



Revista Diadorim
e-ISSN: 2675-1216
v.26, n.1, e63478, 2024
DOI: 10.35520/diadorim.2024.v26n1a63478

Entrevista

A interface entre Linguística Funcional e Gramática de Construções: aspectos teóricos e metodológicos – uma entrevista com Renata Enghels (Ghent University)

Edvaldo Balduino Bispo¹ 

Fernando da Silva Cordeiro² 

Renata Enghels³ 

Editor-chefe

Marcia dos Santos
Machado Vieira

Editores convidados

Edvaldo Balduino Bispo
Fernando da Silva Cordeiro
Renata Enghels

¹Universidade Federal do Rio Grande do Norte. Natal, RN, Brasil.

²Universidade Federal Rural do Semi-Árido. Caraúbas, RN, Brasil.

³Ghent University. Ghent, Belgium.

E-mails: edvaldo.bispo@ufrn.br; fernando.cordeiro@ufersa.edu.br;
renata.enghels@ugent.be

Como citar:

BISPO, Edvaldo Balduino;
CORDEIRO, Fernando da
Silva; ENGHELS, Renata.
A interface entre Linguística
Funcional e Gramática de
Construções: aspectos
teóricos e metodológicos –
uma entrevista com Renata
Enghels (Ghent University).
Revista Diadorim, v.26,
n.1, e63478, 2024. doi:
[https://doi.org/10.35520/
diadorim.2024.v26n1a63478](https://doi.org/10.35520/diadorim.2024.v26n1a63478)

A interface Linguística Funcional e Gramática de Construções já se mostrou produtiva para a análise dos mais variados fenômenos, em diversas línguas. Essas vertentes teóricas compartilham semelhanças no que diz respeito ao lugar do uso nas investigações linguísticas. No Funcionalismo, o uso é a realidade fundamental da língua, uma vez que se entende que a língua se molda pelos usos que dela fazemos nas situações de comunicação em que nos engajamos. Na Gramática de Construções, os construtos – instanciações concretas dos esquemas construcionais – são o meio pelo qual construímos generalizações sobre padrões que emergem da nossa experiência nas práticas interacionais. Funcionalismo e Gramática de Construções se aproximam também ao reconhecerem a estreita relação

entre forma e função, tão cara aos estudos funcionalistas e captada na Gramática de Construções no conceito de construção.

No entanto, esse frutífero diálogo também pode apresentar desafios de ordem teórico-metodológica, o que nos coloca no papel de pensar até que ponto essas vertentes se complementam e se diferenciam, bem como quais precauções se deve ter ao trabalhar com ambas. Nesse sentido, propomos esta entrevista com a Profa. Dra. Renata Enghels (Ghent University), cuja produção ilustra muito bem como a interface entre Funcionalismo e Gramática de Construções pode resultar em uma ferramenta potencial para a análise de fenômenos de natureza diversa. As questões buscam compreender a perspectiva adotada pela pesquisadora nas investigações que tem coordenado e qual o seu posicionamento acerca dos atuais desafios teórico-metodológicos da abordagem funcional-construcionista. Convidamos todos/todas à leitura.

EDVALDO BALDUINO BISPO e FERNANDO DA SILVA CORDEIRO (EBB e FSC): Quais temas você tem investigado e quais bases teóricas e metodológicas norteiam suas pesquisas?

RENATA ENGHELS (RE): Ao longo da minha carreira acadêmica, os pontos focais da minha investigação evoluíram de forma dinâmica, perpassando vários domínios do panorama linguístico. Por isso, não me identifico como especialista numa área específica. Em vez disso, ao abraçar tópicos que podem parecer bastante divergentes, consegui identificar padrões linguísticos subjacentes e adotar uma perspectiva abrangente, global, que gosto de descrever como uma visão de “helicóptero”. Além disso, apesar de ter-me concentrado em fenômenos específicos das línguas românicas, com um enfoque principal no espanhol, estendendo-me ao francês e, em certa medida, ao português, o meu objetivo foi também contribuir para o campo mais vasto da linguística. Os quadros teóricos com que tenho trabalhado podem ser definidos pelo termo abrangente “funcional”, englobando a Linguística Cognitiva e a Gramática de Construções e, mais recentemente, ampliando-se de modo a incluir aspectos da Sociolinguística. De um ponto de vista metodológico, tenho utilizado consistentemente uma abordagem ‘bottom-up’ e ‘data driven’. Inicialmente baseado em estudos de *corpus*, o meu quadro metodológico se desenvolveu, através de colaborações interdisciplinares com colegas da Psicolinguística, Psicologia, Sociologia e Literatura, para uma abordagem “multimétodo”. Essa mudança de paradigma ressalta o meu compromisso com a pesquisa interdisciplinar, refletindo a complexidade e a natureza multifacetada da linguagem.

De modo concreto, comecei com um projeto de investigação sobre a distinção entre nomes contáveis e incontáveis em espanhol e em francês. O meu projeto de doutoramento investigou a estrutura argumental dos verbos de percepção nessas duas línguas, um tópico que se estendeu naturalmente aos meus estudos de

pós-doutoramento, nos quais me concentrei nos verbos causativos em línguas românicas. Embora esses tópicos pudessem inicialmente parecer alinhados com estudos sintáticos bastante “tradicionais”, eles estavam inerentemente interligados às dimensões semânticas e pragmáticas. Guiado pelo princípio da iconicidade, meu objetivo era compreender as motivações subjacentes às variações formais no nível superficial da estrutura linguística. As línguas românicas constituem um laboratório ideal para esse fim, devido à sua capacidade de variação morfossintática em diferentes graus. Estou particularmente interessada nos modos pelos quais os fatores extralinguísticos e conceptuais, incluindo as necessidades comunicativas articuladas pelos falantes, podem explicar as variações na ordem sintática das palavras, nas marcações de caso e nos mecanismos de concordância, entre outros.

