

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).

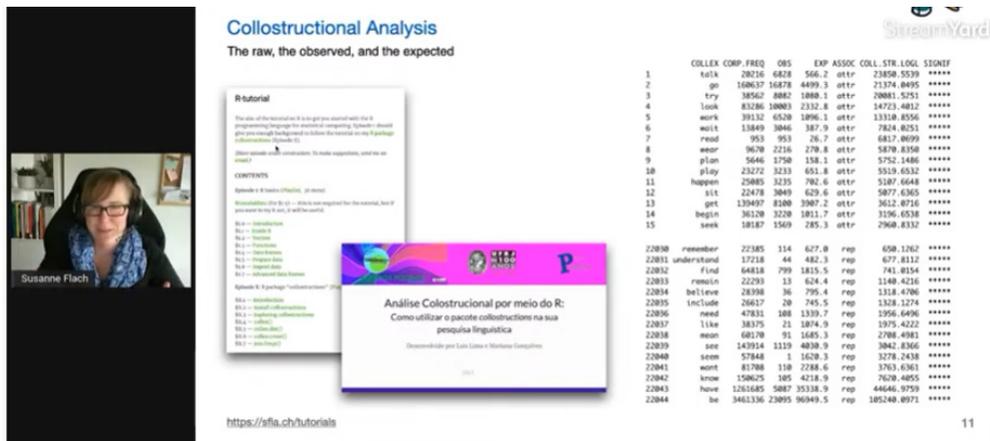
E entendemos que nossa atenção voltada a lidar com a variação construcional por meio de diferentes métodos de tratamento de dados tem enriquecido nossas descrições, viabilizando maior flexibilidade quanto a olhar correlações (entre construções de diferentes dimensões em termos de complexidade, grupos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

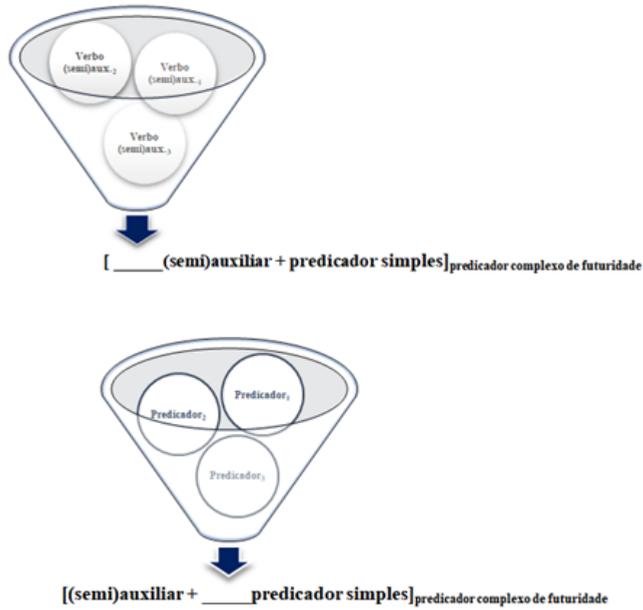
3) O terceiro objeto de estudo é aquele que nos faz entender como se dá a configuração do *design* de análise colostrucional de colexemas covariantes: cf. COSTA, 2022, vídeo.

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:

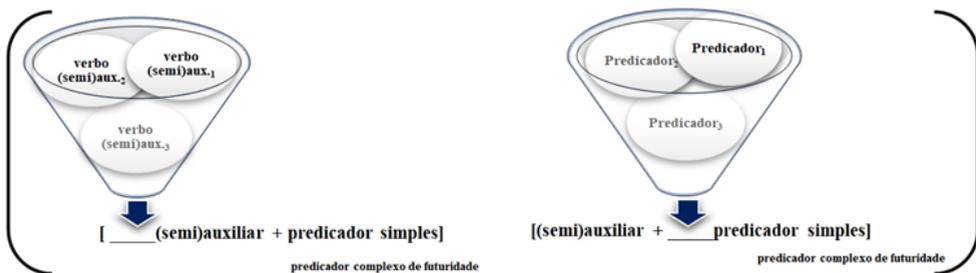
⁹ <https://www.youtube.com/watch?v=itfg2wxyWwk&t=1s>. Acesso em: 23 mar. 2022.

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.



