with information (cf. Diessel 2019, Hilpert and Flach in preparation). The whole of a language user’s grammar and lexicon might thus be conceived of as a network of nodes that are linked with several other nodes. When one node is accessed (because it is selected by the speaker or processed by the hearer) its activation spreads through the network. This is compatible with the idea that pragmatics can be ‘part of’ a construction (Cappelle 2017).

Many more details could be added to this network representation, which is intentionally a simplification of the rather complex linguistic reality. For instance,

different versions of each idiom with a specific adjective (e.g. *have a quick look*, *take a good look*) are likely to be part of a competent speaker's lexicon of conventionalized sequences. The ties that these more specific idioms have with the pragmatic functions may be stronger or weaker than is the case for the two versions without a modifier shown in Figure 20. That there can be two modifiers (e.g. *take a good hard look*) suggests that the network should include even slightly more general schemas in which *look* appears as the head of an NP and in which the preceding (still optional) position is just specified to be semantically a modifier, which can then be syntactically realized by means of more than one adjective, or by an adjective that is itself modified (e.g. *take a real(ly) good look*). The presence of general schemas in the network, however, does not obviate the need for also storing specific instantiations with high usage frequency. It is Langacker (1987) who perhaps most eloquently rejected what he referred to as the 'rule/list fallacy', according to which there would be no reason to separately memorize any individual sequences that are rule-governed. While that would perhaps lead to a more economical, parsimonious representation of the speaker's construct-i-con, this may not be in line with cognition. It is worth quoting Langacker (1987: 42) here in full:

"[S]pecific forms like beads, shoes, toes, and walls can be omitted from the grammar of English if one provides a general rule of plural formation that accounts for them. Listing regular forms thus implies lost generalizations and failure to achieve an optimal analysis.

This line of thought illustrates the rule/list fallacy. It is fallacious because it assumes that one is forced to choose between rules and lists: the options are posed as rules alone vs. lists alone. If these are the only two options, it can be argued that the rules must be chosen, for lists by themselves do not express generalizations. There is in reality a third choice, however, namely both rules and lists. It is plausible to suppose that speakers of English master many regular plural forms as fixed units in addition to learning the general rule of plural formation. Similarly, it is clear that speakers learn as fixed units a large number of conventional expressions that are nevertheless fully analyzable and regular in formation (absolutely incredible, cheap imitation, great idea, turn the page, let the cat out, and so forth). General statements and particular statements can perfectly well coexist in the cognitive representation of linguistic phenomena, just as we learn certain products by rote in addition to mastering general procedures for multiplication. To the extent that this is so, an accurate linguistic description claiming psychological reality must contain both rules expressing generalizations and specific forms learned as fixed units, even if the specific forms accord fully with the rules." (Langacker 1987: 42)

Applied to our constructions, just because *have a quick look* or *take a long hard look* are sequences in full accordance with 'rules' (i.e., schematic patterns) allowing for one or more optional adjectives, this does not mean that these specific sequences cannot be entered in the constructional network as conventional, routinized units.

2.5 IDEAS FOR FURTHER DISCUSSION AND PRACTICE ACTIVITIES

Our demonstration has shown how corpora can be fruitfully used to test predictions based on what we find in dictionaries. Sometimes our intuitions also tell us something about what is probable, and such intuitions, too, can be confirmed or disconfirmed with the help of corpora. Below are some ideas for further corpus exploration that you might find useful to try out if you're a beginning corpus user. Teachers could also use these as class or homework activities. The first two exercises are quite easy; a more challenging task, involving some statistical processing of data, is given at the end.

2.5.1 HAVE GOT A LOOK

Do you think *have got a look* can be used as an alternative of *have (or take) a look*? Use a corpus to find out whether and how often *have got a look* occurs with the meaning of 'have/take a look'. What can you conclude from these quantitative results?

2.5.2 THE DOCUMENTARY HAS/TAKES A LOOK AT...

Based on what you now know about the semantic differences of *have a look* and *take a look*, which of these do you think is more likely to occur: *The documentary has a(n) ADJ look at...* or *The documentary takes a(n) ADJ look at...*? Why?

Do you know the name of the linguistic phenomenon whereby a person (in this case a documentary maker) is referred to indirectly, via the term of a concept closely associated with it (here, the thing made)? Can you find further such examples by means of a corpus?

2.5.3 HAVE/TAKE A LOOK AT GENRES

Which of the two alternatives, *have a look* or *take a look*, do you think is most common in spoken language? Present some data obtained from a corpus in a table and/or a chart that you produce yourself, without copying-and-pasting from a screen. (Hint: in the same way that GloWbE has different language varieties as 'sections', mono-variety corpora such as the BNC and COCA are composed of data from different genres such as spoken language, fiction, magazines, newspapers, academic books and papers, etc. With the Chart function, you can see the frequency of a search string by genre.)

Use R to carry out a statistical analysis of the difference you find between spoken and non-spoken English and the choice of expression.

REFERENCES

- Cappelle, B. 2006. Particle placement and the case for ‘allostructions’. *Constructions* SV1-7/2006, 1–28. www.constructions-online.de, urn:nbn:de:0009-4-6839, ISSN 1860-2010. Last accessed 1 January 2020.
- Cappelle, B. 2017. What’s pragmatics doing outside constructions? In I. Depraetere, & R. Salkie (Eds.), *Semantics and pragmatics. Drawing a line* (115–151). Cham: Springer.
- Cappelle, B. 2020. Looking into visual motion expressions in Dutch, English and French: How languages stick to well-trodden typological paths. In Y. Matsumoto, & K. Kawachi (Eds.), *Broader Perspectives on Motion Event Descriptions* (235–279). Amsterdam & Philadelphia: John Benjamins.
- Cifuentes-Férez, P. 2014. A closer look at Paths of vision, Manner of vision and their translation from English into Spanish. *Languages in Contrast*, 14(2), 214–250.
- Claes, Jeroen. 2019. R Crashcourse for linguists. Available online at http://www.jeroenclaes.be/r_crashcourse_for_linguists/.
- Danlos, L. 1992. Support verb constructions: Linguistic properties, representation, translation. *Journal of French Language Studies*, 2(1), 1–32. doi:10.1017/S0959269500001137
- Davies, M. 2004. *British National Corpus* (from Oxford University Press). Available online at <https://www.english-corpora.org/bnc/>.
- Davies, M. 2008. New directions in Spanish and Portuguese corpus linguistics. *Studies in Hispanic and Lusophone Linguistics*, 1(1), 149–186.
- Davies, M. 2008-. *The Corpus of Contemporary American English* (COCA). Available online at <https://www.english-corpora.org/coca/>.
- Davies, M. 2009. Creating useful historical corpora: A comparison of Corde, The Corpus del Español, and the Corpus do Português. In A. Enrique-Arias (Ed.), *Diacronía de las lenguas iberorrománicas Nuevas aportaciones desde la lingüística de corpus* (137–166). Madrid: Iberoamericana/Vervuert.
- Davies, M. 2013. *Corpus of Global Web-Based English*. Available online at <https://www.english-corpora.org/glowbe/>.
- Davies, M. 2016-a. *Corpus del Español: Web/Dialects*. Available online at <http://www.corpusdelespanol.org/web-dial/>.
- Davies, M. 2016-b. *Corpus do Português: Web/Dialects*. Available online at <http://www.corpusdoportugues.org/web-dial/>.
- Desagulier, G. 2017. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. New York: Springer.
- Diesel, H. 2019. *The Grammar Network: How Linguistic Structure Is Shaped by Language*

- Use. Cambridge: Cambridge University Press.
- Flach, S. 2021. *Collostructions: An R implementation for the family of collostructional methods*. Package version v.0.2.0. <https://sfla.ch/collostructions/>.
- Goldberg, A. E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224.
- Gries, S. Th. 2021. *Statistics for Linguistics with R: A Practical Introduction* (3rd revised ed.). Berlin: Walter de Gruyter.
- Gries, S. Th. & A. Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspectives on ‘alternations’. *International Journal of Corpus Linguistics*, 9(1), 97–129.
- Gropen, J., S. Pinker, M. Hollander, R. Goldberg and R. Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language*, 65(2), 203–257.
- Gruber, J. S. 1967. Look and see. *Language*, 43(4), 937–947.
- Hilpert, M. & S. Flach. In preparation. Modals in the network model of Construction Grammar. In I. Depraetere, B. Cappelle, & M. Hilpert (Eds.), *Models of Modals: From Pragmatics and Corpus Linguistics to Machine Learning*. Berlin: Mouton de Gruyter.
- Kawachi, K. 2020. Should Talmy’s motion typology be expanded to visual motion?: An investigation into expressions of motion, agentive motion, and visual motion in Sidaama (Sidamo). In Y. Matsumoto, & K. Kawachi (Eds.), *Broader perspectives on motion event descriptions* (205–234). Amsterdam & Philadelphia: John Benjamins.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar. Theoretical Prerequisites* (Volume I). Stanford: Stanford University Press.
- Langer, S. 2004. A linguistic test battery for support verb constructions. *Linguisticae Investigationes*, 27(2), 171–184.
- Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago & London: University of Chicago Press.
- look. *LCOED Online*, Longman, June 2021, <https://www.ldoceonline.com/dictionary/look>. Accessed 27 June 2021.
- look, n. *OED Online*, Oxford University Press, June 2021, www.oed.com/view/Entry/110129. Accessed 27 June 2021.
- Matsumoto, Y. 2001. Lexicalization patterns and caused and fictive motion: The case of typological split. Handout for a lecture at SUNY Buffalo, NY.
- Matsumoto, Y., et al. 2021. Linguistic representations of visual motion: A crosslinguistic experimental study.
- Namer, F. 1998. Support verb constructions. In F. van Eynde, and P. Schmidt (Eds.),

Linguistic specifications for typed feature structure formalisms (315–343).
Luxembourg: Office for Official Publications of the European Communities.

R Core Team. 2018. R version 3.5.0. R: A language and environment for statistical computing. R Foundation for Statistical Computing Platform, Vienna, Austria. <https://www.R-project.org/>.

Talmy, L. 1996. Fictive motion in language and ‘ception”. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (211–276). Cambridge, MA & London: The MIT Press.

Talmy, L. 2000. *Toward a cognitive semantics (2 Volumes)*. Cambridge, MA: MIT Press.

Talmy, L. 2018. *Ten lectures on cognitive semantics*. Leiden, The Netherlands: Koninklijke Brill.

CAPÍTULO 3

COLETA E TRATAMENTO DE DADOS NO *CORPUS* DO PORTUGUÊS

Ravena Beatriz de Sousa Teixeira
Jeane Nunes da Penha
(Universidade Federal do Rio de Janeiro)

3.1 INTRODUÇÃO

Com a popularização da internet, assim como dos aparatos tecnológicos, encontramos disponíveis, atualmente, diversas ferramentas capazes de facilitar o pesquisador no que concerne à coleta e ao tratamento de dados linguísticos. Nesse sentido, cabe-nos destacar aqui o papel da Linguística de *Corpus*, subárea da Linguística assim nomeada por ser utilizada, essencialmente, para a coleta, tratamento e análise de *corpora* linguísticos do âmbito escrito produzidos por falantes reais. Na Linguística de *Corpus*, todos os dados são selecionados prévia e criteriosamente por um ou mais estudiosos da área, a fim de facilitar na busca por dados de uma ou mais línguas e/ou variedades (cf. SARDINHA, 2000).

No Projeto PREDICAR (Formação e expressão de predicados complexos e predicações: estabilidade, variação e mudança construcional) algumas investigações já vêm adotando gerenciadores de dados *online*, como o *Corpus* do Português (<https://www>.

corpusdoportugues.org/),¹ que consiste em um banco de dados de *corpora* textuais em língua portuguesa, criado pelos professores Mark Davis e Michael J. Ferreira, das Universidades Brigham Young (BYU) e Georgetown, respectivamente. Desenvolvido para estudos de cunho linguístico, é constituído por uma interface que propicia ao pesquisador buscar por expressões exatas, lemas, trechos de textos e, inclusive, classes gramaticais.

O *Corpus* do Português (DAVIES, 2016) possibilita aos pesquisadores a realização de investigações em um número diversificado de sincronias, desse modo, mostra-se como uma ferramenta relevante para os linguistas que tencionam debruçar-se sobre a análise da língua portuguesa e/ou suas variedades. Com isso, a partir dessa breve apresentação do *Corpus* do Português, objetivamos: (i) mostrar os aspectos positivos e negativos de se trabalhar com bancos de dados online; (ii) ilustrar os diferentes *corpora* que compõem o gerenciador de dados *Corpus* do Português; (iii) explorar os passos metodológicos para a coleta no gerenciador de dados aqui focalizado; e (iv) auxiliar o pesquisador no tratamento dos dados linguísticos.

Dividimos este capítulo da seguinte maneira: após esta seção introdutória, apresentamos uma seção com os aspectos positivos e negativos de se utilizar bancos de dados *online*. Logo após, demonstramos os quatro *corpora* que constituem o *Corpus* do Português (Gênero/Histórico, Web/Dialetos, NOW e WordAndPhrase). Em seguida, detalhamos os passos metodológicos para a coleta de dados (Como se dá a coleta?; Como funciona o menu de busca?; Quais opções o pesquisador encontra na ferramenta de busca?). Por fim, expomos as considerações finais e apresentamos as referências que nortearam este trabalho.

3.2 QUAIS OS ASPECTOS POSITIVOS E NEGATIVOS DOS BANCOS DE DADOS ONLINE?

Reunimos nesta seção alguns pontos positivos e negativos – mais positivos – de se utilizar gerenciadores de *corpora* do meio digital para a coleta de dados linguísticos. Como bônus, o investigador encontra:

1) *Fácil acessibilidade* – os bancos de dados *online* estão disponíveis para pesquisadores do mundo inteiro. Mesmo cobrando uma taxa de cadastro, oferecem a opção de utilizarmos o gerenciador gratuitamente, durante um período de 30 dias, para teste/experiência;

2) *Rápido rastreamento de dados de diversas línguas e variedades* – os bancos de dados digitais processam, selecionam e armazenam milhares de dados linguísticos. Para o linguista que busca analisar o Português, o *Corpus* do Português é uma boa opção, pois disponibiliza dados de diferentes países que possuem a língua portuguesa como idioma oficial. Em contrapartida, se procura investigar outros idiomas, há ou-

1 Teixeira (2020) ilustra complexos verbo-nominais de carga semântica passiva coletados no *Corpus* do Português NOW, como *levar uma pancada, tomar uma pancada, sofrer uma pancada, receber uma pancada e ganhar uma pancada*.

tros gerenciadores de *corpora*, como o *Sketch Engine*, que, além do Português, ainda armazena dados do Inglês, Espanhol, Francês, Italiano, Chinês, dentre outros;

3) *Mais agilidade e menos esforço para a obtenção de dados linguísticos* – diferentemente de outras maneiras usadas por investigadores para a coleta de dados (leituras físicas ou digitais em demasiados textos, procuras incessantes por diferentes fontes, gêneros etc.), os bancos de dados permitem que o pesquisador se depare com milhares de resultados a partir de um *click*;

4) *Fácil seleção das fontes e/ou gêneros dos dados* – por armazenar dados de diferentes fontes e gêneros textuais, os bancos de dados *online* são úteis para o linguista que objetiva compor um *corpus* apenas com dados acionados em um determinado gênero textual, por exemplo, o gênero notícia. Nesse caso, ao invés de pesquisar múltiplos jornais, o linguista tem a opção de selecionar, na página de resultados, somente os dados oriundos de jornais e, assim, buscar os dados mobilizados no gênero em questão; e

5) *Informações sobre os dados* – os gerenciadores de *corpora* digitais oferecem ao investigador informações importantes sobre os dados, como a data de rastreamento, a variedade pertencente, endereço eletrônico da fonte em que o dado foi extraído etc.

Como nem tudo na vida é viver apenas de bônus, com os bancos de dados digitais não é diferente. Como ônus, o linguista defronta-se com:

1) *Páginas indisponíveis* – alguns dados encontrados nos resultados de busca podem não estar mais com o endereço eletrônico disponível. Cabe lembrar que se um dado provém de um *corpus* que foi rastreado três anos antes da ação da coleta é importante se ter em vista que algumas páginas tenham saído do ar/não existam mais; e

2) *Política de privacidade de determinadas fontes* – como mencionado na introdução, os bancos de dados digitais reúnem *corpora* linguísticos do meio escrito produzidos por falantes reais. Sendo assim, o investigador pode notar que algumas páginas apresentam uma política de privacidade seguida de bloqueio do conteúdo, a fim de garantir a não disseminação ou plágio do teor ali apresentado. Os gerenciadores de dados *online*, geralmente, demonstram os dados no contexto discursivo de aplicação, por isso, cabe ao pesquisador decidir se coleta o dado no pequeno trecho apresentado mas não o expõe, ou contabiliza no *corpus*, ou se o ignora.

Dado o exposto, notamos que, embora o procedimento de coleta de dados em bancos digitais apresente alguns pontos negativos, os aspectos positivos são predominantes, o que facilita o investigador na hora de optar por um método mais ágil.

3.3 CORPUS DO PORTUGUÊS: CONHECENDO OS CORPORA

Ao acessar o site do *Corpus* do Português, é possível notar que o banco de dados é composto por duas partes distintas. De acordo com a página de apresentação, encontramos:

um corpus (original e menor) que permite ver as mudanças históricas assim como variações de gênero; um corpus (novo e muito maior) que permite verificar as variações dialéticas (e tem 50 vezes mais dados do português moderno).

Como observado na Figura 1, o *Corpus do Português* conta com quatro abas de pesquisas dos *corpora*: Gênero/Histórico, Web/Dialetos, NOW e WordAndPhrase, descritos a seguir.

Figura 1: *Corpora* disponíveis no banco de dados online *Corpus do Português*.

		Corpus	Tamanho	Criado
1	Info	Gênero / Histórico	45 milhões de palavras	2006
2	Info	Web / Dialetos *	1 mil milhão de palavras	2016
3	Info	NOW (2012 - 2019)	1,1 mil milhão de palavras	2018
4	Info	WordAndPhrase (agora parte do #2)	40.000 palavras principais	2017

Gênero / Histórico

Apresentado como o “original” do *Corpus do Português*, possui duas interfaces (2006 e 2016). É constituído por uma base de dados com 45 milhões de palavras entre os séculos XIII e XX – o que o torna importante para o pesquisador que deseja investigar a Língua Portuguesa desde uma perspectiva diacrônica. No que diz respeito aos dados do século XX, estes se dividem, igualmente, entre os gêneros de estilo falado, ficção, textos acadêmicos e provenientes de jornais digitais. A versão mais recente (2016) permite-nos criar *corpora* virtuais, por exemplo, por um determinado conjunto de fontes, tópicos etc.

Web / Dialetos

Interface desenvolvida em 2016, composta por uma base de dados com cerca de 1 milhão de palavras de páginas digitais oriundas de quatro países de língua portuguesa: Brasil, Portugal, Angola e Moçambique. Os textos foram rastreados entre os anos de 2013 e 2014, por isso permite que o pesquisador realize uma análise sincrônica do *corpus*, além possibilitar a comparação entre as diferentes variedades da língua portuguesa.

News on the Web (NOW)

A aba NOW é a interface mais recente do *Corpus do Português* (agosto de 2018) e está constituída por mais de 1,1 milhão de palavras das distintas variedades do português (Brasileira, Europeia, Africana). De acordo com a página de apresentação, todos os meses são adicionadas ao *corpus* cerca de 35 milhões de palavras oriundas de revistas e jornais digitais.

WordAndPhrase

Essa interface, não mais disponível a partir de 2022, proporcionava ao investigador pesquisar e navegar por 40 mil palavras do português com base na frequência do *corpus*. Para cada palavra buscada, o investigador encontrava disponível diversas informações, como sua definição, sinônimos, gênero, país, colocação, concordância, dentre outras. Além disso, tornava possível inserir e analisar textos completos, destacar palavras-chaves ou frases no texto e realizar uma pesquisa com frases relacionadas em todo o *Corpus* do Português. Assim, considerando a importância de suas ferramentas de busca para a análise linguística, suas funções, em janeiro de 2022, passam a ser incorporadas nos recursos da aba Web/Dialetos.

Agora que nos familiarizamos com todas as abas de pesquisa disponibilizadas no *Corpus* do Português, daremos início aos procedimentos metodológicos – busca, tratamento e análise de amostras linguísticas presentes no banco de dados.

3.4 A PLATAFORMA DE BUSCA E O PROCESSO DE COLETA

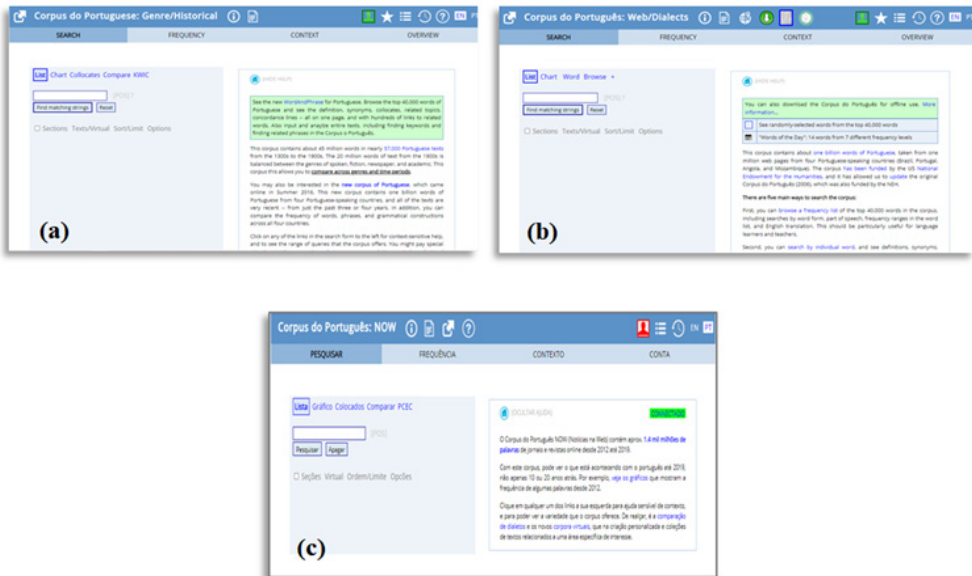
Dentre os aspectos essenciais, ao se tratar do processo de análise linguística, encontra-se a delimitação das características e fonte dos dados a serem considerados pelo pesquisador. A seleção de um (ou mais) banco(s) de dados deve refletir as diretrizes teórico-metodológicas da pesquisa a ser levada a cabo. Entretanto, um dos desafios enfrentados por pesquisadores é a falta de familiaridade com as interfaces de busca propiciadas pelas plataformas associadas aos *corpora online*.

Assim, como indicado anteriormente, nesta seção, buscamos apresentar, de forma sucinta, as ferramentas de busca oferecidas pelas abas do *Corpus* do Português, em especial do *Corpus* do Português NOW (aba mais recente do banco de dados). Inicialmente, apresentaremos os atributos do menu de busca de cada uma das abas, quais são as ferramentas de busca oferecidas e como, de fato, se dá seu funcionamento, para, em seguida, visualizarmos, segundo exemplos, quais são os passos metodológicos a serem considerados no processo de busca.

3.4.1 FAMILIARIZANDO-SE COM A PLATAFORMA DE BUSCA

Em relação à sua configuração, as abas *Gênero/Histórico*, *Web/Dialecto* e *NOW* compartilham a mesma interface de busca, conforme observável nas Figuras 2.a, 2.b e 2.c.

Figura 2: Páginas iniciais da interface de busca das abas do banco de dados *online Corpus* do Português.



Como nosso intuito é, primordialmente, apresentar um tutorial prático do processo de coleta de dados, em especial de predicadores complexos compostos por verbo (semi-)suporte, podemos visualizar, a seguir, uma descrição dos elementos que compõem as janelas de pesquisa das três referentes abas, constituídas por quatro janelas principais: a janela inicial de busca (1); a janela de disposição inicial dos resultados (2); a janela de contexto (3), na qual temos acesso à lista de dados referentes ao resultado da busca; e a janela de ajuda (4).

Figura 3: Janelas referentes ao processo de pesquisa.



Durante o processo de busca, a depender das ações do usuário, as distintas janelas são automaticamente acionadas pela plataforma. Segundo uma ação em cadeia, a partir de uma busca efetuada na janela pesquisa, a janela frequência é utilizada. Consecutivamente, a partir de uma seleção desempenhada em frequência, a janela de contexto entra em atuação. Dessa forma, a *ativação* de cada uma das áreas da plataforma se apresenta interligada.

Além disso, para o acesso aos recursos do banco de dados, ratificamos a importância do processo de cadastro. É apenas por meio da efetivação do cadastro na plataforma que a realização de buscas pode ser concretizada.

3.4.1.1 A janela de pesquisa e a distribuição dos dados no campo de frequência: o menu e suas funções

Figura 4: Menu principal da plataforma de busca do *Corpus* do Português.



Na parte superior do menu principal, em 5, temos as distintas opções de busca oferecidas pela plataforma. A opção *List* possibilita a busca por estruturas simples, como o lexema “soco”, ou expressões complexas, como “levar um soco”, apresentando os resultados da pesquisa efetuada em forma de uma lista. *Chart* assim como em *List*, possibilita uma busca por elementos simples ou expressões; entretanto, o resultado é exposto em forma de gráficos.

A opção *Collocates*, por sua vez, possibilita a busca por combinações, independente do grau de complexidade/tamanho das expressões em jogo. Por meio de seu uso, podemos observar quais elementos ocorrem um com o outro.

Figura 5: Menu de busca da plataforma do *Corpus do Português NOW* segundo a opção *collocates*.

The screenshot shows the search interface for the 'Collocates' option. At the top, there are tabs for 'List', 'Chart', 'Collocates' (which is highlighted with a blue box), and 'Compare KWIC'. Below the tabs, there are two input fields: the first is labeled 'Word/phrase [POS]?' and the second is labeled 'Collocates [POS]'. Below these fields is a horizontal scale of buttons labeled '+ 4 3 2 1 0 0 1 2 3 4 +'. The '0' button on the left is highlighted in blue. Below the scale are two buttons: 'Find collocates' and 'Reset'. At the bottom, there is a checkbox labeled 'Sections Texts/Virtual Sort/Limit Options'.

Contamos, nessa opção, com dois espaços de inserção para busca de elementos. O *Word/phrase* é utilizado para determinar qual item será considerado “principal” na combinação e será o referente para o raio/alcance de busca da segunda expressão, apresentada em *collocates*. Abaixo desse, observamos uma escala que determinará esse alcance – a direção e distância do colocado em relação ao referente. O quadrado totalmente preenchido em azul figurará como a expressão exposta em *Word/phrase* e os números, como a posição do item em *collocates* ao se tratar desse referente. A título de exemplificação, observemos a figura a seguir:

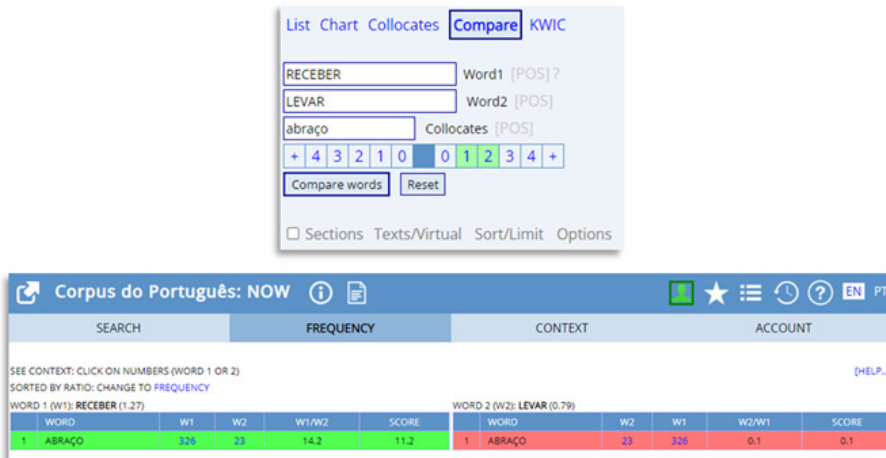
Figura 6: Exemplo de busca na opção *collocates* e de seus respectivos resultados.

The screenshot shows the search results for the combination 'RECEBER' and 'abraço'. The search parameters are 'RECEBER' in the 'Word/phrase [POS]?' field and 'abraço' in the 'Collocates [POS]' field. The scale shows the number 3 highlighted in green. Below the search interface, there is a table with the following columns: SEARCH, FREQUENCY, CONTEXT, and ACCOUNT. The table contains five rows of results, each with a date, time, and a snippet of text containing the search terms. The first row is: 1 19-05-16 PT Hiper-fm disse que o dinheiro era para pagar o funeral de a mulher. Hoje foi receber um abraço de a Cristina Ferreira e um "novo flego" para a vida. The second row is: 2 19-04-27 BR gshow — Foto: TV Globo # "Mentira", exclamou a atriz e o receber um abraço apertado de a Filiza. # "Deste, eu vivi só para. The third row is: 3 19-04-08 PT A Televisão " Os ânimos entre as duas concorrentes só voltam a abrandar depois de Maria receber um abraço de a melhor amiga antes de voltar a casa. # Já par. The fourth row is: 4 19-02-26 PT A Televisão se esquivar de a ex-companheira. Esta segunda-feira, o apresentador recebeu Cristina depois de receber um abraço de Maria em o Vlog em a TV. A. The fifth row is: 5 19-02-14 PT Máxim Mas há quem valente de forma primordial a toque físico, tal como receber um abraço inesperado ou andar de mãos dadas em a rua, sempre.

Ao buscarmos pela combinação “RECEBER” – em *Word/phrase* – mais “abraço” – em *collocates* –, selecionando o número 3 à direita em tal escala, a plataforma me apresentará dados/fragmentos textuais nos quais o lexema “abraçar” aparece até 3 posições após o uso do verbo “receber”, como em “receber abraço”, “receber um abraço” e “receber um grande abraço”.

Em *compare*, pesquisamos por colados associados a dois itens referentes, possibilitando a comparação entre suas configurações estruturais e funcionais. Na Figura 7, temos a busca do colado “abraço” em relação aos lexemas verbais referentes “RECEBER” e “LEVAR”.

Figura 7: Exemplo de busca na opção *compare* e de seus respectivos resultados.



O vocábulo “abraço” vê-se mais empregado em associação ao lexema verbal RECEBER em detrimento a LEVAR. Tal resultado ratifica indícios observados por Teixeira (2020) ao se tratar de uma possível restrição colocacional do uso do verbo *levar* na configuração de predicadores complexos de passividade. Seu uso vê-se mais associado a elementos nominais de natureza negativa.

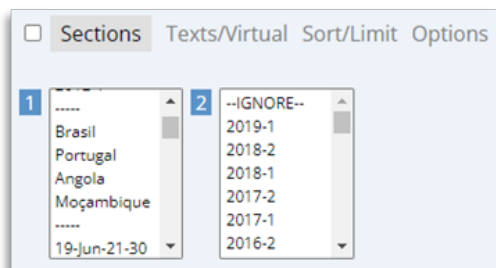
Keyword in Context, denominado KWIC, propicia averiguar os padrões nos quais os itens/expressões ocorrem no *corpus*.

Figura 8: Exemplo de resultados para a realização de busca pelo vocábulo “preciso” na opção KWIC.



A opção *Sections*, por sua vez, viabiliza delimitar em qual(is) seção(ções) – gênero, recorte temporal e/ou dialeto – se deseja efetuar a busca, ensejando a possibilidade de comparar a ocorrência de um objeto em duas seções distintas.

Figura 9: Exemplo de recursos propiciados pela opção *Sections* no menu de busca.



3.4.1.2 Ferramentas de pesquisa

Dentre as ferramentas proporcionadas pela plataforma a fim de otimizar a experiência do usuário/pesquisador, encontra-se a sintaxe específica de busca. Segundo o uso de símbolos especiais, códigos preestabelecidos referentes à distribuição dos elementos em categorias gramaticais (PoS ou *_pos*), a busca por lemas e listas personalizadas de palavras/expressões, podemos acionar comandos de pesquisa de forma rápida, além de ampliar o escopo de resultados.

3.4.1.2.1 Símbolos especiais

No conjunto de itens especiais, contamos com: o asterisco (*), a barra vertical (|), o símbolo de igual (=) e o símbolo de menos (-). O asterisco (*) se figura como um caractere curinga. Estipula que qualquer espaço/*slot* com a sua presença pode ser preenchido por qualquer elemento. Ao ser empregado de forma independente, antes ou depois de uma palavra, determina que o espaço ocupado pelo asterisco, na expressão, pode estruturalmente representar qualquer lexema.

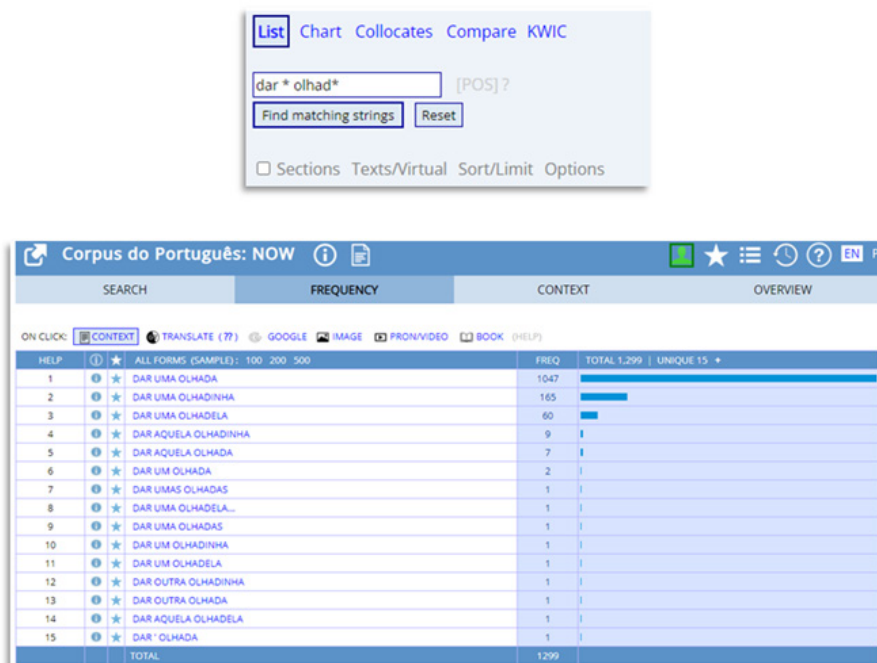
Figura 10: Exemplo de emprego do símbolo asterisco (*) em um comando de busca.

ON CLICK: [CONTEXT](#) [TRANSLATE \(??\)](#) [GOOGLE](#) [IMAGE](#) [PRON/VIDEO](#) [BOOK \(HELP\)](#)

HELP	ALL FORMS (SAMPLE: 100 200 500)	FREQ
1	DAR UMA OLHADINHA	165
2	DAR AQUELA OLHADINHA	9
3	DAR UM OLHADINHA	1
4	DAR OUTRA OLHADINHA	1
	TOTAL	176

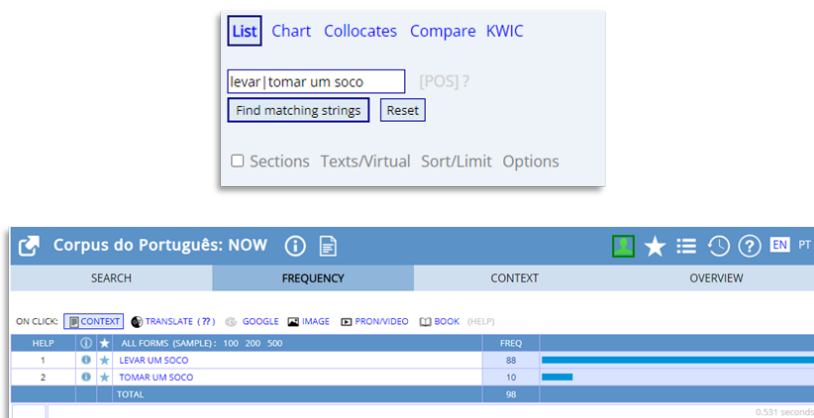
Na Figura 10, ao empregarmos o comando de busca “dar * olhadinha”, a plataforma nos propicia ocorrências de uso com expressões como “dar uma olhadinha” e “dar aquela olhadinha”, nas quais vemos um elemento entre o verbo “dar” e o item “olhadinha”. É importante, nesse caso, realçar a importância do *espaçamento* no processo de pesquisa. O asterisco (*) somente foi lido/compreendido pelo algoritmo de busca como um lexema independente, como uma palavra, devido ao espaço existente entre este e as unidades “dar” e “olhadinha” no comando de pesquisa. Caso esteja acoplado a uma palavra, representará que tal espaço no vocábulo poderá ser preenchido por qualquer caractere ou conjunto de caracteres. Ao delimitarmos, por exemplo, como comando a linha “dar * olhad*”, teremos resultados como “dar uma olhada”, “dar uma olhadinha”, “dar uma olhadela”, “dar aquela olhadinha” e “dar aquela olhada” (cf. Figura 11).

Figura 11: Exemplo de emprego do símbolo asterisco (*), de forma independente e associado a uma palavra, em um comando de busca.



A barra vertical (|), por sua vez, indica alternância. Efetiva uma busca por mais de uma opção de elementos. Ao delimitarmos o comando “levar|tomar um soco”, temos como resultados as expressões “levar um soco” e “tomar um soco”.

Figura 12: Exemplo de emprego do símbolo barra vertical (|) em um comando de busca.



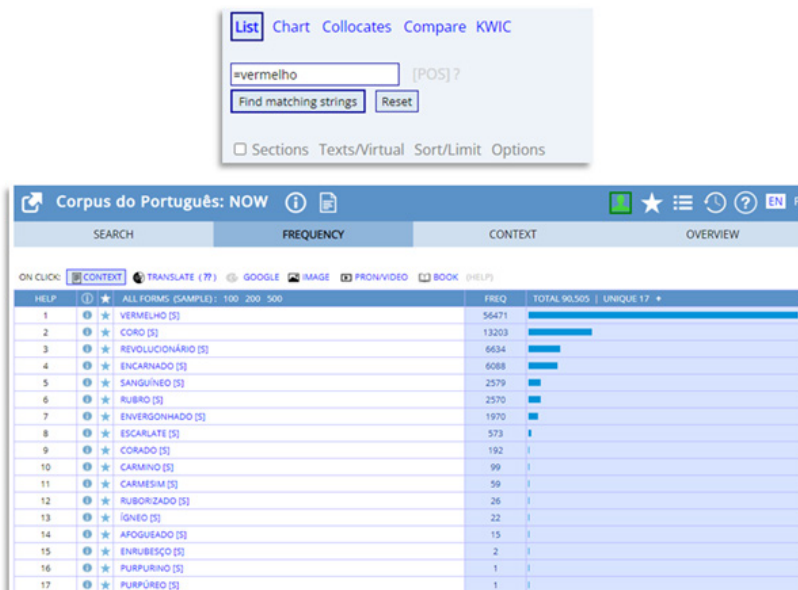
Nesse contexto, chamamos, novamente, atenção para o espaçamento. Não deve haver nenhum espaço entre os elementos de busca alternantes e o símbolo (|). Caso exista, a plataforma indicará que há um erro na sintaxe de busca, pois a expressão não será encontrada dentro o banco de dados (cf. Figura 13).

Figura 13: Exemplo de erro indicado pela plataforma.



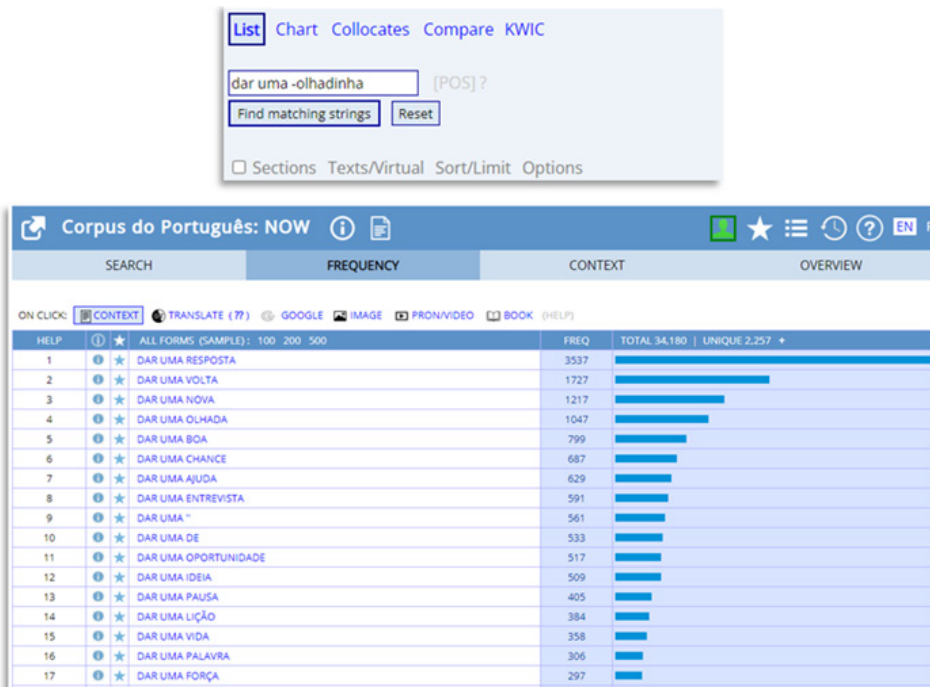
O símbolo de igual (=), ao ser aplicado antes de uma palavra, possibilita a busca por seus sinônimos. Logo, por meio de seu emprego, teremos acesso tanto aos usos de sinônimos dessa palavra no interior do *corpus* quanto de seus usos em si. Ao pesquisarmos por “=vermelho”, obtemos como resultados expressões como “coro”, “rubro” e “escarlate”.

Figura 14: Exemplo de emprego do símbolo de igual (=) em um comando de busca.



Já o símbolo de menos (-) indica uma restrição de busca. Na pesquisa, serão incluídos todos os elementos possíveis, menos aquele precedido pelo sinal de menos. Na busca “dar uma -olhadinha”, mesmo que a expressão “dar uma olhadinha” seja produtiva no *corpus*, a plataforma apresentará somente o uso de expressões compostas por “dar uma” mais elementos que não sejam “olhadinha”.

Figura 15: Exemplo de emprego do símbolo de menos (-) em um comando de busca.



3.4.1.2.2 Codificação de *Part of Speech* (POS e/ou *_pos*)

Uma das características marcantes da organização textual do *Corpus* do Português é o fato de se figurar como um banco de dados de *corpora* anotados. Cada uma de suas unidades vocabulares recebe *tags*/etiquetas referentes às suas características estruturais e funcionais (segundo um ponto de vista (morfo-)sintático). Como exemplo, vemos a segmentação dos lexemas em conjuntos de acordo a sua classe – em verbos, substantivos pronomes, dentre outras –, a sua função sintática na sentença – sujeito ou objeto – e a atributos morfológicos relativos ao seu número, gênero, e no caso dos itens verbais, pessoa – singular ou plural, feminino ou masculino e primeira, segunda ou terceira pessoa verbais. Assim, é possível fazer uso de *códigos/tags* que possibilitam a restrição/especificação dos itens que irão compor nossa busca.

Tais códigos são apresentados na plataforma por meio da função *Parts of Speech* (POS ou *_pos*). Nessa, classificações de partes do discurso/texto são utilizadas como recursos de busca. A seguir, podemos visualizar exemplos de códigos empregados na plataforma segunda essa função:

Tabela 1: Exemplos de códigos empregados no *Corpus* do Português via função *Parts of Speech* (POS e *_pos*)

Códigos PoS (versões longas e abreviadas)	Códigos <i>_pos</i>	Tipo de itens pesquisados	Exemplos de resultados
NOMES			
NOUN	_n	Substantivos comuns em geral	bola, mesa, soco, fome
N			
NMS	_nms	Substantivos comuns masculinos no singular	caderno, carrinho, menino, homem
NMP	_nmp	Substantivos comuns masculinos no plural	cadernos, carrinhos, meninos, homens
NFS	_nfs	Substantivos comuns femininos no singular	borracha, menina, bicicleta, mesa
NFP	_nfp	Substantivos comuns femininos no plural	borrachas, meninas, bicicletas, mesas
O	_o	Substantivos próprios	Vitória, Copa, Maria, Roberto
NAME			
DETERMINANTES			
DET	_d	Determinantes em geral	seu, este
D			
DD	_dd	Determinantes demonstrativos	isto, aquilo, este
DP	_dp	Determinantes possessivos	meu, nosso, seu
ARTIGOS			

ART			
L	_l	Artigos em geral	a, as, um, uns
LD	_ld	Artigos definidos	a, as, o, os
LI	_li	Artigos indefinidos	um, uns, uma, umas
PRONOMES			
P	_p	Pronomes em geral	que, se, ele, ela
PD	_pd	Pronomes demonstrativos	isto, isso, aquilo
PI	_pi	Pronomes indefinidos	nada, algo, alguém
PO	_po	Pronomes pessoais com função de objeto	os, me, lhe
PS	_ps	Pronomes pessoais com função de sujeito	eu, ele, ela
PR	_pr	Pronomes relativos	que, quem, cujo
PREPOSIÇÕES			
PREP	_e	Preposições em geral	para, com, de
E			
CONJUNÇÕES			
CONJ			
C	_c	Conjunções em geral	porque, pois, embora
ADJETIVOS			
ADJ			
J	_j	Adjetivos em geral	maior, grande, melhor
	_jms	Adjetivos masculinos no singular	novo, necessário, bom
	_jmp	Adjetivos masculinos no plural	novos, necessários, bons
	_jfs	Adjetivos femininos no singular	nova, necessária, boa
	_jfp	Adjetivos femininos no plural	novas, necessárias, boas
	_jcs	Adjetivos uniformes no singular	possível, importante, nacional
	_jcp	Adjetivos uniformes no plural	possíveis, importantes, nacionais
ADVÉRBIOS			
ADV			
R	_r	Advérbios em geral	não, também, ainda
INTERJEIÇÕES			
I	_i	Interjeições em geral	caramba, ah, ei
NUMERAIS			

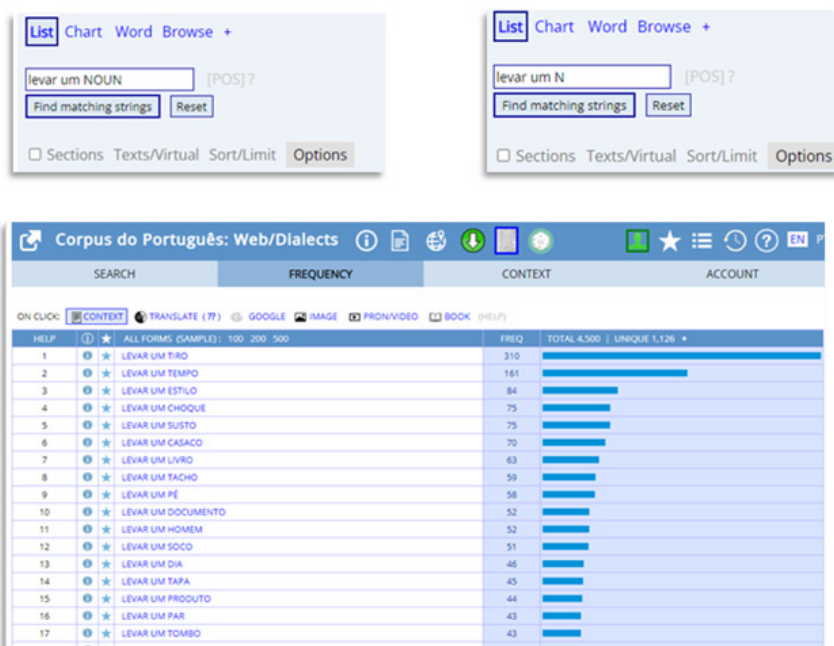
NUM	_m	Numerais em geral	dois, 6, primeiro
M			
MO	_mo	Números ordinais	último, segundo, terceiro
MC	_mc	Números cardinais	3, quatro, 20
VERBOS			
VERB	_v	Verbos em geral	receber, ter, levar
V			
VIP	_vip	Verbo no presente do indicativo	é, está, vai
VIF	_vif	Verbo no futuro do indicativo	levará, estarei, comemos
VIS	_vis	Verbo no pretérito do indicativo	foi, disse, afirmou
VII	_vii	Verbo no pretérito imperfeito do indicativo	estava, eram, queria
VSP	_vsp	Verbo no presente do subjuntivo	seja, possa, tenham
VSJ	_vsj	Verbo no pretérito imperfeito do subjuntivo	fossem, tivesse, ficassemos
VPP	_vpp	Verbo no particípio presente	sendo, incluindo, fazendo
VPS	_vps	Verbo no particípio passado	feito, realizado, publicado
VC	_vc	Verbo condicional	teria, poderia, estaria
VR	_vr	Verbo no infinitivo	ser, ter, estar
CLASSIFICAÇÃO DE PESSOA E NÚMERO VERBAL			
X² -1S	_X-1s	Verbo na primeira pessoa do singular	como, tivesse, andei
X -2S	_X-2s	Verbo na segunda pessoa do singular	tens, pertences, estás
X -3S	_X-3s	Verbo na terceira pessoa do singular	tem, come, anda
X -1P	_X-1p	Verbo na primeira pessoa do plural	comemos, tivéssemos
X -2P	_X-2p	Verbo na segunda pessoa do plural	ficais, moveis
X -3P	_X-3p	Verbo na terceira pessoa do plural	têm, comem, andam

Fonte: Autoral.

- 2 Nesse caso, o elemento “X” não se figura como parte estrutural do código. Apenas representa a possibilidade de preenchimento do *slot* a ele relativo com qualquer um dos códigos previamente expostos referentes ao tempo e modo verbais (*vip*, *vsp*, *vpp*, dentre outros).

Podemos fazer uso de dois tipos de codificação de *Part of Speech*: *POS* ou *_pos*. As *tags POS* são empregadas em letras maiúsculas e fazem referência a tipos de lexemas. Utilizamos códigos *POS* no lugar de vocábulos a fim de efetuarmos uma busca ampla pautada na classificação sintática do tipo de item cuja realização desejamos averiguar no banco de dados.

Figura 16: Exemplo de emprego dos códigos *POS* “NOUN” e “N” em comandos de busca e seus respectivos resultados.



Na Figura 16, temos os comandos “levar um NOUN” e “levar um N”, nos quais empregamos, no *slot* relativo ao item não verbal a compor o padrão composto pelo verbo “levar”, os códigos “NOUN” e “N”. Assim, indicamos ao algoritmo de busca da plataforma uma especificação em relação à configuração do tipo de expressões cuja realização desejamos observar: dados constituídos pelo verbo “levar”, o artigo “um” e qualquer item caracterizado como substantivo, nessa determinada ordem. Como resultados, encontramos “levar um tiro”, “levar um tempo”, “levar um estilo”, dentre outros, em que o terceiro elemento é um substantivo, independente da natureza da expressão em si.

Os códigos *_pos*, em contrapartida, não possibilitam, primordialmente, a substituição de um item de busca por uma codificação pautada nos atributos que desejamos delimitar, mas sim, restringir o campo de atuação de um item predeterminado. Considerando a multifuncionalidade que elementos linguísticos e/ou padrões estruturais podem desempenhar a depender de seu contexto de uso, podemos, segundo o uso de códigos *_pos*, especificar os atributos dos elementos que fazem parte do nosso interesse de pesquisa.

Como exemplo de padrão multifacetado, é possível observar o padrão vocabular “tiro”, que pode se referir, dentro dos *corpora*, conforme a Figura 17, a um elemento nominal no masculino singular (*levar um tiro*), a um elemento verbal na primeira pessoa no singular do presente do indicativo (*Eu tiro o lixo todo os dias*) ou a parte de um nome próprio (*O alistamento é feito em a Junta de o Serviço Militar, em o Tiro de Guerra*). Assim, caso desejemos efetuar uma busca que enseje a análise do uso do le-xema nominal “tiro”, é oportuno realizarmos especificações no nosso comando de busca. Se indicarmos como comando de pesquisa simplesmente o padrão “tiro”, encontraremos tanto usos referentes ao substantivo “tiro” quanto ao verbo “tirar” na primeira pessoa no singular do presente do indicativo.

Figura 17: Exemplo de resultados para o comando de busca “tiro”.

The screenshot shows the 'Corpus do Português: NOW' interface. The search bar contains 'tiro'. The results table is as follows:

HELP	1	2	3	TOTAL	FREQ
	★	★	★		45459
					2724
					1222
					49405

Logo, é possível restringirmos nosso escopo de busca segundo o comando “tiro_nms”, por meio do qual nos serão apresentados resultados referentes somente ao uso do substantivo no masculino singular “tiro”.

Figura 18: Exemplo de resultados para o comando de busca “tiro_nms”.

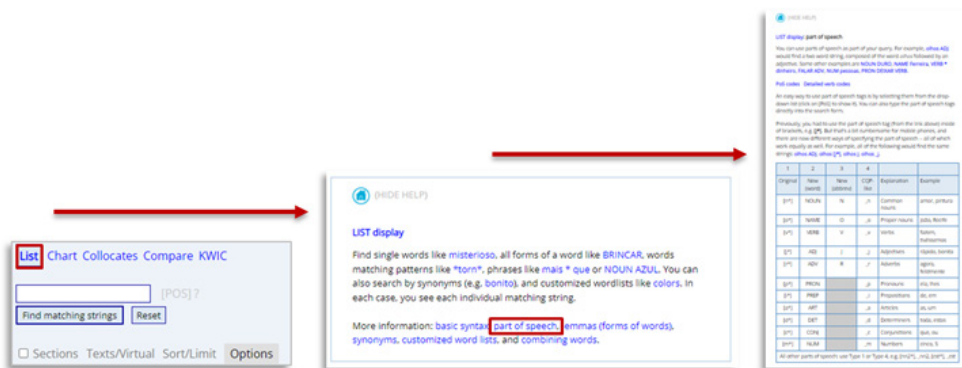
The screenshot shows the 'Corpus do Português: NOW' interface with the search bar containing 'tiro_nms'. The results table is as follows:

HELP	1	FREQ
	★	45459

Nos casos citados, temos conhecimento prévio dos marcadores utilizados para classificar os elementos linguísticos em pauta. Entretanto, nem sempre isso é possível. Logo, é essencial considerarmos meios para identificar a categorização dos itens que nos interessam no interior dos *corpora*. Para isso, contamos com dois caminhos metodológicos:

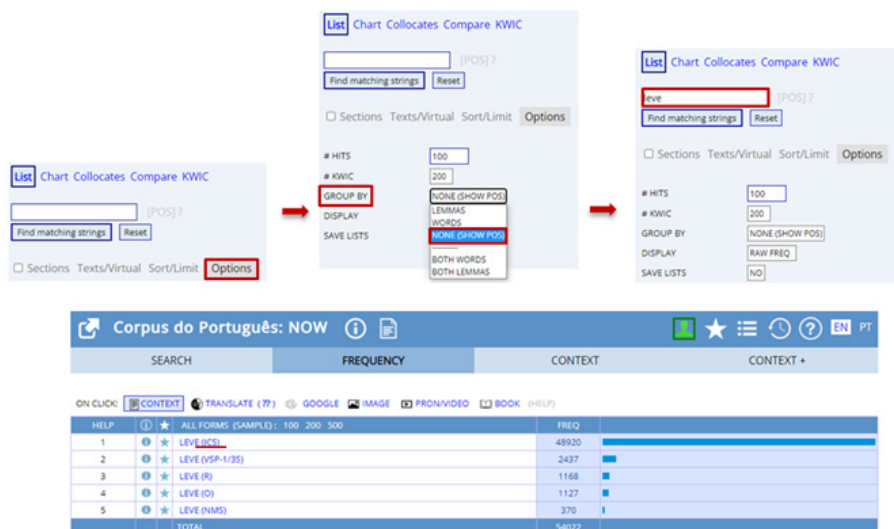
(1) Ir no menu de busca, selecionar a opção “list”, que abrirá a caixa de ajuda referente a essa função na parte direita da tela e, em seguida, clicar na opção “part of speech”, a qual promoverá acesso a uma tabela com os códigos POS e *_pos* básicos utilizados na plataforma, assim como uma breve explicação de seus usos em conjunto com exemplos diversificados.

Figura 19: Procedimento metodológico para acessar a caixa de ajuda referente aos códigos POS e *_pos*.



(2) No menu de busca, clicar em “Options”, referente à forma na qual os resultados são expostos na janela de frequência dos resultados, e, em “GROUP BY”, selecionar a opção “NONE (SHOW POS)”. Em sequência, é necessário apenas digitar o vocábulo que deseja saber a codificação/tag e visualizar os resultados na janela de frequência, que serão organizados com base na classificação do vocábulo no banco de dados. Aparecerá o vocábulo e, ao final, entre parênteses, sua classificação/codificação.

Figura 20: Procedimento metodológico para acessar os resultados de buscas segundo seus códigos POS.

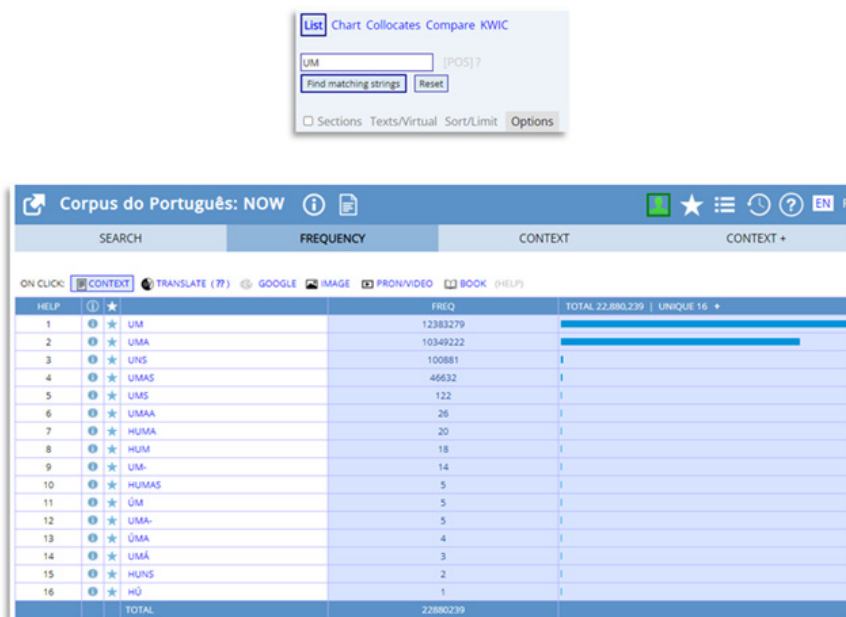


3.4.1.2.3 Uso de lemas

Outra ferramenta é a utilização de uma unidade lexical em letras maiúsculas a fim de efetuar buscas por formas flexionadas que se encontrem associadas a essa unidade base. Todas as unidades linguísticas do banco de dados encontram-se vinculadas a um item lexical que funciona como seu representante, de forma similar a uma entrada de dicionário. Por exemplo, a estrutura verbal no infinitivo “receber”, em um dicionário, atua como representante de todos os itens que se encaixam nesse paradigma verbal, como “recebi”, “recebeu”, “recebido”, dentre outros. Na plataforma, o mesmo ocorre. Temos como efetuar buscas a partir de uma unidade representante, e isso não se restringe a pesquisas centradas em composições verbais.

Ao realizarmos uma pesquisa segundo o comando “UM”, no qual fazemos uso do artigo indefinido masculino em letras maiúsculas, conforme a Figura 21, o determinante será compreendido como um lema. Dessa forma, na janela de resultados, a plataforma nos apresentará dados de usos relativos a todas as unidades que são associadas a esse item representante, como “um”, “uma”, “uns”, “umas”.

Figura 21: Exemplo de busca pelo lema “UM” e seus respectivos resultados.



Para acessar informações referentes aos lemas das unidades linguísticas, contamos com a função “options” disposta no menu de busca principal. É necessário apenas clicar em “options” e, em “GROUP BY”, selecionar a opção “LEMMAS”, que, ao efetuar uma busca por uma expressão, os lemas a que se vê associado serão apresentados entre colchetes como resultados na janela de frequência.

Figura 22: Procedimento metodológico para acessar o(s) código(s) *lemma* associados ao artigo “uma”.

HELP	ALL FORMS (SAMPLE): 100 200 500	FREQ
1	[UM]	10349222
2	[UMA]	321787
3	[]	448
TOTAL		10671457

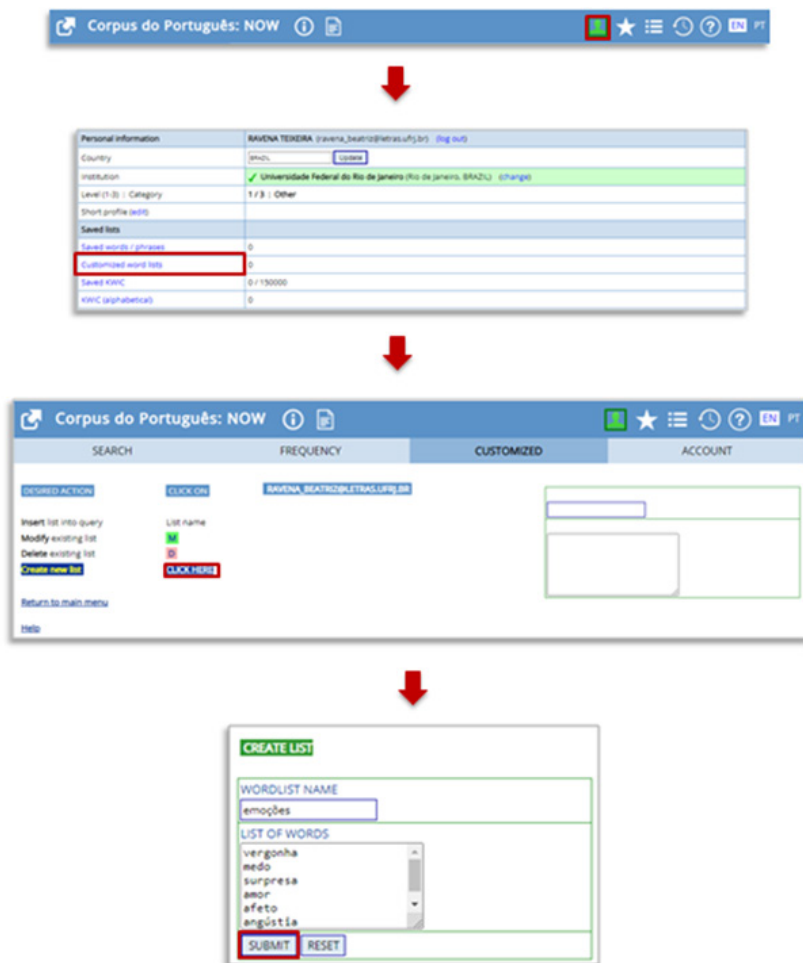
No exemplo exposto na Figura 22, podemos perceber que o artigo indefinido feminino “uma” encontra-se relacionado a mais um lema: “UM”, “UMA” e unidades vazias (casos em que a lematização não foi especificada pelo sistema). Por conseguinte, chamamos atenção para o fato de que uma unidade pode estar associada a mais de um representante. Logo, no uso da categorização de lema no processo de pesquisa, é essencial, a fim de considerar por total os recursos dos *corpora*, verificar, segundo os passos previamente expostos, se os vocábulos a serem focalizados na busca se veem associados a mais de um item base/representante.

3.4.1.2.4 Listas de busca personalizadas

A plataforma também possibilita a criação de listas de busca personalizadas, ou criadas do zero pelo próprio usuário ou salvas segundo resultados de buscas, além daquelas já disponibilizadas como exemplos. Cada uma das listas desenvolvidas será salva no banco de dados do perfil do usuário e poderão ser acessadas independente do momento ou local de uso da plataforma.

Para criar uma lista personalizada, primeiramente é necessário acessar a página relativa ao perfil do pesquisador na plataforma – possível por meio do botão “login”, em verde, na parte superior da tela –, clicar em “Customized word lists”, em seguida, clicar em “CLICK HERE”, inserir, nos quadros apresentados à direita da tela, o nome da lista de busca e os lexemas/vocábulos que deseja incluir na lista e salvar as informações por meio do botão “SUBMIT”, conforme o esquema a seguir:

Figura 23: Procedimento metodológico para criar uma lista de busca personalizada.



Na lista, cada um dos vocábulos a ser considerado deve ser disposto sozinho, um abaixo do outro, em linhas diferentes, de forma que o sistema possa compreendê-los como unidades distintas. Além disso, é importante ratificar que apenas itens simples podem ser adicionados na construção de uma lista (*bola, carrinho, bicicleta*, entre outras). Expressões complexas, formadas por mais de um elemento lexical, estes divididos por espaçamento, não são aceitas pelo sistema.

3.4.1.2.5 Combinações entre ferramentas de busca

Considerando os recursos associados à sintaxe de busca propiciada pela interface do banco de dados, é essencial tratar da combinação entre estes. O usuário pode, no momento de construção do comando de pesquisa, lançar mão de mais de uma ferramenta a fim de alcançar os seus objetivos e otimizar o processo de trabalho. A seguir, temos exemplos de comandos nos quais uma associação entre ferramentas é observada:

Quadro 2: Exemplos de comandos de busca com a presença de mais de uma ferramenta de pesquisa

Comando	Explicação	Exemplos de resultados
LEVAR uma *ada	Vê-se o uso do lema “LEVAR”, possibilitando a busca por todas as formas flexionadas desse paradigma verbal e o emprego do caractere (*), propiciando uma busca por palavras terminadas por “ada”.	levar uma facada, levou uma facada, levei uma bolada, levando uma pancada
levar tomar uma *ada	Utiliza-se a barra vertical, para indicar a possibilidade de busca por um elemento verbal ou outro, e o caractere (*), delimitando uma busca por palavras terminadas em (ada), já que o mesmo viabiliza o preenchimento de sua posição por qualquer conjunto de letras.	levar uma pancada, levar uma bolada, levar uma década, tomar uma gelada, tomar uma goleada, tomar uma pancada
VIP-3P UM TIRO	Faz-se uso do código POS “VIP-3P”, o qual delimita que sua posição pode ser preenchida por qualquer unidade verbal na terceira pessoa no plural do presente do indicativo, e dos lemas “UM” e “TIRO”, indicando que suas posições podem ser ocupadas por qualquer elemento associado aos seus respectivos paradigmas.	dão um tiro, levam um tiro, disparam um tiro, recebem um tiro, mandam um tiro, dão uns tiros
VERMELHO de @emoções	Utiliza-se o lema “VERMELHO”, assinalando que sua posição pode ser preenchida por qualquer unidade associada a esse representante e da lista de palavras “@emoções”, propiciando uma busca na qual sua posição pode ser ocupada por qualquer item pertencente a essa lista.	vermelho de raiva, vermelho de vergonha, vermelha de vergonha, vermelhas de vergonha
TER_vis-3s UM NOUN	Empregam-se: o lema “TER”, para indicar o paradigma verbal em foco, delimitando a flexão verbal de interesse a partir do código _pos “_vis-3s” para indicar que se deseja observar dados em que o verbo ter encontre-se na terceira pessoa do singular do pretérito do indicativo; o lema “UM”, assinalando que sua posição pode ser preenchida por qualquer elemento associado a essa unidade representante e o código POS “NOUN”, o qual sinaliza que essa posição deve ser preenchida por qualquer item etiquetado como substantivo.	teve um papel, teve um aumento, teve uma queda, teve um problema, teve um gol

Fonte: Autoral.

Observamos, logo, que combinações são possíveis, mas também devem ser coerentes. Se procuramos, a título de exemplificação, por expressões compostas pelo verbo “levar” e elementos nominais terminados em “-ada”, caso consideremos a presença de um artigo indefinido na configuração da expressão, este deve concordar com o gênero do item nominal, sendo empregado no feminino. Dessa forma, ao compor um comando de busca segundo as ferramentas de pesquisa disponibilizadas, devemos avaliar quais são as melhores opções para atingir nossa meta de pesquisa e refletir sobre os benefícios e possíveis adversidades que podem se apresentar devido seu uso.

3.4.1.3 As janelas de resultados e acesso aos dados e suas fontes

No que concerne à exposição dos resultados de buscas e ao acesso aos dados, é importante ressaltar a organização das janelas de frequência, contexto e contexto expandido. Na janela de frequência, temos a apresentação inicial dos resultados de uma forma compacta, sua distribuição a depender da opção de busca selecionada – *list*, *chart*, *collocates*, *compare* ou *KWIC*.

3.4.1.3.1 Distribuição dos resultados na janela de frequência

Segundo as opções *list* e *collocates*, como previamente indicado, os resultados são apresentados na janela de frequência na forma de uma lista disposta por meio de uma matriz, na qual cada expressão exposta como resultado funciona como um botão que possibilita o acesso aos dados referentes ao seu uso.

Figura 24: Exemplo de resultado de busca segundo função *list* na janela de frequência.

HELP	ALL FORMS	FREQ	TOTAL L. ABS / UNIQUE TR.
1	LEVOU UM SUSTO	775	
2	LEVEI UM SUSTO	275	
3	LEVARAM UM SUSTO	187	
4	LEVA UM SUSTO	130	
5	LEVAR UM SUSTO	112	
6	LEVAMOS UM SUSTO	34	
7	LEVADO UM SUSTO	30	
8	LEVARA UM SUSTO	28	
9	LEVANDO UM SUSTO	23	
10	LEVO UM SUSTO	11	
11	LEVAR UM SUSTO	10	
12	LEVARÃO UM SUSTO	5	
13	LEVEM UM SUSTO	2	
14	LEVAREM UM SUSTO	2	
15	LEVARA UM SUSTO	2	
16	LEVASSEM UM SUSTO	2	
17	LEVASSE UM SUSTO	1	

Na primeira coluna, temos o número da expressão de acordo com o seu posicionamento dentre os resultados. Na segunda, vemos o botão que possibilita angariar mais informações sobre o(s) elemento(s) que compõe(m) o resultado.

Figura 25: Exemplo de uso de botão referente a informações dos lexemas que compõem os resultados na janela de frequência.

HELP	ALL FORMS (SAMPLE): 300 200 500	FREQ	TOTAL 1,630 UNIQUE 18 +
1	LEVOU UM SUSTO	775	
2	LEVEI UM SUSTO	275	

Já na terceira coluna, temos a opção de salvar³ a expressão disposta como resultado, a fim de visualizá-la ou utilizá-la em um momento posterior. Na quarta e quinta coluna, temos respectivamente, a apresentação dos resultados e sua frequência bruta de ocorrências no *corpus*.

A opção *chart*, por sua vez, apresenta os resultados em forma de gráfico (vertical ou horizontal). As expressões resultantes são expressas segundo segmentações pautadas nas diferentes seções disponíveis nos *corpora*, por data de ocorrência ou por país, no caso da aba *NOW*.

Figura 26: Exemplo de resultados segundo a função *chart* na janela de frequência.



Em *compare*, comparamos a frequência de uso de um item em relação à sua associação com outros dois elementos. Logo, os resultados são expostos em duas matrizes distintas, uma levando em consideração a frequência de ocorrência do item colocado em correlação à expressão 1, e outra considerando sua correlação com a expressão 2, conforme a Figura 27.

Figura 27: Exemplo de resultados segundo a função *compare* na janela de frequência.

WORD 1 (W1): LEVAR UM (1.07)					WORD 2 (W2): TOMAR UM (0.51)						
WORD	W1	W2	W1/W2	SCORE	WORD	W2	W1	W2/W1	SCORE		
1	SUSTO	1808	956	1.9	1.0	1	SUSTO	956	1808	0.5	1.0

A busca segundo a opção *KWIC* não propicia o acesso à janela de frequência, mas sim à visualização do arranjo de elementos no entorno da expressão pesquisada na janela de contexto. Assim, cada elemento cotextual recebe uma demarcação de cor distinta, a fim de possibilitar uma melhor visualização de padrões por parte do usuário.

3 As palavras/expressões salvas são acessáveis segundo o perfil do usuário, em “Saved words / phrases”.

Figura 28: Exemplo de resultados segundo a função KWIC.

The screenshot displays the 'Corpus do Português: NOW' web application. At the top, there is a navigation bar with 'SEARCH', 'CHART', 'CONTEXT', and 'ACCOUNT' tabs. Below this is a search results table. The table has several columns: a numbered list (1-15), a date column (e.g., 18-09-29 BR), a source column (e.g., Jornal Exora), a snippet of text with the search term highlighted in yellow, and a series of icons for actions like 'levar um susto', 'ouvir', 'traduzir', and 'analisar'. The search term 'quando' is highlighted in the snippets. The interface also includes a 'CLICK FOR MORE CONTEXT' button and a 'RE-SORT' button.

3.4.1.3.2 Distribuição dos resultados na janela de contexto

Na janela de contexto, temos acesso aos dados de uso dos resultados observáveis na aba de frequência. Independente da função empregada no processo de busca – *list*, *chart*, *collocates*, *compare* ou *KWIC* –, a janela de contexto apresenta-se uniforme para todos os resultados. É composta por uma tabela na qual temos acesso às seguintes informações:

1ª coluna: número da expressão de acordo com o seu posicionamento dentre os resultados;

2ª coluna: data de publicação do texto fonte do dado em jogo e o país de referência do mesmo;

3ª coluna: botão para ouvir a leitura automática do dado no tradutor do Google;

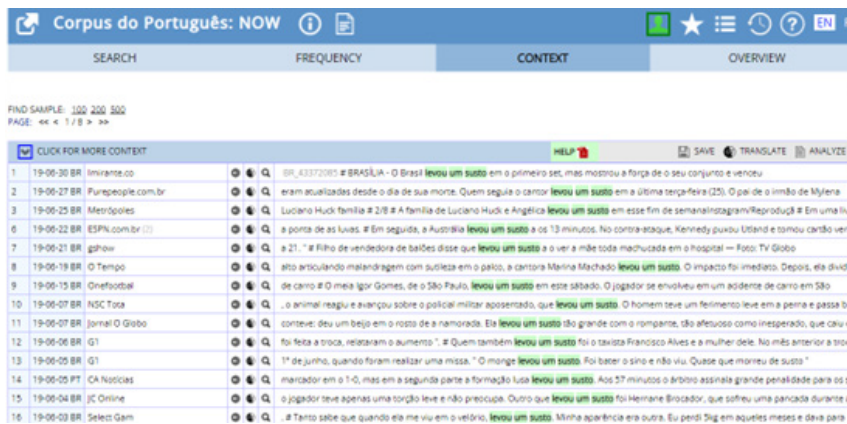
4ª coluna: botão para acessar a tradução do dado no tradutor do Google;

5ª coluna: botão que possibilita a transformação dos vocábulos do dado em *links* para pesquisas no interior do *corpus*; e

6ª coluna: linha de segmento textual no qual o dado do resultado encontra-se em destaque.

A seguir, podemos visualizar a distribuição dos resultados na janela *context*.

Figura 29: Exemplo de resultados na janela *context*.



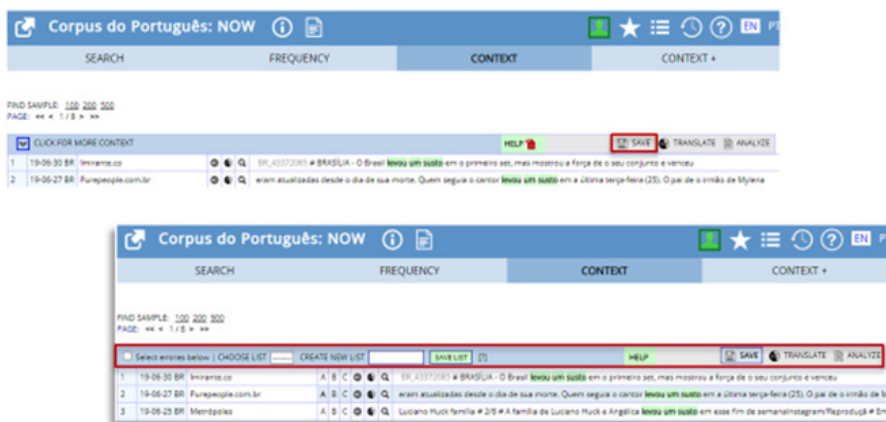
Ao clicarmos na segunda coluna, somos redirecionados para uma janela de contexto expandido, *Context +*, na qual observamos a presença de um trecho expandido do dado selecionado, assim como com informações relativas ao *link* de acesso do texto fonte do dado, e a data de publicação e título do mesmo.

Figura 30: Exemplo de resultados na janela *Context +*.



Já ao clicarmos na terceira coluna da janela de contexto, somos redirecionados para a página eletrônica referente à fonte do dado em questão. Além disso, a janela *context* oferece como recurso a possibilidade de salvar dados selecionados em uma lista (já existente ou nova) que ficará disponível no perfil do usuário para acesso em momento posterior.

Figura 31: Exemplo de uso da função *SAVE* na janela *Context*.



Assim, por meio de pesquisas na plataforma, podemos, a partir dessa ferramenta, criar *sub-corpora* pautados nas necessidades e objetivos dos usuários.

3.4.2 O PROCESSO DE COLETA NO CORPUS DO PORTUGUÊS

A fim de ilustrar o processo de coleta relativo a bancos de dados digitais, em especial na plataforma do *Corpus do Português*, selecionamos tratar da angariação de dados relativos a construções de predicção de passividade compostas por verbo (semi-)suporte no Português Brasileiro (PB), a partir de Teixeira (2020). Por meio de descrição metodológica apresentada pela autora, lidaremos com questões concernentes aos cuidados e passos a serem levados em consideração na coleta de dados em plataformas digitais, assim como apresentar reflexões acerca dos aspectos positivos e negativos que permeiam o processo.

Para ratificar a presença de perífrases compostas por verbo (semi-) suporte na rede de predicadores complexos de passividade do PB, e mapear os atributos (formais e funcionais) daqueles, o estudo desenvolvido por Teixeira (2020) abarca o uso de padrões constituídos pelos verbos *levar*, *tomar*, *sofrer*, *receber* e *ganhar* –, como: *levar um soco*, *tomar uma pancada*, *receber um fora*, *sofrer um gol*, *ganhar um tiro*.

Ao elaborar uma análise de cunho sincrônico, a autora recorre a aba *NOW*, a qual conta em sua base de dados, artigos de revistas, jornais e blogs associados ao Google News cuja publicação se deu a partir do ano de 2012. Em um primeiro, em uma busca exploratória, com o intuito de apreender distinções de colocações relativas ao uso dos elementos verbais em foco, deu-se o uso da opção *compare* do menu de busca. Foram observados os itens aos quais se compatibilizam os verbos (semi-)suporte a serem considerados na análise a fim de visualizar seus sentidos em uso concreto e estabele-

cer uma melhor compreensão dos tipos de dados aos quais se depararia no processo efetivo de coleta.

Após considerar as características configuracionais dos predicaadores complexos, deu-se o seguinte padrão de busca: (i) no campo *Word/phrase* acrescentou os verbos em caixa alta para, assim, dar conta de todas as suas possibilidades de expressão; (ii) em *collocate* – uma das ferramentas que a interface de coleta do banco de dados disponibiliza – empregou os elementos nominais e/ou sintagmáticos que se associam ao elemento verbal; e (iii) definiu-se, como padrão de rastreo para o elemento nominal e/ou sintagmático, sua ocorrência até a quarta posição, levando em consideração o número de palavras que antecedem ou precedem o verbo.

Com o intuito de contemplar as possíveis realizações, fez-se uso, como chave de busca, de escolhas mais específicas e mais vastas do padrão construcional em análise, por exemplo, [LEVAR uma/um + NOUN], que viabiliza observar qual(is) elemento(s) nominal(is) ocorrem em adjacência ao verbo (*LEVAR uma + pancada; LEVAR uma + surra*). Tal procedimento foi repetido com as outras formas verbais em jogo (*tomar, sofrer, receber, e ganhar*). Nessa investigação, vinculada à variedade do Português do Brasil, como parâmetro, leva em consideração somente os textos publicados em sites brasileiros, com isso, em sessões, delineamos o campo de rastreo para Brasil. Finalizado o processo de busca no *corpus*, realizou-se uma triagem dos dados, a fim de verificar se os fragmentos de uso angariados eram relevantes para a investigação.

Dessa forma, para o processo de coleta em banco de dados digital, destacamos quatro passos elementares, conforme o esquema a seguir:

Esquema 1: Passos relativos ao processo de coleta em bancos de dados digitais.



Fonte: Autoral.

Com o intuito de desenvolver um trabalho de coleta coerente e otimizá-lo, o pesquisador deve, em primeiro plano, considerar seu foco de estudo, assim como os recursos que a plataforma/interface de busca pode lhe oferecer. Entretanto, uma compatibilização entre teoria e prática deve ser realizada de modo a evitar, e considerar, possíveis problemas durante o processo de coleta. Logo, é essencial executar pesquisas de caráter exploratório na plataforma para familiarizar-se com as suas ferramentas e suas implicações que concerne ao acesso aos dados. Em seguida, dá-se, em efetivo, o processo de coleta no qual o estudo se pautará e, consecutivamente, a triagem dos dados angariados em tal processo, pois, nem sempre os dados aos que temos acesso são relevantes/adequados para nossos objetivos particulares de análise.

3.5 CONSIDERAÇÕES FINAIS

Buscamos, neste capítulo, introduzir, ao (futuro)pesquisador, o universo de coleta online, em especial a busca de dados por meio do gerenciador de *corpora Corpus* do Português. Com isso, ressaltamos a relevância do uso de *corpora* virtuais para o campo de análise linguística, uma vez que, contribuem para a otimização do processo de coleta de dados. Ressaltamos que, ao recorrer a *corpora* online, o pesquisador encontra: fácil acessibilidade, rápido rastreamento de dados de diversas línguas e variedades, mais agilidade e menos esforço para a obtenção de dados linguísticos, fácil seleção das fontes e/ou gêneros textuais e acesso a informações sobre os dados. Em contrapartida, depara-se também com: páginas indisponíveis e páginas bloqueadas devido à política de privacidade do conteúdo. Por fim, destacamos a importância de delimitar as diretrizes de coleta segundo os recursos dos *corpora* disponíveis, bem como adaptá-las aos objetivos da análise linguística.

REFERÊNCIAS

- DAVIES, M. O corpus do português. *Corpus do Português*. Disponível em: <https://www.corpusdoportugues.org/x.asp>, 2016. Acesso em: 15 mar. 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.
- TEIXEIRA, R. B. de S. *Estruturas com verbo (semi)suporte: a variação sob um prisma construcionista*. Dissertação de Mestrado. Universidade Federal do Rio de Janeiro, Faculdade de Letras, 2020.

CAPÍTULO 4

OS PRIMEIROS PASSOS EM ANÁLISE COLOSTRUCIONAL: PESQUISAS DE PREDICAÇÃO PARA ILUSTRAR UM FAZER

Marcia dos Santos Machado Vieira

Mariana Gonçalves da Costa

Ravena Beatriz de Sousa Teixeira

Pâmela Fagundes Travassos

Pedro Poppolino

Lais Lima de Souza

(Universidade Federal do Rio de Janeiro)

The collocation of a word or a 'piece' is not to be regarded as mere juxtaposition, it is an order of mutual expectancy. The words are mutually expectant and mutually prehended. Firth (1957: 12)

4.1 INTRODUÇÃO

Um tema a mobilizar toda a equipe do Projeto Predicar no final do segundo decênio do século XXI é, sem dúvida, *análise colocacional*, em especial a *colostrucional*. Também as colaborações à engrenagem que moveu a equipe estão mutuamente relacionadas às expectativas de desenvolvimento de pesquisas cada vez mais ricas em (inter)ações.

Procuramos estudar a respeito de métodos de análise colostrucional, para, então, desenvolver pesquisas para mensurar, empiricamente, a força de atração e de repulsa de possibilidades de colocação de lexemas/unidades construcionais em *slot(s)* de construções. E o fizemos com o intuito de investigar, principalmente, problemas relativos ao fenômeno de variação construcional (MACHADO VIEIRA, 2016): alternância de usos verbais na configuração de predicadores complexos (com verbo (semi-) auxiliar ou com verbo (semi-)suporte, cf. TEIXEIRA, 2020; COSTA e LIMA, 2021; TRAVASSOS, 2021) e de predicções (como a predicção transitiva direta com pronome SE, cf. SARAIVA DE PONTES, 2022).

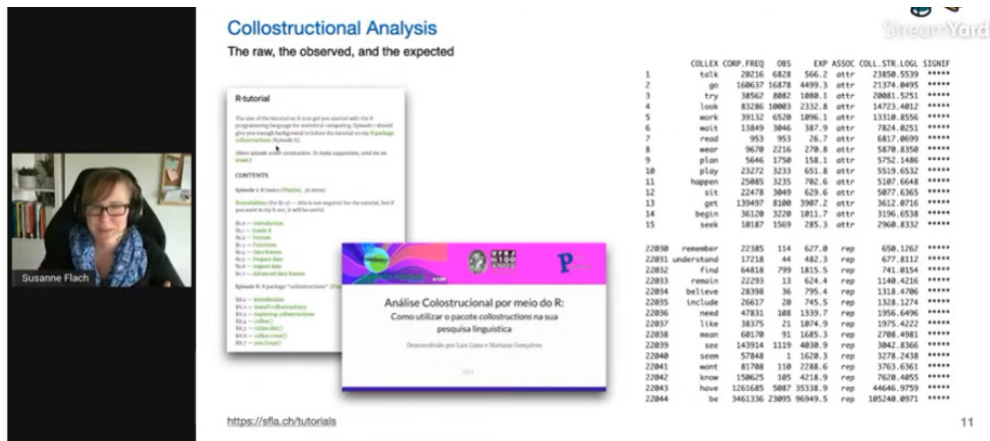
E entendemos que nossa atenção voltada a lidar com a variação construcional por meio de diferentes métodos de tratamento de dados tem enriquecido nossas descrições, viabilizando maior flexibilidade quanto a olhar correlações (entre construções de diferentes dimensões em termos de complexidade, grupos de fatores linguísticos, co-textuais, contextuais).

Em linhas gerais, usamos aqui o termo técnico *análise colostrucional* para designar um método explorado há quase vinte anos, um procedimento empírico de mapeamento estatístico de colocação a partir do ponto de vista da construção. Esse ponto de vista permite identificar instâncias de coocorrência entre lexemas e construções ou entre lexemas e outros lexemas auxiliando na descrição de construções, de unidades construcionais em construções (mais complexas). Diz respeito, na verdade, a uma família de métodos de investigação de coocorrências, que vêm sendo, ao longo dos anos, reconfigurados em razão da ênfase teórico-metodológica e do recorte do objeto de observação com que são perspectivados os fenômenos linguísticos.

Uma boa síntese do que aprendemos no percurso que temos trilhado a partir do momento em que nos mobilizamos para encarar, coletivamente, essa frente de investigação de usos linguísticos de predicções em *corpora*, pode ser vista no minicurso de COSTA; LIMA (2021) que ocorreu durante o Festival do Conhecimento da UFRJ: <https://youtu.be/-lh215zX36U> (Minicurso *Análise Colostrucional através do R: Como utilizar o pacote colostructions na sua pesquisa linguística*). Citado por Susanne Flach (durante sua conferência *Colostructions, contexts, communities*,¹ conforme *print* a seguir), o minicurso configura-se, em alguma medida, como uma tradução em língua diferente do inglês, até porque está centrado no que foi aprendido a partir dos tutoriais de Flach (que podem ser encontrados em <https://sfla.ch/tutorials/>) e na prática de investigação de construções de futuridade (cf. COSTA; SOUZA; MACHADO VIEIRA, 2020).

1 Transmitido ao vivo em 26 de nov. de 2021, no canal [PPG Linguística UFRJ](https://www.youtube.com/watch?v=YNB_rNX1fT0). Disponível em: https://www.youtube.com/watch?v=YNB_rNX1fT0. Acessado em: 11 fev. 2022.

Imagem 1: Print do dispositivo da conferência *Collostructions, contexts, communities*, de Susanne Flach.



O objetivo deste capítulo é introduzir qualquer pesquisador interessado (mesmo iniciante) à prática de análise colostrucional por meio do tratamento quantitativo/estatístico na plataforma R, um software gratuito, via, principalmente, o pacote *collostructions*, de Susanne Flach, assim como no Excel, segundo a abordagem dinâmica de dados a partir de matrizes automatizadas. Então, o básico da linguagem que temos empregado num perfil de passo a passo desse fazer será o cerne deste capítulo. Somamos a esse passo a passo uma breve exposição de pesquisa de colocação de construções em construções textuais-discursivas baseada no sistema, que já nos mobiliza no âmbito do Projeto Predicar. E o fazemos, tendo em vista o olhar sobre combinações de unidades linguísticas em unidades de materialização linguística mais complexas, unidades textuais/discursivas com diferentes configurações, manifesto, por exemplo, em Wiedemer e Machado Vieira (2022).

4.2 PILARES E TIPOS DE ANÁLISE COLOSTRUCIONAL

4.2.1 O QUE APRENDEMOS NA TEORIA E NA PRÁTICA DE ANÁLISE COLOSTRUCIONAL?

You shall know a word by the company it keeps! Firth (1957: 11)

Aprendemos que a observação empírica de colocações já tem uma história de experiências científicas, com perfis e cursos de análise. Um perfil é o do exame estatístico de colocações de palavras. John Rupert Firth, por exemplo, é um dos pesquisadores voltados a uma abordagem de mensuração da colocação/colocabilidade orientada por frequência: associação de palavras (geralmente duas) que cocorram mais fre-

quentemente do que suas respectivas frequências em relação com a extensão do texto em que aparecem poderiam levar a prever.

Outro perfil é o que tem orientado muitos pesquisadores que contemporaneamente dão atenção à relação entre unidades construcionais ou à relação entre lexemas (unidades da gramática) e *slots* construcionais: uma abordagem de mensuração da colocação orientada por significância estatística e pareamento forma-função/significação. Nesse caso, colocação diz respeito, geralmente, à colocação lexical: combinação de uma unidade autônoma/base e outra unidade que é mais ou menos dependente semanticamente (*colocador*). Por exemplo: *dia e bom* > *bom dia* é uma unidade fraseológica no Português, *bom* é um adjetivo atraído ao substantivo *dia* para designar uma fórmula de cortesia/polidez; construção similar, como *ótimo dia* já não revela o mesmo grau de idiomaticidade, embora também possa acompanhar o substantivo e constituir uma expressão de cortesia. Não é raro encontrar o termo colocação associado a situações de “expressões de múltiplas palavras”, combinações resultantes do que é referido como lexicalização na literatura linguística. A análise colostrucional, baseada em coocorrência de itens, é um método de análise colocacional desenvolvido por Stefan Th. Gries e Anatol Stefanowitsch que difere de outros métodos em razão de seu perfil sensível ao conceito linguístico de construção, unidade em que forma e função são sociocognitiva e convencionalmente pareadas. Dá-se fundamentalmente por meio de identificação dos lexemas ou das unidades construcionais que mais ocorrem em *slot* de uma dada construção, em dois *slots* de uma construção ou em *slots* de mais de uma construção.

Também podemos observar coocorrência de palavras em (co)ligação a certos espaços morfossintáticos, gramaticais (categoriais ou funcionais). É o caso, por exemplo, de formas verbais em relação a predicções relacionais: *ser, ficar, virar* em relação a *flamenguista, carioca*; os primeiros ligados à classe dos verbos relacionais (ou de ligação) e os segundos ligados à função de predicativo do sujeito.

Uma vez sintetizadas as frentes de análise de coocorrências, expomos a representação que Proils (2019) tece, recorrendo a exemplos da língua inglesa:

type of cooccurrence	name in linguistics	example
word with word	collocation	<i>dark night, make tea</i>
word with word class	colligation	<i>PREP DET naked eye,</i> <i>POSS true feelings</i>
word with grammatical relation	colligation	<i>attention as direct object</i>
word with construction	collostruction	<i>give + ditransitive,</i> <i>trick + into-causative</i>
word with alternating constructions	collostruction	<i>give + ditransitive vs.</i> <i>give + to-dative</i>
word with word within construction	collostruction	<i>force into making,</i> <i>fool into thinking</i>

Table 1.1: Types of cooccurrences

Fonte: Proils (2019, p. 4).²

*Cooccurrences of various items have been studied under a variety of names, e. g. collocation, colligation or collocation, and are important for identifying multiword units, extracting valency patterns and generally investigating patterns of language use. While there are well-understood and fully worked out statistical models for the analysis of cooccurrences of pairs of words, no such model exists for cooccurrences of larger linguistic structures (PROILS, 2019, p. v).*³

E seguimos, a partir daqui, com breve referência a um outro perfil possível que, por um lado, leva a uma flexibilização de um modelo de estudo de coocorrências mais generalizado em Gramática de Construções e, por outro, revela um desafio de análise ainda a vencer. Afinal, também põe em cena unidades construcionais mais complexas que as sentenciais, além de prever exame de coocorrência de bem mais de duas unidades construcionais/lexemas nessas unidades complexas.

Referimo-nos, especificamente, ao exame de coocorrência orientado pela relação de frequências de unidades linguísticas comparadas entre si e certos textos/discursos em que são acionadas: associação de unidades linguísticas umas com as outras por conta da inclinação de ocorrência que têm em certos ambientes similares, cotexto e/ou contexto. A alternância influenciada por fatores mais amplos de ordem discursivo-pragmática, é variação por padrão ou paradigma discursivo⁴ (LEINO e ÖSTMAN, 2005).

Travassos e Machado Vieira (2019) e Travassos (2021) expõem uma pesquisa sobre o grau de atração de unidades linguísticas a predicadores complexos que designam, em alguma medida, *início* (aspecto inceptivo ou ingressivo) em função de diversos fatores de ordem discursiva e pragmática: *dar saída, dar arrancada, dar entrada, dar partida, dar largada, dar começada, dar iniciada*, por exemplo. Apresentam um estu-

- 2 No quadro: tipos de construção – palavra com palavra, palavra com classe de palavra, palavra com relação gramatical, palavra com construção, palavra com construções em alternância, palavra com palavra dentro de construção; nomes na linguística para o tipo de coocorrência, respectivamente – colocação, coligação, coligação, colocação, colocação, colocação; e os respectivos exemplos do inglês.
- 3 Coocorrências de mais de um item foram estudadas sob uma variedade de nomes, e. g. colocação, coligação ou colocação/colocação, e são importantes para identificar unidades multipalavras, extrair padrões de valência e investigar, em termos gerais, padrões de uso da língua. Embora existam modelos estatísticos bem compreendidos e integralmente elaborados para a análise de coocorrências de pares de palavras, não existe tal modelo para coocorrências de estruturas linguísticas maiores (PROILS, 2019, p. v).
- 4 Leino; Östman (2005, p. 200, tradução nossa): “um padrão de discurso é o correlato cognitivo do tipo de texto definido linguisticamente e o gênero definido socioculturalmente. A compreensão do texto e do discurso ocorre principalmente em termos de padrões discursivos”.

do socioconstrucionista que explora, em termos de distribuição de dados, essa extensão do termo coocorrência, na medida que construções são associadas a domínios particulares de aplicação, como a temática, domínio discursivo, o gênero textual e o ato de fala.

Também Wiedemer e Machado Vieira (2022) expõem e exploram a concepção de combinação de unidades construcionais orientada pela natureza da construção textual-discursiva, ao focalizarem a relação de certos lexemas e certas unidades construcionais aos gêneros receita culinária e resumo acadêmico-científico.

4.2.2 QUAIS SÃO OS TIPOS DE ANÁLISE COLOSTRUCIONAL?

Uma variedade de tipos de coocorrências está abrigada sob o rótulo análise colostrucional, introduzido por Stefanowitsch e Gries (2003). A análise colostrucional lida com coocorrências que envolvem construções, pareamentos forma-função no sentido da Gramática de Construções. E engloba pelo menos esses três tipos:

1) *Análise de colexemas⁵ simples/análise colexêmica simples* – a associação (atração ou repulsa) entre uma construção e lexemas ou outras unidades⁶ num *slot* particular da construção está sob exame – é o caso, por exemplo, de construção de predicação relacional e verbos ([*Participante1 Vrelacional Caracterização do Participante1*]predicação relacional e *ser/estar/ficar*),⁷ em que a força de atração e repulsa dos verbos para o preenchimento do *slot* verbal da predicação é mensurada;

2) *Análise de colexemas distintivos* – a associação entre uma palavra e um par de construções em alternância em que essa palavra ocorre – é o caso de verbo DAR em relação a predicador complexo com verbo suporte (*dar uma palavra*) e predicador complexo com verbo semiauxiliar (*dar de falar*);⁸

3) *Análise de colexemas covariantes* – associação entre pares de palavras que ocorrem em dois *slots* de uma construção sob exame – é o caso, por exemplo, de verbos e elementos não verbais ligados ao predicador complexo com verbo (semi-)suporte (*dar/lançar/deitar + uma olhada/uma espiada/um olhar*).

A lógica central que pode levar o pesquisador a uma dessas análises é a seguinte: O que é típico (em termos configuracionais) para uma construção? O que é (proto)típico de um *slot* construcional? O que é tipicamente atraído para um lugar na construção ou para mais de uma construção? E essa natureza de (proto)tipicalidade não é necessariamente mensurada apenas em termos de frequências brutas/distribuição de

5 **Colexemas são co-localizações/co-location de itens verbais.**

6 **Em certas análises desenvolvidas (como a de Travassos, 2021), ora a colocação de afixos de grau a elemento não verbal de construção com verbo suporte é o foco (*dar uma saidinha/saidela*), ora a colocação de certos sintagmas adjetivos no sintagma nominal ligado ao verbo suporte *dar* é a tônica (*dar uma rápida/breve saída*)..**

7 **Estudo, no Projeto Predicar, a respeito de predicações relacionais é: FERREIRA, 2018.**

8 **Cf., por exemplo, o capítulo *Análise comparativa para detectar auxiliaridade verbal e predicadores complexos* desta obra.**

formas. A informação de frequências brutas nem sempre dá conta de explicar a contingência de uma unidade construcional em presença de uma certa condição (cotextual; e mesmo contextual, como será visto em item específico neste capítulo). Por isso, trata-se de uma relação entre frequência e contingência. E, assim, é viável comparar diferentes tipos de frequências por cotexto, ou contexto de análise. E, então, podemos detectar, a partir do cálculo colostrucional, uma espécie de contínuo que vai desde o que é tipicamente associado a um *slot* construcional até o que é não lhe é tipicamente associado, mas é atípico ou eventual nesse lugar construcional (e, por conseguinte, mais repellido).

A ideia central de coocorrências de lexemas ou unidades construcionais ou deste(a)s em construções é, atualmente, a de que isso se dá em diferentes níveis de manifestação linguística: no nível léxico-gramatical e/ou textual-discursivo (sociointeracional). De acordo com Gries et al. (2005 *apud* Hilpert, 2008: 3), a *performance* do falante tende a ser mais influenciada pela força colostrucional do que pela frequência bruta, ou seja, a força colostrucional ofereceria uma visão mais confiável da realidade psicológica.

4.2.3 EXEMPLIFICANDO OS TIPOS DE ANÁLISE COLOSTRUCIONAL POR MEIO DE TRÊS OBJETOS DE PESQUISA NO PROJETO PREDICAR

1) Um dos objetos de estudo é o que permite visualizar o *design* de análise de colexemas simples: cf. SARAIVA DE PONTES, 2022.

A fim de ilustrar a concepção de uma análise colostrucional de colexemas simples, podemos observar o mapeamento de padrões de predicação direta compostos por pronome –SE empregados no Português do Brasil em prol de diferentes efeitos de impersonalização discursiva, conforme elaborado por Saraiva de Pontes (2022). A autora, ao se voltar para a observação dos atributos (estruturais e funcionais) de predicadores em tais predicações, determina o perfil semântico delas assim como o grau de atração de itens verbais acionados para preencher *slots* nelas:

·o primeiro slot (V1) da construção [Predicador_{TD} + SE]_{predicador};

“Investiga-se o comportamento da variável”

(SARAIVA DE PONTES, 2022, p. 29)

·o primeiro e o segundo slot (V1 e V2) da construção [Verbo_{(SEMI-)AUXILIAR} + SE + VP_{TD}]_{predicador complexo*}

“Pode-se argumentar que existem peritos conveniados”

(SARAIVA DE PONTES, 2022, p. 29).

Assim, conforme a premissa de que verbos relativos a alguma natureza semântica ou discursiva são mais fortemente atraídos para os *slots* dessas construções, como aqueles associados à expressão de processos mentais (cognitivos) do participante con-

trolador (indutor), faz-se uso do programa R, via interface gráfica RStudio Desktop, e do pacote de processamento de dados *collostructions* desenvolvido por Susanne Flach de modo a avaliar as características dos itens verbais em jogo.

Desse modo, com base na proposta de análise colexêmica simples, vê-se que a maioria dos 10 verbos mais atraídos para o *slot* verbal do padrão [Predicador_{TD} + SE]_{predicador} – *saber, observar, perceber, notar, esperar, usar, estimar, destacar, considerar e estabelecer* – revela “noções referentes a processos mentais/cognitivos atrelados a questões de atenção, memória, linguagem, raciocínio, tomada de decisões” (p. 189), enquanto aqueles mais associados aos *slots* V1 e V2 do padrão [Verbo_{(SEMI-)AUXILIAR} + SE + VP_{TD}]_{predicador complexo} – *sendo, respectivamente, poder, dever, pretender, buscar e dizer, observar, afirmar, concluir, ver* – indicam, em contexto de uso, processos cognitivos relativos ao eu-enunciador. Dessa forma, por meio da análise do grau de associação entre itens verbais e *slots* das construções de predicação em pauta, a autora tece delimitações acerca dos atributos funcionais das construções.

2) O segundo objeto de estudo é o que propicia perceber como se configura o *design* de análise colostrucional de colexemas distintivos: cf. TRAVASSOS, 2021, vídeo entre as conferências do I Colóquio Internacional do site do Projeto VariaR.⁹

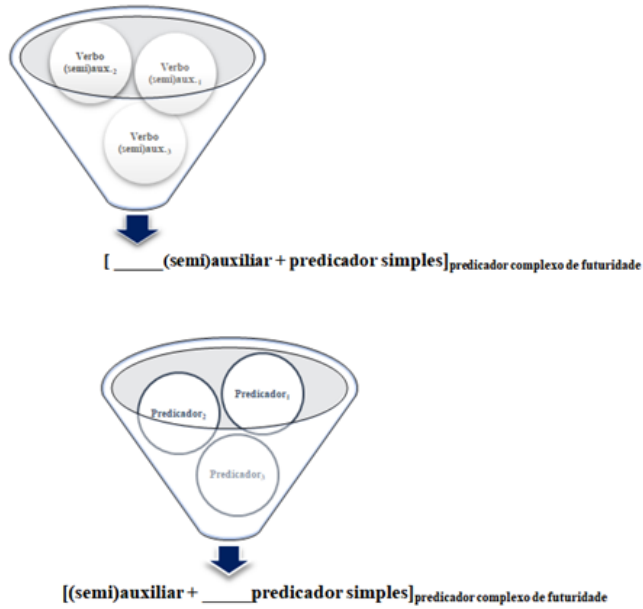
3) O terceiro objeto de estudo é aquele que nos faz entender como se dá a configuração do *design* de análise colostrucional de colexemas covariantes: cf. COSTA, 2022, vídeo.

4.2.4 E COMO PROCEDEMOS À ANÁLISE COLOSTRUCIONAL NA PESQUISA DE VARIAÇÃO DE CONSTRUÇÕES DE PREDICAÇÃO?

Procuramos observar quais formas verbais são atraídas para o preenchimento de *slot* de certas construções:

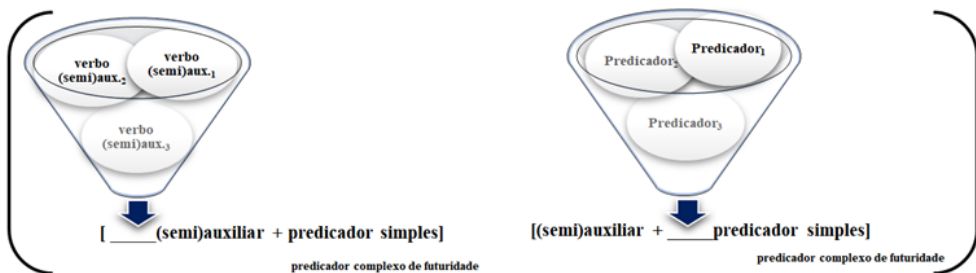
⁹ <https://www.youtube.com/watch?v=itfg2wxyWwk&t=1s>. Acesso em: 23 mar. 2022.

Figura 1: Representação da relação de unidades verbais a *slot* de uma construção de predicador complexo de futuridade, por força de atração ou ao papel de verbo (semi)auxiliar ou ao papel de predicador simples/verbo principal.



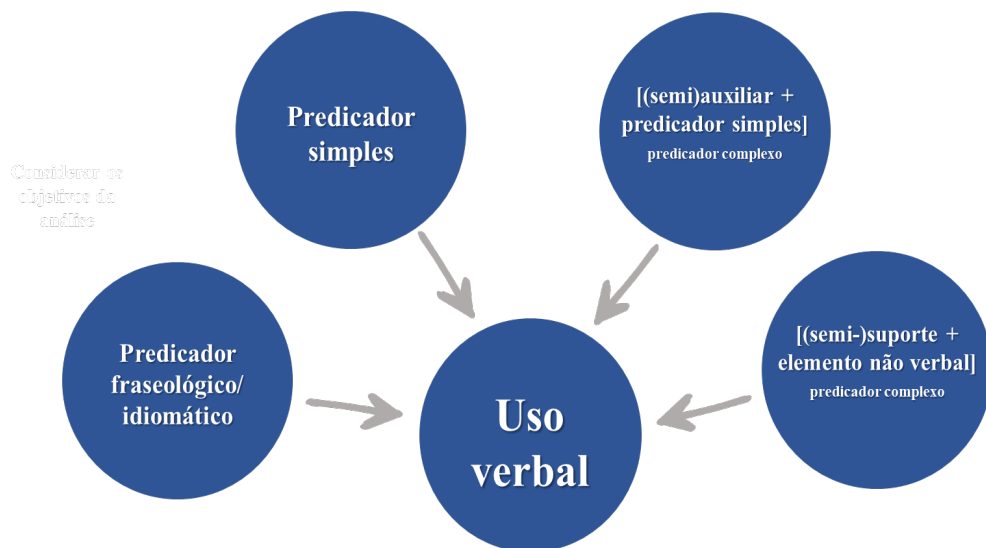
Procuramos investigar quais formas verbais são atraídas para dois *slots* numa construção:

Figura 2: Representação da relação de unidades verbais a *slots* de uma construção de predicador complexo de futuridade, por força de atração ao papel de verbo (semi-)auxiliar e ao papel de predicador simples/verbo principal.



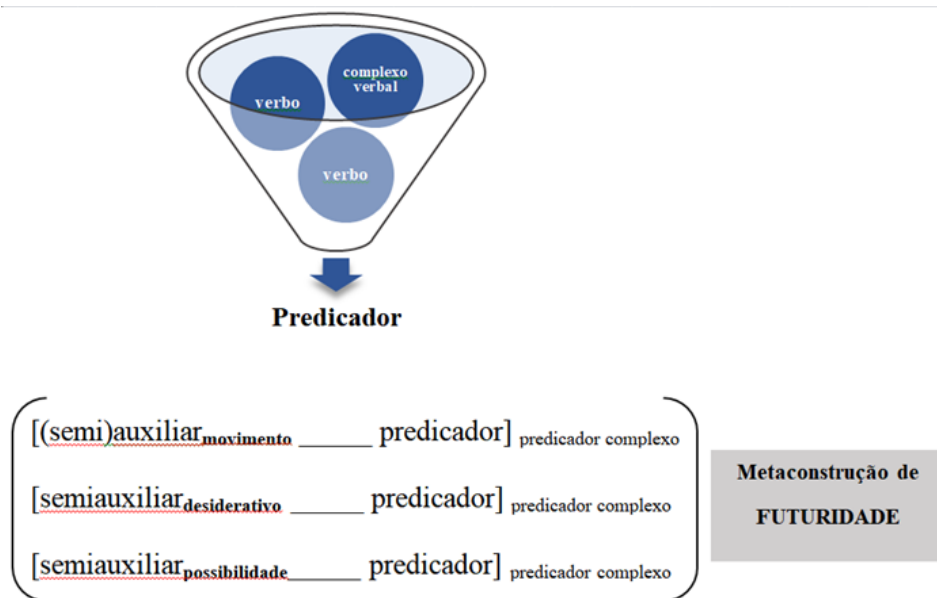
Buscamos examinar a relação entre formas verbais constitutivas das construções que focalizamos com outras construções em amostras de dados:

Figura 3: Representação da relação das unidades verbais atraídas para *slots* de uma construção de predicação ou predicador com unidades verbais atraídas para *slots* de outras construções que licenciam dados no *corpus*.



E normalmente fazemos isso com base na associação de formas verbais a dois padrões construcionais. Por exemplo, no que diz respeito a predicadores complexos que implicam a noção de inclinação ou projeção futura, futuridade, é possível avaliar a relação entre predicadores que são matizados por essa noção temporal verificando sua relação a dois tipos construcionais: o com verbos semiauxiliares desiderativos (*querer, esperar*, por exemplo) e o com verbo (semi-)auxiliar de movimento no tempo (como é o caso do verbo *ir*). Ou podemos considerar outro tipo construcional: o com verbos semiauxiliares modais de possibilidade (*poder*, por exemplo). E assim, examinamos a força colostrucional de atração ou repulsa de itens verbais a *slots* de pelo menos duas construções de predicador complexo de futuridade:

Figura 4: Representação da relação entre padrões construcionais de futuridade em variação no Português.