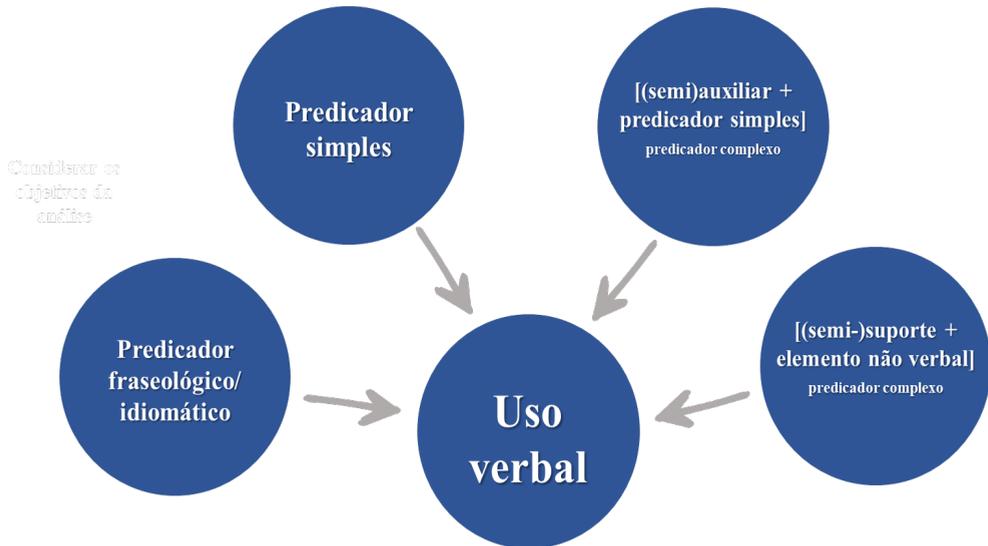
Procuramos investigar quais formas verbais são atraídas para dois *slots* numa construção:

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.



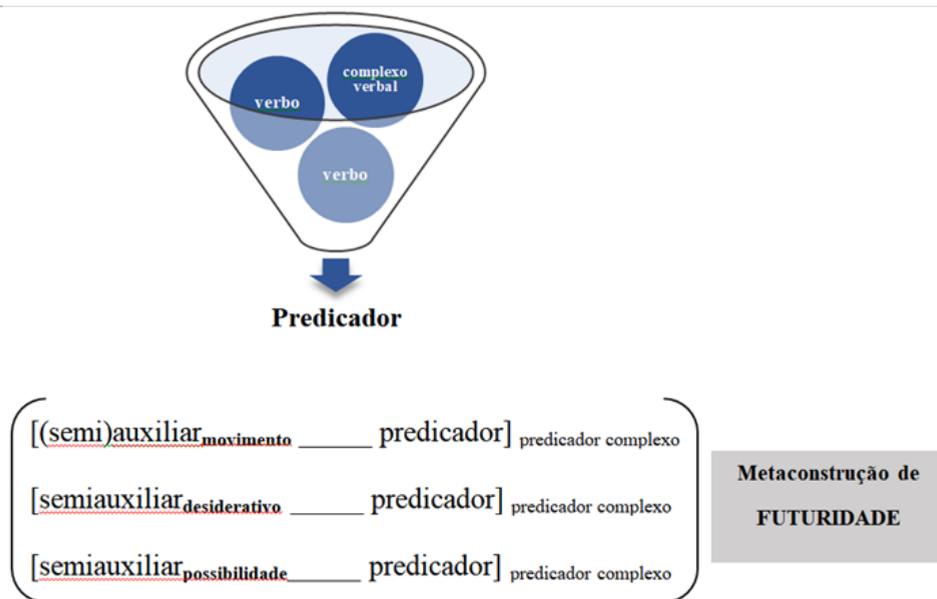
Buscamos examinar a relação entre formas verbais constitutivas das construções que focalizamos com outras construções em amostras de dados:

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.

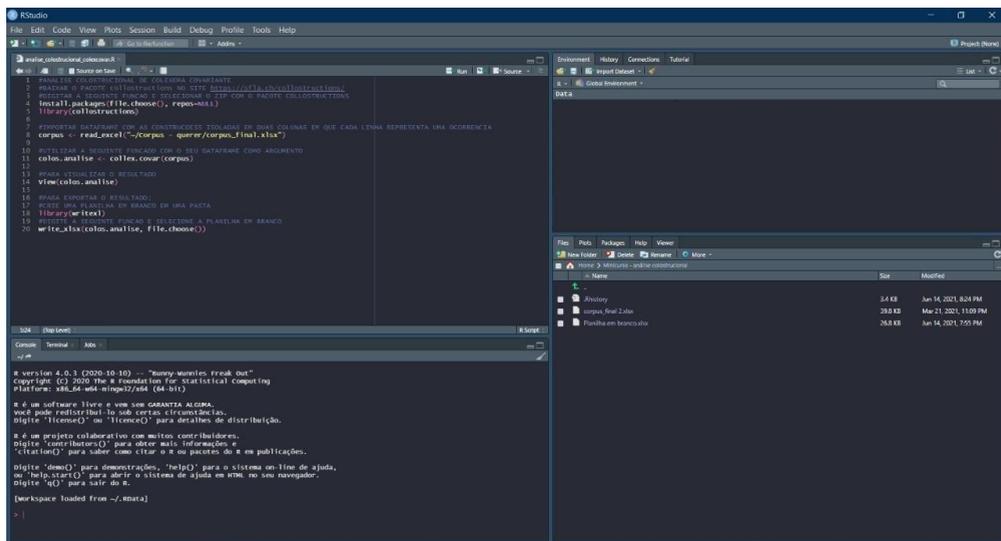
Como está dito e aparece no *print* a seguir, a interface do RStudio apresenta quatro seções de informações: *Code Editor*, para o registro dos *scripts* na linguagem R; *Environment/History*, para a listagem e armazenamento dos objetos e para o histórico de ações executadas no R; *Console*, para registrar 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 ‘writexl’ 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 seleccionar 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.

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

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;

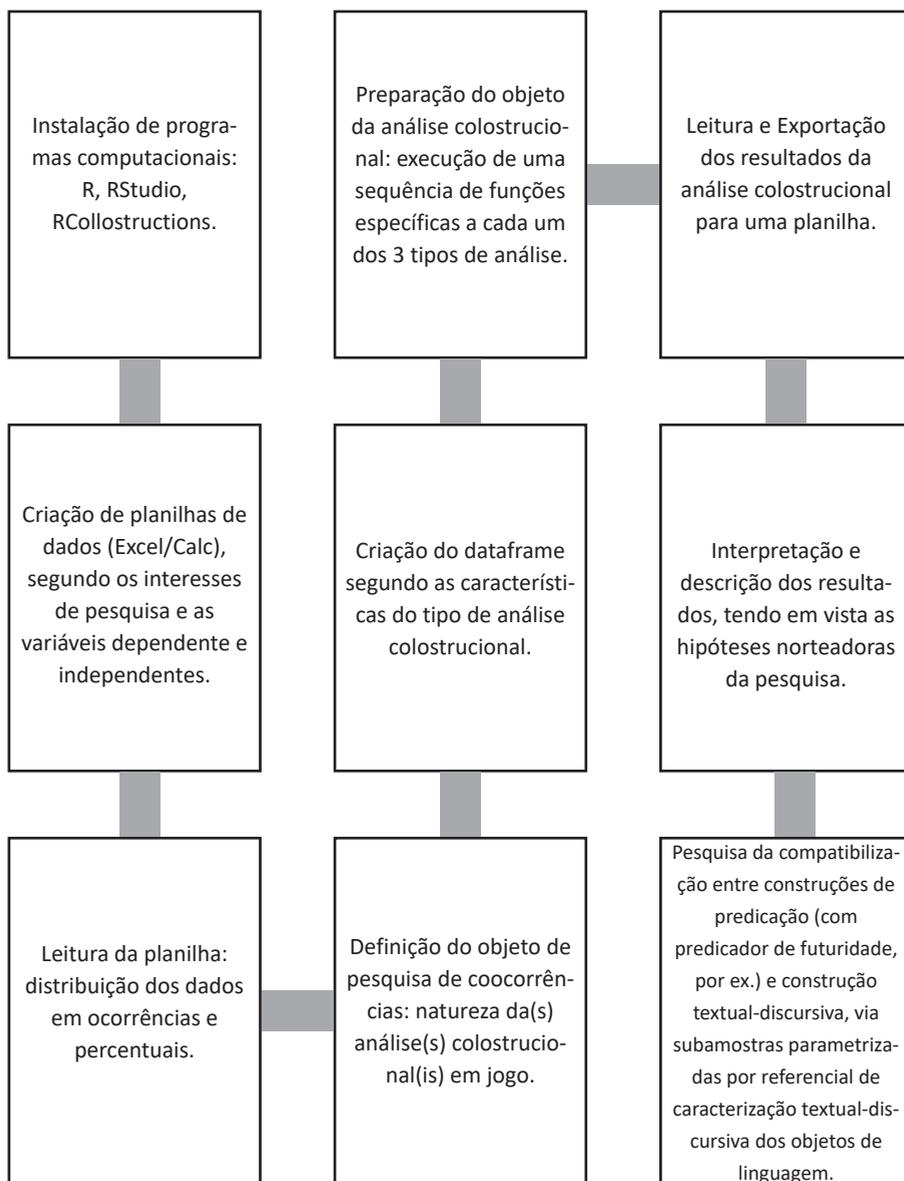
nominais – dados, hipótese, teoria, pressuposto, método, tratamento, quantitativo, qualitativo (WIEDEMER; MACHADO VIEIRA, 2022, p. 258).

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.