4.2.5 POR QUE FAZER USO DA ANÁLISE COLOSTRUCIONAL BASEADA NO ITEM?

Diferentemente da análise colostrucional baseada no sistema, a versão que será demonstrada neste capítulo é feita a partir dos itens estudados (construções) isolados do restante do *corpus*. Isso significa que a análise distributiva é perspectivada quanto a essas construções em si e não ao todo. Considerando as contribuições de Goldberg (2003) à teoria da Gramática das Construções, toda a unidade textual é constituída de diferentes construções de diferentes dimensões combinadas. Em outras palavras, lidamos, nas nossas manifestações linguísticas, com combinação de construções; então, o isolamento de determinadas construções para a análise estatística não seria congruente com essa perspectiva teórica.

Entretanto, optamos pela análise baseada no item. Essa escolha foi feita devido à extensão do trabalho manual envolvido na análise baseada no sistema. Esse trabalho, até o momento, ainda carece de recursos de automatização. Sem a possibilidade de efetuar a análise de maneira (semi-)automatizada, o linguista acaba por se limitar a *corpora* anotados, que facilitam no processo de recolhimento das informações necessárias para a análise baseada no sistema, como: “quantas vezes encontramos o item na construção?”; “quantas vezes encontramos o item no *corpus*?”; “quantas vezes encontramos a construção no *corpus*?”; e “quantas construções há no *corpus*?”. Uma vez que grande parte dos *corpora* em língua portuguesa carece das informações sistematiza-

das necessárias para uma análise como essa, faz-se necessário considerar opções que não impliquem um trabalho manual muito extenso ao linguista e, sendo assim, possibilitem que a análise seja aplicada a *corpora* extensos que não seriam contemplados em análises mais trabalhosas.

Stefanowitsch e Gries (2003) defendem a relevância da análise colostrucional baseada no item, afirmando que os resultados não variam drasticamente daqueles referentes à análise baseada no sistema. A análise baseada no item pode também providenciar resultados preliminares que auxiliem na identificação dos lexemas que apresentaram força colostrucional relevante à descrição da construção. Deve-se salientar, entretanto, que, ao fazer comparações com os resultados disponibilizados em estudos anteriores, o linguista deve ter em mente a distância existente entre esses dois tipos de análise para que estabeleça comparações fundamentadas e coerentes.

Para ilustrar as semelhanças e diferenças entre os resultados das duas análises, seguem as dez primeiras linhas de cada análise:

Tabela 1: Análise colostrucional baseada no item feita através do pacote *collostructions*

Item	Palavra na Construção	Palavra no corpus	Freq. esperada	Associação	LogLikelihood	Significância
dizer	1233263	198203	8498,1	Atraído	959063,1	*****
saber	1138332	72433	7844	Atraído	202978,7	*****
ver	1644220	55236	11330	Atraído	91312,59	*****
fazer	3547611	70096	24445,8	Atraído	60204,51	*****
deixar	668360	20424	4605,5	Atraído	29955,26	*****
mostrar	372608	11864	2567,6	Atraído	18086,94	*****
ir	632720	15664	4359,9	Atraído	17851,04	*****
dar	1408437	23585	9705,2	Atraído	14550,55	*****
continuar	397897	10749	2741,8	Atraído	13613,8	*****
ficar	892208	16198	6148	Atraído	11548,03	*****

Tabela 2: Análise colostrucional baseada no sistema feita através do Excel com base no *corpus* Portuguese Web 2011, disponível no *Sketch Engine*.

Item	Item na construção	Palavra no corpus	Construção	Construções no corpus	Freq. esperada	Associação	LogLikelihood
dizer	198203	1233263	1602800	122600992	16123	Atraído	585.624,39
saber	72433	1138332	1602800	122600992	14882	Atraído	109.329,02
ver	55236	1644220	1602800	122600992	21495	Atraído	35.413,75
deixar	20424	668360	1602800	122600992	8738	Atraído	11.026,03
fazer	70096	3547611	1602800	122600992	46379	Atraído	9.976,74

mos- strar	11864	372608	1602800	122600992	4871	Atraído	6.977,53
ir	15664	632720	1602800	122600992	8272	Atraído	5.100,68
conti- nuar	10749	397897	1602800	122600992	5202	Atraído	4.414,10
ficar	16198	892208	1602800	122600992	11664	Atraído	1.534,87
dar	23585	1408437	1602800	122600992	18413	Atraído	1.298,17

4.2.6 QUAL É A LINHA EXECUTIVA, ENTÃO, DE PESQUISA DE COLOCAÇÕES QUE PERCORREMOS? COMO USAMOS O PACOTE COLLOSTRUCTIONS (DESENVOLVIDO POR SUSANNE FLACH)?¹⁰

O procedimento inicial diz respeito à exploração/instalação de programas computacionais: softwares Excel/Calc e R, pacotes RStudio, RCollostructions. O procedimento seguinte relaciona-se à seleção do perfil de análise colostrucional: baseado no item ou no sistema. Esta e as três próximas subseções focalizam o primeiro perfil, a partir do uso do pacote collostructions (desenvolvido por Susanne Flach). Em seguida, mostramos o perfil de análise *colostrucional* baseado no sistema.

A partir de um software de planilha, como Excel do Windows Office ou Calc do LibreOffice, podemos criar planilhas de dados¹¹ ou ler conjuntos de dados oriundos de acervos de amostras linguísticas, como possibilita o acervo anotado da plataforma *Sketch Engine*. Com tais dados em mãos, podemos categorizar cada ocorrência linguística de acordo com variáveis relevantes ao fenômeno linguístico que estamos investigando. Nos programas de planilha, podemos explorar as distribuições dos dados do *corpus* em relação às variáveis estipuladas (os grupos de fatores a examinar) e até gerar representações gráficas dessas informações levando em consideração o tipo de variável (categórica, numérica ou racional).

O R (<https://cran.r-project.org/>) constitui tanto um *software* quanto uma linguagem de programação, o que permite que ofereça uma gama ampla de recursos para a análise de dados. O R é usado, nacional e internacionalmente, como ferramenta para análise estatística, indo além da Linguística, mas, nesse caso, serve a pesquisas em Sociolinguística, Linguística Funcional-Cognitiva, Gramática de Construções, Linguística de *Corpus* e em outras perspectivas de computação de dados.

What are R and CRAN?

10 Vídeos feitos por Susanne Flach (Acesso em: 18 jan. 2022): <https://www.youtube.com/watch?v=nC-4T7-RGHgE&list=PLIZN-827NSIONkLPWpjaFr0mlKacSLTRy&rt=0s>; , <https://www.youtube.com/playlist?list=PLIZN-827NSIONkLPWpjaFr0mlKacSLTRy>.

11 Arquivos em formato *.xlsx ou *.ods.

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering etc. Please consult the [R project homepage](#) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.¹² (<https://cran.r-project.org/>)

Entretanto, para utilizar o R precisamos do suporte de um ambiente de desenvolvimento integrado (IDE). O RStudio, elaborado por Ross Ihaka e Robert Gentleman, é a IDE mais utilizada pelos pesquisadores por possuir uma interface gráfica amigável e ser de acesso e utilização gratuita. Para baixar o RStudio, acesse <https://www.rstudio.com/> e selecione o arquivo compatível com o seu sistema operacional.

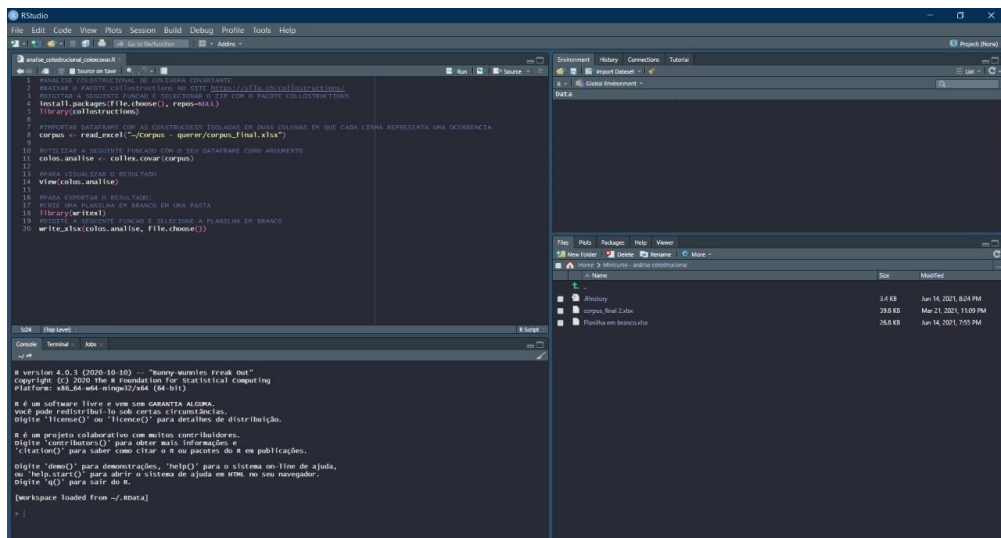
Como está dito e aparece no *print* a seguir, a interface do RStudio apresenta quatro seções de informações: *Code Editor*, para o registro dos *scripts* na linguagem R; *Environment/History*, para a listagem e armazenamento dos objetos e para o histórico de ações executadas no R; *Console*, para registrar efetivamente os comandos *inputs* que serão direcionados à execução pelo programa e para receber os resultados/*outputs* das funções; *Files/Plots/Help/Packages*, para o gerenciamento de pacotes a importar, gravação de arquivos, configuração de gráficos e ajuda.

12 O que são R e CRAN?

R é 'GNU S', uma linguagem e ambiente disponível gratuitamente para computação estatística e gráficos que fornece uma ampla variedade de técnicas estatísticas e gráficas: modelagem linear e não linear, testes estatísticos, análise de séries temporais, classificação, agrupamento etc. Por favor, consulte a página inicial do projeto R para mais informações.

CRAN é uma rede de servidores ftp e web em todo o mundo que armazenam versões idênticas e atualizadas de código e documentação para R. Por favor, use o espelho CRAN mais próximo de você para minimizar a carga da rede (<https://cran.r-project.org/>) Acesso em: 18 jan. 2022.

Imagem 2: A interface gráfica do RStudio.



O pacote relativo ao procedimento computacional de análise colostrucional via R está disponível no site (<https://sfla.ch/collostructions/>) e no canal no YouTube (<https://www.youtube.com/channel/UCKvalxeEetfYQDqgRwBYn5g/featured>) de Susanne Flach. Lá é possível localizar o pacote *collostructions* para versões do R, Mac/Linux e Windows. É importante garantir que o pacote instalado seja compatível com a versão do R instalada em seu computador. Caso o pesquisador esteja utilizando uma versão anterior do R, precisará buscar por versões antigas do pacote disponíveis no site. Para a exportação dos dados no final da análise, também será necessário o pacote 'writlex' ou outro pacote com a mesma funcionalidade.

Tendo uma amostra de dados ou mais constituída(s), procedemos à criação de uma planilha em que cada ocorrência do fenômeno linguístico estudado ocupa uma linha dessa planilha. Isso é feito se não recorrermos a uma plataforma de dados que não nos ofereça essa planilha. Feita a primeira previsão do gerenciamento de informações (categorias formais e funcionais da análise) e ordenação/distribuição delas na planilha, providenciamos que cada coluna da planilha seja voltada à análise de uma variável sob exame: uma coluna é prevista para acolher o resultado da categorização de cada dado quanto a um conjunto de variantes (variável dependente); em sequência, cada coluna subsequente acolhe a categorização de um grupo de fatores potencialmente influenciadores da variação construcional (variável independente) em estudo. Com isso, obtém-se uma planilha com todos os dados categorizados/codificados. Esse material já nos permite enxergar números de ocorrências por categoria de cada variável, percentuais. E, assim, esses fomentam questionamentos e hipóteses de observação importantes ao procedimento de configuração do objeto de análise em termos de colocação construcional ou coocorrências. Obtém-se, assim, um encaminhamento, gestado empiricamente, para a observação do fenômeno linguístico via análise colostrucional.

A análise de colocações inicia-se pela instalação de *softwares* em um computador. Instalamos o *software* R. Se já tiver o R instalado, tenha atenção à versão do R: verifique se é compatível à versão associada ao pacote *collostructions*, conforme configuração indicada no site de Susanne Flach. Uma vez instalado o software R, instalamos o pacote RStudio. Então, passamos à etapa de instalação do pacote *collostructions*. Para tanto, baixamos o pacote <https://sflla.ch/collostructions/>. Há diversos pacotes para uso no R (para manipular dados e proceder a tarefas estatísticas).¹³ Um pacote diz respeito a uma função com que queremos trabalhar no R.

Dito isso, para o trabalho com o pacote *collostructions*, basta importá-lo valendo-nos, para tanto, da função:

```
install.packages(file.choose(), repos=NULL)
```

Ao executar essa função, o R abrirá uma aba para que seja feita a seleção do pacote. Nesse caso, o arquivo importado do site de Susanne Flach poderá ser encontrado na pasta de downloads sob o nome “*collostructions*”. É importante, como ela alerta no site, não tentar descompactar o arquivo zipado que é baixado do site dela.

Instalações no R têm um tempo de duração (podem demorar). É preciso aguardar que o *software* execute a instalação a partir do comando em destaque já apresentado. Apenas execute novos comandos quando a sinalização em vermelho acima do Console tiver sumido.

Sempre que iniciamos uma sessão no RStudio, precisamos abrir os pacotes que vamos usar. Para abrir o pacote *collostructions*, usamos a função:

```
library(collostructions)
```

Outra opção de carregamento de pacotes está disponível na aba *Packages* do RStudio no quadrante inferior direito. Busque o nome do pacote e selecione o quadrado ao lado do nome. Caso busque mais informações sobre o pacote, há descrições sobre as funções do *collostructions* na aba *Help*, que também pode ser acessada clicando no nome do pacote em *Packages* ou digitando *??nomedopacote*.

Para proceder à análise dos **três tipos de análises colostrucionais** que já apresentamos brevemente aqui, vamos lançar mão dessas funções do pacote *collostructions*: função **collex**.

13 O pacote para a análise (sociolinguística) multivariada, por exemplo, é outro. Cf.: Rbrul-Daniel E. Johnson (<https://languagevariationsuite.shinyapps.io/Pages/>). Uma nova versão do Rbrul pode ser acessada a partir de <http://danielezrajohnson.shinyapps.io/Rbrul>.

Primeiramente, é necessário montar um *dataframe* (a ser lido pelo R), ou seja, configurar uma planilha de dados com informações e distribuição delas conforme a seguir: na primeira coluna da planilha, estará cada unidade construcional cujo grau de atração ou repulsa será examinado; na segunda coluna, estará a informação sobre a frequência da unidade na construção; na terceira coluna, estará a informação sobre sua frequência no *corpus*. Para ilustrar, uma planilha possível é a que expomos em seguida:

Tabela 3: Tabela ilustrativa de dados 1

ITEM	Frequência na construção	Frequência no corpus
dizer	1233263	198203
saber	1138332	72433
ver	1644220	55236

Essa planilha será informada, via console, para o R como um *dataframe* na função que R executará, conforme sequência de ações indicada a seguir.

4.3 ANÁLISE COLOSTRUCIONAL DE COLEXEMAS SIMPLES

Em primeiro lugar, é preciso iniciar a seção, abrir o pacote *collostructions*, lançando mão da função `library(collostructions)`, como já apresentado. Lembremos de que, toda vez que vamos começar uma sessão, abrimos os pacotes via console.

Temos de denominar nossos objetos de pesquisa¹⁴ (que armazenam dados, segundo uma estruturação que o pesquisador preferiu) de uma forma que nos permita reconhecê-los ao longo do processo. Por exemplo, se vamos examinar construções de futuridade formadas por predicadores complexos com verbos volitivos (verbo *querer*, *esperar*, por exemplo), nomes descritivos são indicados, como “volitivos” ou “verbos_de_desejo”.

Após abrir o pacote *collostructions*, vamos importar o *dataframe* com a frequência de itens na construção e no *corpus* para dentro de objeto de análise. Para tanto, configuramos um nome para o objeto, por meio da função:

```
data <- read_excel("~/Corpus - querer/corpus_final.xlsx")
```

14 Para saber um pouco mais a respeito, sugerimos outro vídeo (também do Festival do Conhecimento da UFRJ 2020): https://www.youtube.com/watch?v=yD8TVpQ_FxU&t=0s (Minicurso Linguística de Corpus: Introdução ao R para coleta de dados no Twitter) LIMA, COSTA, POPPOLINO, MACHADO VIEIRA, 2020.

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo;

Muda segundo as informações do pesquisador.

Vamos atribuir o elemento para esse objeto, por meio de *menor que* seguido do traço e da função `read_excel`; e, dentro dos parênteses/da função, é preciso informar onde está o *dataframe* no computador (precisando o caminho para esse acesso à planilha). É possível importar manualmente, via RStudio, o arquivo através da opção *Import Dataset* disponível na aba *Environment*. Ao clicar nessa opção, você deverá selecionar *From Excel* e buscar o arquivo em *Browser*, concluindo a ação pelo botão *Import*. Essa forma de importação é possibilitada pela IDE do RStudio que irá gerar e enviar o código automaticamente para o *Console*, por isso é de grande importância instalar a IDE.

Com o *dataframe* importado no R, vamos criar outro objeto específico para a análise colostrucional (afinal, precisamos gravá-la como um objeto). Novamente, vamos escolher um nome descritivo de nossa preferência que nos faça reconhecer esse arquivo como análise colostrucional. No exemplo aqui explorado, denominamos o objeto de *colos.analise*. Então, vamos utilizar a função com o *dataframe* já como argumento:

```
colos.analise <- collex(data, tokens totais do seu corpus)
```

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo;

Muda segundo as informações do pesquisador.

Vamos atribuir (<-) a função *collex* (de análise de colexemas simples) à análise, considerando dois argumentos (dentro será posto o *dataframe* chamado de *corpus*,¹⁵ seguido dos *tokens* totais de seu *corpus* – que pode ser o número de palavras). Essa é a única função que exigirá a informação sobre o número total de *tokens*.

A partir da execução dessa função *collex*, já é possível visualizar resultados de análise colostrucional, bastando, para tanto, recorrer a esta função *View* (obrigatoriamente com letra V maiúscula) seguida do nome do objeto a que foi atribuída a análise colostrucional/*colos.analise*:

¹⁵ A informação dentro dos parênteses tem de ser exatamente igual ao nome com que o objeto-*dataframe* foi designado.

```
View(colos.analise)
```

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo.

A informação dentro dos parênteses tem de ser exatamente igual ao nome com que o objeto-análise colostrucional/tipo 1 foi designado.

Ao visualizar, verificamos se o resultado se alinha ao que era esperado como retorno. Em caso afirmativo, exportamos o resultado. Como? Criamos uma planilha em branco em uma pasta (preferencialmente, no diretório de trabalho do pesquisador). Instalamos o pacote para exportação para planilhas *writexl*:

```
install.packages("writexl")
```

As aspas são necessárias na hora da instalação. Esse pacote para exportação para planilhas pode ser instalado diretamente da listagem de pacotes do R, não sendo necessário importá-lo do site de Susanne Flach.

Uma vez instalado, basta abrir o pacote (neste momento, as aspas já não são necessárias, conforme indicado a seguir):

```
library(writexl)
```

Depois, basta digitar a função abaixo e selecionar a planilha em branco que foi criada, conforme indicado a seguir:

```
write_xlsx(colos.analise, file.choose())
```

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo;

Muda segundo as informações do pesquisador.

Após executar esse comando, o R vai abrir uma pasta (o diretório de trabalho do pesquisador), então escolhemos o arquivo planilha em branco criado.

4.4 ANÁLISE COLOSTRUCIONAL DE COLEXEMAS DISTINTIVOS

Em primeiro lugar, é preciso iniciar a seção, abrir, via console, o pacote *collostructions*, lançando mão da função *library (collostructions)*, como já foi apresentado.

Temos de construir previamente um *dataframe* com outro nome diferente do explorado na análise anteriormente explicitada, com informações relativas aos dados da pesquisa, como: informamos, no caso de estudo de predicador complexo via dois padrões construcionais com verbos de futuridade, na primeira coluna os verbos predicadores que se combinam aos dois padrões com os verbos (semi-)auxiliares que supomos como variantes e, na segunda e terceira colunas, os padrões que se organizam especificamente com verbos (semi-)auxiliares/variantes que terão sua força de atração ou repulsa ao lugar com perfil de auxiliaridade mensurado.

Uma informação que esse procedimento analítico oferece é a verificação a respeito da potencialidade dos itens/lexemas que atuam na qualidade de verbo predicador ocorrerem ligados aos dois padrões sob análise: o com o verbo semiauxiliar *querer* e o com o verbo auxiliar *ir*. E também informa qual dos dois é o preferencial com que predicador/natureza de predicador. A figura a seguir ilustra o que estamos descrevendo aqui:

Tabela 4: Tabela ilustrativa de dados 2

Verbo	querer	ir
dizer	81	8
saber	44	6
ter que	0	24

Após abrir o pacote *collostructions*, vamos importar o *dataframe* com as informações sistematizadas da maneira já indicada para dentro de objeto de análise. Para tanto, configuramos um nome para o objeto, por meio da função:¹⁶

¹⁶ É fundamental dar nomes diferentes aos objetos – *dataframe* e análise colostrucional, ou ela gravará por cima dos arquivos criados na análise colostrucional de colexemas distintivos.

```
datadist <- read_excel("~/Corpus - querer/corpus_final.xlsx")
```

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo;

Muda segundo as informações do pesquisador.

Vamos atribuir o elemento para esse objeto, por meio de *menor que* seguido do traço e seguido da função *read_excel*; e, dentro dos parênteses, é preciso informar onde está o *dataframe* no computador (precisando o caminho para esse acesso à planilha). É possível importar manualmente, via RStudio, o arquivo/*dataframe*.

Com o *dataframe* importado no R (já como argumento), vamos criar outro objeto específico para a análise colostrucional (afinal, precisamos gravá-la como um arquivo). Novamente, vamos escolher um nome descritivo de nossa preferência que nos faça reconhecer esse arquivo como análise colostrucional. No exemplo aqui explorado, denominamos o objeto de *colos.analise*. Então, vamos utilizar a função com o *dataframe* já como argumento:

```
colos.analise.dist <- collex.dist(datadist)
```

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo.

Vamos atribuir (<-) a função *collex.dist* (de análise de colexemas distintivos) à análise, considerando um argumento (dentro será posto o *dataframe* chamado de *corpus-dist*.)¹⁷

A partir da execução dessa função *collex.dist*, já é possível visualizar resultados de análise colostrucional, bastando, para tanto, recorrer a esta função *View* (obrigatoriamente com letra V maiúscula) seguida do nome do objeto a que foi atribuída a análise colostrucional/*colos.analise*:

17 A informação dentro dos parênteses tem de ser exatamente igual ao nome com que o objeto-*dataframe* foi designado.

```
View(colos.analise.dist)
```

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo.

A informação dentro dos parênteses tem de ser exatamente igual ao nome com que o objeto-análise colostrucional/tipo 2 foi designado.

Ao visualizar, verificamos se o resultado se alinha ao que era esperado como retorno. Em caso afirmativo, exportamos o resultado. Como? Criamos uma planilha em branco em uma pasta (preferencialmente, no diretório de trabalho do pesquisador). Instalamos o pacote para exportação para planilhas *writexl*:

```
install.packages("writexl")
```

As aspas são necessárias na hora da instalação. Esse pacote para exportação para planilhas pode ser instalado diretamente da listagem de pacotes do R, não sendo necessário importá-lo do site de Susanne Flach.

Uma vez instalado, basta abrir o pacote (neste momento, as aspas já não são necessárias, conforme indicado a seguir):

```
library(writexl)
```

Depois, basta digitar a função a seguir e selecionar a planilha em branco que foi criada, conforme indicado:

```
write_xlsx(colos.analise.dist, file.choose())
```

Após executar esse comando, o R vai abrir uma pasta (o diretório de trabalho do pesquisador), então escolhemos o arquivo planilha em branco criado.

4.5 ANÁLISE COLOSTRUCIONAL DE COLEXEMAS COVARIANTES

Para essa análise, é preciso preparar um *dataframe* por ocorrências. Isso significa que cada linha da coluna é lida como ocorrência. Cada coluna é um *slot*. E cada *slot* é preenchido por um colexema. Só registramos, nessa planilha, ocorrências: uma a uma, na mesma coluna.

Em primeiro lugar, é preciso iniciar a seção, abrir, via console, o pacote *collostructions*, lançando mão da função **library (collostructions)**, como já foi apresentado.

Após abrir o pacote *collostructions*, vamos importar o *dataframe* com cada ocorrência da amostra para dentro de objeto de análise. Lembre-se de nomear com outro título, diferente dos anteriores. Para tanto, configuramos um nome para o objeto, por meio da função:¹⁸

```
datavar <- read_excel("~/Corpus - querer/corpus_final.xlsx")
```

Vamos atribuir o elemento para esse objeto, por meio de *menor que* seguido do *traço* e seguido da função *read_excel*; e, dentro dos parênteses, é preciso informar onde está o *dataframe* no computador (precisando o caminho para esse acesso à planilha). É possível importar manualmente, via RStudio, o arquivo/*dataframe*.

Com o *dataframe* importado no R (já como argumento), vamos criar outro objeto específico para a análise colostrucional (afinal, precisamos gravá-la como um arquivo). Novamente, vamos escolher um nome descritivo de nossa preferência que nos faça reconhecer esse arquivo como análise colostrucional. No exemplo aqui explorado, denominamos o objeto de *colos.analise*. Então, vamos utilizar a função com o *dataframe* já como argumento:

```
colos.analise.covar <- collex.covar(datavar)
```

18 É fundamental dar nomes diferentes aos objetos – *dataframe* e análise colostrucional, ou ela gravará por cima dos arquivos criados na análise colostrucional de colexemas distintos.

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo.

Vamos atribuir (<-) a função *collex.covar* (de análise de colexemas distintivos) à análise, considerando um argumento (dentro será posto o *dataframe* chamado de *corpusvar*.)¹⁹

A partir da execução dessa função *collex.var*, já é possível visualizar resultados de análise colostrucional, bastando, para tanto, recorrer a esta função *View* (obrigatoriamente com letra V maiúscula) seguida do nome do objeto a que foi atribuída a análise colostrucional/colos.analise:

```
View(colos.analise.covar)
```

Legenda:

Muda segundo o nome que o pesquisador escolher;

Não muda na sintaxe/comando fixo.

A informação dentro dos parênteses tem de ser exatamente igual ao nome com que o objeto-análise colostrucional/tipo 2 foi designado.

Ao visualizar, verificamos se o resultado se alinha ao que era esperado como retorno. Em caso afirmativo, exportamos o resultado. Como? Criamos uma planilha em branco em uma pasta (preferencialmente, no diretório de trabalho do pesquisador). Instalamos o pacote para exportação para planilhas *writexl*:

```
install.packages("writexl")
```

As aspas são necessárias na hora da instalação. Esse pacote para exportação para planilhas pode ser instalado diretamente da listagem de pacotes do R, não sendo necessário importá-lo do site de Susanne Flach.

¹⁹ A informação dentro dos parênteses tem de ser exatamente igual ao nome com que o objeto-dataframe foi designado.

Uma vez instalado, basta abrir o pacote (neste momento, as aspas já não são necessárias, conforme indicado a seguir):

```
library(writexl)
```

Depois, basta digitar a função a seguir e selecionar a planilha em branco que foi criada, conforme indicado:

```
write_xlsx(colos.analise.covar, file.choose())
```

Após executar esse comando, o R vai abrir uma pasta (o diretório de trabalho do pesquisador), então escolhemos o arquivo planilha em branco criado.

Lembramos que a visualização desse passo a passo, bem como dos *scripts* escritos por Susanne Flach, no RStudio, está nos últimos onze minutos do vídeo *Análise Colostrucional por meio do R em pesquisas linguísticas* (COSTA; SOUZA, 2021).²⁰

4.6 TRABALHO COM SOFTWARES DE FUNÇÕES DE PLANILHA PARA UMA ANÁLISE BASEADA NO SISTEMA

Como exposto anteriormente, *softwares* de desenvolvimento/edição de planilhas, como o Microsoft Excel e Google Planilhas, podem ser empregados para a observar a distribuição (qualitativa e/ou quantitativa) de dados no *corpus*, assim como produzir e editar tabelas e gráficos de modo dinâmico. Entretanto, sua funcionalidade, no processo de análise, não se restringe a tais procedimentos. Ao tratarmos do exame quantitativo relativo à análise colostrucional, é importante ressaltar que, além dos recursos oferecidos pelo pacote “*collostructions*”, empregado por meio do R, também contamos com ferramentas disponibilizadas por programas de edição de planilhas contábeis. Através de tais programas, podemos implementar a análise baseada no sistema, citada anteriormente, a partir de medidas de associação como o *LogLikelihood*.

Nesta seção, exploraremos, logo, como empregar os métodos quantitativos para avaliar o grau de atração entre itens construcionais por meio de tipo de *softwares* segundo as diretrizes metodológicas apresentadas por Martin Hilpert²¹.

20 Disponível em: <https://youtu.be/-lh215zX36U>. Acesso em: 01 mar. 2022.

21 As diretrizes do Prof. Hilpert encontram-se disponíveis em seu canal no YouTube. Em especial, nos vídeos intitulados “Análise colostrucional”, “Análise de colexemas distintivos” e “Medidas colocacionais”, disponíveis, respectivamente, em: https://youtu.be/5Mfv_6kzNXo, <https://youtu.be/AdBvI4oR-2Pw> e <https://youtu.be/SZ2RtyKzU6o>.

Diferentemente do observado no tratamento quantitativo delimitado pelo pacote “*collostructions*”, para o desenvolvimento desse perfil de análise, ao considerarmos o grau de associação entre uma construção mais complexa e um item que é acionado a fim de preencher um *slot* em sua configuração, é necessário levantarmos os seguintes questionamentos: qual a frequência com a qual esse item aparece na construção? Quantas vezes esse item aparece na amostra textual como um todo? Quantas vezes essa construção é empregada no *corpus*? Quantos itens construcionais de mesmo perfil do item sob observação aparecem no *corpus*?

É necessário, assim, a elaboração de uma matriz que abarque as seguintes informações numéricas:

- (a) Número de ocorrências do item em análise na construção;
- (b) Número de ocorrências do item no *corpus*/amostra como um todo;
- (c) Número de ocorrências da construção no *corpus*;
- (d) Número de construções no *corpus*.

Uma vez que tenhamos essas informações, basta organizá-las em uma planilha, como observado a seguir:

Tabela 5: Frequência de itens verbais no *slot* de verbo no infinitivo de construção de futuramente com *querer*

	A	B	C	D	E
1	Item em análise	Frequência do item na construção	Frequência do item no corpus	Frequência da construção no corpus	Número de construções no corpus
2	DIZER	237082	1235268	1602800	122600992
3	TER	251078	4041686	1602800	122600992
4	VER	158515	1644220	1602800	122600992
5	SABER	80796	1161188	1602800	122600992
6	SER	83704	10654210	1602800	122600992

Considerando o primeiro exemplo, o pesquisador precisará contabilizar quantas ocorrências há no *corpus* do verbo “dizer” ocorrendo na construção com “querer”, quantas ocorrências há do verbo “dizer” em todo o *corpus*, quantas ocorrências há da construção [*querer* + verbo no infinitivo] e quantas construções há no *corpus* como um todo. A última informação, nesse caso, foi contabilizada em relação às construções envolvendo verbos no infinitivo. Por fim, para calcular a frequência esperada, deve-se seguir esta equação:

$$\text{Frequência esperada} = \frac{\text{Frequência do item no corpus} \times \text{Frequência da construção no corpus}}{\text{Número de construções no corpus}}$$

A maioria dos editores de planilha, como o *GoogleSheets* ou o *Excel*, permitem que essa equação seja executada de maneira automática a partir de funções. A formatação

da função vai variar de acordo com o *software* escolhido e com a localização dos dados nas células da planilha, mas, de maneira geral, deverá ser semelhante a esta configuração: $= (C2 * D2) / E2$.

Em relação ao exemplo com “dizer”, destacamos que a frequência esperada seria de 16149,03 ocorrências enquanto a frequência observada foi de 237082 ocorrências. Isso sinaliza uma atração entre o lexema e a construção em questão. A fórmula é inserida na segunda linha e, para ser aplicada aos demais itens/linhas, deve ser arrastada para baixo ou para cima. Como pode ser observado, cada combinação [letra + número] presente na fórmula corresponde a uma célula da planilha. Em “C2”, letra “C” refere-se a “coluna C” (terceira coluna), enquanto o número “2” indica a linha a ser considerada (a segunda linha). Logo, para nos referirmos ao item na terceira linha, nossa fórmula necessita ser reconsiderada:

$$= (C3 * D3) / E3$$

Ao “arrastarmos” a fórmula para as demais linhas, essa reestruturação é realizada de forma automática. Por conseguinte, é essencial frisar que qualquer reorganização da distribuição das colunas na planilha implica a necessidade de uma reestruturação das fórmulas.

No que concerne à coluna referente ao status do item em relação à construção, podemos automaticamente visualizar a categorização do elemento segundo a fórmula:

➔ $= SE(B2 > F2; "atraído"; "repelido")$

Item em análise	Frequência do item na construção	Frequência do item no corpus	Frequência da construção no corpus	Número de construções no corpus	Frequência esperada	Status do item em relação à construção
DIZER	237082	1235268	1602800	122600992	16149,03369	atraído
TER	251078	4041686	1602800	122600992	52838,18846	atraído
VER	158515	1644220	1602800	122600992	21495,38738	atraído
SABER	80796	1161188	1602800	122600992	15180,56335	atraído
SER	83704	10654210	1602800	122600992	139285,7228	repelido

Assim, se a frequência do item na construção for maior que a frequência esperada, consideraremos que o item é “atraído” para o *slot* construcional. Caso seja menor que o esperado, será categorizado como “repelido”.

Para medir a força de atração, utilizaremos a medida de relevância estatística *LogLikelihood*. Basta criar uma nova coluna e inserir a função:

$$= 2 * (B2 * LOG(B2; EXP(1))) + D2 * LOG(D2; EXP(1)) + C2 * LOG(C2; EXP(1)) + E2 * LOG(E2; EXP(1)) - (B2 + D2) * LOG(B2 + D2; EXP(1)) - (B2 + C2) * LOG(B2 + C2; EXP(1)) - (D2 + E2) * LOG(D2 + E2; EXP(1)) - (C2 + E2) * LOG(C2 + E2; EXP(1)) + (B2 + D2 + C2 + E2) * LOG(B2 + D2 + C2 + E2; EXP(1))$$

Item em análise	Frequência do item na construção	Frequência do item no corpus	Frequência da construção no corpus	Número de construções no corpus	Frequência esperada	Status do item em relação à construção	Grau de atração
DIZER	237082	1235268	1602800	122600992	16149,03369	atraído	766628,247
TER	251078	4041686	1602800	122600992	52838,18846	atraído	354162,7012
VER	158515	1644220	1602800	122600992	21495,38738	atraído	337318,619
SABER	80796	1161188	1602800	122600992	15180,56335	atraído	132751,6066
SER	83704	10654210	1602800	122600992	139285,7228	repeliado	23856,14948

Há possibilidade de alterações precisarem ser feitas para que o *software* reconheça a função. Tais alterações dependerão do *software* utilizado, do idioma selecionado, entre outros fatores.

Dessa forma, por meio de uma planilha do Excel, conseguimos dispor as informações relativas ao perfil dos itens e da construção em correlação ao *corpus* e executar os passos associados à análise quantitativa de forma automática, simplificando e otimizando o trabalho do pesquisador. Assim, o uso de planilhas configura-se como uma alternativa válida e altamente acessível para o processo de análise colostrucional.

4.7 NOTAS SOBRE INTERPRETAÇÃO E DESCRIÇÃO DE RESULTADOS

Retornando aos resultados da análise colostrucional de colexemas covariantes apresentada anteriormente, é possível identificar alguns padrões já previstos nas hipóteses da pesquisa em questão (COSTA, 2022). Uma dessas hipóteses seria a de que as construções “querer saber” e “querer dizer” estariam passando por um processo de construcionalização lexical atuando como marcadores discursivos em contextos como: “Quer saber, esquece” ou “Tu já é assim, quer dizer, em partes”. Os resultados da análise apresentam indicativos favoráveis a essa hipótese já que a força de associação entre os lexemas “querer” e os lexemas “saber” e “dizer” é alta, eles se atraíram e, assim, apresentaram níveis de significância altos. Há também a hipótese de que haveria atração entre os itens “querer” e “ver” devido à construção lexical presente em “Quero ser se volto cá regularmente, embora o tempo não abunde!”, o que foi confirmado pelos resultados da análise colostrucional, porém com nível de significância menor que as demais construções lexicais.

Tabela 6: Análise colostrucional de colexemas covariantes baseada no item feito através do pacote *collostructions*

slot 1	slot 2	freq s1	freq s2	obs	freq esperada	assoc	LogLikelihood	signif
querer	saber	140	8	8	2,2	attr	20,70487	*****
querer	dizer	140	25	17	7	attr	18,25448	****
ir	ter	360	21	20	15,1	attr	7,89027	**
querer	ler	140	3	3	0,8	attr	7,68451	**

querer	ver	140	19	10	5,3	attr	5,31295	*
querer	aprender	140	2	2	0,6	attr	5,11256	*
querer	assistir	140	2	2	0,6	attr	5,11256	*
querer	ir	140	2	2	0,6	attr	5,11256	*
querer	tornar	140	2	2	0,6	attr	5,11256	*
ir	ser	360	31	27	22,3	attr	4,31129	*

Nesse pacote, o nível de significância é medido de acordo com o valor-p seguindo este referencial (também presente na descrição do pacote):

Quadro 1: Tabela de nível de significância.

<p>5 asteriscos = significativo em $p < .00001$, 4 asteriscos = significativo em $p < .0001$, 3 asteriscos em $p < .001$, 2 asteriscos em $p < .01$, 1 asterisco em $p < .05$, ns Não significativo</p>
--

Com o valor-p menor que .05, pode-se rejeitar a hipótese nula, indicando, assim, uma relação entre os itens. Porém, vale ressaltar que, como apontado por Gilquin (2015), a análise colostrucional está baseada em formas e não em sentidos, o que pode gerar ocorrências em que a mesma forma corresponda a diferentes sentidos devido à simplificação inerente a esse tipo de análise. Gilquin propõe uma abordagem mais qualitativa dos dados que envolva a investigação das ocorrências dos itens em contexto e a classificação desses itens em, por exemplo, classes semânticas. Propomos, nesse caso, a classificação baseada na agentividade do verbo. Nos resultados, notamos que há uma força de atração entre os lexemas “ler”, “aprender”, “assistir”, “ir” e “tornar” e o lexema “querer” e entre “ter” e “ser” e o lexema “ir”. Tal distribuição está alinhada à hipótese da pesquisadora de que haveria uma diferença distributiva entre as construções de acordo com a agentividade licenciada pelo segundo verbo.

4.8 ANÁLISE COMPARATIVA POR SUBAMOSTRAS ORGANIZADAS TEXTUAL-DISCURSIVAMENTE

4.8.1 COMO RECONFIGURAMOS A LINHA EXECUTIVA DOS TIPOS DE ANÁLISE COLOSTRUCIONAL ANTERIORMENTE VISTOS PARA A PESQUISA DA COMPATIBILIZAÇÃO ENTRE CONSTRUÇÕES DE PRE-DICAÇÃO E CONSTRUÇÃO TEXTUAL-DISCURSIVA?

Para o desenho de um processo de pesquisa de variação em que perspectivamos a relação de atração ou repulsa de certas unidades construcionais a contextos mais amplos, consideramos a possibilidade de haver regiões funcionais específicas mais claramente atreladas a elas, que podem guiar o uso linguístico.

Determinadas subamostras de textos – seja de domínios discursivos específicos, como o científico, o político, o futebolístico, o jurídico, seja de certos gêneros textuais, como artigos, teses, discursos públicos, crônicas, petições ou, ainda, de contextos, como os de monitoramento de norma culta padrão, registro (in)formal, modalidade expressiva falada/espontânea ou escrita/planejada, faixa geracional/etária, gênero – são exemplos de fatores que podem influenciar na “escolha” preferencial por uma determinada construção, em detrimento de outra.

Projetamos um *design* configuracional com essas fases, com ações e materiais estruturados segundo certos parâmetros de análise:

1) Recurso a amostras de textos já constituídas: *Corpus do Português*, *Sketch Engine*, acervos brasileiros (cf. MACHADO VIEIRA; BARBOSA, 2022)

ou

Constituição de amostras de textos, em linhas gerais via:

Anotação de cada texto segundo parametrização previamente definida;

Reunião de textos em acervo(s);

Quantificação do número/total de textos.

2) Correlações: pontos de partida para a investigação; expectativas do pesquisador (a partir de leituras e experiências prévias) e o que é efetivamente observado em amostra(s) de dados.

Projetamos a potencialidade de haver alguma relação estatística significativa entre certas variáveis. Procuramos mensurar a relação entre o que é esperado e o que é observado em termos da correlação de variáveis (categóricas ou contínuas): qual é a diferença ou a similaridade, a associação/dissociação quanto a certas condições de análise? Por exemplo, a força do acionamento de uma ou outra construção de futuridade está atrelada a algum parâmetro/atributo relativo à organização de construção textual-discursiva que licencia um objeto linguístico-textual quantificável?

3) Da amostra total, compomos diferentes subamostras, com o intuito de averiguar o que é tipicamente associado a um subconjunto de manifestações de construções textuais-discursivas em comparação com outro subconjunto de manifestações de construções textuais-discursivas. É possível nova quantificação guiada por parâmetro(s) que interessem à estruturação de acervos de textos, como modalidade expressiva, gênero textual, domínio sociodiscursivo, entre outros.

E definimos subamostras em função dos parâmetros de análise textual que quisermos averiguar, de correlações que esperamos verificar na observação empírica, de categorizações que provêm disso. Por exemplo, podemos comparar as forças colostru-

cionais de acionamento de uma ou outra construção de predicador de futuridade com base em subamostras de textos diferentes: em subamostra de textos do domínio acadêmico-científico e em subamostra de textos do domínio jornalístico. Essas, por sua vez, podem reunir subamostras de gêneros textuais diferentes (no domínio acadêmico-científico, teses, artigos; no jornalístico, editoriais, notícias, crônicas).

Então, procedemos à análise colostrucional, conforme passo a passo exposto em seção anterior a esta, considerando a comparação de resultados pelas subamostras de dados definidas para análise.

4.8.2 COM BASE EM QUE PARÂMETROS PODEMOS ORGANIZAR SUBAMOSTRAS DE DADOS PARAMETRIZADAS POR REFERENCIAL DE OBJETO TEXTUAL-DISCURSIVO?

É possível considerar diferentes condições de contextualidade nessa parametrização de subamostras: subconjuntos de manifestações concretas de construções textuais-discursivas segundo temática (profissional ou não, familiar ou não), estrutura informacional, domínio discursivo (acadêmico, político, turístico; público ou privado/familiar), gêneros textuais (relativamente estáveis ou híbridos/com contornos difusos), suporte de veiculação (jornal, periódico, programa digital ou não), grau de monitoramento linguístico-estilístico, grau de autoria, natureza do ato de fala mais proeminente, movimento retórico, registro formal-informal, letramento (digital, multimodal/multissemiótico, retórico), distribuição etária ou temporal, distribuição geográfica/comunidade (brasileira, portuguesa, moçambicana, angolana, outra).

Figura 5: Possibilidades de caracterização e organização de subamostras de dados para análises colostrucionais perspectivadas em comparação segundo o parâmetro de contextualidade: relação entre frequências/forças colostrucionais e contingências.



Travassos (2021), por exemplo, chega a resultados bastante interessantes ao tratar da variação de certas construções com verbo suporte mais ou menos associadas à ideia de início por padrão/paradigma construcional discursivo. Com relação à temática/domínio discursivo em que cada padrão construcional aparecia, por exemplo, a pesquisa revelou que “‘dar (a) partida’ está mais fortemente associada ao contexto de Automóvel (orientação, direção, mecânica); ‘dar (a/uma) saída’, ao de Tecnologia/Informática; ‘dar (a) largada’ e ‘dar (a/uma/sua) arrancada’, ao de Esporte; e ‘dar (a/uma) entrada’, ao de Sociedade/Cotidiano/Cidade” (TRAVASSOS, 2021, p. 48).

Já com relação ao valor semântico, ao gênero textual e ato de fala, o estudo mostrou que:

A construção ‘dar (a) partida’ é mais atraída para um contexto em que há uma intenção de orientar/instruir a como ligar o carro, mais presente em

blogs. A construção ‘dar (a/uma) saída’ está mais associada a uma cena de ensino de algo relacionado à Informática no gênero ‘manual de instrução’, com o sentido de ‘extrair dados’. O padrão construcional ‘dar (a) largada’ tem como principal contexto o de um título (ato de fala ‘apresentar/introduzir um tema’) de uma notícia, primordialmente, ao enfatizar a cena de ‘inauguração’ de um evento social. A perífrase verbo-nominal ‘dar (a/uma/sua) arrancada’ apareceu, predominantemente, em títulos (ato de fala ‘apresentar/introduzir tema’) de notícias, com o sentido de ‘melhorar desempenho’. O predicador complexo ‘dar (uma) começada’ ocorreu mais em postagens, com os sentidos de ‘conhecer’ e de ‘iniciar’, tendo, portanto, os atos de fala de ‘convidar’ e de ‘apresentar/introduzir tema’. Já a construção ‘dar uma iniciada’ é mais atrai-da para o gênero debate, com o ato de fala de ‘pedir/solicitar ajuda’, com o sentido de ‘começar’ (TRAVASSOS, 2021, p. 55, adaptado).

Como vemos, um caminho de investigação construcionista que amplia a potencialidade do estudo de coocorrências diz respeito à interrelação entre unidades construcionais cada vez mais complexas. Também Wiedemer e Machado Vieira (2022), valendo-se do conceito de paradigma discursivo (como um rol de construções textuais convencionalizadas), exploram o potencial da relação entre certos lexemas e padrões construcionais com construções textuais-dicursivas dos domínios discursivos da culinária (receitas) e da ciência (resumos acadêmico-científicos):

Em uma consulta rápida quanto às possíveis definições para o lexema “mas-sa”, encontramos pelo menos mais de dez significados (sentidos oriundos do campo da química, gíria, física, uso figurado etc.). Se compararmos a utilização desse lexema num gênero fórmula química e em gênero receita, vemos a atualização do significado a depender do contexto do gênero (WIEDEMER; MACHADO VIEIRA, 2022, p. 252).

Em resumos científicos, tendemos a ver materializado o acionamento de certos lexemas/construções lexicais:

verbais – observar, descrever, explicar, objetivar, verificar, investigar, pesquisar, estudar;

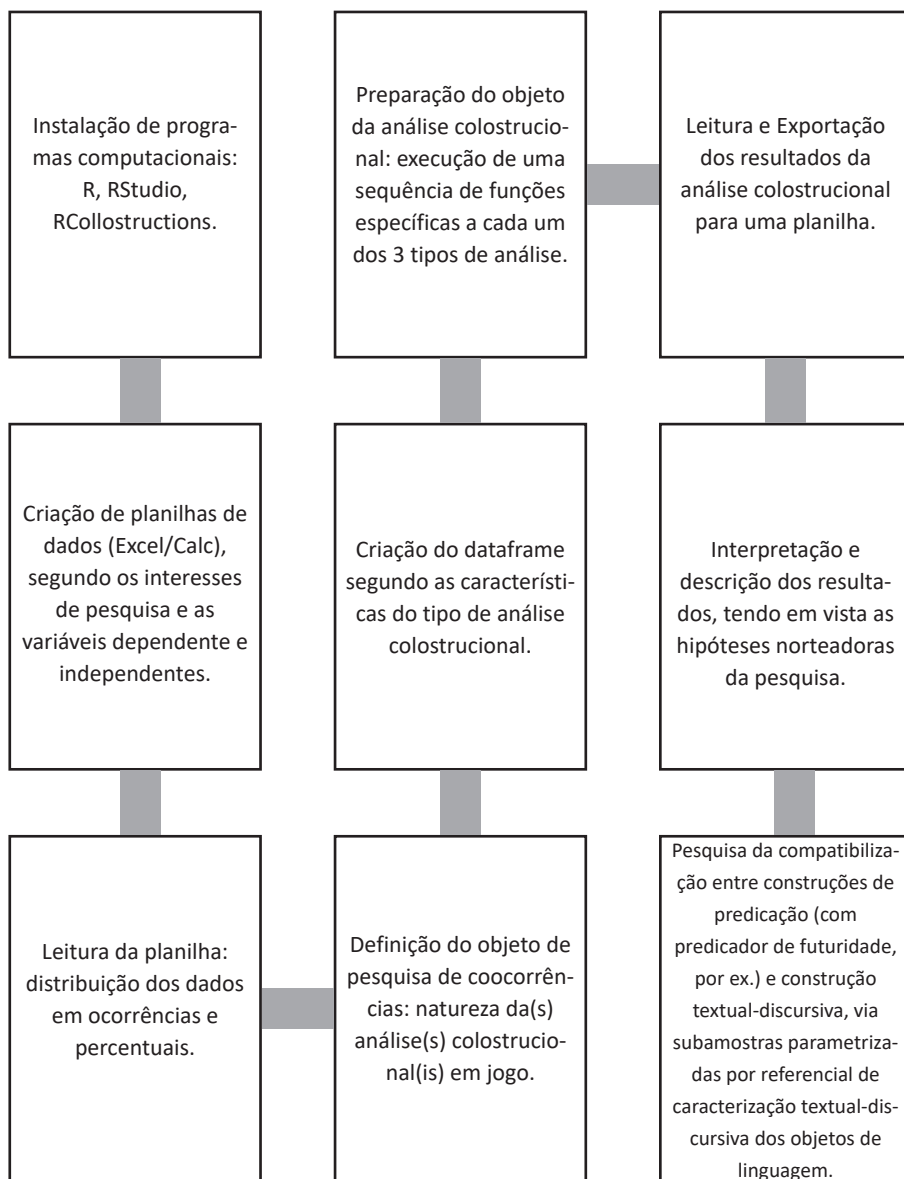
nominais – dados, hipótese, teoria, pressuposto, método, tratamento, quantitativo, qualitativo (WIEDEMER; MACHADO VIEIRA, 2022, p. 258).

O perfil de análise de colocações baseada no sistema, ou melhor, em subsistemas/subamostras dele parametrizado(a)s por condições textuais-discursivas e sociocomunicativas da produção dos dados do uso, enriquece o mapeamento dos efeitos de frequência associados ao acionamento de unidades construcionais e à interrelação delas.

4.9 LINHA EXECUTIVA DE INVESTIGAÇÃO DE COLOCAÇÕES

Uma vez expostos os potenciais perfis de análise de colocações, a seguir expomos, esquematicamente, os principais procedimentos relativos à análise colostrucional.

Figura 6: Design de ações relativas ao método de análise colostrucional.



4.10 CONSIDERAÇÕES FINAIS

Entre fazeres e saberes, há o sabor do processo de aprendizagem em equipe.

Um fazer que desafia, sem dúvida, a análise linguística, relacionado ao fazer socio-linguístico ou socioconstrucionista, diz respeito à reconfiguração de hipóteses de associações entre itens/unidades construcionais/lexemas e destas com construções via perspectivação de uma parametrização de usos linguísticos que dê conta de os observar, quer em termos de conhecimento linguístico centrado na experiência e na prática comunicativa de indivíduos, quer em termos de conhecimento linguístico centrado na experiência e nas práticas sociocomunicativas relativamente estáveis de grupos de indivíduos, comunidades.

Ao observarmos o uso linguístico em contexto, percebemos que determinados fatores podem influenciar decisivamente no acionamento de uma determinada construção, haja vista o exemplo dado de estudos relativos à alternância de construções em função de domínios de aplicação semântico-pragmático-discursiva específicos.

A observação do indivíduo ultimamente passa a mobilizar pesquisadores na área de Gramática de Construções, em razão de críticas ao modelo até então praticado, que normalmente tendeu a perspectivar a língua/gramática como relativamente homogênea, embora via descrições centradas no uso. Na área de estudos funcionais-construcionistas, já há quem defenda esse caminho em que a heterogeneidade ganha atenção. A conferência “Collostructions, contexts, communities” de Susanne Flach baseia-se em pesquisa com esse perfil. Um segundo desafio é o de lidar e mensurar a tipicidade de certas manifestações de unidades construcionais atrelada a certas experiências individuais, certas *personas* que construímos em nossas práticas de linguagem e comunicação (afinal, congregamos uma série de papéis – pesquisadora, docente, gestora, vizinha, parente, amiga, e por aí vai).

Um terceiro desafio para os estudos socioconstrucionistas e/ou funcionalistas-construcionistas é o de lidar com correlações de unidades construcionais a mais de dois *slots* construcionais ou a mais de dois padrões construcionais. Proisl (2019), além de cogitar de análises colostrucionais além do nível léxico-gramatical de sentenças (períodos compostos), expõe, no Capítulo 6 dessa obra, observações e encaminhamento no sentido de se observarem correlações mais complexas (a mais *slots* colostrucionais, a mais construções).

Um quarto desafio tem a ver com a observação centrada em item e não em sistema. Como mencionado anteriormente, os resultados da análise baseada no sistema costumam ser associados a um grau de confiabilidade maior, entretanto, pela falta de recursos de automatização da análise, esse perfil de análise demanda muito trabalho manual do pesquisador e, então, está praticamente limitado a situações em que é possível contar com *corpora* anotados. A análise baseada no item, por outro lado, pode apresentar resultados preliminares que direcionem a análise baseada no sistema. Ambas as análises são recursos importantes para o pesquisador e, tendo em mente as questões relacionadas a cada uma delas, contribuem para a descrição das construções através de uma perspectiva que olha diretamente para o nível construcional.

Sem dúvida, o desafio de lidar com esse método de pesquisa que nos moveu além do que esperávamos inicialmente foi também o que promoveu conquistas significativas a toda uma equipe. Em certa medida, os encaminhamentos basilares de ciência aberta têm sido experimentados nesse processo de lidar com a investigação de associações que têm repercussão nos usos linguísticos que fazemos:

- a) exercício de práticas de pesquisa em (inter)ação (Parceria);
- b) interação de agentes múltiplos, negociação de rumos e proposição de soluções em relação a possíveis conflitos (Paz);
- c) documentação empírica de associações entre unidades linguísticas em línguas ou em suas variedades (Planeta, sociedade e linguagem, mundo da língua portuguesa, salvaguarda, promoção e difusão de nosso patrimônio linguístico).

REFERÊNCIAS

- ALTERNÂNCIA de predicados complexos: comparação de variedades do Português e das línguas românicas, apresentado por TRAVASSOS, Pâmela Fagundes, 2021. [s.l.; s.n.], 2021. 1 vídeo (19min24seg), Publicado pelo I Colóquio Internacional VariaR, 2021. Disponível em: <https://www.youtube.com/watch?v=i-tfg2wxyWwk&t=1s>. Acesso em: 23 mar. 2022.
- ANÁLISE Colostrucional através do R: Como utilizar o pacote collostructions na sua pesquisa linguística, apresentado por COSTA, Mariana Gonçalves da; SOUZA, Laís Lima, 2021. [s.l.;s.n.], 2020. 1 vídeo (1hmin). Publicado pelo Festival do Conhecimento da UFRJ, 2021. Disponível em: <https://www.youtube.com/watch?v=-lh215zX36U&t=294s>. Acesso em: 18 jan. 2022.
- COLLOSTRUCTIONS, Contexts, Communities, apresentado por FLACH, Susanne. [s.l.;s.n.], 2021. Vídeo (1h min). Publicado pelo Programa de Pós-Graduação em Linguística da UFRJ (2021). Disponível em: https://www.youtube.com/watch?v=YNB_rNX1ft0. Acesso em: 18 jan. 2022.
- COSTA, Mariana Gonçalves da. Perífrase volitiva com referência futura no português. Comunicação apresentada na 11a SIAC/UFRJ, fevereiro de 2022. <https://www.youtube.com/watch?v=YxDKvJwzsmM> Acesso em: 17 fev. 2022.
- FERREIRA, Bruna Gois Pavão. A variação na construção relacional de mudança de estado: ficar, tornar-se e virar. *Caderno Seminal Digital*, v. 30, n. 30, 2018, p. 48-80.
- GILQUIN, Gaëtanelle. Contrastive collostructional analysis: Causative constructions in English and French. *Zeitschrift fur Anglistik und Amerikanistik* v. 63.3, p. 253-272, out. 2015.
- GOLDBERG, Adele E. Constructions: A new theoretical approach to language. *Trends*

in Cognitive Sciences, 7(5), 219-224, 2003.

HILPERT, Martin. *Germanic Future Constructions: A Usage-Based Approach to Language Change*. Amsterdam: John Benjamins, 2008.

LEINO, Leino.; ÖSTMAN, Jan-Ola. Constructions and variability. *In*: FRIED, M.; BOAS, H. C. *Grammatical Constructions: back to the roots*. Amsterdam/Philadelphia: John Benjamins Publishing Company. p. 192-213. (Constructional Approaches to Language). 2005.

LINGUÍSTICA de Corpus: Introdução ao R para coleta de dados no Twitter, apresentado por SOUZA, Laís Lima; COSTA, Mariana Gonçalves; POPPOLINO, Pedro; MACHADO VIEIRA, Marcia dos Santos. 2020. [s.l.;s.n.], 2020. 1 vídeo (1hmin). Publicado pelo Festival do Conhecimento da UFRJ (2021). Disponível em: https://www.youtube.com/watch?v=yD8TVpQ_FxU&t=0s. Acesso em: 18 jan. 2022.

KILGARRIFF, Adam; JAKUBÍČEK, Miloš; POMIKALEK, Jan; SARDINHA, Tony B.; WHITELOCK, Pete. *PtTenTen: a corpus for Portuguese lexicography*. Working with Portuguese Corpora, 111-30. https://www.sketchengine.eu/wp-content/uploads/Setting_up_for_corpus_2012.pdf, 2014.

MACHADO VIEIRA, Marcia dos Santos. Variação e mudança na descrição construcional: complexos verbo-nominais. *Linguística*, v. 12, 2016. <https://revistas.ufrj.br/index.php/rl/article/view/5445>.

MACHADO VIEIRA, Marcia dos Santos; SOUZA, Laís Lima; COSTA, Mariana Gonçalves da. *In*: CEZARIO, Maria Maura; ALONSO, Karen Sampaio; CASTANHEIRA, Dennis (org.). *Linguística Baseada no Uso: explorando métodos, construindo caminhos*. Rio de Janeiro: Rio Books, 2021.

MACHADO VIEIRA, Marcia dos Santos; BARBOSA, Juliana Bertucci. Coleções de dados brasileiras para o ensino de Português. *In*: MEIRELES, Vanessa; MACHADO VIEIRA, Marcia dos Santos. *Variação e ensino de Português no mundo*. São Paulo: Blucher, 2022.

PROISL, Thomas. *The cooccurrence of linguistic structures*. FAU Forschungen, Reihe A Geisteswissenschaften Band 12. Erlangen: FAU University Press, 2019. DOI: 10.25593/978-3-96147-201-7.

SARAIVA DE PONTES, Eneile Santos. *Predicação transitiva direta com pronome SE: perfis de impersonalização discursiva em variação*. Tese de Doutorado em Letras Vernáculas. Universidade Federal do Rio de Janeiro, Faculdade de Letras, 2022.

STEFANOWITSCH, Anatol; GRIES, Stefan Th. Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8:2, p. 209-243, 2003.

TRAVASSOS, Pâmela Fagundes. A perífrase verbo-nominal no Português Brasileiro:

um estudo da variação por padrão discursivo entre construções com o verbo-suporte “dar”. *E-escrita: revista do curso de Letras da Uniabeu*, v. 12, n. 1, 2021. p. 41-58.

WIEDEMER, Marcos Luiz; MACHADO VIEIRA, Marcia dos Santos. In: MACHADO VIEIRA, Marcia dos Santos; MEIRELES, Vanessa. *Variação em português e em outras línguas românicas*. São Paulo: Editora Blucher, 2022.

CAPÍTULO 5

A PESQUISA EXPERIMENTAL NA ANÁLISE DE CONSTRUÇÕES COM VERBO SUPORTE

Jeane Nunes da Penha
Pâmela Fagundes Travassos
Ravena Beatriz de Sousa Teixeira
(Universidade Federal do Rio de Janeiro)

5.1 INTRODUÇÃO

Por contribuir com a formação do conhecimento e desenvolvimento humano, a pesquisa vem cada vez mais ganhando espaço na sociedade. Na área da Letras/Linguística, há diversos tipos de pesquisas desenvolvidas/em desenvolvimento, como, por exemplo, a pesquisa observacional, a pesquisa experimental, a pesquisa bibliográfica, a pesquisa de campo, a pesquisa de levantamento/censo etc. Neste capítulo, centralizamos o nosso foco no método de pesquisa experimental, caracterizado pela manipulação das variáveis ligadas ao objeto de estudo, cuja finalidade está relacionada à testagem das hipóteses do pesquisador/linguista via avaliação subjetiva *offline* (cf. KENEDY, 2014). Dito isso, a pesquisa de caráter experimental revela-se de suma importância para os pesquisadores, uma vez que, de modo geral, busca dar conta das

seguintes questões: (i) de que modo determinado fenômeno linguístico é realizado?; e (ii) por que determinado fenômeno linguístico é realizado?

A fim de orientar pesquisadores que, porventura, ainda desconhecem métodos de pesquisas experimentais e/ou que ainda não os tenham explorado, objetivamos apresentar as principais etapas envolvidas na elaboração e aplicação de pesquisas experimentais utilizadas na área da Letras/Linguística, com base em estudos desenvolvidos no âmbito do Projeto PREDICAR (Formação e expressão de predicados complexos e predicções: estabilidade, variação e mudança construcional) entre os anos de 2014 e 2017. Tais investigações culminaram em três trabalhos monográficos (TRAVASSOS, 2016; PENHA, 2018; TEIXEIRA, 2018) e um artigo científico (TRAVASSOS; MACHADO VIEIRA, 2020). Todos eles tinham por finalidade a análise do pareamento de *forma e função/significação* de predicadores complexos verbo-nominais com os verbos suportes¹ DAR e FAZER, oriundos de variadas fontes escritas e orais,² a partir de um padrão construcional mais esquemático do tipo [Verbo suporte *dar/fazer* + det.³ + SN⁴ _{(deverbal)-sufixo indicador de grau}] _{predicador complexo}, como: *dar uma olhada, dar uma olhadinha, dar uma olhadela, fazer uma caminhada, fazer uma caminhadinha*.⁵

As investigações realizadas pelas autoras supracitadas deram-se mediante a elaboração e aplicação de testes de atitude e avaliação subjetiva (cf. FASOLD, 1987; GONZALEZ-MARQUEZ, 2006) do tipo *offline*, assim intitulado por “não envolver medidas de tempo, isto é, não demanda cronometragem, e por aferir comportamentos manifestados em momento reflexivo posterior ao processamento cognitivo da informação linguística” (KENEDY, 2014, p. 32). O intuito era documentar a percepção dos falantes brasileiros com relação aos usos licenciados pela construção mais esquemática com verbo suporte. Para tal, em cada uma das três investigações, foram elaborados dez modelos de formulários – questionários compostos por uma sequência de cerca de 8 a 10 enunciados-estímulo –, apresentados de forma impressa a cada informante.⁶ Isso totalizou uma amostra de 300 materiais respondidos.

- 1 Categoria verbal que opera sobre um elemento nominal (deverbal ou não), formando com este uma unidade complexa predicante. Exemplos: *dar parabéns, fazer uma consideraçãozinha, levar um fora*.
- 2 Como fontes escritas, recorremos ao acervo do jornal online O Globo e às páginas digitais dos jornais Gazeta do Povo, Diário do Nordeste e Em tempo. Foi realizada também uma busca livre na ferramenta online Google, em que cada construção pensada foi pesquisada entre aspas. Com isso, o *corpus* conta com dados advindos de redes sociais, de sites de perguntas e de reclamação etc. Como fontes orais, utilizamos transcrições de entrevistas do Programa do Jô, do projeto Iboruna e do projeto Concordância.
- 3 “det.” faz referência ao determinante, como *um, uma, uns, umas*, por exemplo.
- 4 SN equivale ao sintagma nominal.
- 5 A título de ilustração, seguem dados retirados dos *corpora* utilizados para a elaboração dos testes: “O meu relógio já estava assim um pouquinho fora da moda e *dei uma olhadinha* na Joalheria Imperial”. [PB, Jornal online, www.oglobo.globo.com]; “Ironia à parte, Dilma deu a receita: “É fechar a boca e *fazer uma ginasticazinha, uma caminhadinha*”. [PB, Jornal online, <https://oglobo.globo.com>].
- 6 Os informantes tinham perfil diversificado: homens e mulheres, com ensino médio completo, ensino superior completo ou em curso ou estudantes da Pós-Graduação. De modo geral, a maior parte dos participantes era composta de estudantes de Letras da UFRJ que ainda não tinham cursado a disciplina Morfossintaxe. Dessa forma, provavelmente, não tinham tido, pois, um contato formal com o fenômeno sob análise; desse modo, as respostas eram mais intuitivas.

Este capítulo foi elaborado da seguinte forma: após esta parte introdutória, detalharemos o passo a passo pensado durante a preparação dos testes, a partir de questões como as que seguem: (i) Quais os cuidados a serem tomados pelo pesquisador, em termos de distratores linguísticos e extralinguísticos?; (ii) Como ocorre o processo de seleção dos dados?; e (iii) Qual a extensão ideal para os formulários (quantidade de enunciados-estímulos)? Depois, demonstraremos a análise dos resultados, ou seja, como devemos interpretar os resultados quantitativos e qualitativos. Na sequência, há uma seção sobre o uso de plataformas digitais na pesquisa experimental, na qual apresentamos algumas ferramentas que estão disponíveis para auxiliar o pesquisador em casos em que não é possível a aplicação dos testes de forma presencial, por exemplo. Recursos desse tipo podem ser de grande utilidade quando é necessário verificar a percepção de usuários de variedades em diferentes países do mundo, sem a possibilidade de uma pesquisa local. Por fim, exporemos nossas considerações finais, assim como as referências que embasaram este trabalho.

5.2 ELABORAÇÃO DE TESTES DE PERCEPÇÃO E AVALIAÇÃO SUBJETIVA: ETAPAS E CUIDADOS A SEREM TOMADOS

Para confeccionar um teste de atitudes, antes de mais nada, devemos definir os nossos **objetivos de pesquisa**. Souza Guerreiro (2021), por exemplo, teve como principal objetivo a verificação da percepção dos usuários da língua acerca de usos da construção com verbo suporte (V(ir) Prep.(para) Det.(artigo) SN) em variação, instanciações de, por exemplo, “*ir para o céu*”, “*ir para o inferno*”, “*ir para as cucuias*” e “*ir para o beleléu*”. Machado Vieira e Esteves (2009, p. 239) intentaram “mostrar a interferência de atitudes sistemáticas relacionadas às variantes verbais em estudo sobre o comportamento linguístico”, ao lidarem com alternância entre predicadores complexos e predicadores verbais simples (como em *dar parabéns* e *parabenizar*).

Penha e Teixeira (2015) e Penha (2016a, 2016b e 2017) expuseram, em comunicações, resultados de análises via método experimental, advindas de investigações cujo objetivo primordial era averiguar a percepção dos falantes brasileiros sobre usos da construção com verbo suporte (VS(DAR/FAZER) det.(uma) SN-(afixo de grau)), do tipo *dar uma caminhada*, *dar uma caminhadinha*, *fazer uma parada*, *fazer uma paradinha*. Essencialmente, esperava-se observar se os usos eram associados a: (i) uma leitura aspectual e/ou modal; (ii) oralidade e/ou escrita; (iii) registro formal e/ou informal da língua; e (iv) predicadores simples cognatos (*fazer uma perguntinha* = *perguntar*).

Teixeira (2016a, 2016b, 2017 e 2018), por sua vez, buscou mapear a configuração (formal e funcional) de estruturas com os verbos suportes DAR, FAZER e TER que atuam, em textos escritos e orais do português brasileiro, em conjunto a elementos nominais compostos pelos afixos de grau *-(z)inh(o/a)*, *-(a/i)da*, *-(a/i)dinha*, *-aço* e *-ão* (como *dar um carrinho*, *ter uma implicanciazinha*, *dar uma olhada*, *dar um pedaço*, *fazer um golzão*) a fim de compor um predicador complexo. Averiguaram-se, por meio de experimentos psicolinguísticos, tanto as condições semânticas, discursivas e

pragmáticas referentes aos empregos das distintas formas de predicação, como a possibilidade de predicadores associados a padrões construcionais diferentes (*olhar, dar uma olhada, dar uma olhadinha*) prestarem-se a apresentar um estado de coisas com algum grau de equivalência.

Travassos e Machado Vieira (2019), em uma investigação sobre variação construcional, objetivaram verificar se brasileiros percebem diferenças no polo funcional de predicadores complexos, assim como buscaram observar as motivações para as respostas selecionadas no questionário (opção por uma determinada construção e não outra). Travassos e Machado Vieira (2020) buscaram verificar quais valores usuáries do Português do Brasil associam a usos de perífrases verbo-nominais com o verbo suporte DAR.

Utilizaremos o trabalho desenvolvido nesses estudos (PENHA e TEIXEIRA, 2015; PENHA, 2016a, 2016b, 2017; TEIXEIRA, 2016a, 2016b, 2017, 2018; TRAVASSOS; TRAVASSOS; MACHADO VIEIRA, 2019; MACHADO VIEIRA, 2020) como base para fazer considerações sobre importantes aspectos na organização de uma pesquisa experimental.

Se, por exemplo, queremos contrapor as impressões de que construções com verbo suporte DAR ocorrem somente em gêneros textuais cujo registro está mais voltado ao polo informal, assim como somente estão presentes na modalidade oral da língua, é importante buscarmos dados que estejam materializados em gêneros textuais mais próximos do polo formal. De mesmo modo, é interessante encontrarmos dados escritos do fenômeno, a fim de contra-argumentar, a partir de evidências, as intuições preliminarmente estabelecidas. Uma possibilidade seria a análise de dados provenientes de textos escritos do domínio jornalístico, tal como o jornal *O Globo*, por exemplo (cf. TRAVASSOS, 2019). Textos disponíveis nesse ambiente comunicativo costumam passar por revisão linguística, recebendo atenção de diversos profissionais da área, atentos à forma de expressão. Desse modo, o grau de monitoramento linguístico é maior nesse domínio discursivo.

Tendo os objetivos em mente, passamos a selecionar o **design de pesquisa experimental** capaz de fornecer evidências às hipóteses. É importante pensarmos, entre outras questões, nas sentenças-estímulo, na técnica a ser utilizada e no tipo de teste. Os dados coletados via análise de usos podem servir como estímulos em tarefas de pesquisa experimental. Há diversas técnicas para lidar com o problema da avaliação subjetiva (questionários com respostas de sim/não, de avaliação metalinguística, de escalonamento/escala Likert;⁷ leitura; reação subjetiva, por exemplo) e detectar os

7 Segundo Rodrigues, Assmar e Jablonski (2015, p. 272 e 273), “sendo de simples confecção e aplicação e possuindo boa correlação com outras escalas e critérios de medida de atitudes, a Escala Likert é uma das mais usadas para a mensuração das atitudes. Consiste em uma série de afirmações (em geral entre 20 e 30) relativas a um objeto atitudinal (por exemplo, divórcio, censura, aborto, eutanásia etc.), mais ou menos metade das quais sendo favoráveis ao objeto atitudinal, e a outra metade, desfavorável. A conveniência de dividir o número de afirmações favoráveis e desfavoráveis em duas metades decorre da necessidade de serem evitadas certas tendenciosidades individuais como, por exemplo, a de concordar mais do que discordar. Cada afirmação é seguida de cinco alternativas: concordo plenamente, concordo em parte, não estou seguro, discordo em parte, e discordo totalmente. A cada uma dessas

significados sócio-culturais-discursivos das variantes e suas correlações a certas variáveis/contextos.

No estudo de Travassos e Machado Vieira (2019), a **técnica** foi a de questionário (online) via *Google Forms*, com diferentes configurações de resposta. O teste foi feito não só no modelo fechado (com questões de múltipla-escolha), mas também no modelo aberto (com possibilidade de curtas respostas discursivas, metalinguísticas). Desse modo, é possível ter acesso também aos comentários livres dos informantes, por meio dos quais eles manifestam o que lhes ocorreu ao responderem a cada estímulo. Além disso, as autoras também utilizaram a técnica de preenchimento de lacunas. O teste de reação subjetiva possibilitou observar juízos de valor emitidos pelos participantes acerca das variantes, às correlações socioculturais ou sociodiscursivas que têm em mente.

Um outro aspecto relevante é se o **tempo de reação** aos estímulos será contabilizado (método online) ou não (método *offline*). Essa medida pode ser importante: se for online, podemos ter acesso a indícios do nível de dificuldade de acesso aos elementos da construção acionados no momento de seu reconhecimento cognitivo; por outro lado, se for *offline*, consideramos a potencialidade de haver um processo de reflexão sobre as escolhas feitas pelos informantes. Assim, a seleção de uma técnica em detrimento de outra associa-se aos objetivos investigativos delimitados previamente pelo pesquisador. Nos estudos de base aqui mencionados, o tempo de resposta não foi levado em consideração, de modo que os participantes eram livres para usar o tempo que achassem necessário na resposta ao questionário.

Por meio de registros obtidos via tarefas experimentais, também podemos alcançar, empiricamente, padrões de configuração de *personas*/identidades delineadas discursivamente. Tais definições são, antes de mais nada, elaboradas coletivamente numa espécie de negociação sempre ativa na relação entre indivíduos e destes com construtos histórico-sócio-culturais mais amplos e expectativas de comportamento presentes nas comunidades de que são parte, de que querem fazer parte ou a que querem ter sua imagem associada.

A título de ilustração, a seguir, apresentamos um exemplo do resultado da resposta dos informantes a uma sentença-estímulo de um teste realizado por Travassos e Machado Vieira (2020, p. 14). O objetivo desse teste era verificar qual opção entre as alternativas seria considerada como a mais adequada ao contexto. Além disso, tencionava-se também observar a percepção, a avaliação e a opinião dos informantes sobre a opção feita e/ou sobre as outras alternativas no que diz respeito à intenção comunicativa em jogo.

alternativas são atribuídos valores numéricos de 1 a 5, cabendo ao confeccionador da escala determinar em que direção (positiva ou negativa) ele vai atribuir os valores mais altos”.

Figura 01: Resultado da análise de informantes a sentença-estímulo de um teste de atitude.

9) (Situação: *Governador* dá satisfação à população da utilização da verba pública.)
É uma verba que servirá para retirar os andaimes e escoras, _____ e fazer
pequenos arranjos. (teste 10)

Quadro 10: Respostas de dez informantes para a situação 9 do teste 10 e justificativas

(a) Dar uma pintadinha	1
(b) Uma pintadela dar	0
(c) Dar uma pintadinha	1
(d) Uma pintada dar	1
(e) Dar uma pintadela	1
(f) Uma pintadinha dar	0
(g) Pintar	7
Justificativas de informantes para a marcação da opção (g):	
1) Mais culta, em se tratando de uma figura política	
2) Se explicar com a população	
3) Passar de modo simples a informação	
4) Ser formal	

Fonte: Travassos e Machado Vieira (2020, p. 14).

A opção pelo predicador simples *pintar* foi a preferida. A partir das respostas metalinguísticas dos informantes, percebemos que essa opção se deve ao fato de haver uma autoridade pública envolvida na situação, como locutor da mensagem e, portanto, a expectativa de um maior grau de formalidade, bem como a associação da forma verbal simples, no senso comum, por vezes, a uma norma culta da língua, diferentemente de usos de predicadores complexos. Por outro lado, houve participante que apontou para o público-alvo da mensagem: tendo em vista que a informação é voltada à população, partiu da expectativa de que os recursos linguísticos usados devem ser de fácil acesso; portanto, a forma simples seria preferida nesse caso por esse motivo.

Ainda, dois pontos que merecem destaque e atenção no processo de confecção e realização de uma pesquisa experimental são os **distratores** e os **informantes**. Em uma pesquisa linguística, geralmente, buscamos obter um uso mais próximo possível do real. Dessa forma, não queremos que os participantes da pesquisa saibam qual fenômeno está sendo observado, de modo que sua atitude não se revele artificialmente. Por isso, recorremos a distratores, que consistem em elementos linguísticos ou não linguísticos que cumprem justamente a função de distrair o informante do objeto de análise principal. Os distratores linguísticos podem ser formados, por exemplo, por sentenças, cujo foco recaia sobre outros fenômenos linguísticos, de natureza diversa (fonético-fonológica, morfológica, sintática, semântica, discursiva, pragmática). Já os distratores não linguísticos podem ser tanto objetos presentes no ambiente de resposta ao teste, como sensações físicas do respondente (fome ou estresse, por exemplo), quanto a influência de outras pessoas.

A quantidade ideal de distratores tem relação direta com a natureza do fenômeno em foco. Por vezes, a análise de um fenômeno morfossintático, como o predicador complexo, requer que o contexto esteja expresso na forma de sentença ou até mesmo parágrafo para que sua interpretação seja feita de forma adequada. Sendo assim, selecionar uma quantidade muito grande de distratores, assim como de sentenças-estímulo alvo da investigação, pode tornar o teste extremamente cansativo para o informante; dessa forma, a sua atenção pode ser comprometida. Nesse ponto, é necessário

bom senso do pesquisador, cabendo a ele a decisão pela melhor quantidade de distrações, em função de cada pesquisa e de cada objeto de estudo.

De igual modo, é importante verificar com antecedência o perfil dos informantes, por exemplo, os estudantes da área de Letras (principalmente, os que já tenham cursado a disciplina de Morfossintaxe). Esses costumam ter consciência e domínio dos fenômenos linguísticos, por apresentar um olhar já treinado; assim, geralmente, são excluídos dos testes de percepção e atitude subjetiva, uma vez que esses testes buscam respostas espontâneas, naturais e sem acesso à consciência do que está em foco. Sendo assim, de modo geral, selecionamos pessoas de outras áreas do conhecimento para responder à pesquisa – ou, ainda, estudantes da área de Letras que ainda não tenham cursado Morfossintaxe.

Por fim, ressaltamos um último ponto: o consentimento livre e esclarecido para a realização da testagem, bem como a autorização de uso dos testes na análise linguística. Por lidarmos com seres humanos, ainda que a pesquisa não traga nenhum tipo de prejuízo à pessoa, é imprescindível termos a autorização dos participantes quanto ao uso das respostas disponibilizadas no teste. Isso pode ser feito antes ou ao final do teste. E, ainda, caso o respondente deseje, ele pode ter acesso, posteriormente, aos resultados da pesquisa.

5.3 ANÁLISE CRÍTICA DOS RESULTADOS

Para efeito da análise dos dados obtidos por meio de metodologia experimental, é necessário, em conformidade às questões levantadas na seção anterior, considerar a natureza do tipo/*design* dos experimentos confeccionados e os objetivos por trás de sua configuração. Ao tratarmos de materiais elaborados no âmbito do Projeto Predicar, discorreremos sobre “testes” psicolinguísticos de percepção/avaliação subjetiva, cujos conjuntos de respostas oferecidas pelos participantes são avaliados quantitativa e qualitativamente, respeitando as delimitações da técnica experimental em voga. Considerando que os testes constituem ferramentas voltadas para a averiguação de hipóteses referentes ao conhecimento/comportamento do falante/ouvinte em relação a sua língua, é necessário atentar-se aos questionamentos que se pretende responder por meio da análise, os quais norteiam os objetivos centrais por trás do desenho experimental.

A seguir, observamos um exemplo, de mesmo *design* do considerado no quadro 1, de um dos modelos de teste realizados por Teixeira (2017), para o qual o objetivo central era observar o grau de comparabilidade entre predicadores simples e predicadores complexos com verbo suporte e/ou entre estruturas complexas derivadas de padrões construcionais distintos, assim como mapear a funcionalidade dos predicadores em jogo.

Figura 2: Exemplo de enunciado-estímulo exposto em experimento.

Que expressões cabem nas lacunas abaixo?

1) Obrigado ao ministro Sérgio Amaral. Eu acho que é sempre refrescante a gente ter noção sobre o *timing* disso, mas eu pediria ao ministro Sérgio Amaral, por exemplo, que olhasse ao sul do Brasil, e olhasse para a Argentina, e verificasse se ele não gostaria de talvez seguir a velocidade portenha, que talvez seja um pouco maior do que a velocidade inglesa, e que certamente é muito maior do que a velocidade brasileira, a portenha. Pediria que ele _____ por cima da cordilheira dos Andes e desse uma mirada no Chile, e verificasse como o Chile conduziu velozmente o seu programa de privatização.

a) () olhasse b) () desse uma olhada c) () desse uma olhadinha d) () desse uma olhadazinha

Justificativa:

Se tiver marcado mais de uma alternativa de resposta, indique, apenas com as letras correspondentes às opções, o quanto as expressões marcadas são parecidas entre si.

+ Muitíssima parecida(s)	+ mais parecida(s)	± parecida(s)	- pouco parecida(s)	- Nada parecida(s)

Fonte: Teste aplicado por Teixeira (2017).

Em tal modelo, apresentaram-se, em um formulário impresso, enunciados-estímulos compostos por uma lacuna, em que, no enunciado original, havia um predicador (complexo ou simples). A partir disso, solicitou-se aos participantes que a preenchessem, segundo sua preferência, com uma ou mais das opções ofertadas. Caso o participante selecionasse mais de uma opção, era necessário indicar, em um quadro, o grau de comparabilidade entre as formas marcadas, além de justificar, se possível, a resposta. Dessa forma, temos quatro informações a considerar a fim de apreendermos os dados adquiridos: (i) as opções marcadas em cada um dos enunciados-estímulos apresentados; (ii) o grau de comparabilidade, definido pelos informantes, entre os distintos predicadores expostos como opção em cada um dos enunciados-estímulo; (iii) o perfil dos participantes;⁸ e (iv) as justificativas por estes declaradas.

8 Em primeira instância, devido a limitações enfrentadas em seu contexto de aplicação, não houve um controle rígido do perfil dos participantes, em relação a seu sexo, faixa etária, nível de escolaridade e região de residência. As informações foram solicitadas e atendidas pelos informantes, entretanto, não há uma uniformidade entre os perfis, fator necessário para correlacionar, coerentemente, as variáveis extralinguísticas relativas ao perfil dos respondentes aos padrões de uso detectados nos experimentos. Entretanto, para fins didáticos, tais informações constam em nossa análise.

Entretanto, apenas angariar tais informações não equivale a efetuar, de fato, a análise. Os testes buscam captar indícios da atuação de variáveis (fatores de ordem diversa) no processo de uso da língua. Assim, cada uma das informações supracitadas deve passar por uma meticulosa avaliação qualitativa, respaldada em diretrizes teórico-metodológicas correspondentes à base teórica na qual o estudo se baseia.

A princípio, no processo de análise, é necessário considerar cada enunciado-estímulo e seus respectivos resultados/dados individualmente. Cada contexto de uso do objeto sob observação apresenta suas características individuais. Logo, a análise somente é viável se cada um dos resultados for tratado individualmente. A seguir, observa-se uma matriz com as respostas de dez informantes acerca do exemplo exposto na Figura 2.

Quadro 1: Perfil dos informantes e suas respostas referentes a sentença-estímulo.

Informante	Idade (anos)	Nível de escolaridade	Cidade	Gênero	Opção(ões) selecionadas	Grau de similaridade
1	20	Superior incompleto	RJ	Feminino	desse uma olhada	
2	22	Superior incompleto	RJ	Não binário	desse uma olhada, desse uma olhadinha	Muitíssimo parecidas
3	20	Superior incompleto	RJ	Feminino	olhasse	
4	49	Superior incompleto	RJ	Masculino	olhasse	
5	20	Superior incompleto	RJ	Masculino	olhasse, desse uma olhada	Mais parecidas
6	21	Superior completo	RJ	Feminino	olhasse	
7	19	Médio completo	RJ	Feminino	olhasse	
8	35	Superior incompleto	São Gonçalo	Feminino	olhasse	
9	24	Superior incompleto	RJ	Feminino	olhasse, desse uma olhada	Pouco parecidas

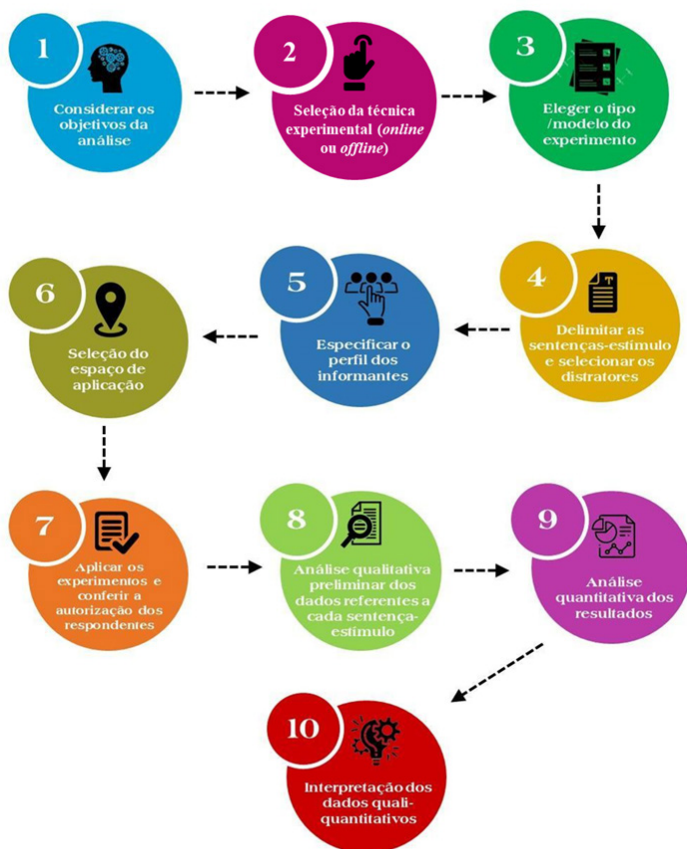
Fonte: Autoral.

Segundo a matriz apresentada, há uma preferência dos participantes da pesquisa pela seleção do predicador simples “olhasse” em detrimento às demais estruturas, o que vai de encontro à configuração original do fragmento textual utilizado como estímulo. Entretanto, as estruturas “desse uma olhada” e “desse uma olhadinha” também foram selecionadas, por mais que com uma frequência inferior. Tal fato indica que, de acordo a determinados contextos e interpretações, predicadores com verbos suportes, oriundos de distintos padrões esquemáticos, podem ser acionados a fim de pregar um estado de coisas com um certo grau de equivalência à predicadores sim-

ples, visto que as perífrases previamente mencionadas foram selecionadas em prol do preenchimento de um *slot* verbal anteriormente composto por uma unidade verbal simples. Além disso, vê-se delimitado um grau de similaridade entre (i) o predicador simples “olhasse” e o predicador complexo com verbo suporte composto pelo afixo *-ada* e (ii) este e o predicador composto pelo afixo *-inha*, diferenciando-se, em ambos os casos, em relação ao grau de semelhança/associação entre as estruturas. Observamos, pois, que apenas os dados referentes a um enunciado tornam possível tecer considerações relevantes sobre a relação de alternância/similaridade entre distintos padrões de predicadores.

Para se obter uma melhor visualização dessas relações e resultados quantitativos que figuram como amostras da realidade gramatical da língua, é necessário, em seguida, considerar as respostas de cada estímulo de forma conjunta. Assim, tratar da análise quantitativa e sua avaliação qualitativa a fim de chegar aos resultados finais, partindo de pequenas porções dos experimentos até podermos angariar uma visão panorâmica dos resultados. Portanto, a seguir, apresentamos os passos realizados ao se tratar da configuração e interpretação dos experimentos psicolinguísticos em pauta.

Figura 3: Etapas relativas à produção, aplicação e análise dos experimentos.



5.4 PESQUISA EXPERIMENTAL EM PLATAFORMAS DIGITAIS

A depender das questões da pesquisa, pode ser importante, por exemplo, comparar diferentes variedades de uma língua, cujos nichos se encontram em diferentes países, como algumas variedades do português: brasileiro, europeu, moçambicano, angolano, entre outros. Nesse ponto, surgem diversas dificuldades não só físicas e geográficas, mas também logísticas e culturais para o desenvolvimento da pesquisa experimental. Uma vez que nem sempre é possível ao pesquisador ir diretamente ao local para aplicar os testes, uma solução viável seria através de plataformas disponibilizadas no meio digital. Por meio de ferramentas como o *Google Forms*⁹ e o *Online-Pesquisa*¹⁰, por exemplo, conseguimos realizar a pesquisa mesmo a longas distâncias.

O *Google Forms*, aplicativo criado pelo Google, pode ser usado tanto para fins investigativos quanto para registro:

Figura 4: Formulário via *Google forms*.

The image shows a Google Form titled "Open Letter Maker". At the top, it says "All of the submissions to this form will be appended into a table as signatures in the style of an open letter." Below this, there are three required text input fields: "Full Name", "Department", and "Position". Each field has a red asterisk indicating it is required. The "Full Name" field has a subtext: "Any text provided here will be used to create a digital signature, including honorifics or degrees". At the bottom of the form is a purple "Submit" button.

Fonte: Google imagens.

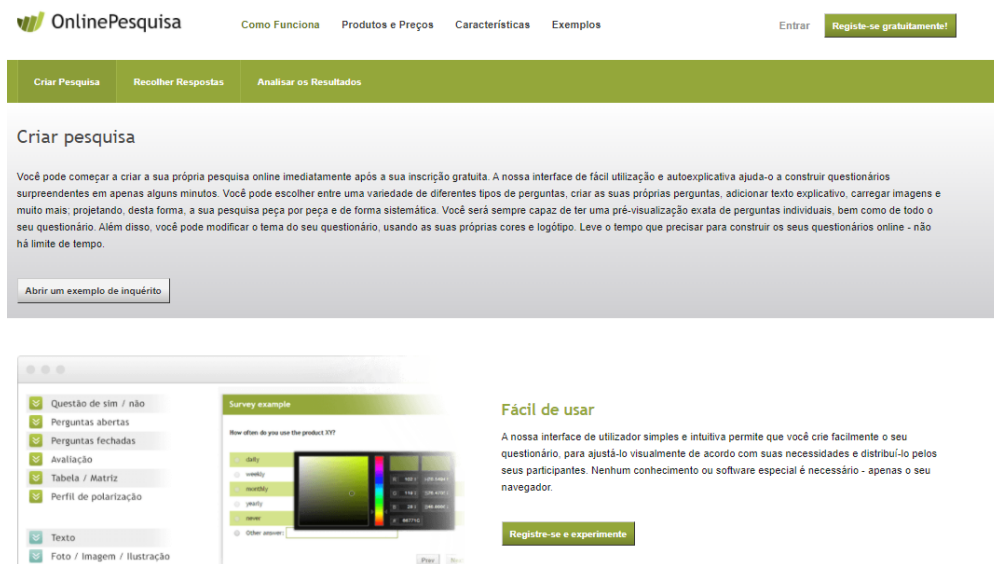
9 O Google Forms pode ser acessado por meio do link [<https://docs.google.com/forms/u/0/?tgif=d>] (acesso em: 10 ago. 2021).

10 O OnlinePesquisa pode ser acessado por meio do link [<https://www.onlinepesquisa.com/>] (acesso em: 10 ago. 2021).

Criar, editar e formatar formulários a partir da ferramenta não requer muita complexidade; além disso, encontramos facilmente um breve guia de ajuda fornecido pelo próprio Google através do *link* [<https://support.google.com/docs/answer/6281888?hl=pt-BR&co=GENIE.Platform%3DDesktop>] (acesso em 17 jan. 2022).

O *OnlinePesquisa*, assim como o *Google Forms*, é uma plataforma que possibilita a criação de formulários:

Figura 5: Criação de formulário via *OnlinePesquisa*.



Fonte: Site *OnlinePesquisa*.

No referido site, encontramos um pacote gratuito para estudantes (*basic*), com um mês de duração, e diferentes pacotes pagos (*personal*, *pro*, *business* e *enterprise*). O que os diferencia? Além do valor, a limitação ou não de perguntas e respostas permitidas por cada pacote, assim como a quantidade de contas de usuários possíveis dentro do pacote.

Para encerrar esta breve seção, ressaltamos que recorrer à utilização de formulários digitais é particularmente importante em um contexto de urgência sanitária, como o que vivemos durante a pandemia de Covid-19, na medida que há restrições de deslocamento e de contato físico.

5.5 CONSIDERAÇÕES FINAIS

Buscamos, neste texto, apresentar um passo a passo na elaboração dos testes de percepção e avaliação subjetiva com ênfase nos cuidados a serem tomados em sua confecção. Os fatores a que os pesquisadores que pretendem adotar a metodologia

experimental devem atentar podem ir além dos aqui propostos. Não tivemos a pretensão de apresentar uma lista exaustiva dos aspectos a serem considerados. Ilustramos apenas alguns elementos que julgamos essenciais na operacionalização dos testes. Além disso, também ressaltamos informações acerca da interpretação de resultados quantitativos e qualitativos e abordamos a pesquisa experimental em plataformas digitais, como o *OnlinePesquisa* e o *Google Forms*, ressaltando a sua importância, principalmente, quando há limitações físicas e geográficas, tal como o contexto pandêmico que vivemos atualmente.

A fim de ilustrar os procedimentos relacionados à produção dos testes, baseamos-nos em estudos sobre a variação entre construções com verbo suporte (TRAVASSOS, 2016; PENHA, 2018; TEIXEIRA, 2018; TRAVASSOS; MACHADO VIEIRA, 2020). Usuários da língua podem associar as variantes a contextos de uso semelhantes, daí a importância do estudo experimental para captar a percepção mediada pela avaliação subjetiva e atitude dos participantes dos testes com relação às perífrases verbo-nominais. Os estudos em questão apresentaram evidências em prol da alternância como uma possibilidade linguística percebida pelos falantes. Investigações outras que possuam objetos de estudo diversos podem repensar o *design* dos testes em função das questões, objetivos e hipóteses de pesquisa.

Outro caminho possível, na linha de revisão do que foi empreendido até agora, é a possibilidade de se recorrer a testes de avaliação subjetiva, os quais contemplem a observação de manifestações/registros de linguagem não verbal, haja vista o fato de a conceptualização do mundo envolver vários canais/modos de expressão (multimodalidade), entre os quais os linguísticos.

REFERÊNCIAS

- FASOLD, R. *The Sociolinguistics of Society*. Vol. I. New York, USA: B. Blackwell, 1987. p. 147-179.
- GONZALEZ-MARQUEZ, M. et al. *Methods in Cognitive Linguistics*. Amsterdam, John Benjamins, 2006.
- KENEDY, E. Nos meandros da experimentação. In: ABRAÇADO, J.; KENEDY, Eduardo (orgs.). *Transitividade traço a traço*. Niterói: Editora da UFF, 2014. p. 29-37.
- MACHADO VIEIRA, M. dos S.; ESTEVES, G. A. T. Metodologia de avaliação subjetiva de usos linguísticos em variação. In: LOPES, C.; REICH, 112 *Uma história de investigações sobre a Língua Portuguesa Uli*. Romania. Variação linguística em megalópoles latino-americanas, 39, p. 237-266, 2009.
- PENHA, J. N.; TEIXEIRA, R. B. S. *O funcionamento de alguns predicadores com verbo suporte FAZER*. XXXVII Jornada Giulio Massarani de Iniciação Científica, Tecnológica, Artística e Cultural da UFRJ, 2015 (Comunicação).
- PENHA, J. N. *Fazer uma investigaçãozinha/pesquisinha/perguntinha: como funciona?*

- XXI Seminário Nacional e VIII Seminário Internacional do Grupo de Estudos Discurso e Gramática, Faculdade de Letras da UFRJ, 2016a (Painel).
- PENHA, J. N. “*Dar uma conferidinha*” e “*Fazer uma modificaçãozinha*”: perífrases com verbo suporte no português brasileiro. 7ª Semana de Integração Acadêmica da UFRJ, 2016b (Comunicação).
- PENHA, J. N. *Predicadores com verbo suporte: vamos “dar uma examinada” em dados e “fazer uma resuminho” dos resultados?* 8ª Semana de Integração Acadêmica da UFRJ, 2017 (Comunicação).
- PENHA, J. N. *O uso de predicadores complexos com os verbos suportes DAR e FAZER na modalização do discurso*. Monografia de fim de curso. Faculdade de Letras. Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2018, 46p.
- RODRIGUES, A.; ASSMAR, E. M. L.; JABLONSKI, B. *Psicologia Social*. Petrópolis, RJ, 32. ed: Vozes, 2015.
- SOUZA GUERREIRO, S. C. G. Estudo experimental das construções com verbo-suporte (“ir para o céu”, “ir para o inferno”, “ir para as cucuias”, “ir para o beleléu”). *e-escrita Revista do Curso de Letras da UNIABEU Nilópolis*, v.12, Número 1, janeiro-junho, 2021.
- TEIXEIRA, R. B. S. *Construções com verbo suporte e verbos plenos cognatos: convivem, fazem competição ou dão uma mudadinha?* 7ª Semana de Integração Acadêmica da UFRJ, 2016a (Comunicação).
- TEIXEIRA, R. B. S. *FAZER UMA CONSIDERAÇÃOZINHA/CONSIDERAR: como operam no PB?* XXI Seminário Nacional e VIII Seminário Internacional do Grupo de Estudos Discurso e Gramática, Faculdade de Letras da UFRJ, 2016b (Painel).
- TEIXEIRA, R. B. S. *Predicadores complexos com verbo-suporte: dando uma olhadinha nos padrões construcionais em jogo*. 8ª Semana de Integração Acadêmica da UFRJ, 2017 (Comunicação).
- TEIXEIRA, R. B. S. *Construções com verbo suporte: mapeamento de predicadores compostos por afixos de grau*. Monografia de fim de curso. Faculdade de Letras. Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2018, 60p.
- TRAVASSOS, P. F. *Construções com verbo-suporte DAR: indicação de aspecto e/ou outro valor?* Monografia de fim de curso. Faculdade de Letras. Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2016, 101p.
- TRAVASSOS, P. F.; MACHADO VIEIRA, M. S. Uma análise construcionista da variação entre construções com verbo-suporte DAR no PB. *REVISTA SOLETRAS*, v. 1, p. 272-298, 2019.
- TRAVASSOS, P. F.; MACHADO VIEIRA, M. S. O que brasileiros dizem acerca de usos de construções com verbo suporte? *Revista (Con)Textos Linguísticos*, Vitória, v. 14, n. 27, p. 198-217, 2020.

CAPÍTULO 6

PROPOSTA DE TESTES EXPERIMENTAIS PARA ANÁLISE DA VARIAÇÃO CONSTRUCIONAL DE PREDICADOR COMPLEXO

Pâmela Fagundes Travassos (*Universidade Federal do Rio de Janeiro*)

A proposta de trabalho aqui delineada visa a expor modelos de testes experimentais para análise da variação entre predicadores complexos com verbo suporte DAR. A seguir, encontramos exemplos dos predicadores complexos ligados à construção aqui sob análise:

[Verbo suporte **DAR** + [(determinante) + (modificador) **X-sufixo** (modificador)]SN]
predicador complexo com verbo suporte – um estado de coisas

*Ex. (1): A fim de **dar um empurrãozinho** e fazer deslanchar de vez a imaginação dos leitores, o Rio Show bolou um guia com sugestões de dez programas*

para dez circunstâncias em que é preciso, no mínimo, de uma dose dupla de criatividade. [PB, Jornal online, www.oglobo.globo.com]

*Ex. (2): “A fila estava muito grande, fiquei quase meia hora esperando”, disse Isadora. “Pedi ao Armando [marido de Luiza] para me **dar uma forcinha**. Então passei na frente”. *Simples assim. [PB, Jornal online, www.oglobo.globo.com]**

*Ex. (3): E apesar de continuar morando em Campinas, quando a filha está no sufoco, corre para **dar uma mãozinha**. [PB, Jornal online, www.oglobo.globo.com]*

*Ex. (4): Neste Natal, o Bukowski resolveu **dar uma ajudinha** para quem quer estender as comemorações. A partir da meia-noite do dia 24 e até às 3h do dia 25, duas vans do bar passarão por pontos estratégicos da cidade, levando os mais animados para curtir uma noite natalina de muito rock - e algumas doses. A van é gratuita. [PB, Jornal online, www.oglobo.globo.com]*

A partir da observação dos exemplos de 1 a 4 apresentados anteriormente, percebemos que as construções “dar um empurrãozinho”, “dar uma forcinha”, “dar uma mãozinha” e “dar uma ajudinha” parecem ter semelhanças não só no nível semântico, sendo equivalentes a “ajudar”, mas também em sua estrutura e em outros aspectos formais e funcionais. Desse modo, em função da possibilidade de alternância, de intercambialidade entre essas expressões, buscamos averiguar, com os testes psicolinguísticos, como usuários da língua enxergam tais construções; se, de igual modo, também veem a possibilidade de variação.

Ao optarmos pela metodologia de pesquisa experimental psicolinguística (FASOLD, 1987; GONZALEZ-MARQUEZ, 2006; DRAGER, 2015, 2018), temos a finalidade de observar, segundo técnicas e testes de avaliação subjetiva do tipo *offline*, a avaliação subjetiva, o comportamento linguístico de usuários do português do Brasil acerca da variação desses complexos verbo-nominais. Com isso, esperamos reunir subsídios sobre os pareamentos forma-função (construção). Tencionamos verificar se os informantes, ao analisarem os contextos disponibilizados, reconhecem sentidos semelhantes semântica e pragmaticamente, embora o “Princípio da não sinonímia”, proposto por Bolinger (1968) e referido por Goldberg (1995, p. 67-68) preveja que, “se duas construções são sintaticamente distintas, elas devem ser semântica ou pragmaticamente distintas”.

Assim, objetivamos observar se usuários da língua detectam, de fato, formas diferentes para função semelhante, se identificam certas estruturas convivendo ou competindo (com uma das formas ganhando mais espaço) e se reconhecem a variação por analogia e por polissemia. Dessa forma, analisamos como se dá a percepção mediada por avaliação subjetiva de usuários da língua acerca dessas construções acerca da possibilidade (ou não) de alternância entre os padrões construcionais nos diversos contextos/situações.

Hipotetizamos que os participantes dos testes de avaliação subjetiva irão reconhecer a influência de aspectos do cotexto estritamente linguístico e do contexto semântico, discursivo e pragmático na marcação dos diversos valores das perífrases, associando essas construções não só ao valor aspectual pontual, mas, principalmente (e mais produtivamente), à indicação de uma atitude de polidez, à perspectiva intersubjetiva, ao valor modal (convencionalizado).

Assim, consideramos orientações da Psicolinguística no que se refere à confecção de testes experimentais. Entendemos, com base em Esteves (2008), “atitude” como “um estado mental de predisposição em relação a formas/estruturas linguísticas em um sistema, que lhes pode ser favorável, desfavorável ou, ainda, neutro, que pode ser relativamente estável e que pode ter origem em comportamentos coletivos”. Segundo Rodrigues, Assmar e Jablonski (2018, p. 161), “podemos definir atitude social como uma organização duradoura de crenças e cognições em geral, dotada de carga afetiva pró ou contra um objeto social definido, que predispõe a uma ação coerente com as cognições e afetos relativos a este objeto”. Para Lent (2010, p. 612-613), a percepção:

para os seres humanos, é a capacidade de associar as informações sensoriais à memória e à cognição, de modo a formar conceitos sobre o mundo e sobre nós mesmos e orientar o nosso comportamento. Isso significa duas coisas: primeiro, que a percepção é dependente mas diferente dos sentidos, isto é, tem um “algo mais” que a torna uma experiência mental particular; segundo, que ela envolve processos complexos ligados à memória, à cognição e ao comportamento (LENT, 2010, p. 612-613).

Desse modo, entendemos que a atitude do usuário da língua pode ser de rejeição, de aceitação ou, ainda, neutra. Já a avaliação subjetiva tem relação com as crenças, conhecimentos, opiniões e ideologias do indivíduo. Por fim, a percepção é algo instantâneo, tem relação com os sentidos humanos. A apreensão (sempre parcial) do processamento ou da percepção linguística se dá em metodologia online e não se circunscreve, apenas, à linguagem verbal. Há uma série de estímulos distratores¹¹ não só linguísticos, mas também não linguísticos (internos ou externos ao indivíduo) que

11 “Sabemos, a partir da experiência de cada dia, que não apenas tarefas-alvo, mas também informação distratora influencia nosso desempenho. Ela é processada até certo ponto; às vezes, fornece informações adicionais que podem ser úteis para a obtenção do que se tem como meta. A natureza exata dos processos cerebrais que ocorrem para lidar com detalhes importantes da tarefa-alvo em uma magnitude esmagadora de distração ainda é altamente debatida” (PRIESTER, L. D. & WISWEDE, D., 2018).

podem influenciar na atenção e na percepção¹² e essas interferências podem ser significativas ou não.

Dentre os elementos externos ao participante que podem interferir, podemos citar os seguintes: (i) tarefa com alta demanda atencional; (ii) “material” demasiadamente extenso da atividade experimental; (iii) ambiente (elementos visuais e auditivos); (iv) interação social; e (v) grau de “atratividade” do objeto de atenção. Como elementos internos ao participante que podem interferir em sua percepção, podemos fazer menção aos que seguem: (i) grau de familiaridade com a tarefa a ser realizada e com os elementos linguísticos em questão; (ii) emocional; (iii) fadiga; (iv) preocupação ou estresse; (v) devaneios; (vi) fome; e (vii) falta de interesse. Assim, entendemos não ser possível captar a percepção por meio de teste *offline*, mas somente através de testes online. Então, dizemos que os testes aqui propostos buscam captar a percepção mediada por avaliação subjetiva, uma vez que passam por algum tipo de avaliação/reflexão com base nas categorias com que o pesquisador formula os testes.

No que diz respeito aos materiais para a análise da avaliação subjetiva de usuários do português do Brasil acerca dos complexos verbo-nominais em estudo, contamos com seis testes, no formato de questionários, voltados para a pesquisa experimental de usos detectados no comportamento linguístico observado em textos, montados em arquivo *word* e adaptados para serem aplicados pela plataforma online *Google Forms*. O objetivo dos experimentos consiste, de modo geral, em fazer previsões comportamentais sobre a avaliação subjetiva dos usos apresentados ao participante de pesquisa, de modo a captar a realidade psicológica do falante nativo, isto é:

Testar empiricamente algum tipo de previsão acerca do comportamento linguístico a ser manifestado por um sujeito (o participante da tarefa), inserido numa situação controlada pelo experimentador, na qual se procura verificar se certas variáveis independentes – isto é, variáveis controladas pelo pesquisador – podem estar relacionadas a alguma variável dependente – isto é, uma variável de resposta, um comportamento (ABRAÇADO; KENEDY, 2014, p. 31)

Baseando-nos na leitura da Resolução CNS nº 510, de 07 de abril de 2016, sobre ética na área de humanas, a proposta de projeto de pesquisa experimental contendo todos os seis modelos de testes desse estudo deve ser submetida ao CEP/CONEP (Conselho Nacional de Ética na Pesquisa) da UFRJ para avaliação do Comitê de Ética. Com a aprovação do conselho, legitimamos a segurança, o bem-estar físico e emocional, o anonimato e todos os direitos dos sujeitos participantes das tarefas.

12 Trabalho desenvolvido e apresentado por Eliete F. Batista, Marcia dos S. Machado Vieira, Mara P. Mariano, Pâmela F. Travassos, Silvia C. Souza e Nahendi A. Mota no evento 2º Seminário do PEUL 2019 (Programa de Estudos sobre o Uso da Língua): Variação, Mudança e Gêneros Textuais-Discursivos em Foco. Título da apresentação: “Estímulos distratores em design de pesquisa experimental off-line”.

A abordagem metodológica da experimentação complementa a metodologia de análise de *corpora*. A operacionalização da primeira será desenvolvida a seguir. Tendo em vista que o método científico é caracterizado, essencialmente, pela transparência, é necessário o fornecimento de cada passo dado, de cada procedimento feito e de cada decisão tomada na formulação e na estruturação dos testes de atitude linguística. Segundo Gonzalez-Marquez, Becker e Cutting (2007, p. 59),

what sets science apart is its dependence on intersubjective verification, the possibility that knowledge can be empirically tested by different researchers. The Scientific Method, in general terms, is an intellectual framework geared at generating the most reliable findings possible as well as at facilitating their verification.

Assim, ao descrevermos o passo a passo dos procedimentos utilizados, estamos possibilitando que novas pesquisas verifiquem/testem os resultados e as descobertas empíricas e, então, possam modificar o que for necessário em estudos futuros com objetivos diversos. Além disso, temos acesso a detalhes de como as hipóteses foram testadas.

Tendo por base elementos essenciais da pesquisa psicolinguística, para a observação de impressões, de opiniões, de atitudes, de interpretações e de percepções mediadas por avaliações subjetivas de usuários da língua portuguesa (variedade brasileira) acerca da variação entre construções com verbo suporte DAR, foram formulados seis testes. Optamos por confeccionar *designs* de pesquisa experimental diferentes, mesmo que o objetivo principal, em todos eles, seja a observação da variação, pensando na possibilidade de algum fator externo ou interno, em algum deles, poder influenciar/enviesar as respostas e, assim, os outros experimentos poderiam reiterar ou não os resultados uns dos outros. E, então, é importante que os resultados (as respostas dos informantes) sejam trabalhados em função de uma análise quantitativa e qualitativa. Os testes de atitude linguística foram confeccionados tendo em vista os objetivos e hipóteses de pesquisa. Assim, atentamos para o tipo de teste, o método, a técnica, o perfil ideal dos participantes, o ambiente de aplicação, o aparato tecnológico e o procedimento a ser utilizado.

Em todos os testes, optamos pela técnica experimental de medida *offline* (método indireto, não cronométrico), de modo que não haja contabilização do tempo de resposta, pois, assim, o informante pode refletir conscientemente sobre a própria língua pelo tempo que achar necessário, bem como sobre as possibilidades mais prováveis de uso em função de sua experiência prévia com a frequência estatística de percepção do fenômeno nos ambientes em que circula. Desse modo, temos acesso aos resultados do processamento linguístico após sua ocorrência.

A técnica usada foi a de questionários, com apresentação de material/formulário. As sentenças-estímulos dos experimentos 1, 2, 3 e 4 são trechos selecionados com dados do acervo do jornal O Globo, já coletados para a formação do *corpus*. Já as ta-

refas experimentais 2.1 e 2.2 são desdobramentos do teste 2 (formato semelhante e sentenças-estímulos semelhantes) que contam também com sentenças que funcionam como distratores. Esses elementos distratores foram acrescentados nesses testes de modo que pudessemos verificar se a “não consciência” do fenômeno linguístico em foco exerceria alguma influência nas respostas.

Ademais, com relação ao tipo de teste, optamos, em todos eles, pelo de autoavaliação. Nesse sentido, a metodologia empregada é introspeccionista, uma vez que são utilizados “julgamentos metalinguísticos extraídos de uma pessoa em particular de modo a identificar a impressão de estranhamento ou de normalidade que determinados estímulos linguísticos podem provocar” (KENEDY, 2015, p. 144). Assim, o indivíduo apresenta a opção que mais lhe convém naquele contexto determinado, pensando em qual seria a opção que escolheria para usar na situação especificada, bem como o motivo dessa escolha (intuitiva). Além disso, também podemos avaliar o quanto o sujeito aceita ou rejeita determinada forma, quais construções ele reconhece, bem como podemos observar comentários relacionados à multifuncionalidade dessas expressões.

Previamente ao enunciado de cada experimento psicolinguístico, há um pequeno cabeçalho, igual em todos eles, com as seguintes sentenças: “*Gostaríamos de contar com sua participação **voluntária** nesse estudo que tem o propósito de mapear a nossa linguagem a partir das percepções de diferentes falantes. Atenção! **Não há** respostas certas ou erradas. **Siga sua intuição**. Queremos **sua opinião**.” Assim, deixamos evidente que a participação no teste é voluntária, ou seja, sem retorno monetário ou de outra natureza beneficente para o informante. O propósito da pesquisa é explicitado de maneira geral, sem termos técnicos que pudessem intimidar os respondentes ou influenciar nas suas respostas. Além disso, fizemos questão de frisar o fato de que não há respostas certas ou erradas; o que se deseja é apenas conhecer a opinião do indivíduo, de modo que ele responda seguindo sua intuição, sem precisar voltar atrás na primeira resposta que lhe ocorreu. Embora os testes de atitude linguística tenham esse nome, preferimos não fazer referência ao termo “teste”, pois essa palavra remete a um momento tenso de avaliação já previamente estabelecido no conhecimento enciclopédico dos indivíduos, que remonta às provas escolares com respostas certas e respostas erradas que devem ser corrigidas. Haveria, pois, nesse sentido, uma preocupação com o julgamento do outro acerca da própria opinião e dos próprios usos. Então, preferimos não utilizar esse conceito na estruturação dos formulários.*

Ao final de cada experimento psicolinguístico, há a pergunta “Autoriza-nos a contar com sua participação?” para o informante expor claramente sua autorização de participação (formulário de consentimento livre e esclarecido) na pesquisa, de modo que os pesquisadores possam contabilizar e analisar sua opinião emitida no decorrer das respostas do teste (ou não). As respostas são anônimas, não há, portanto, a solicitação de identificação com nome pessoal. Em seguida, há um espaço destinado a sugestões e/ou a críticas em relação à tarefa experimental, com o seguinte enunciado: “Se desejar, anote aqui sua sugestão e/ou crítica em relação a esta consulta.”. O preenchimento desse espaço é facultativo e, nele, poderá conter impressões e/ou reclama-

ções dos participantes, as quais podem ser utilizadas para reformulações futuras das pesquisas.

Depois, são solicitadas algumas informações, como idade, sexo/gênero, nível de escolaridade (já concluído) e se possui formação na área de Letras, para se traçar o perfil social dos participantes. Esses dados devem ser coletados e analisados quantitativamente e qualitativamente, para que possamos observar, atentamente, possíveis influências dessas informações sociais no uso de formas alternativas com sentidos aproximados (variação). Com relação ao sexo/gênero, é importante ressaltar que não estamos fazendo distinção entre os dois constructos. Além disso, tanto na indicação da informação sexo/gênero quanto na de nível de escolaridade, disponibilizamos a opção “outro”, que englobaria outras possibilidades não presentes nas já previstas. A pergunta acerca da formação em Letras foi empregada, pois serão descartadas as respostas daqueles que tiverem tal formação, por já possuírem um conhecimento bem sedimentado sobre fenômenos da língua, o que poderia enviesar os resultados da pesquisa. Contaremos apenas com participantes “ingênuos” (não especialistas), isto é, sem conhecimentos da área de linguística.

Por fim, há um agradecimento pela participação: *“Obrigada por colaborar com nossa pesquisa! Grupo de estudos PREDICAR (UFRJ)”*. Optamos por dispor o nome do grupo de estudos, bem como o nome da instituição ao final do teste e não no início para não intimidar o informante com um grau de formalidade grande ao especificar a UFRJ, instituição de grande renome, a palavra “estudos”, que remete a uma atenção/monitoração maior, bem como o nome do grupo: “Predicar”, cujo significado poderia não ser de conhecimento de todos. Evitamos, portanto, o uso de metalinguagem e demos preferência a modelos de testes com linguagem simples, objetiva e clara. A demanda de informações pessoais dos participantes, bem como o agradecimento final seguiu o seguinte modelo:

Preencha os seguintes dados sobre você:

IDADE: _____

SEXO/GÊNERO: () Masculino () Feminino () Outro

NÍVEL DE ESCOLARIDADE (já concluído):

() Fundamental () Médio () Superior () Especialização () Mestrado

() Doutorado

() Outro. Qual? _____

TEM FORMAÇÃO EM LETRAS? () Sim () Não

Obrigada por colaborar com nossa pesquisa!

Grupo de estudos PREDICAR (UFRJ)

A confecção e a aplicação dos testes podem ocorrer por meio virtual, através da plataforma online de formulários *Google Forms*. Então, será necessário que os participantes tenham acesso à internet. Por um lado, com relação ao ambiente de resposta, não será possível ter controle de possíveis fatores que possam distrair os participantes da tarefa no momento de sua realização, uma vez que cada pessoa realizará a pesquisa em um ambiente diferente. Porém, por outro lado, também não haverá interferência de abordagem, pois não há uma pessoa específica que entrará em contato diretamente com os indivíduos, uma vez que haverá apenas o contato direto com a tela do computador.

Os sujeitos que respondem os testes podem ter perfil diverso e ser de ambientes variados. É importante buscarmos ter a mesma quantidade de pessoas para responder cada um dos experimentos, partindo da quantidade mínima de dez participantes por tarefa experimental (totalizando 60 informantes) e, assim, mantendo um equilíbrio na amostra. Sujeitos que tiverem formação em Letras podem ser excluídos, por já possuírem conhecimento, ainda que mínimo, sobre a temática. No que diz respeito à distribuição dos participantes, pode ser utilizada a abordagem “*between subjects*”,¹³ na qual cada participante é exposto a apenas uma condição experimental, isto é, cada pessoa irá responder a somente um modelo de teste.

Com esses dados, tencionamos observar se há tendências predominantes de uso e de escolhas (intuitivas) de formas linguísticas em função da região e de outras informações sociais (idade, sexo/gênero, nível de escolaridade e se possui formação em Letras) e textuais (gênero textual); ou seja, se há indícios de variação por influência de fatores extralinguísticos. Ademais, objetivamos investigar se usuários da língua percebem (ou não) divergências funcionais entre as diferentes perífrases verbo-nominais ou entre estas e predicadores verbais simples (cognatos ou não); se percebem uma forma convivendo ou competindo com outra e também se tendem a associar as construções em jogo às noções de aspectualidade, modalidade e intersubjetividade.

Acreditamos que, com relação à análise das variáveis linguísticas, os resultados apontarão para a reiteração dos valores predominantes (aspectualidade, modalidade e intersubjetividade) indicados pelas construções complexas (cf. resultados presentes em TRAVASSOS, 2019), principalmente, com a marcação de modalidade. Esses valores, majoritariamente, não são alcançados pelas construções verbais simples somente. Portanto, nesse caso, haveria indícios de variação por competição, com as construções com verbo suporte ganhando mais espaço para melhor indicação desses valores específicos.

Baseamo-nos na hipótese de que há variação por convivência e por competição não só entre predicadores complexos entre si, mas também entre estes e predicadores simples. Com relação ao tipo de predicador verbo-nominal em função do X-afixo, acreditamos que aquele formado pelo esquema [DAR (uma) X-adela]_{predicador complexo} será percebido e avaliado como diferente dos demais, uma vez que, intuitivamente, como usuários da língua, acreditamos que eles sejam pouco frequentes no português do Brasil atual.

13 “entre participantes ou interparticipantes” (tradução nossa).

No que se refere à análise das variáveis extralinguísticas, hipotetizamos que não haverá qualquer tipo de influência determinante das seguintes informações relacionadas ao informante: nível de escolaridade, sexo/gênero e idade. A informação da formação em Letras serve à finalidade de excluir as respostas dos participantes que tiverem essa formação, por poder enviesar os resultados, uma vez que podem possuir conhecimento consciente, já estabelecido/sistematizado sobre o fenômeno em estudo. Por fim, com relação ao gênero textual, acreditamos que haverá mais dados no gênero notícia, por ser o gênero mais comum do domínio jornalístico.

O teste 1 foi pensado de modo que pudéssemos verificar se usuários da língua associariam formas diferentes a um mesmo significado (semântico, discursivo, pragmático, cognitivo, social) no mesmo contexto, isto é, se há variação por convivência, situação em que formas seriam escolhidas como semelhantes, em harmonia, sem a percepção de quaisquer diferenças funcionais entre elas ou por competição, caso em que uma ou mais opção(ões) ganharia(m) destaque na escolha intuitiva pelo uso, uma vez que indicaria(m) melhor um determinado valor.

Nesse experimento psicolinguístico, estamos nos referindo a formas diferentes àquelas construções com verbo suporte *DAR* com a mesma base X do elemento não verbal, mas com mais de um sufixo. Com base nesse critério, selecionamos, no *corpus* dados que funcionariam como sentenças-estímulos que possuísem instanciações de construções com a mesma base, mas com sufixos diferentes. Desse modo, constituímos quatro grupos formados, cada um, por dados que, no contexto original, eram formados por perífrases verbo-nominais com configurações formais diferentes. Então, organizamos os quatro grupos dessa forma: (i) “dar uma escapada”, “dar uma escapadinha” e “dar uma escapadela”; (ii) “dar uma olhada”, “dar uma olhadinha” e “dar uma olhadela”; (iii) “dar uma espiada”, “dar uma espiadinha” e “dar uma espiadela”; e (iv) “dar risada”, “dar uma risadinha” e “dar um risinho”, resultando em um total de doze sentenças-estímulos. Assim, para esse experimento, optamos por não utilizar estímulos distratores, de forma que o informante tivesse consciência do fenômeno em foco e, desse modo, pudesse refletir sobre ele. Ademais, preocupamo-nos com o tamanho do teste, o qual teria sua extensão bem maior se tivesse distratores, o que poderia comprometer a atenção do indivíduo, ao deixá-lo “fatigado” com a tarefa.

O comando do teste era o seguinte: “Marque, com X, a resposta que se relaciona, com maior precisão de significado, a cada situação. Se achar que mais de uma alternativa é possível, assinale mais de uma.”. O questionário é, em parte, fechado, pois apresenta opções (a, b, c e d) no modelo de múltipla-escolha para preenchimento de lacuna em cada uma das doze situações. Caso o informante achasse que nenhuma das alternativas seria adequada ou, ainda, se encontrasse outra opção não disponível previamente, poderia escrever em qual resposta pensou no espaço destinado a essa possibilidade (“Outra opção? Qual?”).

Em seguida, no caso de a pessoa ter marcado mais de uma opção, teria de indicar, na escala de parentesco de significado (esquema de categorização/diferenciação no modelo de escala psicométrica Likert – com base em LIKERT, 1932), o quão parecidas são as construções selecionadas (“totalmente”, “muito”, “mais ou menos”, “pouco” ou

“nada”). Tal escala foi pensada de modo que houvesse uma simetria contrastiva em lados opostos (“totalmente” x “nada”; “muito” x “pouco” e “mais ou menos” no centro). Além disso, a paleta de cores gradientes reforça o grau indicado, ou seja, “totalmente” (preto), “nada” (branco), “muito” (cinza escuro), “pouco” (cinza claro) e “mais ou menos” (cinza médio). Dessa maneira, podemos captar, em um *continuum*, a intensidade da opinião, da avaliação subjetiva e do comportamento do respondente, levando em conta as nuances possíveis entre os extremos. O ponto central da escala indica neutralidade ou indecisão.

Acreditamos que os informantes irão indicar, predominantemente, as duas primeiras alternativas (com os sufixos –ada e –adinha) como “totalmente parecidas” ou “muito parecidas” e as duas últimas opções (com os sufixos –adela e –adelazinha) como “mais ou menos parecidas” em relação às outras, porque acreditamos, com base na introspecção/intuição e na experiência de usuário da língua (e de pesquisador), que estas últimas opções são menos frequentes de modo geral. Abaixo da escala de parentesco de significado, disponibilizamos um espaço para o informante justificar, se possível, a(s) escolha(s) intuitiva(s) feita(s). Assim, embora o questionário apresentasse uma parte fechada, também havia essa parte aberta, em que era possível apresentar a opinião. Seguem os primeiros comandos do teste 1 (primeira sentença-estímulo) a título de ilustração:

1) Os motivos geralmente envolvem alguma necessidade de mentir ou enganar alguém, seja para matar o trabalho, desmarcar um encontro ou _____ da vigilância do marido ou namorado.

a) () dar uma escapada

b) () dar uma escapadinha

c) () dar uma escapadela

d) () dar uma escapadelazinha

Outra opção? Qual? _____

Se marcou mais de uma alternativa, indique, na tabela abaixo, como se apresentam (DE ACORDO COM O SIGNIFICADO). Basta usar as letras que correspondem às respostas.

TOTALMENTE parecidas	MUITO parecidas	MAIS OU MENOS parecidas	POUCO parecidas	NADA parecidas

Justificativa: _____

Já o enunciado do teste 2 era o seguinte: “Substitua a forma em negrito por outra(s) que exprima(m) sentido equivalente/parecido. Se acredita que não há forma com sentido equivalente/parecido, marque a opção correspondente.”. Assim, deixamos em aberto para que a pessoa completasse com a(s) forma(s) que achasse que possuía(m) o mesmo significado (ou parecido) da construção com verbo suporte em negrito. Portanto, esse questionário é do tipo aberto, pois não há opções fechadas/obrigatórias, o informante pode apresentar livremente sua opinião. Além disso, embora não haja espaço para justificativa da escolha intuitiva logo embaixo de cada situação, há um espaço para “Comentário(s)/Justificativa(s)” ao final do teste caso o informante queira se expressar melhor sobre as opções feitas nos diversos contextos. Ademais, optamos por pôr, no experimento 2, sentenças-estímulos diferentes daquelas utilizadas no teste 1, a fim de observar a maior variedade de dados possível. A seguir, para ilustrar, há a primeira situação do teste 2:

Situação	Forma(s) com sentido equivalente
1) Antes de fechar os olhos, dou uma olhadela nas primeiras páginas do livro de Sá Pereira.	1)
	Não há ()

Na tarefa experimental 2, também buscamos analisar o ponto de vista de informantes/usuários da língua sobre a possibilidade de haver ou não a alternância de estruturas com perífrase verbo-nominal. No entanto, nesse teste, na seleção das sentenças-estímulos, diferentemente do teste 1, consideramos como formas diferentes aquelas construções com verbo suporte, cuja base X do elemento não verbal era diferente. Com esse pensamento em mente, buscamos, no *corpus*, sentenças com perífrases verbo-nominais com base X diferentes, mas que, a nosso ver, revelavam valores alinhados. Queríamos, portanto, verificar se a avaliação subjetiva dos respondentes estava de acordo com a nossa. Acreditamos que os dados apontarão para um resultado positivo com relação à hipótese postulada, ou seja, cremos que os participantes do experimento também irão reconhecer a variação (principalmente, a variação estável ou variação por convivência) como um fato linguístico ligado ao fenômeno em estudo, com o reconhecimento de que estruturas diferentes podem representar o mesmo valor, estando no mesmo contexto semântico, discurso, pragmático, cognitivo, social.

Então, selecionamos usos de construções que formaram quatro grupos de formas: (i) “dar uma dormidinha” e “dar uma cochiladinha”; (ii) “dar uma olhadela” e “dar uma espiadela”; (iii) “dar uma saidinha”, “dar uma fugidinha”, “dar uma voltinha”, “dar uma esticadinha” e “dar uma escapadinha”; e (iv) “dar um empurrãozinho”, “dar uma forcinha”, “dar uma mãozinha” e “dar uma ajudinha”. Acreditamos que os dois primeiros grupos de formas, nos seus contextos específicos, estão a serviço, predominantemente, da indicação de aspecto não durativo, com significados, respectivamente, de “dormir brevemente” e “olhar rapidamente/superficialmente”. Ademais, associamos os dois últimos grupos ao valor modal, com significados, respectivamente e predominantemente, de “passar” e “ajudar”, com ênfase intersubjetiva. Sendo assim, imaginamos que os respondentes associarão a outras construções verbais complexas (ou simples) sentidos semelhantes a esses.

A ordem com que as instanciações dessas construções foram dispostas no teste foi aleatória/misturada. Desse modo, formas com sentidos aproximados (de acordo com nossa percepção, aquelas que pertenciam a cada um dos quatro grupos) não ficaram imediatamente antes ou depois uma da outra. Assim, acreditamos evitar que o informante utilize a forma anterior como resposta para a sentença posterior. Optamos por não utilizar estímulos distratores nesse teste, de modo que o informante tenha consciência do fenômeno linguístico em jogo.

Os testes seguintes são desdobramentos do teste 2, apresentado anteriormente, e se diferenciam deste, na medida que apresentam estímulos distratores. Caso tivéssemos acrescentado todos os distratores somente no teste 2, ele ficaria com uma extensão muito grande, o que poderia comprometer a atenção e a disposição dos participantes ao respondê-lo. Sendo assim, decidimos desmembrá-lo em dois, utilizando, aproximadamente, a metade (sete) das sentenças-estímulos do teste 2 em um (2.1) e, aproximadamente, a outra metade (seis) em outro (2.2).

Então, o teste 2.1 contemplou instanciações das seguintes construções, agrupadas, a seguir, de acordo com o valor/sentido predominante (de acordo com a nossa intuição, que buscamos confirmar com a avaliação subjetiva dos informantes): (i) “dar uma saidinha”, “dar uma fugidinha”, “dar uma voltinha” (valor modal e sentido de “passear”); (ii) “dar uma olhadela” (valor aspectual não durativo e sentido de “olhar brevemente”); (iii) “dar uma dormidinha” (valor aspectual não durativo e sentido de “dormir brevemente”); e (iv) “dar um empurrãozinho” e “dar uma forcinha” (valor modal e sentido de “ajudar”). Já o teste 2.2 englobou as seguintes: (i) “dar uma mãozinha”, “dar uma ajudinha” (valor modal e sentido de “ajudar”); (ii) “dar uma esticadinha” e “dar uma escapadinha” (valor modal e sentido de “passear”); (iii) “dar uma espiadela” (valor aspectual não durativo e sentido de “olhar brevemente”); e (iv) “dar uma cochiladinha” (valor aspectual não durativo e sentido de “dormir brevemente”).

Mantivemos exatamente a mesma estrutura de questionário aberto, com o mesmo cabeçalho, enunciado, espaço para “Comentário(s)/Justificativa(s)”, espaço para sugestão e/ou crítica, solicitação das mesmas informações sociais e encerramento do teste 2. Os estímulos distratores são provenientes do *Google* e possuem natureza linguística (pertencente ao mesmo nível linguístico do fenômeno na tarefa-alvo): metade (quatro) focaliza predicador verbal simples e a outra metade (quatro) enfatiza adjetivo adverbializado, totalizando oito estímulos distratores em cada experimento. Assim, essa quantidade de distratores foi selecionada para cada tarefa experimental, tendo em vista o objetivo de distrair os participantes, sem perder de vista o quão “cansativo” poderia ficar. A nosso ver, ter uma quantidade um pouco maior (uma sentença, no teste 2.1, ou duas, no teste 2.2) do que a quantidade de sentenças-alvos/estímulos seria o ideal para alcançarmos um equilíbrio dos prós e contras relacionados à quantidade/qualidade da pesquisa.

O teste 2.1 apresentou um total de quinze situações/contextos, enquanto o teste 2.2 apresentou catorze. Os distratores foram usados, de modo que os participantes não tivessem consciência de qual objeto de estudo linguístico, marcado em negrito em cada situação, realmente estava sob estudo (se predicador verbal simples, se predica-

dor verbal complexo, se adjetivo adverbializado). Assim, seria mais difícil para o participante identificar padrões de recorrência e, então, conseguimos evitar respostas automáticas/viciadas. A ordem de disposição das sentenças na pesquisa foi randomizada, misturando os três fenômenos em jogo.

Desse modo, objetivamos verificar se haveria diferença nas respostas em função da consciência ou não do fenômeno linguístico em questão. Hipotetizamos que, no teste 2 (sem distratores), haja mais respostas com predicadores complexos, por influência das situações do entorno (cada um dos contextos da pesquisa de opinião, tendo a construção com verbo suporte em destaque: -negrito e sublinhada-), enquanto, nos testes 2.1 e 2.2 (com distratores), haja respostas variadas, contemplando não só predicador complexo, mas também predicador simples em quantidade maior do que a esperada no teste 2. A seguir, podemos observar as primeiras sentenças dos testes 2.1 e 2.2, respectivamente, como exemplo de suas estruturas:

Situação	Forma(s) com sentido equivalente
1) “Falem baixo, por favor”, João Gilberto morreu. O “maior artista de todos” deixou-nos no sábado e as influências são confessadas por muitos.	1)
	Não há ()
2) Antes de fechar os olhos, dou uma olhadela nas primeiras páginas do livro de Sá Pereira.	2)
	Não há ()
3) O ex-ministro José Dirceu está na capa da revista Veja desta semana, que também destaca o julgamento do mensalão. A reportagem revela que ele chegou a pensar em fugir do país.	3)
	Não há ()
Situação	Forma(s) com sentido equivalente
1) Fique calmo e responda tudo o que te perguntarem. Em seguida, você será liberado e basta ir na direção indicada pelas placas até chegar à área de retirada de bagagens. Procure em qual esteira foram colocadas as malas do seu voo e aguarde a sua.	1)
	Não há ()
2) Enquanto Cláudia dá entrevistas, sua mãe remexe antigas fotografias da filha. Aos 8 anos, aos 13, na hora do embarque, aos 15 anos em traje de debutante e as várias fotos já como profissional. Está feliz. E apesar de continuar morando em Campinas, quando a filha está no sufoco, corre para dar uma mãozinha.	2)
	Não há ()
3) É possível que, em algum momento teve a tentação de espiar o WhatsApp seu cônjuge ou de seus filhos, mas você sabia que se você fizer isso, você pode estar cometendo um crime?	3)
	Não há ()
4) Neste fim de semana, quando você for visitar o Barra Shopping, dê uma esticadinha até Santa Monica e conheça o outro grande sucesso da Barra. Um bairro planejado para o novo estilo de viver que surge na cidade.	4)
	Não há ()

5) "Algumas pessoas tímidas, retraídas, podem não se sentir confortáveis em uma situação social, não ser boas oradoras, falar baixo demais". Muitas pessoas que falam baixo ou gostam de resmungar, psicologicamente, não são capazes de projetar suas vozes alto o suficiente.	5) Não há ()
---	------------------

Ao observarmos os textos completos (do *corpus*) dos quais retiramos trechos com dados de construções com verbo suporte, verificamos que, nas sentenças originais, por vezes, havia alternância entre predicadores verbais simples e predicadores verbais complexos no mesmo contexto. Assim, para a construção do teste psicolinguístico 3, selecionamos (dos textos originais) seis trechos, nos quais havia as seguintes construções simples e complexas no mesmo contexto: (i) "piscar" e "dar uma piscadela"; (ii) "dar uma arrumadinha" e "arrumar"; (iii) "dar uma pedalada" e "pedalar"; (iv) "engrossar" e "dar uma engrossada"; (v) "dar uma lavadinha" e "lavar"; e (vi) "dar uma saidinha" e "sair". Os usos selecionados para esse teste foram diferentes daqueles que formaram as sentenças-estímulos dos testes anteriores (1, 2, 2.1 e 2.2). Essa escolha foi feita para que pudéssemos ampliar nosso escopo de análise da variação baseada em usos concretos diversos da língua.

Na confecção da tarefa experimental, no lugar dessas expressões, deixamos disponíveis quatro opções (a, b, c e/ou d) para que o informante marcasse a(s) forma(s) que acha possível/eis nesse contexto, em função do sentido que apresenta(m). O respondente poderia marcar mais de uma opção ou, ainda, todas, caso achasse possível, como o comando do teste prevê: "Marque, com X, a resposta que se relaciona, com maior precisão de significado, a cada situação. Se achar que mais de uma alternativa é possível, assinale mais de uma.". Três das opções consistiam em predicadores complexos com sufixos diferentes (-ada, -adinha e -adela) e a quarta opção correspondia a um predicador verbal simples cognato. Por exemplo, na primeira situação do experimento, as opções para cada "lacuna" são "(a) dar uma piscada (b) dar uma piscadinha (c) dar uma piscadela (d) piscar", como podemos observar a seguir:

1) Além de terem que ficar pelo menos 24 horas sem um cochilo, os quatro participantes não vão poder nem (a) dar uma piscada (b) dar uma piscadinha (c) dar uma piscadela (d) piscar. Aquele que conseguir ficar mais tempo sem (a) dar uma piscada (b) dar uma piscadinha (c) dar uma piscadela (d) piscar, vence a prova.

Outra opção? Qual? _____

Justificativa (se possível, diga o motivo que o levou a marcar a(s) opção/opções que escolheu OU diga por que não marcou as outras): _____

Algumas sentenças sofreram pequenas adaptações, de modo que pudessem se adequar às possíveis escolhas intuitivas do respondente. Como no segundo contexto do teste, por exemplo: “(...) a) dar uma arrumada b) dar uma arrumadinha c) dar uma arrumadela d) arrumar **(n)a** casa sem música é tipo o Ôh (...)”, em que colocamos “(n) a”, prevendo ou não o uso da preposição no complemento em função da marcação de predicador simples ou complexo. O questionário pode ser considerado do tipo aberto, pois, embora apresente opções de múltipla-escolha, não se restringe a elas, nem obriga o participante a escolher uma das alternativas previstas, uma vez que apresenta espaço para que explicita outra opção que tenha pensado e que não esteja contemplada anteriormente. Ademais, há também um espaço para justificar sua(s) escolha(s) intuitiva(s), caso queira apresentar sua opinião. Optamos por não apresentar estímulos distratores nesse teste para que tivéssemos acesso à reflexão explícita do usuário da língua sobre o uso da construção em foco.

O objetivo do teste 3 é verificar se usuários da língua sinalizam a possibilidade de marcação de formas diferentes para funcionalidades alinhadas, focalizando, principalmente, a variação entre esses dois tipos de predicadores (simples e complexos), como aparecem nos contextos originais, mas sem desconsiderar a variação só entre predicadores complexos ou só entre predicadores simples entre si. Desse modo, poderemos comparar o uso com a avaliação subjetiva e, com as justificativas dos informantes, poderemos observar como se deu a reflexão sobre o próprio uso e sobre a própria língua. Além disso, tencionamos verificar se os respondentes associam mais fortemente as construções com verbo suporte a valores (como modalidade, (inter) subjetividade, aspecto não durativo) e efeitos de sentido específicos, não alcançados com a forma simples.

Acreditamos que os resultados serão compatíveis com a hipótese de que usuários do português do Brasil verão semelhanças funcionais entre as quatro opções fornecidas em cada “lacuna”, no entanto, considerando, especialmente, as perífrases verbo-nominais como “equivalentes” tanto em termos semânticos, quanto discursivos e pragmáticos na indicação de valores diversos. Sendo assim, tal percepção empírica contrariaria o Princípio da Não Sinonímia (GOLDBERG, 1995, p. 67), o qual prevê que construções com formas diferentes ou serão distintas no nível da semântica ou no da pragmática.

Para a construção do teste 4, utilizamos as mesmas situações do teste 3, porém, mantivemos as construções verbais simples e complexas tal como no contexto original (sem as alternativas, como no teste 3) e as colocamos em destaque (negrito e sublinhadas) para que os informantes pudessem expor sua opinião acerca do motivo pelo qual foram usadas formas diferentes em cada situação. Em seguida, perguntamos se o respondente percebia alguma diferença de sentido entre essas formas e, se sim, para dizer qual. Trata-se, portanto, de questionário do tipo aberto com respostas discursivas curtas. Devido à demanda de escrita que esse tarefa exige, optamos por não utilizar estímulos distratores, de modo que tivéssemos o fenômeno em estudo em foco na reflexão e atenção dos participantes, sem que houvesse um esforço cognitivo extra com os distratores. Esse experimento foi feito com o objetivo de analisar a reação, o

juízo inconsciente e a opinião de usuários do português do Brasil com relação à variação entre predicadores verbais simples e complexos em usos já dados. A hipótese para esse teste é a de que os informantes não verão distinções de sentido entre as formas. A seguir, há a primeira situação do teste 4, a título de ilustração:

Leia, com atenção, cada uma das situações a seguir e responda às perguntas.

1) Além de terem que ficar pelo menos 24 horas sem um cochilo, os quatro participantes não vão poder nem ***piscar***. Aquele que conseguir ficar mais tempo sem ***dar uma piscadela*** sequer vence a prova.

a) Por que motivo você acha que foram usadas formas diferentes (em negrito)?

b) Há diferença de sentido entre essas formas? Se houver, diga qual.

Com as respostas dos informantes registradas, podemos proceder à etapa seguinte de análise quantitativa/estatística dos dados dos testes que permitirem essa abordagem e análise qualitativa das respostas de todos os experimentos, à luz das orientações teóricas, a fim de observar se as previsões de pesquisa se cumpriram, descobrir a natureza dos mecanismos cognitivos/psicológicos por trás das percepções e mapear os perfis de comportamento linguístico.

REFERÊNCIAS

- ABRAÇADO, J.; KENEDY, E. (orgs.). Nos meandros da experimentação. In: *Transitividade Traço a Traço*. Rio de Janeiro: Editora da UFF, p. 29-37. 2014.
- BOLINGER, D. Entailment and the Meaning of Structures. *Glossa* 2(2). p. 119-27, 1968.
- DRAGER, K. *Linguistic Variation, Identity Construction and Cognition*. Berlin: Language Science Press. 2015.
- DRAGER, K. *Experimental Research Methods in Sociolinguistics*. London: Bloomsbury. 2018.
- ESTEVES, G. A. T. *Construções com DAR + Sintagma Nominal: a gramaticalização*

- desse verbo e a alternância entre perífrases verbo-nominais e predicadores simples. Rio de Janeiro: Faculdade de Letras, UFRJ. Dissertação de Mestrado. 2008.
- FASOLD, R. *The Sociolinguistics of Society*. Vol. I. New York, USA: B. Blackwell. p. 147-179. 1987.
- GOLDBERG, A. E. 1995: *Constructions*. A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.
- GONZALEZ-MARQUEZ, M. *et al. Methods in Cognitive Linguistics*. Amsterdam, John Benjamins, 2006.
- GONZALEZ-MARQUEZ, M.; BECKER, R. B. e CUTTING, J. E. *An introduction to experimental methods for language researchers*. p. 53-86. 2007.
- KENEDY, E. Psicolinguística na Descrição Gramatical. In: Maia, M. *et al.* (org.) *Psicolinguística, psicolinguísticas: uma introdução*. São Paulo: Contexto, p. 143-155. 2015.
- LENT, R. *Cem bilhões de neurônios? Conceitos fundamentais de neurociência* (2. ed.). São Paulo, SP: Atheneu. 2010.
- LIKERT, R. A technique for the measurement of attitudes. *Archives of Psychology*. n. 140, p. 44-53, 1932.
- RODRIGUES, A.; ASSMAR, E. M. L.; JABLONSKI, B. *Psicologia Social*. 18. ed. Rio de Janeiro: Vozes, 2018.
- TRAVASSOS, F. P. *Variação e mudança construcional: um olhar funcional-cognitivo sobre usos de construções com verbo-suporte DAR no PB*. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2019.

CAPÍTULO 7

DISCURSO DE PARLAMENTARES: UM PERFIL METODOLÓGICO DE ANÁLISE DE *CORPUS* DE DOMÍNIO POLÍTICO

Fábio Rodrigo Gomes da Costa (*Universidade Federal do Rio de Janeiro*)

7.1 INTRODUÇÃO

Neste capítulo, apresentaremos um percurso metodológico de análise de *corpus* relativo ao domínio discursivo político. O objeto de estudo são as construções de predicação formadas com o verbo *trazer* inseridas no discurso de parlamentares da Câmara dos deputados federais. Nossa meta consiste em apresentar um caminho de estudo das expressões com verbo *trazer* como mecanismo enunciativo que visa a uma estratégia argumentativa. Para uma melhor compreensão do tema exposto neste capítulo, iniciaremos, na seção 1, com algumas reflexões acerca do discurso político. Na seção 2, trataremos da constituição do *corpus* no discurso político, conforme Charau-deau (2011). Na seção 3, traçaremos o caminho metodológico utilizado para tratar de

nosso objeto de pesquisa, em que um conjunto de variáveis relativas a atributos de forma e função/significação é arrolado na comparação de dados de predicação; e, nas considerações finais, uma breve reflexão em torno do papel enunciativo da construção de predicação com verbo *trazer*.

7.2 DISCURSO POLÍTICO: ALGUMAS REFLEXÕES

As pesquisas linguísticas sobre discurso político tendem a escolher como seu escopo a análise de discursos produzidos por políticos profissionais. Essa escolha decorre de uma compreensão que tem como base os atores políticos, considerando o âmbito do discurso e o contexto da comunicação.

Segundo Van Dijk (1997), embora se reconheça que a diversidade de atores políticos vai desde profissionais da área a cidadãos engajados, o discurso político deve realizar uma ação notadamente política (governar, legislar, protestar, votar etc.) e estar inserido em uma situação comunicativa bem definida (debates parlamentares, entrevistas na mídia, publicações nas redes sociais etc.). Desse modo, o discurso político limita-se ao discurso de profissionais inseridos em um contexto institucionalizado, deixando de fora, por exemplo, as discussões sobre o tema apresentadas nos veículos de informação ou proferidas pelo cidadão comum.

Outros autores admitem que qualquer discurso ou qualquer ação humana pode ser potencialmente política (PELINKA, 2007). A partir daí, é possível discutir se o discurso político está restrito somente ao contexto institucional, como os debates parlamentares e discursos proferidos em conferências partidárias, ou se abrange os conflitos cotidianos, como entre homens e mulheres, empresários e trabalhadores, policiais e jovens negros etc.

Adotamos em nossa pesquisa o ponto de vista de Charaudeau (2011a; 2011b), segundo o qual qualquer enunciado pode ter um sentido político a depender da situação de comunicação. O autor afirma que o discurso político é fabricado na própria interação comunicativa e pode ser entendido como sistema de pensamento (ou o resultado de uma atividade discursiva cuja posição política norteia os diferentes posicionamentos e opiniões), como ato de comunicação (ou uma construção discursiva por meio de procedimentos retóricos em uma cena de comunicação política), como comentário (ou uma discussão que se dá fora do campo de ação política). Dessa maneira, o discurso político é manifestado tanto intragoverno, no qual a ação política tem o propósito de poder governar, quanto extragoverno, no qual a palavra política não tem poder de decisão.

Os diferentes lugares de fabricação do discurso político não estão separados um do outro, pois, assim como todo discurso social, circula no interior dos grupos sociais e também atravessa outros grupos de modo a se reconstruir de forma diferente. O discurso político está ligado à organização da vida social tanto como governo quanto como discussão. É no discurso político que o sujeito se engaja e se posiciona de modo a exercer influência no outro.

A partir das reflexões em torno do discurso político, abordaremos na seção seguinte a constituição do *corpus* no discurso político.

7.3 A CONSTITUIÇÃO DO *CORPUS* NO DISCURSO POLÍTICO

Quando falamos de discurso, não estamos falando de texto. Conforme Charaudeau (2011a, p. 06), “o discurso é um percurso de significância que se acha inscrito num texto, e que depende de suas condições de produção e dos locutores que o produzem e o interpretam”. Da mesma maneira que o discurso perpassa diferentes textos, pode haver diferentes discursos em um mesmo texto. O discurso depende do texto para poder ganhar significado. Dessa forma, o *corpus* para a análise de uma pesquisa linguística é formado por textos e não por discursos.

O sentido do discurso é resultado de um conjunto de fatores contextuais, como os participantes da cena enunciativa, o gênero textual, a tipologia textual etc. Podemos citar como exemplo o enunciado “Venho trazer um convite”. Fora de uma situação comunicativa, o enunciado pode indicar a presença de um actante (aquele que fala) que se dirige a seu interlocutor com o objetivo de entregar-lhe um convite. Nesse caso, podemos imaginar o convite como um elemento concreto e, assim, o verbo “trazer” expressa a ação realizada pelo actante: transferir o objeto (convite) para o outro com quem ele fala.

No entanto, se considerarmos o enunciado “Venho trazer um convite” em seu contexto de produção, um parlamentar, que, ao fazer uso de sua fala em plenário, comunica a seus pares o desejo de que todos participem de um ato político. Nesse caso, temos a palavra “convite” não como um objeto e sim como o próprio ato de “convidar”. O sujeito da ação encarrega-se de transmitir a informação aos demais sem demonstrar ser o autor da ação. Ao lançar mão desse recurso de predicação (em lugar de um verbo pleno “convidar”), mobiliza o objetivo de diminuir/mitigar o seu papel quanto à responsabilidade pela ação que foi encaminhada. Portanto, para entendermos a funcionalidade da expressão “trazer um convite”, é preciso recorrer ao contexto interacional, a fim de observar a intencionalidade discursiva por detrás dos enunciados.

Em se tratando de discurso político, um aspecto que deve ser levado em consideração é a noção de contexto situacional, que consiste nos pressupostos de posicionamentos interdiscursivos, bem como na natureza da identidade dos participantes da cena enunciativa e na finalidade da situação interacional. A esse conjunto de fatores, dá-se o nome de “contrato de comunicação”, cuja finalidade é instruir discursivamente o sujeito falante a como proceder em um determinado ato comunicativo. Desse modo, além do contexto estritamente linguístico e do entorno textual, é necessário levar em conta o contexto situacional para formar o *corpus* de análise do discurso político.

O *corpus* de discurso político constitui, portanto, um agrupamento que, segundo Charaudeau (2011a), pode ser de três tipos: *corpus* segundo o paratexto, *corpus* segundo a situação e *corpus* segundo o interdiscurso. O primeiro tem como constituinte as

palavras, os enunciados e os modos de enunciação; o segundo, os locutores, a identidade e a finalidade; e o terceiro, a ideologia e os saberes de conhecimento e de crença. Os três tipos de *corpus*, conforme o autor, são determinados respectivamente pelas seguintes problemáticas (ou conjunto de proposições hipotéticas):

Problemática cognitiva e categorizante: neste caso, é preciso observar a combinação de mecanismos discursivos no interior de uma produção discursiva, bem como descrever o seu funcionamento dentro do discurso. O *corpus*, nesse caso, não precisa ser estabelecido a partir de uma única situação comunicativa e pode limitar-se aos contextos linguísticos nos quais é possível identificar os mecanismos em análise.

Problemática comunicativa e descritiva: a análise parte de determinado comportamento recorrente de um indivíduo em seus atos de troca social. Nesse caso, pode-se formular a hipótese de que tal manifestação está submetida a uma condição de realização. A partir disso, são descritas as características discursivas que se relacionam a tal comportamento, como a identidade dos locutores, a finalidade comunicativa etc. Desse modo, o sujeito da comunicação é definido por sua identidade psicológica e social, por seu comportamento submetido na interação discursiva e por suas intenções diante do seu interlocutor.

Problemática representacional e interpretativa: a análise parte das hipóteses de representações sociodiscursivas que são supostamente dominantes em determinado período histórico e que caracterizam um grupo social. O *corpus*, nesse caso, varia desde aqueles textos com valor emblemático de discurso dominante até aqueles que representam valores ou ideologias.

A título de ilustrar os rumos de pesquisa a que tais problemáticas podem levar o cientista/linguista, passamos a delinear, brevemente, algumas rotas a explorar em transcrições do discurso de parlamentares. Com relação à problemática cognitiva e categorizante, é possível constituir uma coleção de discursos com base em parâmetros como: período de tempo, temática, siglas de partidos dos parlamentares e inclinações à esquerda ou à direita (em função da reunião dos vínculos a partido político). A partir desse material, é possível analisar o conjunto de predicadores (simples e complexos) com que são organizados os discursos, procurando capturar a relação entre predicadores e conceptualização do mundo.

Quanto à problemática comunicativa e descritiva, uma outra rota é organizar coleção de discursos por parlamentares, com a finalidade de traçar o perfil de predicação e conceptualização de mundo que se revela nos subconjuntos de discursos, na associação de cada subconjunto com o locutor/parlamentar. Esse recurso permite, via comparação, traçar as características discursivas dos interactantes, bem como a intenção comunicativa.

Quanto à problemática representacional e interpretativa, é possível traçar um perfil dos participantes do discurso de modo a identificar sua concepção ideológica, seu grupo social, sua visão de mundo e observar a que valores ou ideologias o conjunto de predicadores estão associados. Naturalmente, outros temas linguísticos, além dos relacionados à predicação, poderão ser explorados nessas rotas.

Feita a ilustração dos diferentes rumos de uma pesquisa de *corpus* político, devemos ressaltar que a problemática cognitiva e categorizante é a que interessa à nossa pesquisa. Na seção seguinte, iremos apresentar as rotas trilhadas para a análise dos dados.

7.4 AMOSTRA DE PESQUISA: ANÁLISE DE *CORPUS* POLÍTICO

Conforme Charaudeau e Mainguenu (2002, p. 148), “Um *corpus* é uma coleção de dados de linguagem que são selecionados e organizados de acordo com critérios linguísticos explícitos para servir como amostra de linguagem”. Em nossa pesquisa, o *corpus* constitui um conjunto de enunciados que são tomados como o próprio objeto de análise e que servem de base para a descrição linguística.

Nesta seção, exploramos o perfil de pesquisa quali quantitativa de dados linguísticos em discurso político que delineamos com base numa coleção de discursos parlamentares. A amostra de dados que serve para ilustrar esse perfil foi organizada a propósito de uma investigação que consiste na análise de expressões predicantes formadas com o verbo suporte *trazer*, presentes em proferimentos de parlamentares da Câmara dos deputados. O conjunto de dados para análise advém de transcrições oficiais de discursos proferidos na Câmara dos deputados federais no período de 2020 e 2021 as quais estão disponíveis no site eletrônico <https://www.camara.leg.br>. Trata-se de discursos orais/oralizados cuidadosamente preparados por políticos profissionais e apresentados em plenário aos demais deputados. Conforme o Portal da Câmara dos deputados, o plenário é assim definido:

O Plenário é o órgão máximo de deliberação da Casa. Nele, os representantes do povo, reunidos em sua totalidade, discutem e votam soberanamente as proposições em tramitação, no cumprimento da função constitucional conferida ao Poder Legislativo de elaboração do ordenamento jurídico e de fiscalização financeira e orçamentária (<https://www2.camara.leg.br/a-camara/conheca/como-funciona>).

No site do Portal da Câmara dos deputados, após clicar na opção “Atividade Legislativa” e depois em “Discursos e debates”, é possível realizar a coleta dos dados, conforme a ilustração a seguir:



Fonte: <https://www.camara.leg.br/>.

Em uma nova aba do site, após a escolha da data inicial e da data final do período para a coleta de dados, são listados em ordem decrescente todos os proferimentos dos deputados no período escolhido. É preciso clicar em cada um deles para visualizar o discurso e, assim, realizar a busca por expressões formadas com o verbo *trazer*.

A partir da observação dos dados, partimos da hipótese de que os predicadores verbais com o verbo *trazer* se encarregam de apresentar uma configuração discursiva que visa a abrandar o papel do enunciador como participante da cena enunciativa, pois tem como objetivo a preservação da face do sujeito enunciator. Segundo Brown e Levinson (1978, p. 61), a face é “a auto-imagem pública que todo membro deseja reivindicar por si mesmo”.

Em conformidade com Costa e Machado Vieira (2021), partimos do princípio de que a minimização ou atenuação da força enunciativa constitui uma estratégia argumentativa que opera com a intenção de buscar a aceitação do outro. Entendemos que o discurso político favorece o uso de tais predicadores verbais, pois compreende um espaço voltado para o exercício retórico e para a arte do convencimento. Desse modo, as hipóteses que norteiam a pesquisa são:

- As construções com o verbo suporte *trazer* associam-se à funcionalidade de atenuar a responsabilidade do enunciador quanto à ação proposta;
- A atenuação do discurso é um mecanismo do exercício retórico, o que está em proeminência no discurso político; e
- O contexto político é, então, espaço propício ao uso da construção de predicação formada com o verbo suporte *trazer* como mecanismo de atenuação/mitigação.

Conforme Charaudeau (2006), se o objetivo é investigar o uso da linguagem quanto às estratégias de persuasão no discurso político, é necessário considerar a condição de credibilidade, que, segundo o autor, está relacionada à imagem de si (*ethos*), que serve como suporte de identificação de valores comuns desejados. A partir disso, tentamos observar o uso da construção com verbo *trazer* como configuração discursiva com vistas a estratégias persuasivas.

O objeto da pesquisa são os predicadores complexos formados com o verbo suporte *trazer* seguido de elemento não verbal, que funcionam à semelhança de um predicador verbal simples. Por exemplo:

a) “Sr. Presidente, em segundo lugar, eu quero aqui **trazer uma reflexão importante** que está à margem deste debate, mas que está no centro da retomada do desenvolvimento do Brasil, que é sobre a questão energética.” (Câmara dos deputados, 22/04/2020)

b) “Sr. Presidente, Srs. Parlamentares, nós estamos vivendo, como eu disse no meu pronunciamento anterior, um caos no nosso País. O Deputado Enio Verri, que foi Líder da nossa bancada, **trouxe aqui a informação** sobre a inflação do mês de agosto. A inflação do mês de agosto foi de 0,87%, e o acumulado dos últimos 12 meses é de quase 10%. Contudo, precisamos, de fato, comparar isso com o custo da cesta básica.” (Câmara dos deputados, 09/09/2021)

O objetivo é mapear a configuração formal e funcional das expressões de predicação formadas com o verbo suporte *trazer*: os diferentes elementos não verbais atraídos à construção predicante e o papel do enunciador na configuração dos enunciados em que a construção tem lugar. A concepção teórico-metodológica é a perspectiva funcional-cognitiva, bem como os pressupostos da Gramática de Construções (TRAUGOTT; TROUSDALE, 2013), cujo princípio básico é de que a unidade básica da língua é a construção. Essa é entendida como pareamento de forma e função/significado interconectado em uma rede construcional.

A pesquisa consiste em, a partir de um conjunto de dados reunidos que permitam a observação quanto ao nível morfossintático, léxico, pragmático e discursivo, investigar a configuração dos predicadores complexos com o verbo *trazer*, bem como os níveis de composicionalidade, produtividade e esquematização construcionais que eles implicam. Conforme Traugott e Trousdale (2013), a composicionalidade está relacionada à transparência entre forma e significado; a produtividade está relacionada à frequência de uso; e a esquematicidade, à captura de padrões mais abstratos em uma série de construções específicas.

Para realizar a triagem do material reunido via coleção de discursos de parlamentares, o primeiro passo é a análise dado a dado que nos leve à categorização dos dados, a conhecer sua natureza em termos de forma e funcionalidade. São traçadas, então, as variáveis que julgamos relevantes para mapear os atributos formais e funcionais da construção analisada e dos enunciados em que ela opera. Então, operamos com variável que contém fatores e que se delinea com base em uma hipótese. Na pesquisa em questão, as variáveis delineadas até o momento são estas:

(i) Formas verbais acionadas (trazer, traz, trouxe, trago etc.):

Motivação, a partir de observação empírica preliminar: há enunciados em que o sujeito da ação expressa pelo predicador ou está relacionado à primeira (1) ou à tercei-

ra pessoa discursiva (2), via, respectivamente, marcação gramatical de primeira pessoa (singular) ou das demais.

Hipótese: sendo o acionamento de predicador complexo com *trazer* um recurso de atenuação discursiva, supomos que ele seja um expediente a servir principalmente a predicções com predicador complexo relativo à primeira pessoa do discurso. E predicções com outros usos de *trazer* sejam mais associadas a outras pessoas discursivas.

(1) “...hoje eu **trago uma planilha**, de 2015 até 2021, período em que exerço meu mandato, das minhas emendas individuais, emendas de bancada, da qual faço parte...” (Câmara dos deputados, 05/10/2021)

(2) “Diante desse quadro, ontem fomos surpreendidos com um relatório que **traz consigo muitas chagas do passado** (...)” (Câmara dos deputados, 05/10/2021)

(ii) Predicador do tipo verbo pleno, Predicador com verbo suporte e expressão cristalizada/não composicional e idiomática (chunk):

Motivação, a partir de observação empírica preliminar: podemos perceber, por meio dos dados coletados, que há predicadores formados com o verbo *trazer* mais cristalizados ou mais composicionais, como “trazer um fato” (3), em comparação a outros menos composicionais como “trazer de volta” (5), podendo ser mais repetidos e, então, sociocognitivamente entrincheirados (4) ou menos entrincheirados (3).

Hipótese: supomos que o uso de estruturas com o verbo suporte *trazer* seja mais acionado em vista de ser um recurso mais característico da estratégia da atenuação do discurso.

(3) “Ela **traz um fato que jamais nós deveríamos permitir que acontecesse na vida de qualquer ser humano**.” (Câmara dos deputados, 05/10/2021)

(4) “A emenda **traz duas sugestões muito importantes**.” (Câmara dos deputados, 13/10/2021)

(5) “Portanto, esta Comissão Geral denuncia o nome do responsável por **trazer de volta** o Brasil ao Mapa da Fome.” (Câmara dos deputados, 05/10/2021)

(iii) Estrutura de argumento em que o predicador se combina (intransitiva ou transitiva):

Motivação, a partir de observação empírica prévia: na triagem dos dados, deparamo-nos com dados em que o predicador complexo se liga/combina a uma estrutura sintática com valor intransitivo (6) ou transitivo (7).

Hipótese: supomos que seja predominante o uso de predicadores complexos com verbo suporte ligados ao procedimento discursivo de atenuação em estrutura transitiva, diferentemente do que normalmente se diz quanto a esse recurso. De todo modo, não se perde de vista a potencialidade de a associação desse recurso de predicação à elipse de detalhes, à inclinação a uma conceptualização do evento via perspectivação genérica (como nos exemplos (8) e (9)), sem muitos detalhes (conforme destacado em MACHADO VIEIRA, 2018), se sobrepor ao esperado, em razão da condição de atenuação em proeminência.

(6) “Eles **trouxeram as reivindicações**, elas foram atendidas, e agora cabe aos Parlamentares decidir.” (Câmara dos deputados, 13/10/2021)

(7) “A, B, C, D e F, entre outros, **traz uma proteção ao mal advogado** por criar dificuldades para que haja a busca e apreensão no escritório ou no local de trabalho deles.” (Câmara dos deputados, 13/10/2021)

(iv) Configuração formal do elemento não verbal (plural ou singular, acompanhado de determinantes e/ou modificadores):

Motivação, a partir de observação empírica inicial: a partir da observação dos dados coletados, percebemos que o elemento não verbal pode apresentar-se no singular (8), no plural (9) e/ou acompanhado de elementos determinantes (10) ou modificadores (11). Isto é, na literatura (MACHADO VIEIRA, 2018), tido como elemento a colaborar para desfazer a coesão que caracteriza um complexo verbo-nominal.

Hipótese: supomos que o predicador complexo seja mais acionado sem a presença de elementos adjuntos ou, quando muito, com a presença de determinantes e/ou modificadores mais específicos (intensificadores, pronomes possessivos/a reforçarem a condição do enunciador-sujeito) e marcação de plural no sentido de evitar detalhes da cena que é objeto da predicação.

(8) “O trabalho também **traz dignidade**.” (Câmara dos deputados, 05/10/2021)

(9) “Aqui quero **trazer esclarecimentos** e somar-me às palavras do Deputado Alex Manente.” (Câmara dos deputados, 05/10/2021)

(10) “...aos artesãos e artesãs do Rio de Janeiro eu queria **trazer o meu abraço e a minha satisfação de ver essa instituição fortalecida**.” (Câmara dos deputados, 13/10/2021)

(11) “Gostaria de agradecer a todos aqueles que participaram desta sessão. Gostaria de dizer que esta medida provisória **traz avanços significativos** para o nosso País (...)” (Câmara dos deputados, 16/09/2021)

(v) Natureza semântica do elemento não verbal (*dicendi* ou não):

Motivação, a partir de observação empírica preliminar: deparamo-nos com dados em que o elemento não verbal é formado por deverbal de verbo *dicendi* (12) ou não *dicendi* (13).

Hipótese: é possível prever que os deverbais de verbo *dicendi* são mais acionados ao slot do elemento não verbal em comparação aos deverbais de verbos não *dicendi*. Essa é uma condição de identificação de predicador complexo com verbo suporte.

(12) “Sr. Presidente, colegas Parlamentares, eu quero, em breves palavras, **trazer um comunicado importante.**” (Câmara dos deputados, 06/10/2021)

(13) “Estamos construindo esse texto atendendo praticamente todos os pedidos, para que, primeiro, **possamos trazer** nesse projeto **um alento** para a grávida.” (Câmara dos deputados 05/10/2021)

(vi) Inserção de elementos adverbiais/circunstanciais no predicador complexo:

Motivação, a partir de observação empírica prévia: há enunciados que permitem a inserção de elementos adverbiais de tempo, modo ou lugar, como em (14).

Hipótese: podemos prever a maior presença de elementos adverbiais de lugar que se inserem na construção formada com o verbo *trazer* como marca do lugar de fala/ interação do enunciador, justamente o que é a tônica que mobiliza atenuação discursiva.

(14) “(...) quero aqui parabenizá-lo por **trazer aqui um assunto que trata exatamente da fome e das mazelas que têm atingido o nosso País.**” (Câmara dos deputados, 05/10/2021)

(vii) A relação entre o papel de remissão textual do predicador verbal e o estatuto verbal:

Motivação, a partir de observação empírica preliminar: é possível observar que o predicador formado com o verbo *trazer* exerce a função de remissão textual ou contextual, conforme expresso em (15).

Hipótese: supomos que a presença de predicadores que exerçam a função de retomada de informações ditas anteriormente ou de localização do enunciador frente a um público colaborem para evidenciar o papel do predicador complexo como recurso de atenuação discursiva, assim como o papel do verbo como suporte.

(15) “Por isso, ainda que defendamos a permanência da empresa, nós não abriremos mão de defender o respeito ao meio ambiente, em especial às comunidades qui-

lombolas das localidades que eu citei. Eu **trago esta discussão** ao conhecimento de todos, da tribuna da Câmara dos Deputados, exigindo da empresa respeito às leis da Bahia, às leis ambientais.” (Câmara dos deputados, 15/09/2021)

(viii) Presença de auxiliares modalizadores:

Motivação, a partir de observação empírica prévia: deparamo-nos com dados em que o predicador complexo é auxiliado por verbos modalizadores, como “ir”, “poder”, “querer”, como em (16), em que vemos um movimento de projeção, de futuridade em jogo, ao mesmo tempo que a atitude do enunciador se torna proeminente em relação ao dito.

Hipótese: supomos haver um número considerável de estruturas com verbo auxiliar que modaliza a predicação, a fim de contribuir com o propósito enunciativo. Supomos que predicadores complexos com *trazer* sejam atraídos para contextos de modalização; sendo assim, há combinação de forças quando verbos modais operam sobre esses predicadores.

(16) “Sr. Presidente, colegas Parlamentares, **eu quero**, em breves palavras, **trazer um comunicado importante.**” (Câmara dos deputados, 06/10/2021)

(ix) Grupos semânticos:

Motivação, a partir da observação empírica: por meio da triagem dos dados, observamos que as construções formadas com o verbo *trazer* apresentam diferentes padrões semânticos, como resultado negativo (17), resultado positivo (18), esclarecimento (19), exteriorização de sentimento (20), aproximação (21), exteriorização de pensamento (22), notícia (23), localização na interação como porta-voz de um povo (24), encaminhamento de proposta (25), conversão emotiva (26), transferência física (27), mudança (28), responsabilização de outrem (29).

Hipótese: é possível prever que os diferentes padrões semânticos estejam relacionados com a tomada de opinião ou defesa de uma ideia. Supomos que atos de localização no espaço discursivo como porta-voz de um povo, de esclarecimento e de encaminhamento de proposta, por exemplo, estejam mais associados ao domínio sociopolítico em que se dá a atenuação via predicador complexo com *trazer*. Já contextos de transferência física farão emergir o papel de verbo pleno/predicador de *trazer*.

(17) “Se demoramos no tratamento, com certeza a doença avança, o que **traz sofrimento** para a pessoa, para a família (...)” (Câmara dos deputados, 07/10/2021)

(18) “Estou muito convicto de que, com esse projeto de lei, **estou trazendo um benefício** à nossa Nação num momento totalmente diferente daquele anterior.” (Câmara dos deputados, 05/10/2021)

(19) “Eu estou lhe dizendo que a sua questão de ordem é improcedente e eu a indefiro, porque eu **trouxe um esclarecimento**, não estou debatendo a matéria.” (Câmara dos deputados, 14/10/2021)

(20) “... eu quero **trazer aqui a minha indignação**, a minha insatisfação.” (Câmara dos deputados, 15/09/2021)

(21) “Pensar no futuro o que é? Permitir que os entes federados possam, através de processos específicos, **trazer gente nova**, gente com outra cabeça, que domine as tecnologias que já estão correndo o mundo inteiro...” (Câmara dos deputados, 15/09/2021)

(22) “... vou ter que **trazer hoje um assunto do Acre** para o debate nacional, porque percebo que é necessário.” (Câmara dos deputados, 15/09/2021)

(23) “E eu queria **trazer uma notícia** que nos preocupa...” (Câmara dos deputados, 11/03/2021)

(24) “Portanto, Sr. Presidente, eu **trago a solidariedade do povo do Maranhão...**” (Câmara dos deputados, 05/08/2020)

(25) “Portanto, eu **trago aqui um requerimento**, Presidente Rodrigo Maia...” (Câmara dos deputados, 09/06/2020)

(26) “Paulo Gustavo, com sua arte, era muito responsável em nos **trazer alegria**.” (Câmara dos deputados, 05/05/21)

(27) “... e autorizamos as empresas a comprarem e **trazerem vacina** para o Brasil.” (Câmara dos deputados, 06/04/2021)

(28) “... que se empenhou incansavelmente nessa grande solução que vai **trazer a reversão** de todo o atraso que a cidade de Magé sofreu ao longo do tempo.” (Câmara dos deputados, 17/03/2021)

(29) “Eu queria **trazer a responsabilidade** deste Parlamento, Sr. Presidente.” (Câmara dos deputados, 08/04/21)

(x) Movimento retórico (nível proposicional ou argumentativo):

Motivação, a partir da observação empírica: podemos perceber que as construções formadas pelo verbo *trazer* se inserem em dois movimentos do discurso: no nível proposicional, em que se apresenta a ideia central do proferimento (30); e, no nível argumentativo, em que se apresenta uma justificativa para a ideia defendida (31).

Hipótese: partimos da hipótese de que as expressões formadas com o verbo *trazer* estão mais associadas ao movimento retórico de introdução da ideia principal.

(30) “Aqui **quero trazer esclarecimentos** e somar-me às palavras do Deputado Alex Manente.” (Câmara dos deputados, 05/10/2021)

(31) “O grande problema é que essa exposição pode **trazer riscos**.” (Câmara dos deputados, 13/10/2021)

A amostra de variáveis apresentada anteriormente constitui o resultado a que chegamos na etapa inicial de categorização dos dados coletados. Na etapa seguinte, será possível atestar os níveis de composicionalidade, de produtividade e de esquematização construcional de dados de predicções envolvendo o verbo *trazer*, seja como verbo pleno seja como verbo instrumental integrado num complexo mais ou menos não composicional. Nessa etapa, por meio do tratamento quantitativo e qualitativo dos dados, da análise multivariada alcançada com a categorização de cada dado por variável e do exame descritivo baseado em frequências (*token e type*, de ocorrências e de tipos construcionais), será preciso confirmar as hipóteses que ora norteiam a pesquisa e, possivelmente, até identificar novas questões, motivações, hipóteses e variáveis no desenrolar do trabalho. Afinal, um caminho de investigação tem o poder de levar a novas trilhas.

7.5 CONSIDERAÇÕES FINAIS

Neste capítulo, apresentamos uma amostra inicial de uma pesquisa em andamento com *corpus* de discurso político. Por meio da observação empírica dos dados em análise, podemos afirmar que a expressão formada com o verbo *trazer* constitui procedimento que visa a atenuar o papel do enunciador e contribuir para a fabricação do *ethos*. Conforme Charaudeau (2011b), dentre os procedimentos enunciativos no discurso político, está a enunciação elocutiva, que permite que o depoente se coloque em cena. Advogamos que as construções formadas com o predicador complexo com verbo *trazer*, apesar de admitirem a presença em cena de um sujeito causador e focalizarem a ação implementada, permitem que o enunciador seja desfocalizado de modo a atenuar o seu papel quanto à responsabilização da ação na cena enunciativa.

Desse modo, o objetivo do capítulo foi apresentar uma proposta metodológica de pesquisa que possa dar conta da articulação de mecanismos enunciativos em um domínio discursivo. Para isso, apresentamos uma pesquisa dos predicadores com verbo *trazer* que, inseridos no discurso político, atenuam o papel do enunciador e se constituem como estratégia argumentativa.

REFERÊNCIAS

- BROWN, P.; LEVINSON, S. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press, 1978.
- CHARAUDEAU, P.; MAINGUENAU, D. *Dicionário de análise do discurso*. Paris: Éditions du Seuil, 2002.
- CHARAUDEAU, P. “Dize-me qual é teu corpus, eu te direi qual é a tua problemática”. *Revista Diadorim*, Volume 10, Dezembro de 2011a.
- CHARAUDEAU, P. *Discurso político*. São Paulo: Contexto, 2011b.

- CHARAUDEAU, P. O discurso político. In: *Análise do discurso: gêneros, comunicação e sociedade*. EMEDIATO, W., MACHADO, I. L., MENEZES, W. (orgs.). Belo Horizonte: Núcleo de Análise do discurso, Programa de Pós-Graduação em Estudos Linguísticos, Faculdade de Letras da UFMG, 2006.
- COSTA, F. R. G. da; MACHADO VIEIRA, M. dos S. A construção de atenuação do discurso com verbo suporte trazer: uma abordagem construcional baseada no uso. *SOLETRAS online*, n. 41, 2021. p. 116-144.
- ISBN: 9788580393088, DOI 10.5151/9788580393088-06<https://openaccess.blucher.com.br/article-details/06-21011>.
- MACHADO VIEIRA, M. dos S. Predicar com construção com verbo suporte. In: DE PAULA *et al.* *Uma História de Investigações sobre a Língua Portuguesa: Homenagem a Silvia Brandão*. São Paulo: Blucher, 2018, 91-112.
- PELINKA, A. Language as a political category: The viewpoint of political science. *Journal of Language and Politics* 6(1). 129-143, 2007.
- PERREZ, J., RANDOUR, F., REUCHAMPS, M. De l'uniformité du discours politique : analyse bibliométrique et linguistique de la catégorisation des discours politiques. *Revue de l'association française de linguistique cognitive*, Volume 19, 2019.
- TRAUGOTT, E. C. & TROUSDALE, G. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press, 2013.
- VAN DIJK, T. A. What is political discourse analysis? *Belgian Journal of Linguistics* 11, 11-52, 1997.

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 predicações 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

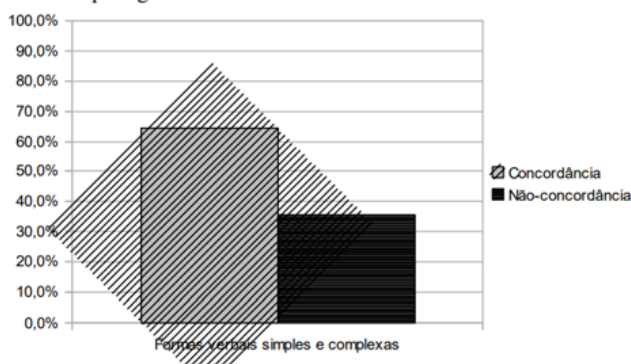
7 Para mais informações sobre como proceder no *Goldvarb X*, sugerimos a leitura de: BERLINCK, R. A.; BIAZOLLI, C. C. Ferramentas metodológicas para análises (sócio)linguísticas. In.: *Estudos Linguísticos*, 2011, p. 260-273. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/2047/1363>.

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:

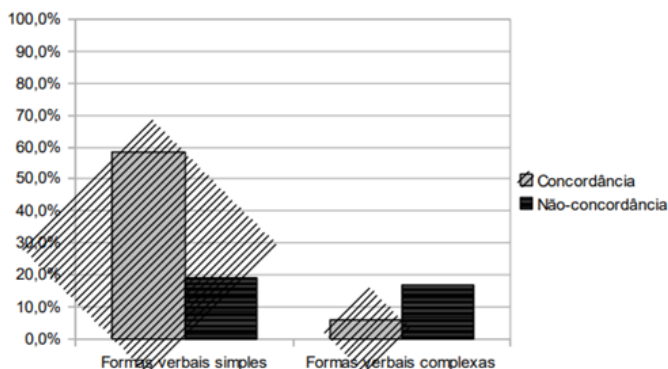
Gráfico 1: Distribuição percentual das variantes concordância vs. não concordância em estruturas passivas pronominais formadas de predicadores simples ou sequências verbais seguidos de SN plural.

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 impessoalização 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.

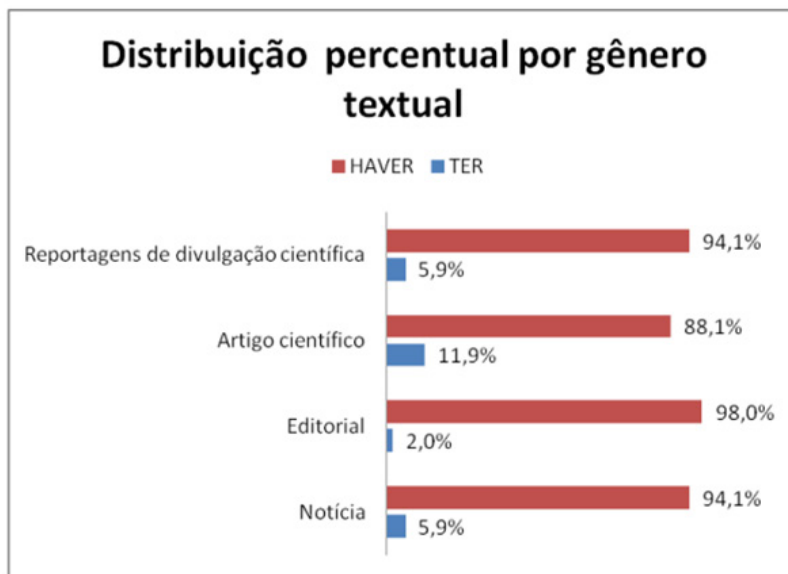
A constituição do *corpus* de análise aconteceu em 3 fases. Na primeira,

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.

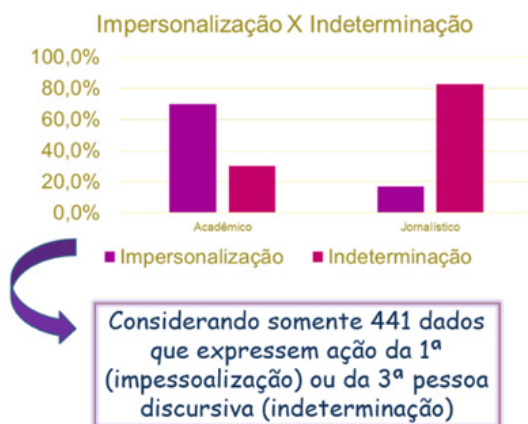
Mais uma vez, foram consideradas as 20 páginas iniciais das teses e dissertações, a partir da introdução e buscou-se um equilíbrio entre o número de palavras dos textos científicos e jornalísticos, pois importa realizar a análise colostrucional (cf. STEFANOWITSCH; GRIES, 2003) para verificar a força 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. Disponível em: <https://doi.org/10.15406/rv.2018.14.0001>. Acesso em: 03 jan. 2022.

- 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.

CAPÍTULO 9

ANÁLISE COMPARATIVA PARA DETECTAR AUXILIARIDADE VERBAL E PREDICADORES COMPLEXOS

Marcia dos Santos Machado Vieira (*Universidade Federal do Rio de Janeiro,
CNPq/Faperj*)

9.1 INTRODUÇÃO

Uma questão frequente nas aulas de Português diz respeito à dificuldade para definir a que categoria alguns verbos pertencem. Verbo principal? Verbo de ligação? Verbo auxiliar? Existem verbos *mais ou menos* auxiliares? Há outros subtipos de verbos instrumentais? Em alguns casos – como nas sequências *ter/haver + particípio passado*, em estruturas de tempo composto, e *ser + particípio*, em estruturas de voz passiva –, geralmente os alunos não hesitam quanto à classificação das unidades verbais. Suas dúvidas, de fato, começam a aparecer em sequências de verbos como estas:

(1) Uma mãe consciente da importância do aleitamento materno

tenta amamentar seu filho.

quer amamentar seu filho.

consegue amamentar seu filho.

costuma amamentar seu filho.

tem de amamentar seu filho.

há de amamentar seu filho.

fica a amamentar, em lugar de tem de amamentar

esforça-se por amamentar seu filho.

Nesses enunciados (construídos a título de ilustração), há um período simples ou um período composto? As formas sublinhadas exercem papel de *verbo principal*, cujo complemento interno é preenchido por um sintagma oracional com predicador verbal no infinitivo? Ou exercem função associada à de *verbo auxiliar*, que opera sobre um predicador verbal (simples, *amamentar*, ou complexo, *dar de mamar*) no infinitivo formando com este uma unidade verbal complexa (Vauxiliar + Vauxiliado), uma perífrase verbal?

Com que contornos semânticos a predicação se materializa? O estado de coisas pode ser conceptualizado pelo viés da tentativa ou conação/intencionalidade (*tenta/esforça-se* por amamentar), do desejo ou de futuridade (*quer/há* de amamentar), da consecução (*consegue* amamentar), da repetição ou iteratividade (*costuma* amamentar), da obrigação (externa ou interna) ou modalidade deôntica, do dever (*tem* de/que amamentar ou tirar, no exemplo a seguir), da continuidade (*fica* a amamentar).

1 Predicador verbal é sinônimo de verbo principal, verbo pleno, verbo nocional ou unidade verbal predicante (Cf. item 1.3.1).

(2)



2

Temos que tirar deu para acreditar

(3)



começa a melhorar



3

O estado de coisas pode, ainda, ser retratado por seu aspecto inceptivo (*dar para acreditar*, no cartoon de Clóvis Lima, ou *começa a melhorar*, no tweet do Sensacionalista). E, assim, vemos como o verbo se presta à marcação de uma perspectiva ou um matiz na expressão de um estado de coisas num texto.

2 Clóvis Cartoon, 27 de março de 2022, <https://www.facebook.com/profile.php?id=100060760133843>, <https://www.facebook.com/notaterapia/photos/a.1395172920517894/4906731969361954/> / Acesso em: 28 mar. 2022.

3 Sensacionalista, 29/03/2022, <https://blogs.oglobo.globo.com/sensacionalista/post/educacao-esta-sem-ministro-de-bolsonaro-e-ja-comeca-melhorar.html> Acesso em: 29 mar. 2022.

Para além dessas dúvidas, cabe lembrar, ainda, de intuições relacionadas a outros tipos de sequências verbais, ou melhor dizendo, a sequências verbo-nominais. Muitas vezes, o analista (professor e aluno) tem a percepção de que certos verbos chegam a constituir com elementos não verbais unidades funcionais de predicação. Entretanto, como raramente encontra amparo em descrições gramaticais (nem mesmo entre expressões idiomáticas registradas em verbetes de obras lexicográficas) que valide esse tipo de impressão, trata esses verbos como verbos principais. É o que acontece a sequências como:

(4) Há razão para ter medo/preocupação à morte? [ter medo a/de, temer] ou [ter preocupação, preocupar-se]

(5) Por que adolescentes fazem guerra a quem só lhes quer bem? [fazer guerra a, guerrear com]

(6) É preciso pôr freio à corrupção na administração pública. [pôr freio a, (re)frear, conter, reprimir]

Exemplo desse tipo de sequência verbo-nominal em texto pode ser visto a seguir: *dar respeito*, *respeitar*; *prezar*, *ter apreço* ou *estima* por.

(7)



Mesmo Rocha Lima (2005 [1972]: 250) – que menciona as três sequências verbo-nominais como exemplos de “conglomerados constituídos de *verbo* + *objeto direto*, dos quais depende o indireto” – acaba conferindo uma descrição a itens verbais como esses que os aproxima muito mais dos itens da categoria verbo principal. É bem verda-

4 Armandinho, 25 de março de 2015. Disponível em: <https://www.facebook.com/tirasarmandinho/> <https://www.facebook.com/tirasarmandinho/photos/np.21015315.100005065987619/948291325216174/>. Acesso em: 28 mar. 2022.

de que tal descrição também encontra, na literatura linguística, defesas nesse sentido, haja vista, a título de ilustração, a classe dos chamados verbos leves, com “propriedades predicativas” na perspectiva de Gonçalves et al. (2010), dentre outros (Raposo et al., 2013).

Essas observações ilustram como o tema da auxiliaridade ainda carece de um tratamento que propicie tanto ao aprendiz quanto ao professor uma iniciação teórica e metodológica que os torne capazes de reconhecer propriedades das construções com verbos auxiliares e, então, descrevê-las nas aulas de Português (língua materna e não materna) ou explorá-las conscientemente nos diversos contextos de interação e produção textual, bem como em contextos em que seja necessária a categorização.

9.2 O TRATAMENTO DA AUXILIARIDADE

9.2.1 O ENFOQUE TRADICIONAL

O assunto tem sido tratado como se as dúvidas a ele relacionadas pudessem ser sanadas de duas maneiras basicamente. Uma delas é simplesmente recorrer a uma obra de consulta lexicográfica ou gramatical, que fornece verdadeiras listas fixando as categorias a que pertence um verbo, como se uma lista de unidades verbais com ou sem função (semi-)gramatical estivesse inequívoca e rigidamente definida *a priori*, ou, ainda, como se nessas unidades houvesse uma espécie de caráter gramatical “imane”nte”. A outra maneira é contar com critérios meramente morfossintáticos, que colaboram para a identificação dos diferentes comportamentos de um verbo: a ideia de que o *verbo auxiliar* é aquele que (i) é empregado para formar locução verbal, uma espécie de combinação com infinitivo, gerúndio ou particípio de verbo principal, com ou sem intermediação de preposição, e (ii) recebe as flexões de pessoa, número, tempo e modo. É, por exemplo, o que se percebe em descrições como a seguinte:

A fim de melhor se expressarem certos aspectos especiais não traduzíveis pelas formas simples já estudadas, possuem os verbos alguns tempos compostos, nos quais uma das formas nominais (infinitivo, particípio, ou gerúndio) é acompanhada de outro verbo, chamado AUXILIAR (ROCHA LIMA, 2005 [1972]: 134).

Em muitas obras didáticas, apenas se observam esses dois tipos de informação morfossintática compondo as descrições relacionadas ao tema da auxiliaridade.

Nesse caso, permanece a dúvida sobre como, de fato, se podem distinguir ocorrências de locução verbal de ocorrências em que dois verbos principais aparecem dispostos lado a lado em enunciados em que, à primeira vista, se encontra a estrutura de verbo flexionado nas categorias morfológicas supracitadas seguido de verbo na forma de infinitivo, gerúndio ou particípio. Por exemplo:

(8)



5

Esse tipo de descrição tem feito com que, na tentativa de resolver a dúvida, se busque a solução para a categorização em listas de verbos auxiliares. Só que a delimitação dos elementos que pertencem a essa classe não é unânime, nem se dá fora de contexto. Consequentemente, as listagens que podem servir de referência à consulta diferem de obra gramatical para obra gramatical. Além disso, em algumas obras, distribuem-se, de modo disperso, em diferentes seções (em descrição de verbo ou de colocação pronominal, por exemplo). Na gramática de Rocha Lima (2005 [1972]), há, por exemplo, estas passagens:

“São numerosos os auxiliares em português: querer (quero sair), estar (estou escrevendo), ficar (fiquei a contemplá-la), ir (a tarde ia morrendo), etc.

Estudemos a conjugação dos auxiliares fundamentais que formam tempos compostos (ter e haver) e a voz passiva (ser).” (p. 134).

Em português, esses verbos auxiliares que mais habitualmente REGEM OUTRO VERBO são os seguintes:

5 https://www.rtp.pt/noticias/mundo/quase-meio-milhao-de-pessoas-voltaram-a-ucrania-desde-o-inicio-da-guerra_v1395343. Acesso em: 31 mar. 2022.

poder
saber
querer
dever

chamados 'auxiliares modais' (p. 413)

Jamais deixei de ajudar-te.

Começou a ensinar-lhe português. (p. 454) [deixar de, começar a]

“O morcego vem te chupar o sangue.” (ALENCAR) (p. 455) [vir]

Diante de um quadro de categorização como esse, associado, muitas vezes, a uma descrição que privilegia alguns aspectos morfossintáticos – em detrimento de outros aspectos (morfossintáticos, bem como semânticos) envolvidos na questão da auxiliabilidade –, restará ainda a dúvida sobre como classificar as estruturas de enunciados em que haja verbos que aparecem em umas e não em outras listagens, como QUERER, TENTAR, BUSCAR, PRECISAR (no exemplo a seguir), COSTUMAR, CONSEGUIR, VIVER: estruturariam perífrases verbais ou não?

(9)



6

Outro procedimento nesse enfoque que também chama a atenção é o fato de que, normalmente, a tradição escolar assume a existência de duas grandes classes de verbos: a dos auxiliares e a dos principais.

A classe dos *verbos auxiliares* contempla, tradicionalmente, as unidades linguísticas consideradas não nocionais e de comportamento gramatical, que se combinam com outro verbo (sendo este principal ou mesmo outro verbo auxiliar seguido de verbo principal) formando uma locução, em torno da qual, por sua vez, se organiza uma predicação, uma oração. Esse é o caso, por exemplo, do auxiliar *ter/tinha* em estruturas de tempo composto como *tinha comprado* ou como predicadores explorados no excerto de uma seção de “dicas de Português” de jornal digital brasileiro a seguir:

6 Armandinho, 14 de março de 2022 [tira de 2016], <https://www.facebook.com/tirasarmandinho/photos/a.488361671209144/5281634095215187> Acesso em: 29 mar. 2022.

(10) Ele **DISSE** ou **TINHA DITO** que chegaria cedo, mas chegou às 5h? A diferença entre **DISSE** e **TINHA DITO** é o tempo verbal: **DISSE** está no pretérito perfeito e **TINHA DITO**, no pretérito mais-que-perfeito do indicativo. O pretérito perfeito indica uma ação concluída no passado: “Ele disse, saiu, fez...”; o pretérito mais-que-perfeito indica uma ação anterior a outra ação que já está no passado: “Quando eu cheguei (pretérito perfeito = ação já passada), ele já tinha dito ou dissera ou havia dito, tinha saído ou saíra ou havia saído, tinha feito ou fizera ou havia feito (pretérito mais-que-perfeito = ação anterior à ação já passada)”. Assim sendo, quanto à pergunta do nosso leitor, o mais adequado é: “Ele tinha dito que chegaria cedo, mas chegou às 5h”. A ação de “dizer” é anterior a ação de “chegar”. O pretérito mais-que-perfeito é o passado do passado.⁷

O interessante nesse excerto usado como exemplo é também o fato de ficar evidente que, conforme dito em Machado Vieira (2020, p. 46), há variação linguística manifesta em lugares de prescrição (textos jornalísticos delineados a partir de norma culta padrão, obras gramaticais normativas, por exemplo). A descrição feita alude às variantes lexicais *ter* e *haver* como itens atraídos ao *slot* de auxiliar em construção de pretérito mais-que-perfeito composto, além de às variantes morfossintáticas predicador complexo (*tinha dito*) e predicador simples (*dissera*) que se prestram à sinalização desse tempo verbal.

A classe dos *verbos principais*, por sua vez, compreende as unidades linguísticas nocionais de comportamento lexical, responsáveis pela projeção dos papéis participantes de uma predicação e da configuração semântica e sintática básica desta. É o caso de *comprar/comprado* no exemplo anterior ou o de *dizer, querer, enviar* no exemplo a seguir:

(11)



O enfoque tradicional contempla, ainda, os *verbos de ligação* entre os de inclinação auxiliar/instrumental, que se relacionam a um *predicativo do sujeito*, quer dizer, um elemento adjetival ou nominal com função predicante. É o caso, por exemplo, do verbo *ser/é*, que, na tirinha, relaciona o sintagma nominal “o critério da verdade” como predicativo do termo de natureza nominal com função de sujeito “a prática”, a caracterizar este. Porém, em geral, trata dessa categoria ainda mais superficialmente do que

⁷ Sérgio Nogueira, O Globo, 15/08/2012, Conheça os sete pecados da crase, <https://g1.globo.com/educacao/blog/dicas-de-portugues/31.html> Acesso em: 29 mar. 2022.

⁸ Armandinho, 16 de março de 2022, <https://www.facebook.com/tirasarmandinho/photos/a.488361671209144/5285908058121124/> Acesso em: 29 mar. 2022.

aborda a de verbo auxiliar. Não é raro encontrar o assunto apenas mencionado quando a obra apresenta a possibilidade de configuração de predicados nominais. Então, há alusão a esses verbos na meta de explicitar a configuração destes. E, geralmente, muitas descrições limitam-se a explorar adjetivos ou sintagmas adjetivais nesse lugar de predicativo.

Nesse sentido, esse enfoque não contempla os casos intermediários entre verbos principais e verbos auxiliares: os verbos semiauxiliares, que se situam em um *continuum* de gramaticalização entre as categorias de verbo predicador e verbo auxiliar (casos de semigramaticalidade). Muito menos aborda os casos de verbos cuja função é operar gramaticalmente sobre elementos auxiliados de *natureza não verbal*, constituindo com estes uma unidade complexa (*verbo + elemento não verbal* ou, abreviadamente, *V+ñV*), para a qual pode haver uma forma simples correspondente. Para exemplificar isso, há estas perífrases com verbo suporte: “fazer menção” (“mencionar”), “ter respeito” (“respeitar”), “dar voto” (“votar”), “dar uma freada” (“frear”), “pôr medo” (“amedrontar”), ou ainda “tomar conta”, “passar carão”, “passar fogo”, “pegar confiança”, “levar um fora”, “levar uma fechada”, entre tantas outras. É inegável que, nessas construções, os verbos têm um papel auxiliar, uma vez que servem de reforço ao sentido da predicação definido principalmente pelo elemento não verbal da estrutura perifrástica e de suporte ou apoio para a expressão das categorias gramaticais de tempo, modo, número e pessoa.

A disparidade entre as classificações de verbos instrumentais existentes em livros didáticos e obras gramaticais pode ser consequência de fatores diversos, como: (i) analisar conjuntos de dados diferentes; (ii) examinar a amostra de dados com base em múltiplos critérios nem sempre coincidentes (visto que foram definidos a partir de princípios teóricos distintos) ou, muitas vezes, pouco rigorosos (já que nem sempre se aplicam exaustivamente aos dados); e (iii) não considerar um *continuum* intercategorial, ou seja, graus intermediários entre as classes de verbos estudadas. O quadro que aqui brevemente se delineou mostra que a questão é complexa e carece de um tratamento que se dê noutros moldes.

9.2.2 O ENFOQUE DE GRAMÁTICAS DESCRITIVAS

Também a consulta a gramáticas descritivas do Português evidenciará diferenças quanto aos verbos citados como auxiliares, como se pode visualizar no quadro a seguir, que procurou reunir exemplos desse tipo de item linguístico, encontrados em quatro obras:

Mateus et al (2003)	Ilari e Basso (2008)	Castilho (2010)	Raposo (2013)
Ter DO Haver DO Ser DO Estar DO/a R Ficar DO/NDO Ficar a/por R Começar a/por R Começar NDO Ir R Continuar a R Continuar NDO Chegar a R Acabar de/por R Deixar de R Tornar a R Poder R Dever R Haver de R Ter de/que R, entre outros verbos.	Ter DO , Haver DO, Ser DO, Ser de R, Ficar DO/NDO/ a R, Começar NDO/ a/por R, Pôr-se a R, Pegar a R, Deitar a R, Desandar a R, Desatar a R, Passar a R, Pegar a R, Cair a R, Ir R/NDO, Vir (a) R/NDO, Estar NDO, Viver NDO, Andar NDO, Continuar a R/NDO, Seguir NDO, Permanecer NDO/DO, Chegar a R, Terminar de/por R, Acabar de/por R, Parar de R, Deixar de R, Cessar de R, Poder R, Dever R, Haver de R, Ter de/que R, Conseguir R, Querer R (perífrase de volição), Desejar R, Pretender R, Costumar R (de iteratividade), Viver a R, Habituar-se a R, Tentar R (perífrase de incerteza), Saber R (perífrase de certeza), Esforçar-se por R (de sentido conativo), entre outros verbos.	Ter DO , Haver DO, Ser DO, Estar DO, Ficar DO/NDO/ a R, Principiar a R, Começar NDO / a/por R, Pôr-se a R, Pegar a R, Deitar a R, (A)garrar (a) R, Desandar a R, Desatar a R, Passar a R, Pegar a R, Cair a R, Ir R/NDO, Vir (a) R/NDO, Estar NDO / a/por R, Viver NDO, Andar NDO, Continuar NDO / a R, Seguir NDO / a R, Chegar a R, Terminar de/por R, Acabar de/por R, Deixar de R, Cessar de R, Poder R, Dever R, Haver de R, Ter de/que R, Querer R (perífrase de volição), Desejar R, Pretender R, Costumar R (de iteratividade), Viver a R, Habituar-se a R, Tentar R (perífrase de incerteza), Saber R (perífrase de certeza), Esforçar-se por R (de sentido conativo), entre outros verbos.	Ter DO, Ser DO, Haver de R, Poder R, dever R e Ter de/que R, Estar NDO/ a R e Continuar NDO/ a R, Começar a R, Acabar de R e deixar de R, Ficar NDO/ a R e passar a R, Andar a R, Voltar a R e tornar a R, Chegar (a) R, Ir R/NDO e Vir (a) R/NDO, Querer e parecer R, Ameaçar R _{meteorológico} entre outros verbos.

Quadro 1. Elenco de verbos (semi)auxiliares citados em obras gramaticais descritivas do Português.

Fonte: Machado Vieira (2015).

Naturalmente, o tratamento do assunto nessas obras é completamente diferente. Em linhas gerais, pode-se destacar que, além de haver nelas uma preocupação com o detalhamento dos critérios de auxiliabilidade que servem à análise e à categorização dos verbos e até com a maneira com que se opera com eles nesse processo (cf. ILARI & BASSO, 2008), tais obras lidam com os mais diversos aspectos envolvidos em complexos verbais ou verbo-nominais.

9.2.3 A PROPÓSITO DE OUTRA POSSIBILIDADE DE TRATAMENTO DIFERENTE DA TRADICIONAL

Quanto mais se vasculha o tema, mais se descobre que ele se vincula a um território profícuo em fragilidades, sutilezas e indeterminações, que, não raras vezes, causam aos interessados no assunto um sentimento de desconforto em relação ao instrumental metalinguístico de análise e descrição do tópico.

Assim sendo, para o tratamento rigoroso desse tópico linguístico, cabe àquele que pretende enveredar em tal território primeiramente conhecer um conjunto de propriedades que possibilitem uma coerente categorização de usos de formas verbais, separando os de caráter gramatical daqueles de caráter semigramatical e daqueles de

caráter lexical, ou seja, respectivamente verbos auxiliares, semiauxiliares e predicadores. Com base nesse conhecimento, resta-lhe explicitar as propriedades de *auxiliaridade* contempladas em sua análise e descrição.

Desse tratamento, certamente resultarão classificações ora convergentes ora divergentes, em decorrência da natureza (semântica e/ou morfossintática) dos critérios considerados e/ou priorizados, das opções teórico-metodológicas quanto à seleção e à análise da amostra de dados, das decisões descritivas e explicativas tomadas nos casos de dúvida/fronteira entre uma e outra (sub)classificação estabelecida. Com certeza, resultará uma categorização pautada em um maior nível de consciência a respeito dos critérios pelos quais se pode pautar um estudioso da sintaxe dos verbos (professor ou aluno) e, em consequência disso, uma classificação provavelmente (mais) rigorosa e pertinente/abrangente, além de empiricamente delineada via comparação de usos atestados na experiência (se possível, a mais diversa).

Importa lidar com o assunto de uma maneira que permita ao interlocutor – no caso da sala de aula, ao aluno: (re)conhecer o instrumental de análise e interpretação de (sub)categorias verbais; perceber e explicitar as relações de similaridade e dessemelhança entre definições, categorias e unidades linguísticas; examinar dados da língua em uso; e refletir sobre os limites e a pertinência das descrições que são levadas a cabo. Enfim, importa tornar o espaço dedicado ao tratamento do assunto convidativo ao raciocínio e à atividade metalinguística e, assim, criar um contexto propício a benefícios pedagógicos.

9.3 PARA O TRATAMENTO DA AUXILIARIDADE

9.3.1 PARA INÍCIO DE “CONVERSA”: SETE PONTOS DE PARTIDA PARA A DESCRIÇÃO E A CATEGORIZAÇÃO DE FORMAS VERBAIS

9.3.1.1 A relação entre Sintaxe e Semântica/Função

Não se pode perder de vista a relação entre Sintaxe e Semântica. Nem sempre essa relação está livre de problemas, mas nem por isso deixa de ser fundamental para o tratamento de um assunto dessa natureza. O pareamento de atributos de forma e função (semântica, discursiva, pragmática, social, cognitiva) é basilar à concepção de língua como uma gramática de construções interconectadas.

A categorização morfossintática de um verbo tem de ser definida em virtude do enunciado em que ocorre, do significado que emerge neste. A atualização de determinados empregos de um verbo vincula-se ao contexto discursivo-pragmático, como se demonstra nos exemplos a seguir:

(12a) É tamanha a especialização que há um funcionário na seção de testes desta fábrica de brinquedos só para abrir mão de bonecos. (abrir = verbo predicador)

(12b) No Natal, esse garoto não abre mão de bonecos. (abrir = verbo suporte / [abrir mão]predicador complexo)

(13a) Pedro deixou o colega aborrecido. (deixar = verbo suporte (causativo))

(13b) O professor deixou de sair com os colegas. (deixar = verbo semiauxiliar (com noção de fase final de um estado de coisas, aspecto terminativo))

(13c) O professor deixou correr solto o debate. (deixar = verbo semiauxiliar (causativo))

(13d) Daniel deixou a casa dos pais. (deixar = verbo predicador (movimento))

(14a) Essa menina anda preguiçosa. (andar = verbo de ligação ou verbo relacional)

(14b) Essa menina anda estudando pouco. (andar = verbo (semi-)auxiliar (duração))

(14c) Essa menina já anda sozinha. (andar = verbo predicador (movimento))

A permuta de um verbo (semi-)gramatical por outro acarreta efeitos de sentidos distintos. Logo, atende a finalidades discursivas distintas, como se nota nos exemplos a seguir:

(15) Pedro deu/fez parte da equipe de futebol.

(16) Essa menina é/está/continua/anda/vive preguiçosa.

(17) O professor deixou/parou/acabou/tem de sair com os colegas.

É importante levar o aluno a ter consciência da relação entre o componente semântico-discursivo e as escolhas por determinadas estruturas morfossintáticas e por um ou outro verbo auxiliar/suporte/relacional ou de ligação, explorando-se textos orais e escritos.

Entre outros procedimentos, o professor pode trabalhar com textos que revelem isso também via múltiplas linguagens, multissemióticas. Lembramos, por exemplo, o caso de um anúncio publicitário de cerveja, veiculado na TV, que explorou, além de um olhar machista da mulher na publicidade, os matizes semânticos proporcionados pelos verbos PODER e IR: possibilidade e probabilidade futura praticamente certa. A cena passa-se na praia e envolve um homem e uma mulher, ambos figuras da TV conhecidas. A mulher pede uma Boa [cerveja], que é posta no balcão. Ao tentar alcançar um copo de cerveja no balcão de um quiosque, sua saia fica presa a uma cadeira. Então, tem dificuldade de alcançar o copo. Ao ver a cena, o homem, movimenta-se em direção ao copo e lhe diz que “assim sua saia **pode** rasgar”. Acreditando que ele a ajudará a alcançar o copo de cerveja e a evitar que sua saia rasgue, ela espera, mostrando, por linguagem verbal e não verbal, comoção com a gentileza. Em seguida, o homem afasta ainda mais o copo de cerveja da mulher e, então, diz: “assim, ela **vai** rasgar, com certeza”.⁹ Ressaltamos o jogo entre “poder” e “ir” na esfera expressiva de futuridade via predicador complexo.

9 Cf. LIMA, Juliana Acioly. A imagem do feminino na publicidade contemporânea: estudos de casos.

O professor pode trabalhar, por exemplo, finalidades discursivas para as quais a opção por perífrases pode ser considerada mais indicada, como as seguintes: reforço de um determinado sentido do evento descrito; obtenção de um determinado efeito de sentido, não ou dificilmente alcançado mediante outro mecanismo; ou simplesmente necessidade de variar a expressão linguística (de versatilidade na construção textual). Com a opção por enunciados, como *O professor pôs/deixou a turma apavorada* e *A imprensa fez saber a verdade ao povo*, é possível reforçar o caráter causativo, efeito que não é tão transparente com variantes simples (“apavorou” ou “informou”). Em *O professor deu uma olhada/conferida no texto do colega*, sobressai o aspecto superficial e ligeiro do evento, que não é recuperado por uma forma simples equivalente. Nesse caso, possivelmente o recurso a um adjunto adverbial (*rapidamente, por alto*) é uma alternativa de configuração morfossintática no sentido da sinalização desse tipo de actualidade.

9.3.1.2 Uma proposta de categorização escalar dos verbos com caráter instrumental: de membros com configuração prototípica a membros menos exemplares da categoria

Auxiliaridade é um processo gradual de gramaticalização a que se submetem determinados empregos primariamente lexicais dos verbos: de verbo com comportamento lexical (verbo principal ou predicador) a verbo com comportamento semigramatical ou gramatical (verbo semi-instrumental ou instrumental). *Auxiliaridade* é a condição (estrutural e conjuntural) resultante desse processo. Isso ocorre quando itens verbais se combinam a *slots* em construções procedurais por serem normalmente atraídos para esses lugares ou por serem eventualmente coagidos ao comportamento nesses lugares. Nesse processo, o verbo que assume comportamento (semi-)gramatical tem enfraquecida ou perde sua função de predicar e passa a exercer o papel de marcador gramatical de tempo, modo, modalidade, aspecto e/ou voz, além de indicar número e pessoa. Em alguns casos, a identificação de verbos com comportamento gramatical é mais nítida. Em outros, a identificação é muito difícil ou até não é possível tomar decisão quanto ao caráter do verbo. Alguns verbos são produtivamente usados na língua como auxiliares. Outros, só ocasionalmente. Assim sendo, esse é um terreno em que o analista da língua nem sempre tem muitas certezas. Isso explica a existência de pontos de vista às vezes totalmente distintos quanto à classificação dos verbos.

Ao lidar com um assunto complexo como esse, é natural haver dúvidas na análise e fragilidade descritiva. Mascara essa situação prejudica tanto o aluno quanto o professor, porque os engana, não lida com a complexidade da língua, não conduz à reflexão sobre as (sub)categorias do verbo baseada em critérios linguísticos, nem confere a esse trabalho maior rentabilidade didático-pedagógica.

9.3.1.3 Predicação: predicador e construção

Os estados de coisas são expressos via predicções. E os predicadores (simples ou complexos) são os elementos que projetam uma estrutura de papéis participantes que se podem ligar, por atração ou coerção, a uma estrutura de argumentos com perfil gramatical impessoal ou pessoal, de um lado, e intransitivo ou transitivo, de outro.

Em muitas propostas funcionalistas, por exemplo, a de Dik (1997), a predicação é considerada uma peça fundamental da organização textual, a base da construção semântica de uma expressão linguística, de um enunciado. Outra consideração também evidente nessas propostas é a centralidade do verbo/predicador na configuração de uma oração e o acionamento de uma estrutura de participantes/termos.

Predicação é, geralmente, concebida como um fenômeno que lida com a sentença como estrutura semântica e estrutura gramatical.¹⁰ Assim, a uma estrutura semântica de papéis temáticos corresponde uma estrutura sintática de argumentos, que, por sua vez, é viabilizada mediante a projeção de sintagmas, unidades linguísticas ou arranjos linguísticos de estruturação.

O termo “predicador” (e, muitas vezes, “predicado”) é utilizado para indicar uma unidade estrutural que tem a função de predicar: (i) designar propriedade/atributo de um participante (no caso de verbos que exprimem estado) ou uma relação semântica entre participantes (no caso de verbos transitivos diretos de ação, entre outros); e (ii) especificar uma configuração argumental básica (o número de argumentos/participantes, a relação gramatical/sintática entre eles e as restrições de seleção semântica dos termos que ocuparão as posições argumentais). Em outras palavras, corresponde a uma unidade (simples ou complexa, quando constituída por verbo suporte, por exemplo) que faz exigências quanto à estruturação da predicação, da sentença. Observe-se, a título de ilustração, o verbo *sorrir/dar um sorriso/dar uma sorrída*, como aparece no exemplo:

10 E ainda pode ser entendida como lidando com a sentença também como uma estrutura discursiva dos participantes da enunciação.

(18)



11

Nesse tipo de uso, o predicador “deu uma sorrída” requer apenas um argumento com o papel semântico de agente, com a relação sintática de sujeito e a configuração semântica de referente humano e, sobre este, predica um evento ativo (caracterizado pelas propriedades de [+ controlado], [+ dinâmico] e [+ momentâneo]), segundo a tipologia de predicados de Dik (1997)). Machado Vieira (2018) expõe mais informações sobre predicadores como esse.

No enfoque tradicional da gramática, normalmente são descritos três tipos de predicado: (i) verbal; (ii) nominal; e (iii) verbo-nominal. Essa classificação põe em evidência que, na verdade, tanto elementos de natureza verbal quanto elementos de natureza não verbal podem ser centrais no fenômeno semântico-sintático da predicação e, então, fazer exigências quanto à estruturação de argumentos da sentença.

9.3.1.4 O caráter multifuncional de verbos do Português

Um ponto negativo de certas metodologias de ensino/descrição diz respeito a um tratamento dos verbos em que se perdem as noções de variação e de polifuncionalidade das unidades linguísticas, dos vocábulos, que se combinam a diferentes padrões construcionais de predicador (simples ou complexos).

O falante conta com elementos da língua para comunicar-se. Em virtude de uma necessidade de comunicação, para criar um determinado efeito de sentido, ele recorre aos itens linguísticos e pode conferir-lhes distintos papéis a depender da ligação que façam a uma ou outra construção gramatical de predicador e predicação. Observem-se alguns exemplos a seguir:

11 <https://twitter.com/mallmkt/status/593002542243643392/photo/1>. Acesso em: 31 mar. 2022.

Extensões semântico-sintáticas de alguns verbos: ter, virar, viver, ir, levar, querer, ser, ficar, fazer.			
Vpredicador	Vinstrumental		
	de ligação	suporte	(semi)auxiliar
A criança tem muitos brinquedos.		A criança tem necessidade de brincar.	Essa criança tem de brincar. Essa criança tem aprendido muitas travessuras.
O menino virou a moeda.	O medo virou esperança.	Ele virou a casaca. Agora é Botafogo.	
João vive em Natal.	José vive insatisfeito.		Seu carro vive enguiçando.
Ela vai ao cinema.			Ela vai encontrar o namorado.
A mãe levou o menino ao hospital.		O menino levou um tombo/susto.	Sua atitude levou -me a pensar no assunto.
Quero sorvete.			Quero pensar no assunto. Hoje quer chover.
Era uma vez...	Carlos é estudante.		A carta foi escrita por ele.
Ficou a esperança de dias melhores. O menino ficou lá.	O menino ficou calado (permaneceu como estava).	O menino ficou calado (calou-se).	O professor ficou de avaliar o caso. Ficou pensando no assunto.
Marta fez um bolo. Fiz 120 km/h.	O medo fez-se esperança.	Marta fez menção ao bolo.	O vestido fez realçar a beleza de Marta.

É importante observar o contexto de uso das formas verbais para, então, categorizá-las. Afinal, sua funcionalidade emerge em (con)texto. Uma determinada unidade verbal pode ser empregada para diversas funções na língua: verbo predicador plenamente configurado (“Marta fez um bolo.”) ou não (extensão de sentido de comportamento lexical, “Fiz 120 km/h.”) ou verbo (semi-)gramatical (verbo auxiliar típico ou não; neste caso, verbo semiauxiliar, verbo de ligação ou verbo suporte). De acordo com o princípio de divergência descrito por Hopper (1991), uma extensão de sentido/uso gramatical pode coexistir com um emprego lexical de uma unidade linguística. O verbo predicador *fazer* (forma autônoma com valor referente à atividade/ação de “criar ou dar existência a”), por exemplo, convive com seus empregos que são associados às categorias verbo suporte (espécie de “formante derivacional” que promove a formação de um predicador complexo) e verbo semiauxiliar, além de coexistir com seu uso de predicador não pleno (em suas extensões semânticas, como no segundo exemplo de *fazer* da primeira coluna do quadro anterior).

9.3.1.5 O conceito de auxiliaridade

Norteiam essa descrição basicamente duas noções de *auxiliaridade/auxiliarização*. Entende-se *auxiliarização* como um processo de gramaticalização que atua sobre formas verbais e que implica a associação de usos dessas formas ligadas normalmente à categoria lexical (verbo predicador) a uma categoria gramatical (verbo auxiliar, semiauxiliar, de ligação ou suporte), sob certas condições semânticas e morfosintáticas.

Em sentido lato, toma-se, então, o termo *auxiliaridade* para nomear o comportamento instrumental que uma unidade verbal pode desempenhar num *continuum* de empregos na língua. A *auxiliaridade* tanto pode incidir sobre um verbo gramatical que se junta a um outro verbo (este de caráter lexical), formando uma sequência de verbos (*Vauxiliar + Vauxiliado*),¹² quanto sobre um verbo gramatical que se alia a um elemento de natureza não verbal (sintagma nominal, adjetival, preposicional, adverbial), formando com este uma sequência verbo-nominal (*Vauxiliar + ñVauxiliado*): *dar queixa, dar ruim, tornar disponível, levar em conta, fazer mal, ser mais* (no sentido de *preferir*). Então, qualquer uma das quatro classificações verbais com que se trabalha aqui (auxiliar, semiauxiliar, de ligação ou suporte) tem, em sentido amplo, comportamento auxiliar (ou instrumental).

Em sentido estrito, o termo *auxiliaridade* é utilizado para designar os casos em que as unidades verbais demonstram comportamento gramatical mais típico e atuam sobre elementos verbais, constituindo com estes locuções verbais. É por isso que o rótulo “auxiliar” é empregado em duas das quatro classificações de verbos instrumentais supracitadas: verbos auxiliares e semiauxiliares.

Isso se deve ao fato de se ter optado pela manutenção dos rótulos com que as obras gramaticais e/ou didáticas costumam denominar as classes de verbos: (semi-)auxiliar, de ligação. Na classificação aqui feita, esses termos são utilizados estritamente para nomear classes gramaticais de verbos específicas (o primeiro para as que operam sobre outra forma verbal (+ *Vauxiliado*), o segundo e, ainda, o termo verbo suporte, para as que se ligam a uma forma não verbal (+ *ñVauxiliado*)), embora todos eles, em sentido lato, revelem, em maior ou menor grau, comportamento auxiliar/instrumental.

9.3.1.6 A abrangência do termo “perífrase verbal”

Uma estrutura perifrástica formada de *elemento verbal auxiliar + elemento (verbal ou não) auxiliado* pode apresentar diferentes graus de integração. A ordem das unidades que compõem essas estruturas não é, apesar de obras gramaticais lançarem mão

12 A sequência em *auxiliação* pode ser composta por mais de um V(semi-)gramatical, haja vista a possibilidade de predicções, como: “Pedro deve ter estado estudando até tarde”, “Ele vai ser médico”, “Ela deve fazer limpeza na casa”. No primeiro exemplo, a sequência é composta de quatro verbos, três dos quais têm comportamento gramatical (semiauxiliar, auxiliar, semiauxiliar). A segunda é composta de verbo semiauxiliar e de verbo relacional. E a terceira, de verbo semiauxiliar e verbo suporte.

dela, um critério de rendimento funcional inequívoco: nem sempre a aproximação de dois ou mais verbos constitui uma locução verbal.

Construções, como *A filha partiu chorando* e *As crianças falam sonhando* apresentam-se na mesma ordem e não contêm um verbo auxiliar; a primeira forma flexionada e a segunda não flexionada mantêm sua natureza lexical, de verbos principais. Soma-se a isso o fato de que até uma estrutura perifrástica prototípica admite a dissociação de seus componentes por conta da inserção de determinados elementos discursivos (*Tinha sempre feito tudo que estava ao seu alcance*). Segundo Bechara (1999:112), a intenção do falante é que determinará a existência de uma locução em casos de aproximação de dois ou mais verbos. Isso pode ser percebido pela comparação de enunciados citados pelo gramático, empregados para a obtenção de efeitos de sentido distintos, como: *Queríamos [colher rosas (e não fazer outra coisa)]* e *Queríamos colher [rosas (e não outra flor)]*.

Partindo-se do pressuposto de que um verdadeiro complexo verbal perifrástico possui *unidade semântica, sintática e funcional* (cf. Barroso, 1994: 61), reserva-se o termo *locução verbal/verbo-nominal* para denominar as estruturas perifrásticas exemplares (com configuração prototípica), cujos componentes apresentam maior grau de integração e significação (praticamente) indissolúvel. Atribuem-se outras designações mais genéricas – como, *complexo verbal* e *unidade verbal complexa/verbo-nominal* – às estruturas perifrásticas cujos componentes têm menor grau de coesão e possibilidade de significação disjunta. O termo perífrase, ou ainda *predicador complexo (verbal/verbo-nominal)*, é, portanto, usado em ambos os casos. O que diferencia esses casos é o estatuto morfossintático de seus componentes, bem como a natureza (mais ou menos integrada) e o efeito de sentido do vínculo entre eles, de seu grau de não-composicionalidade.

9.3.1.7 Parâmetros básicos para a identificação de sequências em auxílio (Elemento auxiliar + Elemento auxiliado (este de natureza verbal ou não))

O ponto de partida para a identificação e a delimitação de verbos em relação aos quais é possível cogitar algum potencial de *auxiliaridade* está nos seguintes parâmetros de análise e descrição, a serem destrinchados na próxima seção:

(i) *as estruturas em que ocorrem tais verbos são instâncias de períodos simples, ou podem ser interpretadas como tal;*

(ii) *o principal elemento responsável pela predicação (com função predicante, que determina o esquema predicativo¹³ e designa propriedade de um participante ou relação entre entidades participantes) pode ter natureza verbal, nominal ou adjetival;*

13 **Esquemas predicativos:** padrões de organização de predicações, que fornecem informação sobre (i) número de argumentos do predicador (Arg_n), (ii) relação gramatical dos argumentos (Sujeito, Objeto), (iii) condições semânticas necessárias para a inserção dos termos que ocuparão as posições argumentais (Agente/Paciente; Animado/Inanimado), entre outras.

[predicador_[Verbal, Nominal ou Adjetival] (Arg_n: termo¹⁴)]_{Predicação}
 Plantar/Destruir_V (Arg₁)_{Sujeito Agente/Força} (Arg₂)_{Objeto Tema}
 Plantação/Destruição_N (Arg₁)_{Agente/Força} (Arg₂)_{Tema}
 Plantado/Destruído_A (Arg₁)_{Tema} (Arg₂)_{Agente/Força}
 Alegre_A (Arg₁)_{Tema}

(iii) quanto mais depende de um constituinte adjacente (“predicante” ou com tal função em potencial), tanto mais um verbo se afasta de seu caráter lexical e assume função gramatical;

(iv) quanto menos colabora para impor restrições de seleção semântica e/ou atribuir papel temático (de agente, tema, paciente, entre outros)¹⁵ ao constituinte que tenha relação gramatical de Sujeito com o núcleo da predicação, maior o caráter auxiliar da forma verbal; e

Exs.: FAZER_[Vpredicador] (Arg₁: animado)_{Sujeito Agente} (Arg₂: inanimado)_{Objeto Tema}
 Joana fará caixas de presente.”
 DAR_[Vpredicador] (Arg₁: animado)_{Sujeito Agente} (Arg₂: inanimado) Objeto Tema
 (Arg₃: animado)_{Objeto Beneficiário}
 “Pedro deu flores à mãe.”
 LEVAR_[Vpredicador] (Arg₁: animado)_{Sujeito Agente} (Arg₂: animado)_{Objeto Paciente}
 (Arg₃: inanimado)_{Objeto Origem} (Arg₄: inanimado)_{Objeto Destino}
 “Pedro levou Maria do gabinete do diretor ao auditório.”

- 14 **Termo/Argumento:** elemento cujo papel é indicar um participante (função referencial). Num exemplo como *O pedreiro Augusto construiu casas bonitas*, “O pedreiro Augusto” e “casas bonitas” são os termos com os quais são preenchidos os lugares sintáticos projetados pelo predicador *construir*.

João e casas = “núcleos de termos distintos” – função referencial

Fazer = predicador verbal – função predicante

Bonitas = atributo – função atributiva

- 15 Exemplos de **PAPÉIS TEMÁTICOS** ou **FUNÇÕES SEMÂNTICAS** (relações semânticas entre predicadores (simples ou complexos) e seus Argumentos – sujeito e/ou complemento(s):

Agente – entidade-participante responsável (intencional) pela ação/atividade (Ex.: Rita correu em direção à mãe.)

Paciente – entidade-participante (normalmente animada) que sofre uma ação/atividade (Ex.: Rita foi mordida pelo cão.)

Tema – entidade-participante primariamente envolvida num estado (numa circunstância) ou que é efetuada/deslocada por uma operação, sob o controle de um Agente (Exs.: O quadro foi pintado por Rita. / Rita é estudante.)

Força – entidade-participante responsável (não intencional) pelo evento/estado de coisas (Ex. A tempestade causou muito pânico.)

Experienciador – entidade-participante que experimenta algum estado (psicológico) (Ex.: Rita escutou seu pedido de auxílio.)

Posicionador – entidade-participante que controla/escolhe uma posição (Rita ficou em casa.)

Beneficiário/Receptor – entidade-participante para a qual algo é transferido e que, desse modo, se beneficia da ação expressa no evento. (Exs.: Rita ganhou/recebeu muitos livros. / Minha mãe preparou uma festa surpresa para Rita. / Enviei uma mensagem à Rita.)

(v) quanto maior a extensão de seu território de atuação, ou seja, a possibilidade de se combinar com elementos auxiliados que apresentem qualquer configuração semântica (verbos que expressem qualquer tipo de evento – ação, processo, estado –, sintagmas não verbais concretos ou abstratos, animados ou inanimados, entre outros) e/ou sintática (verbos principais de estado, atividade, processo, posição; intransitivos ou transitivos; sintagmas nominal, adjetival, preposicional, adverbial), maior seu caráter instrumental.

9.4 PARA A ANÁLISE COMPARATIVA DOS (SUB)TIPOS DE VERBOS GRAMATICAIS E DE CONSTRUÇÕES DE PREDICADORES COMPLEXOS (VERBAIS OU VERBO-NOMINAIS)

9.4.1 VERBOS (SEMI-)AUXILIARES

9.4.1.1 Propriedades de *auxiliabilidade*

Para a identificação de sequências verbais em auxílio e a caracterização morfo-sintática e semântica de uma forma verbal como verbo auxiliar, há diversos fatores a serem considerados. É da multiplicidade de critérios que advêm os diferentes inventários de verbos auxiliares correntes em materiais didático-pedagógicos. À luz de alguns dos importantes estudos sobre o assunto – especialmente dos trabalhos de Lobato (1975), Gonçalves (1992) e Gonçalves e Costa (2002) –, são apresentados aqui onze critérios de *auxiliabilidade*, que, tomados em conjunto, na comparação de usos verbais possibilitam ao professor e ao aluno delimitar com rigor a classe de verbos auxiliares prototípicos (em sentido estrito) e contribuem para que sejam identificadas subclasses de verbos semiauxiliares (também “auxiliares”, em sentido lato):

9.4.1.1.1 Atuação sobre outro(s) verbo(s) num só domínio de predicação

Um verbo auxiliar ocorre em construções em que existe apenas um domínio predicativo, ou seja, uma predicação cujo núcleo é uma locução/perífrase formada de um ou mais verbo(s) (semi-)auxiliar(es) que precede(m) um verbo predicador (verbo auxiliado), ou seja: uma sequência de pelo menos dois verbos em *auxiliação* – V(semi-)auxiliar e Vauxiliado – que constituem uma unidade sintática. O verbo predicador, que pode ocorrer na forma de infinitivo, gerúndio ou particípio, é o responsável por: projetar o número de argumentos da predicação, estabelecer a relação gramatical dos argumentos com o núcleo predicador, impor as condições semânticas necessárias para a inserção dos termos que ocuparão as posições argumentais, expressar um estado de coisas sobre a(s) entidade(s)-participante(s) – uma propriedade de uma entidade (ou mais entidades) ou a relação entre entidades.

Pode-se observar essa característica em enunciados como:

(19) As chuvas intensas [**têm destruído**] plantações no interior de São Paulo.

(20) Certos políticos [**haverão lembrado**] de velhas reivindicações do povo.

[Vauxiliar + Vpredicador] = unidade sintática

Um verbo auxiliar também pode ocorrer numa sequência formada por duas ou mais unidades com comportamento gramatical e uma unidade com comportamento lexical de verbo predicador.

(21) Quando meu filho nascer, certamente [**vou ter escrito**] muitas páginas deste diário sobre minha gravidez. Assim, se tiver interesse, ele [**poderá ficar sabendo**] como foi a gestação.

[V(semi-)auxiliar + V(semi-)auxiliar + Vpredicador]

A sequência em que ocorre(m) verbo(s) (semi-)auxiliar(es) pode ser ainda constituída de perífrase com verbo suporte ou com verbo de ligação:

(22) Este artigo **poderá vir a fazer parte** da obra que, em breve, **vai estar pronta**.

[V(semi-)auxiliar + V(semi-)auxiliar + Predicador verbo-nominal]¹⁶

9.4.1.1.2 Alteração semântica

Ao se submeter à gramaticalização ou convencionalização como recurso gramatical, um verbo tem sua acepção primária de verbo predicador alterada e, em decorrência de um processo de reconfiguração semântica, sua área de atuação semântico-sintática amplia-se quando passa a desempenhar papel instrumental na língua: matizar semanticamente o verbo principal ao qual se liga. É o caso, por exemplo, do verbo TER nestes exemplos, em que marca estados de coisas que se repetem no tempo (aspecto iterativo):

(23) [**Tenho** escrito] cartas. (emprego diferente do em “*Tenho [cartas escritas]*”.)

(24) O professor Pedro, diretor da Faculdade de Direito, **tem** participado de todas as reuniões do sindicato.

É o caso também de VOLTAR e IR em sentenças como *Voltei a/Vou ler ‘O nome da Rosa’*, em que *voltar* e *ir* perdem o vínculo com a acepção de ‘movimento no espaço geográfico’ (que, por exemplo, aparece no exemplo (8) oriundo de notícia jornalística da RTP, em que também há o predicador variante *regressar*). E, nessa construção, indicam ‘movimento no tempo’ e sinalizam: “repetição/retomada de um estado de coisas sem a ideia de frequência”, no caso do primeiro verbo; e “evento futuro ou prestes a começar”, com o segundo.

É o caso ainda de DAR em sentenças como *Do nada ele deu de falar mal de todos na reunião*, em que *dar* deixa de ser associado à acepção de ‘transferência’ e passa a indicar “os primeiros momentos de um estado de coisas” (também no exemplo (2), [*dar para Vinfinitivo*]).

16 *Fazer parte* e *estar pronta*, em que *fazer* é verbo suporte e *estar* é verbo de ligação ou verbo relacional.

Em geral, o verbo auxiliar típico é caracterizado como uma unidade da língua que não possui valor lexical no mundo biossocial (valor referencial), mas instrumental no sistema linguístico. É a extensão de uso que se especializa na marcação de categorias gramaticais: pessoa, número, tempo, modo, modalidade, voz e aspecto. Por conta disso, há itens verbais que se especializam na indicação:

(i) do momento do evento descrito relativamente à situação de enunciação, situando o momento da ocorrência do evento que se predica em relação ao momento da fala como anterior (passado), simultâneo (presente) ou posterior (futuro) a esse momento – *Ele tinha escrito o artigo. Ele está escrevendo o artigo. Ele vai escrever o artigo;*

(ii) da duração ou do desenvolvimento ou completamente de um estado de coisas – *Ele vive escrevendo artigo. Ele chegou a escrever o artigo.*

Apesar de muito produtivo, há restrições quanto à aplicação desse critério de *auxiliaridade*, pelo fato de nem sempre viabilizar a categorização inequívoca dos verbos auxiliares.

Há determinados usos verbais que, apesar de citados em compêndios como (semi-)auxiliares, não se submetem da mesma maneira a um processo de desgaste semântico. É o caso de unidades, como COMEÇAR (A), TERMINAR (DE), QUERER, TENTAR, CONSEGUIR – que mantêm seu significado primário (quase) inalterado.

Há, ainda, alguns verbos rotulados tradicionalmente de auxiliares que, a depender da finalidade discursiva, revelam, ou até congregam, mais de um valor distinto. É o que ocorre, por exemplo, com PODER no enunciado *O professor Pedro pode ser porta-voz de nossas reivindicações no sindicato*. PODER contempla diferentes significados, que só serão determinados no contexto discursivo: (a) possibilidade (É possível que o Professor Pedro seja nosso porta-voz, já que é o Diretor de nossa unidade, num contexto de dúvida quanto à participação de Pedro nas reuniões do sindicato); (b) permissão (*Permite-se que o Professor Pedro assuma a função de porta-voz de nossas reivindicações no sindicato*); e (c) capacidade (*O Professor Pedro é capaz de ser porta-voz de nossas reivindicações no sindicato*, pois é o único que tem participado de todas as reuniões).

Essa constatação conduz inevitavelmente a dois questionamentos. Qual o significado original de um verbo? Como determinar isso? A fim de resolver tal questão, é possível recorrer a obras lexicográficas ou a estudos diacrônicos para verificar que acepção pode ser considerada provavelmente primária e a natureza de sua ligação com outras acepções ou, ainda, para verificar a produtividade das extensões de sentido de um verbo e os caminhos de convencionalização e entrincheiramento do que é associado entre atributos de forma e funcionamento. De qualquer modo, esse é um problema com que se terá de lidar, ao se considerar esse parâmetro nocional de gramaticalização para a categorização de verbos auxiliares.

(I) Desligamento semântico entre verbo auxiliar e sujeito gramatical

Um verbo auxiliar típico não seleciona o termo que ocupará a posição sintática de sujeito na predicação. Não lhe atribui papel semântico, nem lhe impõe restrições de seleção semântica. É o verbo predicador (verbo principal) que determina as condições semânticas dos termos que podem preencher o lugar sintático de sujeito. Isso pode ser notado pelos exemplos a seguir:

(19) [Os meninos deste orfanato]_{Paciente} têm/haviam **recebido** muito afeto.

(20) [Meu filho]_{Agente} tem/havia **ido** ao colégio.

(21) [O colégio]_{Locativo} tem/havia **abrigado** várias pessoas.

(22a) [As árvores]_{Tema/não humano} já tinham **florescido** nesta época.

(22b) [As árvores]_{Tema/não humano} já **floresceram** nesta época.

(22c) *[Pessoas]_{Tema/humano} têm **florescido** nesta época.¹⁷

(23a) [Cientistas]_{Agente/Controlador} têm **descoberto** novos medicamentos nos últimos tempos.

(23b) [Cientistas]_{Agente/Controlador} **descobrem** novos medicamentos nos últimos tempos.

(23c) *[As árvores]_{Não controlador/Inanimado} têm **descoberto** novos medicamentos nos últimos tempos.

(24a) [A gestante]_{Experienciador/animado} havia **enjoado** muito.

(24b) *[O carro]_{Inanimado} havia **enjoado** muito.

(25a) [O carro]_{Tema/inanimado} havia **enguiçado**.

(25b) *[A gestante]_{Animado} havia **enguiçado**.

Esses exemplos evidenciam que *ter* e *haver* auxiliares de tempo composto podem ocorrer com argumentos sujeitos que desempenhem qualquer função semântica (paciente, agente, locativo, tema, experienciador). São os verbos principais “receber”, “ir”, “abrigar”, “florescer”, “descobrir”, “enjoar” que determinam as condições semânticas para a seleção do termo a exercer a função de sujeito na predicação, e não os verbos “ter” e “haver”. Assim, verbos como “florescer” e “enguiçar” restringem a seleção de argumentos a termos sem o traço humano ou com traço inanimado, ao passo que “descobrir” e “enjoar”, por exemplo, já exigem um termo com traço animado (ou apresentado como tal).

Quando uma forma verbal colabora em algum grau para a seleção semântica do constituinte sujeito, diz-se que ela já não pertence à classe dos verbos auxiliares típicos. Se ainda assim formar com o verbo predicador uma unidade complexa, diz-se

17 Utiliza-se, aqui, asterisco antes de determinados exemplos para sinalizar uma estrutura sintaticamente malformada ou o baixo grau de gramaticalidade e/ou aceitabilidade do enunciado. Em alguns casos, também se usa o ponto de interrogação para indicar os casos em que se tem dúvida quanto a esses julgamentos.

que ela se insere no rol dos verbos semiauxiliares. Nesse caso, outras propriedades vão direcionar o analista para a detecção de uma unidade complexa de funcionamento.

(26a) Vândalos têm destruído monumentos públicos.

(26b) Vândalos querem/desejam destruir monumentos públicos.

(26c) Depredações têm destruído monumentos públicos.

(26d) *Depredações querem/desejam destruir monumentos públicos.

(27a) Agricultores sem-terra descontentes têm destruído plantações no Sul do país.

(27b) Agricultores sem-terra descontentes querem/desejam destruir plantações no Sul do país.

(27c) As chuvas intensas têm destruído plantações no Sul do país.

(27d) *As chuvas intensas querem/desejam destruir plantações no Sul do país.

Em outras palavras, o papel participante que se apresenta como sintagma nominal (SN) sujeito é selecionado pelo verbo auxiliado. Assim, SNs com diferentes papéis temáticos podem ocorrer como sujeito de frases com verbos auxiliares (*A chuva/criança vai manchar a parede.*). É o verbo principal o responsável por inviabilizar determinados sintagmas na posição de sujeito (**A chuva vai estudar.*).

Cabe aqui uma ressalva quanto a esse critério. Há verbos predicadores (como *existir, pertencer*) que não impõem restrições semânticas ao constituinte sujeito. Assim sendo, apesar de válida, a extensão combinatória do elemento “auxiliar” não é suficiente para a apreensão do conjunto de auxiliares. Na verdade, nenhum critério por si só o é. É a coocorrência de certas propriedades que conduz o estudioso a essa categorização: quanto mais propriedades associadas à auxiliaridade coocorrem num uso verbal, mais auxiliar ele se mostra.

(II) Fusão semântica com um verbo predicador numa única predicação

Por revelar (algum grau de) desgaste semântico em relação à aceção que tem quando usado como predicador, um item verbal é associado à categoria de verbo auxiliar por passar a constituir com o verbo predicador uma unidade que possui significação conjunta: o primeiro elemento (Vauxiliar) desempenha (principalmente) função gramatical, a de indicar as categorias gramaticais do verbo e nuances de aspecto, modalidade e intencionalidade e a de matizar o verbo predicador; e o segundo elemento (Vauxiliado) exerce função lexical, a de estabelecer o estado de coisas em que se encontra(m) projetado(s) o(s) participante(s) no mundo biossocial ou referencial.

(28) Maria [quer passar] no Vestibular.

(29) Maria [começou a estudar] para o Vestibular.

(30) Repentinamente Maria [desatou a rir].

(31) Maria [terminou/acabou de estudar] para o Vestibular.

(32) O professor [deixou/parou de sair] com os amigos.

A interpretação volitiva do primeiro enunciado lhe é conferida pelo verbo “querer”.¹⁸ A possibilidade de focalizar o início ou o fim de um evento/estado de coisas é obtida com o uso de COMEÇAR A/POR e DESATAR A ou TERMINAR DE/POR, ACABAR DE e DEIXAR/PARAR DE, respectivamente. As duas primeiras unidades verbais expressam o aspecto inceptivo do evento (isto é, seu início), ao passo que as três últimas colaboram para exprimir o aspecto conclusivo. Para expressar aspecto inceptivo, conta-se ainda com PASSAR A, DAR DE, PÔR-SE A, DESANDAR A, METER-SE A, por exemplo.

Outras interpretações poderão ser obtidas com a permuta do verbo auxiliar: valor permansivo com CONTINUAR, ESTAR A, FICAR A; sentido resultativo (noção de consecução de um estado de coisas) com CONSEGUIR, CHEGAR A ou VIR A; significado reiterativo com COSTUMAR, VIVER A, entre outros.

Um dos problemas de aplicação desse parâmetro para a identificação de verbos auxiliares ocorre em enunciados em que o valor semântico (de volição, por exemplo) é obtido também com usos de verbo predicador, como é o caso em *Maria quer [que o filho passe no vestibular]/[um carro]*.

(III) Impossibilidade de substituição da estrutura sintagmática introduzida pelo verbo predicador por oração completiva finita

Como o verbo auxiliar e o verbo predicador não finito (Vauxiliado) ocorrem num único domínio predicativo, o verbo predicador e os sintagmas argumentais ou adjuntos que a ele se relacionarem não podem comutar com orações subordinadas completivas finitas (precedidas de “que” ou “se”). Por exemplo:

(33a) Meu filho tem [ido ao colégio].

(33b) *Meu filho tem [que (meu filho) vai/vá ao colégio].

(33c) *Meu filho tem [que [o colega] vai/vá ao colégio].

(34a) O colégio havia [abrigado várias pessoas].

(34b) *O colégio havia [que (o colégio) abriga/abrigava várias pessoas].

(34c) *O colégio havia [que [o diretor] abriga/abrigava várias pessoas].

(35a) Pedro está [lendo muito nos últimos dias].

(35b) *Pedro está [que (Pedro) lê/leia muito nos últimos dias].

(35c) *Pedro está [que [o filho] lê/leia muito nos últimos dias].

Como esses são casos claros de períodos simples, a construção a partir do verbo principal (entre colchetes) naturalmente não pode ser substituída por uma oração completiva, ainda que se mantenha o referente-sujeito do verbo auxiliado.

18 Sugere-se leitura a respeito da relação entre desejo e futuridadade em dados com o verbo *querer*: MACHADO VIEIRA; SOUZA; COSTA, 2021.

Caso haja essa possibilidade, as ocorrências verbais não serão consideradas membros exemplares da categoria de verbo auxiliar; no máximo revelarão comportamento semiauxiliar. É o caso dos dois tipos de ocorrências verbais a seguir:

(A)

(36a) Pedro fez/deixou [ver a importância do estudo ao filho].

(36b) Pedro fez/deixou [(com) que seu filho visse a importância do estudo].

(37a) Pedro mandou [dormir seu filho].

(37b) Pedro mandou [que seu filho dormisse].

Em enunciados com FAZER, DEIXAR e MANDAR causativos, em que o segundo verbo já não partilha identidade referencial com o primeiro verbo, é possível a co-ocorrência com completivas finitas.

(B)

(38a) Maria quer amamentar seu filho. (*querer volitivo*)

(38b) *?Maria quer [que (Maria/ela) amamente seu filho].

(38c) Maria quer [que [alguém] amamente seu filho].

(39a) Maria conseguiu amamentar seu filho. (*conseguir resultativo*)

(39b) *?Maria conseguiu [que (Maria/ela) amamente/amamentasse seu filho].

(39c) Maria conseguiu [que [alguém] amamente seu filho].

Verbos como QUERER, CONSEGUIR, DESEJAR, TENTAR, entre outros, podem coocorrer com completivas finitas, particularmente se um referente-sujeito diferente for introduzido.

A diferença entre os tipos (A) e (B) deve-se, principalmente, à existência ou não de correferência entre as posições de sujeito dos verbos da sequência em auxílio, próximo parâmetro a ser tratado.

(IV) Uma só posição sintática de sujeito com um só referente-sujeito para as formas verbais da unidade complexa

(40a) Meu filho tem ido ao colégio.

(40b) Meu filho tem [meu filho] ido ao colégio.

(41a) Maria vai amamentar seu filho.

(41b) Maria vai [Maria] amamentar seu filho.

(42a) Maria começou a amamentar seu filho.

(42b) Maria começou a [Maria] amamentar seu filho.

(43a) Vândalos querem destruir monumentos públicos.

(43b) Vândalos querem [Vândalos] destruir monumentos públicos

Os dois verbos do complexo apresentam o mesmo referente-sujeito. O verbo predicador (no particípio passado, no primeiro exemplo, e no infinitivo, nos demais exemplos) expressa o evento ativo praticado pelo sujeito e o verbo que o precede e se liga ao mesmo referente-sujeito determina apenas a modalidade – de aspecto – da ação expressa pelo predicador, conferindo-lhe as seguintes nuances: continuidade da ação, intento futuro, início do evento ou volição, respectivamente.

(V) Impossibilidade de substituição da estrutura sintagmática introduzida pelo verbo predicador por uma forma pronominal demonstrativa

Um verbo auxiliar típico não se articula a um verbo predicador que, juntamente com o(s) argumento(s) interno(s) que projetar, possa ser substituído por um pronome demonstrativo como “isso” ou como o pronome átono “o”.

(44) Meu filho tem ido ao colégio, mas minha filha não *o tem / tem **isso**.

Essa restrição sintática já não ocorre com verbos de comportamento semiauxiliar, conforme evidenciam os exemplos a seguir:

(45) O jornalista queria publicar o texto na íntegra, mas seu editor-chefe não o quis / quis **isso**.

A verificação dessa propriedade pode ficar comprometida pela tendência ao apagamento de clíticos de terceira pessoa no Português oral do Brasil. Nessa variedade, é bastante frequente a ocorrência de enunciados como:

(46) Marta vai/quer devolver as provas de seus alunos, mas eu não vou/que-ro.

(47) Tenho estudado, mas meu filho não tem.

(VI) Ocorrência de pronomes átonos complementos em adjacência ao verbo auxiliar.

Em construções nas quais a perífrase é constituída de um verbo auxiliar típico, o esperado é que os pronomes átonos ocorram mais produtivamente em torno deste (em próclise, ênclise ou mesóclise, a depender de fatores condicionadores da ordem pronominal).

As estruturas perifrásticas com TER ou HAVER + particípio ilustram justamente os casos mono-oracionais em que esse critério terá aplicação total – em função, talvez, das especificidades da própria forma participial – ou quase total, se considerarmos a possibilidade de variante (cf. ex. 48d), ainda que esta seja pouco produtiva.

(48a) As chuvas intensas têm destruído plantações no interior de São Paulo.

(48b) As chuvas intensas têm-**nas** destruído.

(48c) As chuvas intensas (não) **as** têm destruído.

(48d) *?As chuvas intensas têm destruído-as.

As estruturas perifrásticas com gerúndio ou infinitivo já não terão comportamento tão regular: ocorrência de complementos clíticos em torno do verbo auxiliar ou do verbo auxiliado. Nesses casos, haverá também a possibilidade da variante pós-complexo verbal.

(49a) A geada está destruindo plantações.

(49b) A geada (não) **as** está destruindo.

(49c) A geada está destruindo-as.

(50a) A professora irá enfrentar a Direção sozinha.

(50b) A professora (não) **a** irá enfrentar sozinha.

(50c) A professora irá enfrentá-**la** sozinha.

Em construções “bioracionais”, ou seja, com dois domínios predicativos, os pronomes átonos ligam-se mais produtivamente à segunda parte da estrutura em auxílio. Nos casos intermediários de semiauxiliarização, oscilam entre a posição adjacente ao primeiro verbo e a posição adjacente ao segundo da sequência.

(51a) A professora (não) deve enfrentar a Direção sozinha.

(51b) A professora (não) deve enfrentá-**la** sozinha.

(51c) A professora (não) **a** deve enfrentar sozinha.

(52a) O jornalista (não) tinha de apresentar o texto na íntegra.

(52b) O jornalista (não) tinha de apresentá-**lo** na íntegra.

(52c) O jornalista (não) tinha de **o** apresentar na íntegra.

(52d) O jornalista (não) tinha-**o** de apresentar na íntegra.

(52e) O jornalista (não) **o** tinha de apresentar na íntegra.

A rentabilidade desse critério para a definição do estatuto de um verbo como auxiliar ou semiauxiliar fica prejudicada por fatores, como: (i) o caráter variável da ordem dos pronomes clíticos em complexos verbais nos diferentes dialetos nacionais da língua portuguesa; (ii) a dificuldade em se definirem como aceitáveis ou não determinadas alternativas de colocação testadas numa perífrase verbal, tendo em vista a sua baixa ou nula produtividade no uso em variedades brasileiras; e (iii) a possibilidade de não correspondência entre a posição dos pronomes oblíquos no enunciado e sua ligação prosódica com o verbo auxiliar ou o verbo principal na perífrase, nos casos de variantes intracomplexo verbal.

(VII) Impossibilidade de incidência de advérbio de negação sobre o domínio do verbo predicador

Um verbo auxiliar típico ocorrerá numa construção que só admita um advérbio de negação, preferencialmente posicionado à esquerda da sequência em auxílio para que a negação possa modificar toda a predicação.

(53a) Meu filho **não** tem ido ao colégio.

(53b) *Meu filho tem **não** ido ao colégio.

(54a) Pedro **não** está lendo muito nos últimos dias.

(54b) * Pedro está **não** lendo muito nos últimos dias.

Já em enunciados “bioracionais”, cada predicação poderá ser afetada por um advérbio de negação frásica:

(55a) Pedro **não/nunca** fez/deixou ver a importância do estudo ao filho. (A negação incide sobre *deixar/fazer e ver*.)

(55b) Pedro fez/deixou seu filho **não** ver a importância do estudo. (A negação só incide sobre *ver*; *deixar/fazer* tem caráter afirmativo.)

(56a) Maria **não/nunca** quer/deseja/consegue amamentar seu filho. (não querer/desejar/conseguir, não amamentar)

(56b) Maria quer/deseja/consegue **não/nunca** amamentar seu filho. (querer/desejar/conseguir, não amamentar)

(57a) A professora **não** pode comparecer à festa do colégio. (não poder, não comparecer à festa do colégio)

(57b) A professora pode **não** comparecer à festa do colégio. (poder/ser possível/capaz, não comparecer à festa do colégio)

(VIII) Impossibilidade de circunstância temporal (adjunto adverbial de tempo) que afete apenas o domínio do verbo predicador

Uma predicação contendo verbo auxiliar típico seguido de verbo predicador não admite dois adjuntos adverbiais de tempo que representem circunstâncias totalmente distintas para os eventos, pelo fato de uma estrutura mono-oracional (um período simples) só ter um domínio temporal, no qual ocorre *Vauxiliar+Vauxiliado*. Se existirem dois domínios de expressão temporal, não se estará diante de uma, mas de duas predicções, de uma estrutura bi-oracional. Observem-se os exemplos a seguir:

(58) ***Hoje** o colégio tem abrigado várias pessoas **amanhã**.

Entretanto,

(59) **Ontem** Pedro queria ir ao teatro **amanhã**, hoje já mudou de ideia.

(60) **Ontem** Pedro podia ir ao teatro **amanhã**, hoje não pode mais.

(61) **Ontem** Maria ia tentar [começar a] amamentar seu filho **hoje**, mas hoje já mudou de ideia.

A agramaticalidade/inaceitabilidade do primeiro enunciado deve-se ao fato de os dois advérbios de tempo, incompatíveis semanticamente, afetarem o mesmo domínio predicativo. O domínio do verbo predicador infinitivo é temporalmente dependente do domínio do verbo auxiliar, razão pela qual não são admitidos modificadores com valores distintos no eixo temporal.

Já a gramaticalidade dos demais decorre da possibilidade de os circunstantes de tempo relacionarem-se a domínios de predicação distintos (“querer ontem/ir ao teatro amanhã”, “poder ontem/ir ao teatro amanhã”, “ir/ter a intenção ontem”/“[começar a] amamentar seu filho hoje”). Num domínio, há a marcação de uma modalidade ou intencionalidade; e noutro, a indicação de um evento/estado de coisas.

Embora citado por diferentes autores, esse critério acarreta problemas quanto à aceitabilidade dos dados linguísticos que se testam. Mesmo em construções com verbos menos gramaticais, nem sempre é aceitável mais de um domínio temporal, como se nota em **? O delegado, ontem, não quis/conseguiu divulgar o nome do suspeito hoje*. Até onde foi testado, parece contribuir pouco para a categorização dos verbos semiauxiliares, já que é difícil encontrar sequências em auxiliação com as quais se verifique a possibilidade de mais de um domínio temporal. Ao que parece, sua aplicação fica restrita a determinadas relações temporais entre os domínios predicativos que podem figurar numa unidade complexa, a casos em que o domínio do verbo no infinitivo pode ter ou tem certa independência temporal do domínio do primeiro verbo, do verbo semiauxiliar. Em *Pedro decidiu (ontem) falar (hoje) com os colegas*, a localização temporal da situação apresentada por *falar com os colegas* é posterior à da situação introduzida por *decidir*, ao momento em que Pedro tomou a decisão. Semelhante é o caso de *Pedro ontem queria ir ao teatro amanhã* (cf. ex. 59).

(IX) Comportamento em bloco do verbo auxiliar e do verbo auxiliado diante de transformações sintáticas (passiva e interrogativa)

A ‘apassivação’ e a ‘interrogação’ não afetam cada um dos verbos individualmente, mas os dois como se de um só se tratasse. Por exemplo:

(62a) O arquiteto tem estudado o projeto.

(62b) O projeto tem sido estudado pelo arquiteto.

(62c) O que o arquiteto tem estudado?

(62d) *O que o arquiteto tem?

(63a) Estou escrevendo este trabalho.

(63b) O trabalho está sendo escrito por mim.

(63c) O que você está escrevendo?

(63d) *O que você está?

No caso da transformação interrogativa, o domínio do verbo principal (no participio, no primeiro exemplo; no gerúndio, no segundo) normalmente não pode ocorrer como resposta a uma pergunta que apenas tenha o auxiliar, como: *O que o arquiteto tem? Estudado o projeto. O que está? Escrevendo este trabalho*. Isso revela que os dois verbos funcionam como uma unidade sintática. Em perífrases cujos componentes têm menor integração, essa propriedade já não se mantém:

(64a) O arquiteto quer/tentou examinar o projeto.

(64b) O que quer/tentou o arquiteto? Examinar o projeto.

(65a) O professor deixou usar o laboratório.

(65b) O que o professor deixou? Usar o laboratório.

Os usos de unidades verbais que exibem todas as onze propriedades anteriormente descritas são considerados membros prototípicos da categoria AUXILIAR, do polo extremo do *continuum* de gramaticalização. A esse pertencem os temporais TER e HAVER, quando usados com comportamento gramatical em estruturas de tempo composto.

Em geral, há unanimidade quanto ao caráter auxiliar desses dois verbos. Não obstante, alguns linguistas vinculam a essa classe outras unidades verbais, a depender dos critérios de *auxiliaridade* contemplados. LOBATO (1975) e MATEUS *et al.* (2003), por exemplo, estão entre os que consideram outros verbos como membros da categoria de auxiliar típico. Aquela insere nessa classe SER auxiliar de voz passiva e ESTAR. Essas incluem, além desses dois, os verbos auxiliares aspectuais ANDAR, FICAR, IR e VIR seguidos de gerúndio.

O importante é mostrar, com isso, o caráter fluido da delimitação de algumas unidades verbais. Não se nega a relevância dessa delimitação, porém mais relevante do que fazer uma classificação é o conhecimento que se pode obter com a análise dos fatos linguísticos fundamentada na determinação e na aplicação de critérios de *auxiliaridade*, que sejam observados considerando-se a plasticidade e a dinamicidade da relação forma-função em (con)texto.

9.4.2 VERBOS SEMIAUXILIARES: UMA PROPOSTA DE CATEGORIZAÇÃO ESCALAR DE VERBOS EXEMPLARES DE DIFERENTES CONTORNOS DE GRAMATICALIDADE E, PORTANTO, DE NÃO COMPOSICIONALIDADE A OUTRA UNIDADE VERBAL (NO PARTICÍPIO, INFINITIVO OU GERÚNDIO)

Os elementos que não revelam todas as onze características são considerados semiauxiliares (“auxiliantes”, segundo Pottier *apud* LOBATO, 1975) – classe que, por sua vez, abrange membros com diferentes comportamentos: alguns resultam de um processo de auxiliarização mais completo; outros, ainda são associados a estágios intermediários ou híbridos de comportamento.

Para ilustrar a última fase de gramaticalização a que um item verbal pode chegar, cita-se o caso de HAVER (*habere*), em contextos como os de *cantare habeo* (> *cantarei*): uma de suas extensões de sentido, empregada com papel instrumental (comportamento gramatical), passou de um significante livre a um significante preso (morfe-ma de futuro), ou seja, submeteu-se a um processo diacrônico de gramaticalização/‘morfologização’. São raros, entretanto, os casos de verbos que alcançam tal estágio. Em geral, há casos incompletos de gramaticalização de verbos,

que, obviamente, geram dúvidas quanto à categoria a que pertencem. Além disso, nem tudo ocorre conforme as generalizações convencionalizadas socialmente a partir do já experienciado e, quando produtivo, mais esperado em novas enunciações. Conta-se, na linguagem e na língua viva que se (re)configuram por ação dos indivíduos e comunidades que as exploram, com combinações compatíveis às condições construcionais já entrincheiradas cognitivamente/*matches* e com incompatibilidades/*mismatches*.

Pode-se traçar para esses casos uma escala de *auxiliaridade*, conforme o seu grau de afastamento da categoria de verbo predicador e de aproximação da categoria de verbo auxiliar. Tal escala compreende desde extensões de uso/sentido mais próximas da categoria de verbo auxiliar típico (descrita no item anterior) e, por conseguinte, mais afastadas da de verbo predicador até extensões mais afastadas da categoria verbo auxiliar e bem mais próximas da de verbo predicador. A escala pode ser definida por meio de 5 subclasses de verbos semiauxiliares.

1º grau de afastamento do polo de <i>auxiliaridade</i>
SER, ESTAR e FICAR seguidos de participio (em construções de voz passiva analítica).

Propriedades que os fazem diferentes dos membros típicos da categoria Auxiliar:

É possível substituir o participio passado que os acompanha por um pronome átono demonstrativo.

(66) Ana foi elogiada pelo professor, mas sua irmã não o foi.

Percebe-se neles comportamento similar ao de verbo de ligação, que, por sua vez, opera sobre elementos não verbais: não só pelo fato de, em certos contextos, as construções em que aparecem possibilitarem leitura estativa (cf. exemplos (68) e (69)), mas também por conta da variabilidade da forma de participio, que se manifesta pela relação de concordância com o constituinte sujeito.

(67) As atletas foram expulsas da competição pela comissão organizadora do evento.

(68) Os bandidos ficaram cercados pelos policiais.

(69) A reforma da escola está prejudicada por falta de dinheiro.

Os verbos auxiliares típicos TER/HAVER operam sobre um participio invariável. Em “As meninas têm escrito cartas”, não se estabelece relação de concordância entre *escrito* e qualquer um dos argumentos.

Têm a restrição de basicamente auxiliarem Vprincipais com estrutura transitiva direta, embora, eventualmente, sejam empregados como auxiliares de alguns verbos de estrutura diferente – como em “As leis são obedecidas.” ou:

“Entretanto, ele não tem bem a certeza se é preciso ser um entendido em amor, para escrever um livro destes. Mas, para os devidos efeitos, este homem casado há 17 anos (dedica o livro à mulher, Alexandra), garante que ama, que é amado, que já gostou e não foi correspondido, que já foi gostado e não correspondeu; que já traiu e que

já foi traído. Se não passou por todos os estádios do amor, pelo menos já fez escala em apeadeiros suficientes na linha dos afetos para ter experiência no que escreve: com conhecimento de causa da dor de corno, dos remorsos ou das noites em branco por causa de um telefonema.” [Diário de Notícias, Notícias Magazine, *O último dos românticos*]¹⁹

2º grau de afastamento do polo de <i>auxiliaridade</i>
ESTAR, VIR, IR, FICAR, ANDAR aspectuais em construções com Vpredicador no gerúndio. IR, VIR, HAVER (<i>de</i>) temporais em construções com Vpredicador no infinitivo.

Apesar de nessas extensões verbais se notar a perda significativa dos semas caracterizadores de seus empregos como verbo predicador, há características que as afastam do polo da *auxiliaridade*.

Propriedades que os fazem diferentes dos membros típicos da categoria Auxiliar:

Não se juntam ao particípio passado invariável, forma nominal do verbo considerada tipicamente de natureza não oracional (não frásica), pelo fato de não poder ser retomada por uma forma nominal demonstrativa.

São permitidos complementos clíticos em adjacência ao verbo predicador no infinitivo ou no gerúndio.

(70) O professor ficou corrigindo os testes. / O professor ficou corrigindo-**os**.

(71) Pedro vem estragando muitas camisas no futebol. / Pedro vem estragando-**as** no futebol.

(72) O professor vai corrigir os testes. / O professor vai corrigi-**los**.

(73) A diretora há de considerar minhas inquietações. / A diretora há de considerá-**las**.

3º grau de afastamento do polo de <i>auxiliaridade</i>
PODER, DEVER modais ²⁰ em construções com Vpredicador no infinitivo. ESTAR, FICAR, ANDAR, VOLTAR, TORNAR, COSTUMAR, CONTINUAR, PERMANECER, COMEÇAR, PASSAR, PÔR-SE, METER-SE, CHEGAR, PEGAR aspectuais (acurativos) seguidos da preposição <i>a</i> em construções com Vpredicador no infinitivo. CONTINUAR aspectual seguido de Vpredicador no gerúndio.

Embora se especializem na indicação de determinados matizes semânticos (aspectuais ou modais) que se mostram produtivos na língua em uso, apresentam propriedades que os afastam da configuração prototípica de verbo auxiliar.

Propriedades que os fazem diferentes dos membros típicos da categoria Auxiliar:

¹⁹ <https://www.noticiasmagazine.pt/2013/o-ultimo-dos-romanticos/estilos/comportamento/1881/>. Acesso em: 30 mar. 2022.

²⁰ Os verbos modais exprimem uma atitude do locutor em relação ao conteúdo do enunciado veiculando valores, como: obrigatoriedade, probabilidade, possibilidade, permissão.

O critério de desgaste semântico não se verifica da mesma maneira nessa subclasse de verbos. Apesar do papel instrumental que assumem e do valor de aspecto (duração ou fase de desenvolvimento do estado de coisas), intencionalidade ou modalidade que marcam, mantêm muitos dos traços de seu valor lexical, de seu significado objetivo.

A depender do efeito de sentido pretendido pelo usuário, verbos como DEVER e PODER, por exemplo, podem assumir, e até condensar num mesmo contexto enunciativo, mais de um significado distinto: (i) probabilidade/possibilidade; e (ii) obrigação/permissão. Observem-se os exemplos a seguir:

(74) O clube **pode** participar do próximo campeonato. (Leituras possíveis: “É possível que o clube participe do próximo campeonato.” ou “O clube tem permissão [do Tribunal de Justiça Esportiva] para participar do campeonato.”)

(75) O réu **deve** dizer a verdade. (Leituras possíveis: “O réu é obrigado a dizer a verdade.” / “É provável que o réu diga a verdade.”)

O advérbio de negação pode atuar sobre o verbo predicador no infinitivo.

(76) Os meninos podem **não** temer o escuro.

(77) Os meninos voltaram a **não** oferecer doces à irmã.

(78) Os professores continuam a **não** manifestar suas opiniões na reunião.

São permitidos complementos clíticos em adjacência ao verbo predicador no infinitivo.

(79) Os meninos podem temê-**lo**.

(80) Os meninos voltaram a oferecer-**lhe** doces. / Os meninos voltaram a oferecê-**los** à irmã.

(81) Os professores continuam a manifestá-**las** na reunião.

4º grau de afastamento do polo de <i>auxiliaridade</i>
TER (<i>de/que</i>) e HAVER <i>de</i> modais em construções com Vpredicador no infinitivo. ESTAR, FICAR, DEIXAR, ACABAR, PARAR, TERMINAR, CESSAR, DAR aspectuais seguidos de preposição <i>de</i> , <i>para</i> ou <i>por</i> em construções com Vpredicador no infinitivo.

Propriedades que os fazem diferentes dos membros típicos da categoria Auxiliar:

Os clíticos complementos de verbos predicadores no infinitivo mais raramente passam ao seu domínio (ao domínio da “frase matriz”). Gonçalves e Costa (2002) chegam a afirmar, com base na norma de uso dos pronomes átonos da variedade europeia, que esses verbos semiauxiliares não permitem a extração de pronomes átonos complementos do domínio do infinitivo para o seu domínio.

Algumas extensões de uso impõem restrições de seleção quanto à natureza aspectual dos verbos predicadores sobre os quais operam, limitando a classe de verbos predicadores. Um verbo como DEIXAR limita a classe de predicados verbais que com ele se articulam (?*No ano passado, Pedro deixou de nascer/morrer.*). Não é qualquer verbo principal que pode ocorrer com ele.

Além de permitirem que o advérbio de negação incida sobre o verbo predicador com que se combinam, aceitam mais de uma instância de negação frásica (*O professor (não) ficou de não interferir na decisão dos alunos*).

É tênue o limite entre o terceiro e o quarto grau. Para a delimitação desses dois níveis, colabora também o tipo de elemento/preposição que liga verbo semiauxiliar e verbo principal: o baixo peso fonético-fonológico (pouca quantidade de material) e o esvaziamento de sentido da preposição “a” versus o maior peso fonético-fonológico e a maior colaboração semântica das outras preposições. Outra opção de categorização é desconsiderar as diferenças aqui apontadas e reunir num só nível de *auxiliaridade* esses usos semiauxiliares.

5º grau de afastamento do polo de <i>auxiliaridade</i> (mais próximo do polo de <i>lexicalidade</i>)
TENTAR, QUERER, ESPERAR, DESEJAR, GOSTAR (de), LOGRAR, CONSEGUIR, OUSAR, ATREVER-SE, PARECER, PRETENDER, TENCIONAR, e tantos outros citados em gramáticas escolares (PRECISAR, entre estes), seguidos de Vpredicador no infinitivo.
MANDAR, FAZER, DEIXAR, LEVAR causativos em construções com Vpredicador no infinitivo.
VER, OLHAR, OUVIR, SENTIR, SABER perceptivos/sensitivos em construções com Vpredicador no infinitivo.

Nesse grau, estão reunidos os verbos cujo emprego semiauxiliar se situa na fronteira com seu emprego lexical, de verbo principal. O comportamento híbrido desses verbos leva muitos estudiosos a questionar a pertinência de sua inclusão no rol dos semiauxiliares, principalmente a dos semiauxiliares causativos e perceptivos/sensitivos.

Propriedades que os fazem diferentes dos membros típicos da categoria Auxiliar:

Muitas das unidades verbais supracitadas praticamente não revelam desgaste semântico e mantêm sua aceção primária (quase) inalterada.

A forma verbal flexionada e a não flexionada, mantendo cada uma o seu valor lexical, ocorrem em estruturas “bi-oracionais”, já que têm “complemento” de natureza frásica e não de natureza verbal/não oracional. As construções formadas com esses verbos sujeitam-se a duas interpretações: (i) é possível considerar que há dois domínios predicativos relacionados por subordinação, logo um período composto; e (ii) é possível entender que, em estrutura superficial, esses verbos se submetem a um processo de reestruturação que, ao promover sua integração a um verbo predicador, gere, pelo menos no nível da representação final, comportamento de construção “mono-oracional” ou uma unidade complexa.

A barreira frásica existente entre eles e os verbos predicadores não flexionados aos quais se articulam pode ser percebida em enunciados em que cada um dos verbos da sequência pode se relacionar a noções de tempo diferentes (*João ontem queria/tencionava/prendia ir ao cinema hoje*; hoje já mudou de ideia). Essa barreira pode ser enfraquecida em casos em que houver uma única cadeia temporal (**?João ontem quis ir ao cinema hoje*) e, assim, eles formarem complexos verbais de menor integração.

É possível comutar os domínios dos verbos predicadores não flexionados que os seguem por uma completiva finita (introduzida por conjunção integrante *que* ou *se*). Logo, tais verbos têm “complementos” de natureza frásica, visto que estes podem ser substituídos por completivas finitas. (João deseja [encontrar-se com a namorada no cinema]. João deseja [que (eles/ele e a namorada) se encontrem no cinema].)

No caso da transformação interrogativa, o domínio do verbo principal pode ocorrer como resposta a uma pergunta que apenas tenha o auxiliar (Que deseja João? Encontrar-se com a namorada no cinema.)

O marcador de negação frásica pode ocorrer apenas sobre o “domínio frásico encaixado”. (João deseja não encontrar a namorada no cinema. Porém: *João tem não encontrado a namorada no cinema.)

Além de permitirem que o advérbio de negação incida sobre o verbo predicador com que se combinam, aceitam mais de uma instância de negação frásica.

(82) Os professores não conseguiram não manifestar suas opiniões na reunião.

(83) Não gostaria de não me manifestar sobre o assunto.

(84) A professora não mandou não sair.

Esses verbos contribuem para definir as condições semânticas de preenchimento da posição de sujeito. Por exemplo, o emprego de MANDAR é incompatível com um sujeito não humano; o de VER, com um sujeito inanimado. O emprego semiauxiliar de QUERER ora restringe os termos que podem ocupar o lugar sintático de sujeito, ora parece não impor restrições semânticas a esse constituinte oracional, conforme se nota em certos enunciados produzidos no português brasileiro: “ESTAR + QUERER no gerúndio + Vpredicador no infinitivo” (*O tempo está querendo mudar, O carro está querendo enguiçar* ou *Está querendo chover*), estrutura em que QUERER parece indicar que “algo está prestes a ocorrer/começar” .

Com alguns desses verbos, o critério de um só referente-sujeito não se mantém (*João viu começar o tumulto*).

A diferença entre alguns dos verbos listados nesse quinto grau de afastamento do polo de *auxiliaridade*, que até poderia acarretar a delimitação de um sexto grau, diz respeito ao fato de que complexos verbais com QUERER, CONSEGUIR, TENTAR, entre outros, têm comportamento diferente do comportamento de complexos verbais com semiauxiliares causativos e perceptivos/sensitivos, se for considerada a possibilidade de substituição do domínio do verbo principal por completiva finita associada à manutenção da identidade referencial do sujeito dos componentes da perífrase verbal.

(85) Eu quero *amamentar meu filho*/ ?*que (eu) amamente meu filho* / *que (alguém) amamente meu filho*. (com o mesmo referente-sujeito, enunciados similares; com sujeitos diferentes, enunciados semanticamente distintos)

(86) Pedro mandou *dormir (a) o filho* / *que o filho dormisse*. (com sujeitos diferentes, enunciados similares)

Na perífrase com MANDAR, o referente-sujeito de cada verbo é diferente. Na perífrase com QUERER, o referente-sujeito se mantém.

O sistema de graus de (semi-)gramaticalização verbal aqui estabelecido com base em onze critérios de *auxiliaridade* poderá, a depender do nível de escolaridade com que o professor trabalhe, ser substituído por um sistema simplificado que seja definido a partir de um conjunto mínimo de propriedades reconhecidas num verbo auxiliar típico.

Por exemplo, a partir da exigência de que a sequência em auxiliação pertença a um período simples (um domínio de predicação) e o verbo auxiliar seja uma extensão que se especializa na indicação de tempo, voz, aspecto ou modo/modalidade, as condições minimamente envolvidas na definição de graus de gramaticalização verbal são estas: (i) um só referente-sujeito para Vauxiliar e Vauxiliado; (ii) um único domínio de negação; e (iii) impossibilidade de substituir por uma oração subordinada finita a estrutura sintagmática a partir do verbo predicador. Com esses parâmetros, o conjunto de verbos essencialmente auxiliares amplia-se. Incluirá, além de TER e HAVER temporais, os verbos dos graus 1 e 2. O conjunto de verbos semiauxiliares incluirá os verbos dos demais graus, subdividindo-se, por sua vez, em três subconjuntos: (i) o dos verbos mais gramaticais (que reúne os verbos dos graus 3 e 4); (ii) o dos verbos menos gramaticais (os do grau 5 que mantêm a condição de identidade referencial do sujeito com o verbo auxiliado); e (iii) o dos verbos ainda menos gramaticais (os verbos causativos e sensitivos no grau 5, que não atendem a qualquer dessas condições, restando-lhes apenas a propriedade de sinalizarem ou realçarem valor de causalidade e percepção). Como foi dito, é até possível excluir esse terceiro subconjunto da escala de (semi-)auxiliaridade.

No estudo da categorização verbal, o importante é levar o aluno a ter condições de (re)conhecer e discutir as definições e características das categorias auxiliar e semiauxiliar com base num exame criterioso de dados e numa visão escalar dos vínculos categoriais do usos verbais (uns mais exemplares das categorias, outros mais periféricos).

9.4.3 VERBOS SUPORTES

O usuário da língua portuguesa conta, ainda, com perífrases verbo-nominais constituídas a partir da atuação de um verbo suporte sobre um elemento não verbal (em geral, um constituinte nominal – substantivo ou adjetivo). Em determinados enunciados, o centro semântico-sintático da predicação não está no constituinte verbal, mas no componente não verbal que se alia àquele na formação da perífrase. Por exemplo:

(87a) Pedro fez queixa do cão ao vizinho.

(88a) Pedro tem preocupação com o filho.

(89a) Pedro deu uma passada na reunião.

(90a) Diogo levou/tomou um susto.

(91a) Carlos deixou seu filho assustado.

(92a) Carlos ficou entusiasmado com a nova proposta de emprego.

É muito produtiva nas línguas a construção de predicador complexo: [Vsuporte + elemento não verbal_{unidade predicante}]-predicador complexo. O deslocamento do núcleo irradiador da predicação do verbo para o elemento não verbal dá-se em virtude da: (i) convencionalização do verbo como recurso gramatical gerador de verbos a partir de elementos que não têm esse estatuto (embora, em alguns casos possam ter alguma verbalidade inerente); e (ii) conseqüente integração semântica e sintática desse verbo ao elemento não verbal que redundava numa locução/perífrase verbo-nominal, que é percebida como unidade de funcionamento similar ao de um predicador simples.

Uma das manifestações do caráter perifrástico dessas construções é a existência, em muitos casos, de verbos predicadores (Vplenos/Vprincipais) com significado equivalente ao do predicador complexo.

(87b) Pedro queixou-se do cão ao vizinho.

(88b) Pedro preocupa-se com o filho.

(89b) Pedro passou na reunião.

(90b) Diogo assustou-se.

(91b) Carlos assustou seu filho.

(92b) Carlos entusiasmou-se com a nova proposta de emprego.

Nesses exemplos, verifica-se a possibilidade de substituição das perífrases por verbos predicadores cognatos ao elemento nominal incorporado. Porém, nem sempre há esse tipo de correspondência, conforme revelam os enunciados a seguir.

(93) A frota de ônibus não dá vazão à demanda de passageiros.

(94) Ana faz ginástica todos os dias.

Ainda assim, percebe-se a integração semântica e sintática entre os componentes (verbal e nominal), o comportamento gramatical do verbo decorrente de seu esvaziamento semântico e de seu emprego instrumental sobre elementos nominais de qualquer classe semântica, propriedades responsáveis pela interpretação da estrutura *V + ñV* (*fazer queixa, deixar assustado, deu bom*)²¹ como uma unidade complexa com função predicante.

Diz-se, portanto, que um verbo atua como *Vsuporte* toda vez que se associa a um elemento não verbal de tipo “especial” (com função predicante (secundária)²² ou ao

21 “Não é que essa roupa deu bom pra apresentação de dança?” <https://twitter.com/ParkJMDancer/status/1508174314994118659> Acesso em: 30 mar. 2022.

22 De acordo com DIK (1997), as categorias básicas de predicado (Verbo, Nome e Adjetivo, com determinadas funções primárias na língua: Verbo com função predicante, Nome com função referencial e Adjetivo com função atributiva) podem, a depender do contexto, ter seus empregos prototípicos convertidos em usos derivados (secundários) – coincidentes, em maior ou menor grau, com usos

qual se quer atribuir tal função), formando uma expressão sintática que tem significação particular e (relativamente) “indissolúvel”, não composicional.

O verbo suporte é um recurso da língua usado com frequência e sistematicidade pelos falantes para a formação de unidades lexicais. Apresenta regularidade de função: operador/marcador de verbalização de elementos não verbais (cf. MACHADO VIEIRA, 2001, 2018). Verbo responsável pela atribuição de função predicante a elemento não verbal, o verbo suporte manifesta algum grau de alteração semântica em comparação com o verbo predicador; assume valor mais genérico e abstrato que o deste. Passa a ter, portanto, significado “instrumental/gramatical”, já que funciona como instrumento morfossintático na formação de predicador complexo derivado. Forma com o elemento com que se combina uma unidade semântica e sintática.

Um verbo suporte tem comportamento “léxico-gramatical”: contribui para a formação semântica do predicado verbo-nominal, apesar de o item nominal ser o principal responsável pelas propriedades semânticas do predicado. Nota-se sua colaboração para o efeito semântico do predicado complexo, comparando-se os seguintes exemplos:

(95a) O menino levou/tomou uma surra/um esbarrão do colega.

(95b) O menino deu uma surra/um esbarrão no colega.

(96a) O professor tem confiança em seus alunos.

(96b) O professor dá/passa confiança a seus alunos.

(97a) Dei força para ele estudar para o concurso.

(97b) Fiz força para passar no concurso.

(97c) Fiz força para abrir essa porta pesada.

(98a) O professor pôs medo nos alunos. / O professor fez medo aos alunos.

(98b) Trovões/Pessoas vestidas de Papai Noel dão medo em algumas crianças.

(98c) Muitas pessoas têm medo de trovão.

Extensão semântico-sintática mais ou menos afastada da que tem como verbo predicador, o verbo suporte mantém, em maior ou menor grau, traço(s) do significado lexical básico daquele. Em virtude disso, contribui para definir o tipo de situação que a predicação designa: uma situação de evento agentivo e/ou causativo (*Pedro fez queixa do cão ao vizinho*), uma situação de estado de posse (de entidade abstrata) (*Pedro tem preocupação com o filho*), uma situação de evento de tipo transferencial (*Pedro deu sua contribuição à equipe*) e uma situação de evento resultativo (*Pedro levou um susto*). Já em *deu bom*, conforme ocorre no exemplo posto em nota rodapé anterior, *dar* contribui para pôr em proeminência um resultado, uma situação de evento resultativo (*resultar/dar bom* ou *funcionar, servir*).

O verbo suporte tem implicações semântico-discursivas relevantes, que devem ser exploradas no tratamento didático-pedagógico e/ou acadêmico-descritivo do proces-

so de produção e leitura de textos.²³ A opção por um ou outro operador de verbalização pode conferir à predicação, por exemplo, diferentes valores: *resultativo* (*O menino tomou uma surra do colega*), *agentivo/causativo* (*O menino deu uma surra no colega* ou *Trovão dá medo nas pessoas*) ou *estativo* (*Muitas pessoas têm medo de trovão*); mais causativo (*O professor pôs medo nos alunos*, em que o causador (agente) é o responsável intencional pelo evento) ou menos causativo (*Trovões/Pessoas vestidas de Papai Noel dão medo em crianças*, em que a entidade-causadora (força/tema) responsável pelo estado de coisas não o causa intencionalmente); evento do tipo “transferencial” (*Dei força para ele estudar para o concurso*) ou “não transferencial” (*Fiz força para passar no concurso*).

Assim sendo, o verbo suporte contribui para a apresentação de um evento, que é especificado, principalmente, pelo elemento predicante ao qual se associa. Como revela desgaste semântico em decorrência do processo de gramaticalização ao qual se submete, atua na lexia verbo-nominal, principalmente, para codificar as categorias gramaticais do verbo e, assim, dar suporte gramatical ao elemento nominal ao qual se alia. Partilha com o elemento não verbal (sintagma nominal, sintagma adjetival, sintagma preposicional) a função de atribuir papel temático ao(s) argumento(s). Isso pode ser percebido pelos diferentes papéis temáticos dos termos que preenchem a posição de sujeito dessas perífrases: em construções com verbo suporte FAZER, por exemplo, encontram-se termos com as funções de agente (*Pedro fez bagunça no quarto*), força (*O temporal fez um estrago na lavoura*), paciente (*O paciente deste leito fez uma cirurgia*), tema (*Uma pedra fez um buraco no vidro do meu carro*).

Diferentemente do emprego de verbo predicador, um verbo suporte típico (quase) não determina condições semânticas para o constituinte sujeito. Essas condições lhe são impostas principalmente pelo elemento não verbal, núcleo semântico da perífrase verbo-nominal.

O primeiro elemento da unidade (o Vsuporte) desempenha função (léxico-)gramatical (reforço do sentido do predicado e apoio para a expressão das categorias morfossintáticas de tempo, modo, número e pessoa); e o segundo (elemento nominal, frequentemente), a função lexical (especificar a atividade expressa pela unidade). O verbo suporte (ou “verbo de apoio”) comporta o valor categorial de predicado verbal (assim como um componente de aspecto) e o complemento predicativo/predicante supre o resto do significado. O conteúdo semântico de <fazer + elemento predicante> é o de impor ao(s) argumento(s) complemento(s) um evento de determinada natureza.

A perífrase verbo-nominal em questão dispõe de outra propriedade que caracteriza uma estrutura perifrástica prototípica: a correspondência entre o referente/sujeito gramatical do verbo suporte e o referente/sujeito do elemento não verbal do tipo predicado nominal, à semelhança do que ocorre com Vauxiliar e Vauxiliado. Partilha com o elemento com que se combina o referente-sujeito (“um todo funcional”).

23 Vale conferir um exemplo desse potencial no artigo: *Jornal da USP. Maratona da preguiça: um neologismo que não sai do sofá*. Texto de Marcelo Módolo e Henrique Braga. *Saense*. <https://saense.com.br/2020/05/maratona-da-preguica-um-neologismo-que-nao-sai-do-sofa/>. Publicado em 29 de maio (2020). Acesso em: 30 mar. 2021.

As predicções com perífrases verbo-nominais exibem, não obstante, outras propriedades que são responsáveis pelo estatuto semigramatical do verbo suporte. Destacam-se aqui quatro dessas propriedades.

(I) É possível dupla análise dos sintagmas preposicionais complementos.

(99a) É [*entre o homem e a mulher profissional*] que alguns empregadores **fazem discriminação**.

(99b) É **discriminação** [*entre o homem e a mulher*] que alguns empregadores **fazem**.

Nem sempre há sintagmas preposicionais com estatuto de complemento nas predicções com verbo suporte. Inclusive, isso está vinculado ao fato de as perífrases verbo-nominais com verbo suporte serem estruturas linguísticas com as quais os usuários da língua podem contar quando têm a finalidade de obter um predicado de valência reduzida que lhes permita apresentar genericamente um evento (sem entrar em detalhes), prescindir de complementação.

(100) O governador fez uma declaração.

(101) O professor deu uma sugestão.

(102) Tenho preocupações.

Uma vez que o complemento de um predicador nominal é menos exigido discursivamente do que o complemento de um predicador verbal, o falante pode optar por não declarar o(s) argumento(s) interno(s) projetados pelo predicador nominal (*declaração, sugestão, preocupações*).

Porém, quando ocorre, o sintagma preposicionado, que é argumento interno, pode ser entendido como complemento da perífrase $V_{\text{suporte}} + \tilde{n}V$ ou simplesmente do verbo.

(II) O elemento não verbal incorporado tem comportamento semelhante ao de complemento de um verbo predicador, sob extração, em estruturas interrogativas ou clivadas.

(103a) Pedro fez queixa do cão ao vizinho.

(103b) O que é que Pedro fez ao vizinho? Queixa do cão.

(103c) Foi queixa do cão que Pedro fez. / ?Foi queixa que Pedro fez do cão.

(104a) Pedro tem preocupação com o filho.

(104b) O que é que Pedro tem? Preocupação com o filho.

(104c) É preocupação com o filho que Pedro tem. / É preocupação que Pedro tem com filho.

(III) Assim como ocorre com certos verbos semiauxiliares, é possível a reativação do elemento não verbal que se integra ao verbo suporte mediante substituição por um pronome (clítico ou não) ou repetição.

Confirmam-se os exemplos a seguir:

(105) Pedro fez queixa do cão ao vizinho. Pedro fê-la porque já não suportava mais o cão. O vizinho ouviu com atenção a queixa.

(106) Pedro deu sua contribuição à reunião. Ela foi decisiva para o bom andamento da reunião.

(107) Alguns empregadores fazem discriminação entre o homem e a mulher. Fazem-na porque sabem que não haverá maiores repercussões. Essa discriminação tem sido o motivo de muitos processos trabalhistas.

A retomada do elemento não verbal atende, na verdade, a pelo menos duas das finalidades discursivas vinculadas ao emprego de predicado complexo com verbo suporte: coesão semântico-sintática (conforme ocorre nos exemplos 105 e 106) e versatilidade textual (diversificação dos enunciados, enriquecimento do repertório de possibilidades para estruturação de um texto).

(IV) O verbo suporte pode ser apagado em proveito de uma estrutura do tipo “de + SN argumento externo/sujeito (+ SN argumento interno)”. É possível formar grupo nominal que conserva o(s) complemento(s) do “elemento não verbal predicante”, mediante a redução do verbo suporte no domínio de uma oração relativa. Já um verbo predicador (principal) não pode ser apagado, sem prejuízo semântico-discursivo.

(108a) Alguns empregadores **fazem discriminação** entre o homem e a mulher.

(108b) **A discriminação que** alguns empregadores **fazem** entre o homem e a mulher...

(108c) **A discriminação (de alguns empregadores)** entre o homem e a mulher... (enunciado com [Redução de Vsuporte], o apagamento de “fazer”)

(109a) O menino **deu uma surra** no colega.

(109b) **A surra que** o menino **deu** no colega...

(109c) **A surra (do menino)** no colega... (enunciado com [Redução de Vsuporte], o apagamento de “dar”)

Além dessas quatro propriedades, há restrições quanto à configuração do elemento não verbal que se incorpora ao verbo suporte que contribuem para delimitar os casos de perífrases verbo-nominais. O elemento incorporado não é um termo pleno, ou seja, item ou expressão com a função de estabelecer referência, designar especificamente uma entidade do mundo biossocial. Seu significado é mais genérico. Sua função, na verdade, é especificar o evento geral (estado, processo, posição ou ação) expresso pelo verbo suporte. Ocorre, tipicamente, em sua forma básica, ou seja, não é acompanhado de qualquer determinante (artigo, pronome) ou modificador (adjetivo, advérbio de intensidade), nem é flexionado em número. Ocupa, preferencialmente, a posição depois do verbo suporte. Por exemplo: *fazer contato*, *dar continuidade*, *tomar medo*, *pegar sol*.

O não atendimento a essas condições morfossintáticas para o elemento não verbal e para o verbo suporte conduz à hesitação quanto ao grau de integração de $V + \bar{n}V$ num predicado complexo.

9.4.4 VERBOS (SEMI-)SUPORTES: UMA PROPOSTA DE CATEGORIZAÇÃO ESCALAR DAS CONSTRUÇÕES DE PREDICADOR COM VERBO SUPORTE DE DIFERENTES CONTORNOS DE NÃO COMPOSICIONALIDADE

Com base na pesquisa descrita em Machado Vieira (2001, 2014), também é possível estabelecer uma escala para as perífrases com verbo suporte. Essas são basicamente de dois tipos. Um deles é o das expressões completamente cristalizadas e de sentido composicional nulo ou opaco, correspondentes às chamadas expressões idiomáticas normalmente listadas em obras lexicográficas: “*fazer vaquinha*”, “*fazer hora*”, “*fazer cara feia*”, “*fazer vista grossa*”, “*fazer uma horinha*”, “*fazer uma fezinha*”, “*fazer a caveira (de alguém)*”, “*fazer gato e sapato*”; “*ter (o) topete*”, “*ter modos*”; “*dar as caras*”, “*dar zebra*”, “*dar uma geral*”, “*dar bandeira*”, “*dar a mão à palmatória*”; *entre tantas outras*. Elas são caracterizadas por baixa ou nula possibilidade de alteração da forma da perífrase e de mobilidade de seus componentes, bem como pela fixação de um significado sociocultural específico, que normalmente não é recuperado a partir do significado de suas partes. O outro é o das expressões com verbo suporte não lexicalizadas e de sentido mais composicional transparente, cujos componentes apresentam diferentes graus de integração a depender de características anteriormente descritas.

O maior grau de integração é o dos predicadores complexos cujos componentes têm todas as propriedades de configuração prototípica e, inclusive, são formados por nomes deverbais (*fazer concessão*, *dar movimentação*, *ter consideração*, *pegar confiança*). Nesses casos em que a verbalidade é inerente à forma nominal, é mais fácil detectar a perífrase verbo-nominal, seu comportamento semelhante ao de um verbo principal simples.

Num segundo grau de integração já estão os predicadores complexos cujo elemento incorporado não é nome deverbal (*fazer barba*, *levar um susto*, *pôr medo*, *passar apuros*, *dar um murro*, *dar conselho*). Nesses casos, o analista já poderá hesitar quanto a considerar ou não a perífrase verbo-nominal.

Um terceiro nível de enfraquecimento da integração é o dos predicadores complexos cujo componente não verbal, apesar de ser um predicador nominal, é antecedido de determinante (*fazer sua opção*, *fazer a narração*, *fazer essa inserção*), fator que contribui para tornar mais referencial o elemento incorporado. Ainda mais enfraquecida é a coesão entre os componentes de predicadores complexos cujo elemento não verbal, além de não ser um nome deverbal, é antecedido de determinante (relativamente) definido (*fazer o desenho*, *fazer a feira*, *dar essa ordem*, *ter o trabalho*).

Outro grau que também revela coesão enfraquecida, mas por outro motivo, é o dos predicadores complexos cujo elemento não verbal é um predicador nominal antecedi-

do de modificador intensificador (*fazer muito movimento, ter imenso gosto*) ou *quantificador* (*fazer duas previsões, dar três declarações*).

E ao menor grau de integração pertencem os predicadores complexos cujo elemento não verbal é um predicador nominal acompanhado de outro tipo de modificador (*fazer alegações falsas, dar declarações absurdas*). Esses são os que mais dúvida causam, em virtude do maior nível de recuperação da função referencial do sintagma incorporado (a função primária dos nomes na língua), já que este remete a uma entidade no mundo mais específica e identificável/recuperável. Fazem fronteira com as predicções constituídas de “Vpredicador pleno + termo/sintagma complemento” (formas independentes, com autonomia e (certa) diversidade de estruturação/distribuição).

9.4.5 VERBOS SUPORTES: FINALIDADES DISCURSIVAS PARA O EMPREGO DE PREDICADOS COMPLEXOS COM VERBO SUPORTE

As perífrases com verbo suporte são empregadas com diversas finalidades discursivas (conforme descreve NEVES, 2000). Entre as mais recorrentes nas amostras de dados já analisadas no âmbito do Projeto Predicar, estão as seguintes:

Expressão de um significado especial do nome não alcançado pela forma verbal simples correspondente.

(110) “Há pouco tempo deu um vazamento aqui no escritório, chamei lá o, o bombeiro, e o sujeito pra arrebentar ali a parede... **fazer uma soldazinha**... que trabalhou o quê? Acho que uma meia-hora. Ele apresentou a conta, cem cruzeiros. (...)” (PB oral, NURC, inq. 127)

(111) “E digo uma coisa para você: eu, nem de graça, eu queria um carro daquele. (...) Se **der uma batidinha** na frente, já era! O carburador vai logo tudo embora (...)” (PB oral, PEUL, inq. 25)

Construções como essas citadas permitem ao usuário da língua focalizar ou acentuar, na apresentação do estado de coisas, um determinado sentido (mais pontual e “ligeiro” que durativo (como no primeiro exemplo); mais superficial (no último); ou vice-versa) da predicção.

Indicação no predicado (complexo) do valor reiterativo do evento por meio da pluralização do nome predicante.

(112) “**Fizeram-se viagens** pelo país à procura de projectos que se adaptassem ‘como uma luva’ ao espírito programático da Polis.” (PE escrito, Jornal Público, 24/05/2000, “Programa Polis – Os fins justificam os meios?”, p. 9)

Alternativa à estrutura com pronome clítico (“ter interesse”/“interessar-se” por; “levar um susto”/“assustar-se”; “tomar banho”/“banhar-se”; “ficar acostumado”/“acostumar-se” a).

(113) “eu vou comprar o título não sei de quê já que estamos tratando de... mercado de capitais... eu não **tenho essa preocupação** ...” (PB oral, NURC, inq. 355) [“me preocupo com isso”]

Obtenção de maior adequação comunicativa de registro/linguagem (formal/informal; técnica/não técnica).

(114) “faca nele assim ... mata ele [peixe] e **dá umas porrada** nele ... aí bota na panela já limpinho né? ...” (PB oral, APERJ, inq. MAC169)

(115) “O governo, quando lhe parece, entra pelos direitos dos cidadãos e, decidindo ora em um, outra em outro sentido, **faz tábuas raso** de tudo quanto solenemente prometera.” (PB escrito, Jornal do Brasil, 16/09/1909. “Revolucionários...”, p. 2)

As finalidades discursivas é que devem ser o alvo de ensino de um tema como esse. Já até se encontra em dicionários e em livros/materiais didáticos do Ensino Médio²⁴ referência à categoria *verbo suporte*. Mais importante, porém, que fazer o estudante conhecer uma subcategoria verbal, é fazê-lo ter consciência das implicações sintáticas, semânticas, pragmáticas, discursivas e sociais que a opção por perífrases verbo-nominais como predicadores pode ter num texto.

9.4.6 VERBOS DE LIGAÇÃO OU VERBOS RELACIONAIS

Também designados, na literatura linguística, como *predicativos, relacionais e copulativos*, são verbos que se relacionam a elementos não verbais com função atributiva, para conferir a estes elementos o papel de projetar sintática e semanticamente predicções (estruturas de papéis participantes com certa relação gramatical), ou seja, com função predicante. Por exemplo:

(116) O bebê está feliz.

(117) Maria é estudante.

(118) Meus amigos continuam com esperança de passar no Vestibular.

(119) A UFRJ fica longe.

Para explicar o comportamento semigramatical que verbos, como os destacados nos exemplos anteriores demonstram, os estudiosos têm recorrido à interpretação de que tais usos se articulam a uma predicção que não é projetada por um predicador verbal. O núcleo semântico do enunciado não reside no verbo, mas no elemento não verbal que àquele se relaciona. Então, o verbo é entendido basicamente como marcador morfossintático de número, pessoa, tempo, modo e aspecto e como o constituinte que sinaliza o comportamento gramatical de oração da estrutura sintagmática.

De acordo com Mateus *et al.* (2003), tais verbos, que, superficialmente, ocorrem numa estrutura “Sujeito V_{COPULATIVO} PREDICATIVO_{Sujeito}” (ou seja, de sujeito, seguido de verbo de ligação e predicativo do sujeito), são “predicadores sintaticamente primários” que selecionam, na verdade, apenas um argumento interno – uma “oração pe-

quena”. O núcleo predicante dessa “oração pequena”, que constitui o “predicador sintaticamente secundário” da frase, não é verbal; pode ser adjetival (*feliz*), nominal (*estudante, gente*), preposicional (*com esperança de passar no Vestibular*) ou adverbial (*sempre*). Por exemplo:

Cada instante é sempre.²⁵

Metade de mil. Nossa, isso é bastante gente, é como se uma escola toda me seguisse.²⁶ [intensificador ____] ou [elemento de comparação_{como/tal qual/tipo} se ____]; com *slot* aberto a muitas formas de preenchimento

Esse elemento não verbal predicante atribui uma propriedade/qualificação ou caracterização ao participante argumento que projeta: um termo (*O bebê/Maria/Meus amigos/A UFRJ/Cada instante/isso*) com relação semântico-gramatical de sujeito. Em casos de predicadores não verbais de dois lugares (por exemplo, “*O álcool é prejudicial ao fígado*” e “*Esse material é resistente ao calor*”), haverá um segundo termo com relação semântico-gramatical de complemento interno (*ao fígado/ao calor*).

Essa interpretação remete-nos à interpretação de Dik (1997) de que os elementos não verbais com função ou interpretação atributiva, que compõem normalmente um enunciado que exprime estado (ou seja, um estado de coisas [- controlado] e [- dinâmico]), podem assumir papel predicante sob a operação de um *verbo cópula* (*suporte*), que é assim chamado, por esse linguista, devido à sua similaridade com um “operador semigramatical de verbalização”.

Para explicitar a relação entre um elemento não verbal predicante e seu argumento externo, o falante conta com um operador gramatical chamado de *verbo de ligação*. Por meio de uma regra de expressão linguística – regra de “verbo cópula (*suporte*)” –, associa-se o verbo de ligação a qualquer elemento diferente de verbo ao qual se deseja atribuir função predicante. Dessa operação resultará uma perífrase (*verbo de ligação + elemento não verbal predicante, V_{de ligação} + ñV*), um predicador complexo, que se articula semântica e sintaticamente a um papel participante que se manifesta como argumento com relação gramatical de sujeito na estruturação morfossintática.

Evidência favorável à interpretação perifrástica aqui descrita reside nos seguintes fatos:

(i) Não é o verbo o principal elemento a determinar as condições semânticas sob as quais os termos preenchem as posições argumentais previstas, conforme se nota na comparação entre os exemplos

(120) Marta é ruiva. Porém: *A pedra é ruiva.

(121) A flor está despedaçada. Porém: *A farinha está despedaçada.

“Ruiva” e “despedaçada” são os elementos responsáveis por inviabilizar o preenchimento da posição de sujeito por termos como “A pedra” ou “A farinha”, e não os

25 <https://www.pensador.com/frase/MjYwNDk3OQ/>. Acesso em: 30 mar. 2022.

26 <https://twitter.com/VacaoComTil/status/1508473279245430789/photo/1>. Acesso em: 30 mar. 2022.

verbos SER e ESTAR. É, portanto, o predicativo do sujeito que restringe semanticamente as possibilidades de preenchimento do constituinte SN sujeito.

(ii) Entretanto, o verbo de ligação contribui para a significação da estrutura perifrástica, uma vez que a permuta de um operador semigramatical por outro altera o significado da expressão, conforme mostram as alternativas testadas no enunciado a seguir: o sentido de permanência, em SER, *versus* transitoriedade, em ESTAR, da propriedade atribuída (preguiçosa); o caráter permansivo decorrente do uso de CONTINUAR; o matiz de hesitação/aparência (“modalização”) da predicação em virtude da opção por PARECER.

(122) Marta é/está/continua/parece preguiçosa.

Outro aspecto que parece contribuir para o estatuto semigramatical do verbo de ligação e também para essa interpretação é o fato de a escolha desse verbo estar, em alguns casos, vinculada ao tipo de propriedade que o constituinte com relação gramatical de predicativo do sujeito expressa:

(123) O retângulo é um quadrilátero. / ?*O retângulo está/anda um quadrilátero.

(124) João está contente. / ?*João é contente.

Um quadrilátero indica propriedade permanente de espécie, característica incompatível com verbos de ligação que exprimem fase/transitoriedade, servem à expressão de uma propriedade temporária.

9.5 CONSIDERAÇÕES FINAIS

O exame de qualquer forma verbal, com perfil mais ou menos polissêmico, num espectro que vai do polo da lexicalidade ao polo da *auxiliaridade/gramaticalidade*, ou seja, da extensão de sentido menos gramatical e mais lexical à extensão mais gramatical leva à descoberta de diversas (sub)categorias e põe em evidência diferentes graus de integração entre os componentes da estrutura perifrástica, que advêm das propriedades que estes assumem nos contextos em que são postos em uso. O trabalho de categorização verbal baseado no exame dessas propriedades constitui uma valiosa oportunidade a quem se debruça sobre o assunto de exercitar sua reflexão crítica e sua argumentação com relação aos conceitos e à classificação que estabelece ou com que se depara na descrição tradicional. É um exercício de categorização centrada em movimentos de associação e dissociação na organização dos objetos da experiência do mundo e, assim, na concepção de que há membros exemplares e membros mais ou menos próximos ou periféricos na interrelação com as categorias tomadas como pontos de referência para esse processo de organização e representação em descrições linguísticas.

Mesmo no mundo fora da gramática, as categorizações não são ideais ou fechadas, como a alguns ainda podem parecer. Há fronteiras que se diluem a depender do ponto de vista tomado, há pontes entre categorias. As categorias não se estabelecem com base em membros tomados em listas feitas por critérios necessários e suficientes, mas

a partir de membros que se associem a representantes médios que revelam mais ou menos atributos de um centro prototípico.

Uma categorização rigorosa e cientificamente fundamentada dos verbos de comportamento (semi-)gramatical do Português deverá visar a fornecer subsídios para o tratamento do tema da *auxiliaridade* como um tópico gramatical cuja exploração didático-pedagógica encaminhe atividades que: (i) colaborem para a construção e/ou sedimentação de um conhecimento explícito da língua; e (ii) fomentem a manipulação consciente e segura de estruturas da língua portuguesa com verbos auxiliares (em sentido lato), assim como o entendimento de sua funcionalidade e de suas condições de uso. Com base naquele conhecimento, o usuário da língua poderá desvendar as implicações semântico-pragmático-discursivas e sociais das escolhas gramaticais que faz e perceber as vantagens envolvidas no emprego de uma ou outra possibilidade estrutural (forma perifrástica ou simples), na opção por um ou outro operador/marcador gramatical (verbo auxiliar, verbo semiauxiliar, verbo (semi-)suporte, verbo de ligação).

Vale lembrar que esse é um tema que repercute no trabalho com outro tópico gramatical bastante frequente nos livros didáticos: a delimitação de períodos compostos por subordinação. É importante saber reconhecer, quando, num texto, se esbarra numa sequência de verbos, se se está diante de um período simples com perífrase verbal ou de um período composto com dois predicadores, para prosseguir com uma atividade, por exemplo, a de classificação ou de conexão de orações. Para tanto, na tarefa de dar instrumentos ao aluno, para que este assuma uma postura de analista/pesquisador dos fatos linguísticos com que se depara, caberá ao professor articular diferentes conhecimentos gramaticais e estes com a construção de unidades textuais, bem como fornecer um conjunto de informações que torne esse aluno mais seguro e crítico (i) no uso da língua, (ii) no raciocínio e controle epilinguísticos, (iii) na representação e conscientização metalinguísticas e, finalmente, (iv) na evolução de habilidades e competências metalinguísticas e automatização dos metaprocessos implicados no modelo de desenvolvimento metalinguístico (Gombert, 1992).

O bom aproveitamento do tema da *auxiliaridade* nas aulas de Português centra-se em: (i) articulação entre expedientes morfossintáticos de auxiliarização/verbalização, seus efeitos de sentido na constituição de uma perífrase, propósitos comunicativos e exigências situacionais; (ii) exploração do conjunto de propriedades semânticas e sintáticas que caracterizam verbos predicadores, verbos (semi-)auxiliares, verbos (semi-)suportes, verbos relacionais/de ligação e, com base nelas, identificação/classificação do(s) comportamento(s) dos verbos e detecção do tipo de predicação/período (simples ou complexo); e (iii) exercícios que, organizados pelo tipo de recurso linguístico e pelo tipo de instrução de sentido vinculado aos recursos, impliquem a manipulação de textos, bem como a formulação de generalizações quanto ao comportamento lexical, léxico-gramatical e gramatical dos verbos.

Generalizações acerca das propriedades das diferentes categorias de verbo e perífrases verbais podem contribuir para a compreensão de sua funcionalidade discursiva na fala e na escrita. Quem sabe, ao testar os parâmetros de *auxiliaridade/gramaticali-*

dade aqui delineados no exame de outros dados, possa o professor aprofundar essas generalizações e organizar material didático para abordar mais apropriadamente a questão da formação e da interpretação de predicadores complexos, perífrases verbais ou verbo-nominais, na construção de textos, para além de promover o processo de categorização linguística em termos mais apropriados. Afinal, a unidade do conhecimento é permanentemente construída e reconstruída, com base no interesse que se tem em (i) resolver e (re)descobrir problemas, (ii) administrar tensões decorrentes de propostas teóricas, descritivas e/ou metodológicas diferentes e (iii) alcançar soluções descritivas e refinar a análise e a interpretação dos dados, bem como o instrumental metalinguístico de trabalho. Esse desafio merece ser encarado pelo professor de língua (materna ou não materna).

Este capítulo é uma contribuição descritivo-metodológica fundada em experiências de análises linguísticas do Projeto Predicar e uma versão ampliada e modificada de MACHADO VIEIRA (2004).

REFERÊNCIAS

- BARROSO, H. *O aspecto verbal perifrástico em português contemporâneo: visão funcional/sincrónica*. Porto: Porto Editora, 1994.
- BECHARA, E. *Moderna gramática portuguesa*. 37^a ed. revista e ampliada. Rio de Janeiro: Lucerna, 1999.
- DIK, S. C. *Theory of functional grammar*. Editado por Kees Hengeveld. Berlin: Mouton de Gruyter. 2 v., 1997.
- GOMBERT, J. E. *Metalinguistic development*. Great Britain: Harvester Wheatsheaf/Chicago: The University of Chicago Press, 1992.
- GONÇALVES, A. P. M. *Para uma sintaxe dos verbos auxiliares em português europeu*. Lisboa: Universidade de Lisboa. Dissertação de Mestrado. Mimeo, 1992.
- GONÇALVES, A.; COSTA, T. da. *(Auxiliar a) Compreender os verbos auxiliares: descrição e implicações para o ensino do Português como Língua Materna*. Lisboa: Colibri, 2002.
- GONÇALVES, A. *et al.* Propriedades predicativas dos verbos leves *dar, ter e fazer*: estrutura argumental e eventiva. In: Actas do XXXIX Simpósio Internacional de la Sociedad Española de Lingüística (SEL), Santiago de Compostela, 1-4, Fevereiro 2010.
- HOPPER, P. J. On some principles of Grammaticalization. In: TRAUGOTT, E. C.; HEINE, B. (eds.). *Approaches to grammaticalization*, Vol. 1. Amsterdam/Philadelphia, John Benjamins Company, 1991. p. 16-35.
- LOBATO, L. M. P. Os verbos auxiliares em português contemporâneo. Critérios de auxiliabilidade. In: LOBATO, L. M. P. *et al.* *Análises Linguísticas*. Petrópolis: Vozes,

1975. p. 27-91.
- MACHADO VIEIRA, M. dos S. *Sintaxe e semântica de predicções com verbo fazer*. Rio de Janeiro: UFRJ, Faculdade de Letras, 2001. 362 fl. mimeo. Tese de Doutorado.
- MACHADO VIEIRA, M. dos S. Perífrases verbais: o tratamento da auxiliaridade. In: VIEIRA, S. R.; BRANDÃO, S. F. *Morfossintaxe e ensino de Português: reflexões e propostas*. Rio de Janeiro: Faculdade de Letras/UFRJ, 2004, p.65-96.
- MACHADO VIEIRA, M. dos S. Idiomaticidade em construções com verbo suporte do Português. *SOLETRAS*, [S.l.], n. 28, p. 99-125, dez. 2014. ISSN 2316-8838. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/soletras/article/view/14200>. Acesso em: 01 abr. 2022. doi:<https://doi.org/10.12957/soletras.2014.14200>.
- 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, marzo 2015, p. 210-230.
- MACHADO VIEIRA, M. dos S. Predicar com construção com verbo suporte. In: DE PAULA, Alessandra *et al.* *Uma história de investigações em língua portuguesa: homenagem a Silvia Brandão*. São Paulo: Blucher Open Access, 2018, p. 91-112, cap.6. <https://openaccess.blucher.com.br/article-list/9788580393088-389/list#undefined>.
- MACHADO VIEIRA, M. dos S. Língua, sociedade e relações de poder: a produção escrita de surdos. In: FREITAS JR., R. de; SOARES, L. A. A.; NASCIMENTO, J. P. da S. (orgs.). *Aprendizes surdos e escrita em L2: reflexões teóricas e práticas*. Rio de Janeiro: UFRJ, 2020.
- MACHADO VIEIRA, M. dos S.; SOUZA, L. L.; COSTA, M. G. da. In: CEZARIO, M. M.; ALONSO, K. S.; CASTANHEIRA, D. (orgs.) *Linguística Baseada no Uso: explorando métodos, construindo caminhos*. Rio de Janeiro: Rio Books, 2021. https://discursoegramaticablog.files.wordpress.com/2020/11/linguistica-baseada-no-uso_explorando-metodos-construindo-caminhos.pdf.
- MATEUS, M. H. M. *et al.* *Gramática da língua portuguesa*. Ed. revista e aumentada. Lisboa: Caminho, 2003.
- NEVES, M. H. M. *Gramática de usos do português*. São Paulo: UNESP, 2000. p. 25-65.
- RAPOSO, E. B. P. *et al.* *Gramática do Português*. Portugal: Fundação Calouste Gulbenkian, vol. I e II, 2013.

POSFÁCIO

IL PROGETTO PREDICAR E LO STUDIO DELLE COSTRUZIONI

Carla Valeria de Souza Faria (Università Ca' Foscari Venezia)

Nell'ambito delle ricerche sull'uso del linguaggio, lo studio delle costruzioni trova applicazione in campi diversi come la linguistica diacronica (teoria della costruzionalizzazione e ricostruzione sintattica), la variazione linguistica (studi consolidati sulla variazione interlinguistica e quelli emergenti sulla variazione sociolinguistica – intralinguistica e intraparlante), l'acquisizione del linguaggio (acquisizione di L1 e di L2), la didattica delle lingue (tecniche di insegnamento della L2), la lessicografia, l'analisi del discorso, la psicolinguistica, la neurolinguistica e, infine, la linguistica computazionale (implementazioni della grammatica delle costruzioni, elaborazione di metodi e analisi, modellizzazione computazionale), come indicato da Masini (2016).

Il progetto Predicar, che compie vent'anni e che si riflette in questo libro tramite gli argomenti trattati, si fonda su un'ottica di ricerca collaborativa e coinvolge linguisti di altri paesi e ambiti di studio per costruire un'interlocuzione sulla lingua portoghese e sulla sua diversità socioculturale e linguistica. Tale diversità si evince non solo dal suo status linguistico nei differenti spazi in cui è utilizzata (lingua madre, lingua non materna, lingua di accoglienza, lingua veicolare, lingua minoritaria, lingua etnica o

d'origine) ma anche dai diversi scenari di vita (territoriali e di confine) presenti in tutto il mondo e legati alla mobilità, alla migrazione, all'insegnamento, alla traduzione, all'accessibilità, alle politiche linguistiche nonché alle tecnologie digitali.

In questo libro, ci si sofferma principalmente sulle questioni di metodologia, raccolta, trattamento e analisi dei dati che riguardano le costruzioni a verbo supporto, l'analisi collostruzionale, i predicati complessi, le collocazioni (in base alle loro preferenze combinatorie) e gli idiomatismi, utilizzando, a seconda del fenomeno analizzato, l'analisi multifattoriale, l'analisi sperimentale e l'analisi collostruzionale tramite il linguaggio R, l'applicazione RStudio ed Excell.

Parimenti, si pone l'attenzione sull'elaborazione di criteri finalizzati all'individuazione di alcuni fenomeni costruzionali come pure sull'utilizzo di *corpora* annotati nelle banche dati disponibili online e sulla valutazione dei loro aspetti positivi e negativi. Ulteriormente, ci si focalizza anche sull'elaborazione di modelli di test con lo scopo di analizzare sia la variazione tra predicati complessi a verbo supporto sia la percezione dei parlanti del portoghese riguardo tale fenomeno.

Tutto ciò mantenendo saldo l'obiettivo di riflettere sulle questioni di metodologia di analisi delle co-occorrenze linguistiche volte allo sviluppo dell'analisi collostruzionale senza, però, perderne di vista i limiti e le implicazioni nonché l'articolazione con altre prospettive (sociolinguistiche, socio-funzionaliste, socio-costruzioniste o costruzioniste diasistematiche) che continuano a guidare questo gruppo di ricerca.

MASINI, Francesca, *Grammatica delle costruzioni. Un'introduzione*, 1a ed., Roma, Carocci, 2016.

POSFÁCIO

LA NÉCESSAIRE ARTICULATION ENTRE UNE CERTAINE MANIÈRE DE FAIRE DE LA RECHERCHE SUR LES PRÉDICATS ET LES PRÉDICATIONS ET L'ENSEIGNEMENT DE LA LANGUE

Liliane Santos¹ (Université de Lille)

L'idée selon laquelle l'enseignement de la langue – et plus spécialement de la grammaire – ne peut se passer de la description et de l'analyse des usages linguistiques, des rapports entre la langue et la société et, par conséquent, des interactions sociales, fait l'objet d'un large consensus entre les linguistes et les didacticiens depuis déjà plusieurs décennies. De ce consensus découle l'idée que la matière de cet enseignement ne sera pas l'étude d'une grammaire ou d'une langue directement à partir des recherches en linguistique, mais plutôt les moyens qui pourront aider l'apprenant à devenir un usager conscient des nombreuses ressources et stratégies que la langue met à sa disposition – qu'il s'agisse ou non de sa langue maternelle.

1 Univ. Lille, CNRS, UMR 8163 – STL – Savoirs Textes Langage, F-59000 Lille, France. liliane.santos@univ-lille.fr.

Les travaux réunis dans ce volume – qui portent sur la description de prédicats et prédictions complexes et sur des méthodes de recherche y afférentes – fournissent des éléments qui peuvent servir de base à la réflexion de l’enseignant sur la nécessaire articulation entre la recherche linguistique et ses applications en classe de langue. À partir de ces éléments, le professeur pourra, par exemple, proposer des travaux pratiques qui permettent d’appréhender la langue comme un objet hétérogène, dynamique et à multiples facettes.

Dans l’enseignement d’une langue maternelle, on peut partir des connaissances des élèves pour les aider à expliciter les règles implicites – pas seulement celles qui concernent la grammaire, mais surtout celles liées à l’usage de la langue – qui, comme on le sait, l’école a souvent du mal à reconnaître. Pour ce faire, l’appel à la perception des élèves (usagers de langue) devient un puissant instrument quand on travaille sur la variation des constructions, leurs (non-)équivalences et leurs conditions d’utilisation.

Dans l’enseignement d’une langue non maternelle, que ce soit en situation d’immersion ou par l’intermédiaire d’enquêtes réalisées à distance, l’apprenant peut assumer le rôle du chercheur qui analyse le sentiment des usagers natifs à propos de la langue qu’il apprend.

Dans les deux cas, les méthodologies didactiques qui font appel à la participation active des apprenants, en les mettant en position de réfléchir critiqueusement sur les relations entre la langue qu’ils apprennent et la (les) société(s) qui parle(nt) cette langue, produisent leurs meilleurs fruits lorsqu’elles sont utilisées avec des apprenants plus expérimentés. Plus spécifiquement, pour le Portugais Langue Maternelle, ce travail pourrait être développé avec des lycéens (c’est-à-dire, au cours des trois dernières années de *l’ensino médio* au Brésil), tandis que pour le Portugais Langue Non Maternelle, ce travail devrait être accompli avec des apprenants de niveau C (« utilisateur expérimenté ») du *Cadre Européen Commun de Référence pour les Langues* (Conseil de l’Europe, 2000 ; Conselho da Europa, 2001).

Dans un cas comme dans l’autre, l’objectif est non seulement « l’apprentissage » de langue : il s’agit surtout de la prise de conscience d’un certain nombre de faits à propos de la langue. Par exemple, (a) le fait que la langue est constitutivement hétérogène, (b) que cette hétérogénéité naît et de la diversité des groupes sociaux qui la parlent et des rapports qu’ils entretiennent entre eux et avec la langue, mais aussi avec l’histoire – celle de la langue et celle desdits groupes sociaux –, (c) que cette hétérogénéité ne veut pas dire désordre, chaos ou manque de repères, mais, au contraire, qu’elle est déterminée par des facteurs de nature variée et qu’elle obéit à des règles, et (d) que c’est en ce sens que la langue est un système en perpétuel mouvement, bien loin du caractère statique souvent associé à cette idée.

Dans la mesure où la nature polycentrique (cf. *pluricêntrica*) de la langue portugaise ne fait plus de doute, des travaux comme ceux ici présentés acquièrent une importance fondamentale, non seulement pour la description du fonctionnement de la langue, mais, bien au-delà, pour la nécessaire réflexion sur la langue par ses usagers, natifs ou non natifs.

CONSEIL DE L'EUROPE (2000). *Cadre Européen Commun de Référence pour les Langues. Apprendre, Enseigner, Évaluer*. Paris : Didier.

CONSELHO DA EUROPA (2001). *Quadro Europeu Comum de Referência para as Línguas. Aprendizagem, Ensino, Avaliação*. Lisboa: Edições Asa.

AS AUTORAS E OS AUTORES

BERT CAPPELLE

Bert Cappelle is an Associate Professor in English Linguistics at the University of Lille (Research Unit UMR 8163 STL– Savoirs Textes Langage, Univ. de Lille, and CNRS, France). He investigates how we use constructions to put our thoughts into words. In line with Usage-based Cognitive Construction Grammar, he defends the view that much of our linguistic knowledge is exemplar-based. This means that our grammar is constructed piecemeal, in a bottom-up way. From the concrete utterances we are being exposed to as speakers, we gradually detect recurring sequences and extract more abstract patterns. Schematic templates (constructions) and several of the lexical sequences that gave rise to them thus co-exist in a single large mental storage space, which we call the ‘construct-i-con’.

E-mail: bert.cappelle@univ-lille.fr

Curriculum Lattes: bert.cappelle@univ-lille.fr, <https://pro.univ-lille.fr/ber-cappelle/>

ORCID: : <https://orcid.org/0000-0002-4779-6259>

ENEILE SANTOS SARAIVA

Doutora em Língua Portuguesa do Programa de Pós-Graduação em Letras Vernáculas da Faculdade de Letras da UFRJ. Desenvolve pesquisa sobre a variação de usos de construções transitivas diretas com pronome SE, sob a orientação da Professora Doutora Marcia dos Santos Machado Vieira. É pesquisadora do Projeto Predicar e atua também como professora da educação básica de língua portuguesa e literaturas do estado do Rio de Janeiro.

E-mail: eneilesaraiva@letras.ufrj.br

Curriculum Lattes: <http://lattes.cnpq.br/5196279772369306>

Projeto de pesquisa: <https://projeto-predicar.wixsite.com/predicar>

FABIO RODRIGO GOMES DA COSTA

Doutorando em Língua Portuguesa pelo Programa de Pós-Graduação em Letras Vernáculas na Universidade Federal do Rio de Janeiro (UFRJ). É mestre em Estudos Linguísticos pelo Programa de Pós-Graduação em Letras e Linguística (PPLIN-Uerj/FFP) e especialista em Língua Portuguesa pela UERJ/FFP. Possui licenciatura em Letras-Português/Literatura pela Universidade Salgado de Oliveira. Integra o projeto Predicar (Formação e expressão de predicados complexos: estabilidade, variação e mudança construcional), coordenado pela Prof.^a Dr.^a Marcia dos Santos Machado Vieira. É professor de Língua Portuguesa da Secretaria Estadual de Educação do Rio de Janeiro (Seeduc). Desenvolve pesquisa sobre predicadores complexos formados com o verbo suporte *trazer* e sua atuação como atenuadores do discurso.

E-mail: fabiorodrigogc@letras.ufrj.br

Curriculum Lattes: <http://lattes.cnpq.br/3881158262910747>

ORCID: <https://orcid.org/0000-0003-1831-2443>

Projeto de pesquisa: <https://projeto-predicar.wixsite.com/predicar>

JEANE NUNES DA PENHA

Doutoranda do curso de Língua Portuguesa pelo Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro. Mestre em Língua Portuguesa pela UFRJ. Possui licenciatura em Letras pela UFRJ, habilitação: Português-Espanhol. Foi aluna de Iniciação Científica entre 2015-2017, sendo bolsista FAPERJ entre 2016-2017. No âmbito acadêmico, concentra-se em estudos de Morfossintaxe sob o enfoque da Linguística Funcional-Cognitiva, da Sociolinguística e da abordagem da Gramática de Construções, seguindo a Linha de Pesquisa: Língua e Sociedade: variação e mudança. Integra o projeto PREDICAR (Formação e expressão de predicados complexos e predicções: estabilidade, variação e mudança construcional), coordenado pela Prof.^a Dr.^a Marcia dos Santos Machado Vieira. Possui estudos sobre (i) o uso de perífrases verbo-nominais com os verbos suportes DAR e FAZER na modalização do discurso, (ii) o pareamento forma-função de construções com os verbos suportes DAR, FAZER, PASSAR, POSAR e TIRAR seguidos de elemento nominal preposicionado e, atualmente, (iii) a construção [ELEMENTO VERBAL + DET./PREP. + ELEMENTO NOMINAL] representação através de uma abordagem construcional diassistêmica.

E-mail: jeane.nunes@letras.ufrj.br, jeanee07@gmail.com

Curriculum Lattes: <http://lattes.cnpq.br/5562543165718460>

ORCID: <https://orcid.org/0000-0002-3889-0251>

Projetos de pesquisa:

<https://projeto-predicar.wixsite.com/predicar>

<https://variatar.wixsite.com/variatar>

LAIS LIMA DE SOUZA

Graduanda em Licenciatura em Letras – Português/Inglês pela Universidade Federal do Rio de Janeiro. Integrou o Projeto PREDICAR, coordenado pela Prof.^a Dr.^a Márcia Machado Vieira, entre 2018 e 2021, onde desenvolveu pesquisa sobre construções de desejo/futuridade sob a perspectiva da Linguística Funcional-Cognitiva. Atualmente, é bolsista de Iniciação Científica sob orientação da Prof.^a Dr.^a Daniela Cid de Garcia, do Departamento de Letras Anglo-germânicas e do Programa de Pós-Graduação em Linguística, dedicando-se, agora, ao estudo do processamento da leitura e bilinguismo. Também atua como aluna extensionista no Projeto de Extensão Linguística no Básico, que visa a criar um diálogo entre a universidade pública e escolas da educação básica. É graduada em Direito pela Universidade Federal do Rio de Janeiro (2016).

E-mail: lais@letras.ufrj.br

Curriculum Lattes: <http://lattes.cnpq.br/7512318752609723>

ORCID: <https://orcid.org/0000-0002-0911-5205>

MARCIA DOS SANTOS MACHADO VIEIRA

Professora-pesquisadora da Faculdade de Letras da Universidade Federal do Rio de Janeiro. Foi professora de Língua Portuguesa no Ensino Fundamental da Prefeitura do Rio de Janeiro. Doutora e mestre pela UFRJ, tem experiência na investigação de: variação e mudança de fenômenos fonéticos (pretônicas) e morfossintáticos (construções de referenciação, predicação, modalização, intensificação e atenuação discursivas). Coordena: o Projeto brasileiro PREDICAR – Formação e expressão de predica-dos complexos e predicações; o Projeto franco-brasileiro VariaR – Variação em Línguas Românicas – em parceria com Vanessa Meireles (UPVM), bem como um desdobramento dele intitulado Projeto Portal digital de estados de coisas em Português e em línguas românicas a variar e ensinar (CNPq e Faperj/Cientista do Nosso Estado); o Projeto brasileiro CAPES PrInt – Vozes e escritas nos diferentes espaços da língua portuguesa –, do PPGLEV da UFRJ. Participa do grupo de estudos Discurso & Gramática. É membro da Diretoria da Associação Nacional de Pós-Graduação e Pesquisa em Letras e Linguística no biênio 2021- 2023. Integra, desde 2014, a coordenação do eixo 1 do GT de Sociolinguística da ANPOLL. Coordena esse GT desde 2018. Coordena a Comissão Científica da Área de Sociolinguística da ABRALIN. É mem-

bro do Comitê de Ética em Pesquisa do IESC/UFRJ e de NDE de cursos da Faculdade de Letras/UFRJ. É editora-chefe da *Diadorim*: revista de estudos linguísticos e literários da UFRJ. É pesquisadora do CNPq e Cientista do Nosso Estado/Faperj. Áreas de interesse: Sociolinguística, Linguística Funcional-Cognitiva, Gramática de Construções, Ensino de Língua Portuguesa. Sua produção bibliográfica inclui outras obras publicadas pela editora Blucher.

E-mail: marcia@letras.ufrj.br

Curriculum Lattes: <http://lattes.cnpq.br/0796977308756789>

ORCID: <https://orcid.org/0000-0002-2320-5055>

Projetos de pesquisa:

<https://projeto-predicar.wixsite.com/predicar>

<https://variav.wixsite.com/variav>

MARIANA GONÇALVES DA COSTA

Graduanda em Letras – Português/Inglês (Licenciatura) pela Universidade Federal do Rio de Janeiro. É bolsista de iniciação científica através do PIBIC/CNPq por sua pesquisa sobre a construcionalização gramatical do verbo volitivo querer como indicador de futuridade, tendo como base a Gramática das Construções. Integra o Projeto PREDICAR: Formação e expressão de predicados e predicções: estabilidade, variação e mudança construcional, onde colabora no estudo interdisciplinar de metodologias de análise de língua em uso. Paralelamente, atua como aluna extensionista no Projeto de Extensão Linguística no Básico (PELB), que visa a criar um diálogo entre a universidade pública e escolas da educação básica.

E-mail: marianag.costta@gmail.com

Curriculum Lattes: <http://lattes.cnpq.br/148569423565537>

ORCID: <https://orcid.org/0000-0002-8088-0794>

Projeto de pesquisa:

<https://projeto-predicar.wixsite.com/predicar>

PÂMELA FAGUNDES TRAVASSOS

Doutoranda do curso de Língua Portuguesa do Programa de Letras Vernáculas da Universidade Federal do Rio de Janeiro (UFRJ). Integra o projeto PREDICAR (Formação e expressão de predicados e predicções: estabilidade, variação e mudança construcional), coordenado pela Prof.^a Dr.^a Marcia dos Santos Machado Vieira. De março a agosto de 2021, realizou estágio de Doutorado Sanduíche (bolsa CAPES PrInt) na *Université de Lille* (França), no laboratório “*Savoirs, Textes, Langage*”, sob a supervisão do Prof. Dr. Bert Cappelle. Integra, como estudante colaboradora, o Pro-

jeto VariaR – Variação em Línguas Românicas –, coordenado pelas pesquisadoras Vanessa Meireles (*Université Paul Valéry - Montpellier 3*) e Marcia dos Santos Machado Vieira (UFRJ). É membro do GT de Sociolinguística da ANPOLL (Associação Nacional de Pós-Graduação e Pesquisa em Letras e Linguística). Concentra-se em estudos de Morfossintaxe à luz das perspectivas Sociofuncionalista, Funcional-Cognitiva e da abordagem da Gramática de Construções, seguindo a Linha de Pesquisa: Língua e Sociedade: variação e mudança. cursou Mestrado em Língua Portuguesa (Letras Vernáculas) na UFRJ (bolsa CAPES em 2017 e bolsa FAPERJ Nota 10 em 2018). É especialista em Língua Portuguesa pela Universidade Federal Fluminense (UFF). Possui licenciatura em Letras (UFRJ), habilitação: Português-Literaturas (bolsa CNPq no período de Iniciação Científica – 2014 a 2016). É professora da rede municipal do Rio de Janeiro.

ORCID ID: orcid.org/0000-0002-0683-9742

E-mail: fagundespamela@letras.ufrj.br, fagundespamela@hotmail.com

Curriculum Lattes: <http://lattes.cnpq.br/9943814815338406>

Projetos de pesquisa:

<https://projeto-predicar.wixsite.com/predicar>

<https://variari.wixsite.com/variari>

PEDRO IVORRA ORDINES

Pedro Ivorra Ordines es investigador postdoctoral en la Universidad de Santiago de Compostela en el seno del Departamento de Inglés y Alemán dentro del Programa Postdoctoral Margarita Salas. Recientemente, obtuvo su doctorado en Traducción e Interpretación en la Universitat Pompeu Fabra. Asimismo, ha participado en conferencias nacionales e internacionales y ha sido beneficiario de una beca de investigación de la Fundació Mercè Rodoreda (2018), auspiciada por el Institut d'Estudis Catalans. Es miembro del grupo de investigación FRASESPAL, en el que se está llevando a cabo el proyecto de investigación Gramática de Construcciones y Fraseología. Las construcciones fraseológicas en alemán y español en contraste a través de *corpus*.

Email: pedro.ivorra@usc.es

ResearchGate: <https://www.researchgate.net/profile/Pedro-Ivorra-Ordines>

ORCID: <https://orcid.org/0000-0002-2353-1002>

Research site: <http://frasespal.com/pedro-ivorra/>

PEDRO POPPOLINO

Técnico em Informática pelo Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ, 2018), atualmente é graduando em Ciência da Computação pela Universidade Federal do Rio de Janeiro (UFRJ). Integra os projetos “PRE-

DICAR - Formação e expressão de predicados complexos e predicacões: estabilidade, variação e mudança construcional” e “VARIAR - Variação nas línguas românicas”, ambos coordenados pela Prof.^a Dr.^a Márcia dos Santos Machado Vieira (UFRJ). É pesquisador bolsista do CNPq. Desenvolve pesquisa no âmbito do Projeto “Portal digital de estados de coisas em Português e outras línguas românicas a variar e ensinar”. Tem interesse na relação entre Computação, Linguagem e Sociedade.

E-mail: poppolinopedro@gmail.com

Curriculum Lattes: <http://lattes.cnpq.br/0545797101363643>

ORCID: <https://orcid.org/0000-0002-6888-4186>

Projetos de pesquisa:

<https://projeto-predicar.wixsite.com/predicar>

<https://variatar.wixsite.com/variatar>

RAVENA BEATRIZ DE SOUSA TEIXEIRA

Doutoranda do Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro (UFRJ). Mestre em Língua Portuguesa (Letras Vernáculas) e Graduada em Letras Licenciatura em Português-Espanhol pela UFRJ. No âmbito acadêmico, concentra-se em estudos de Morfossintaxe sob enfoque da Gramática de Construções Baseada no Uso, seguindo a Linha de Pesquisa “Língua e Sociedade: variação e mudança”. Atua como pesquisadora no Projeto PREDICAR – Formação e expressão de predicados complexos e predicacões: estabilidade, variação e mudança construcional, sob coordenação da Prof.^a Dr.^a Marcia dos Santos Machado Vieira (UFRJ), e como docente de Língua Portuguesa, Redação e Espanhol na educação básica, assim como em preparatórios voltados para concursos públicos. No momento, estuda a disposição formal e funcional de complexos verbo-nominais compostos por verbos (semi-)suporte que, no português brasileiro (PB), projetam um cenário segundo uma perspectiva passiva – tendo, por exemplificações, “levar um tiro”, “tomar um tiro”, “sofrer um tiro”, “receber um tiro” e “ganhar um tiro”.

E-mails:

ravena_beatriz@letras.ufrj.br

ravenabst@gmail.com

ravena_beatriz@hotmail.com

Curriculum Lattes: <http://lattes.cnpq.br/4077682631814647>

ORCID: <https://orcid.org/0000-0002-4816-6654>

Projetos de pesquisa:

<https://projeto-predicar.wixsite.com/predicar>

<https://variatar.wixsite.com/variatar>



In questo libro, si pone l'attenzione principalmente sulle questioni di metodologia, raccolta, trattamento e analisi dei dati riguardanti le costruzioni a verbo supporto, l'analisi collostruzionale, i predicati complessi, le collocazioni (in base alle loro preferenze combinatorie), nonché gli idiomatismi, sfruttando, a seconda del fenomeno studiato, l'analisi multifattoriale, l'analisi sperimentale e l'analisi collostruzionale, tramite il linguaggio R, l'applicazione RStudio ed Excell.

Carla Valeria de Souza Faria (Università Ca' Foscari Venezia)

Dans la mesure où la nature polycentrique (cf. pluricêntrica) de la langue portugaise ne fait plus de doute, des travaux comme ceux ici présentés acquièrent une importance fondamentale, non seulement pour la description du fonctionnement de la langue, mais, bien au-delà, pour la nécessaire réflexion sur la langue par ses usagers, natifs ou non natifs.

Liliane Santos (Université de Lille)



openaccess.blucher.com.br



Blucher Open Access