Esse interesse na interação dinâmica entre forma e função e, especificamente, no modo como os falantes escolhem um item/elemento entre diferentes membros de um ‘esquema’ abrangente continuou em projetos de pesquisa mais recentes. Embora possam parecer bastante diferentes dos tópicos da “primeira onda”, eles estão, de fato, intrinsecamente relacionados por um fio condutor: a exploração das escolhas linguísticas e da variação no quadro das teorias funcionais. Em termos concretos, esses temas dizem respeito:

- i. ao exame de fenômenos específicos do espanhol coloquial informal. Esse domínio oferece um terreno fértil para a observação de escolhas linguísticas espontâneas, incluindo, mas não se limitando a, marcadores pragmáticos, bem como estratégias de intensificação e atenuação;
- ii. à exploração de padrões que influenciaram o desenvolvimento histórico de categorias altamente produtivas nas línguas românicas, como as perífrases verbais e os marcadores pragmáticos. Essa investigação recorre às teorias da gramaticalização e da construcionalização para desvendar as trajetórias evolutivas desses elementos linguísticos, contribuindo para nossa compreensão da mudança e do desenvolvimento linguísticos;
- iii. a uma investigação sobre o “significado” de itens linguísticos específicos em socioletos relacionados à faixa etária, com especial incidência na linguagem dos jovens. Essa linha de investigação visa descrever os significados multifacetados implicados nas práticas linguísticas das gerações mais jovens, revelando como a inovação linguística pode refletir mudanças sociais e culturais mais amplas;
- iv. à análise da “gramática” subjacente que rege os padrões de alternância de código.

Nesses projetos, concentramo-nos na variação observada tanto entre indivíduos quanto no seio das comunidades, com o objetivo de delinear a natureza sistemática da alternância de código como fenômeno linguístico.

O fio condutor entre esses tópicos é a exploração do modo como as variações sintáticas e as variações formais como um todo estão intrinsecamente ligadas ao “significado” que um falante pretende expressar, entendendo-se “significado” num sentido lato. Para além do significado lexical, semântico e pragmático, considero que as escolhas linguísticas são simultaneamente orientadas por significados conceituais e sociais. Tais significados refletem a percepção que um indivíduo tem do mundo exterior, os traços de sua personalidade, o seu posicionamento em contextos comunicativos específicos e a identidade que pretende projetar por meio de suas escolhas linguísticas. Reflexões recentes levaram-me a abraçar a perspectiva de que “está tudo conectado”, no sentido de que fenômenos que são tipicamente estudados pela Sociolinguística, pela Linguística Histórica ou pela Pragmática, para citar apenas alguns, parecem ser conduzidos por princípios linguísticos comparáveis, transversais. Esses princípios incluem considerações de economia (a necessidade de eficiência linguística), de iconicidade (o espelhamento entre forma e significado), de criatividade (a capacidade de criar expressões inovadoras), de expressividade (a necessidade de exprimir emoções e posições) e de produtividade (a capacidade geradora dos sistemas linguísticos de produzir variações de um padrão).

(EBB e FSC): Como você avalia, teórica e metodologicamente, a aproximação (convergência?) entre a Linguística Funcional e a Gramática de Construções?

(RE): A Linguística Funcional e a Gramática de Construções estão, naturalmente, muito próximas no que diz respeito a seus objetivos primários e fundamentos teóricos. Ambas as teorias partem de um forte enfoque na explicação do uso da língua em contextos (sociais). Enquanto a Linguística Funcional evoluiu através do desenvolvimento de uma série de modelos, teorias e estudos de caso, partindo do papel das funções comunicativas no uso da língua, a Gramática de Construções (Cognitiva) enfatizou uma visão da língua como um *constructicon* complexo, ou rede de pares forma-significado, isto é, construções.

Como consequência, há muitos pontos de convergência teórica entre os dois modelos/paradigmas. Ambos se comprometem a abordar a língua principalmente como um instrumento de comunicação, com uma forte ênfase no significado e no uso. A língua é vista como um *continuum*, rejeitando-se a separação estrita entre níveis linguísticos, como o léxico, a semântica, a sintaxe etc. Assim, o ‘significado’ é definido de forma ampla, estendendo-se muito além das fronteiras tradicionais do conteúdo semântico, com integração recente de elementos de significado socialmente relevantes no polo funcional das construções na Gramática de Construções. Para além de outras premissas básicas partilhadas, as perspectivas também se alinham na sua visão da “gramática emergente” e na importância do contexto e do uso para compreender a essência da língua(gem).

Por conseguinte, ambos os modelos se baseiam fortemente em abordagens metodológicas semelhantes, incluindo dados empíricos representativos do uso real da língua (por exemplo, dados de *corpus*, mas também os resultados de experimentos *online* e *offline*). No entanto, embora não haja uma divisão rigorosa do trabalho, existem algumas diferenças quanto aos métodos preferidos para analisar esses dados. Na Linguística Funcional, o foco pode estar mais nas avaliações qualitativas do discurso, observando de perto os padrões linguísticos nas interações, ao passo que a Gramática de Construções (Cognitiva) implica frequentemente uma modelagem quantitativa que fornece informações sobre o modo como as construções são produzidas e processadas pelos falantes.

Penso que a convergência entre ambas as abordagens é muito valiosa e faz todo o sentido. Quando observo estudos de casos concretos, tenho por vezes dificuldade em distinguir entre elas. O que me parece particularmente atraente na abordagem construcional é a premissa de que a língua se organiza em torno de ‘esquemas’ mais ou menos abstratos, correspondentes a diferentes necessidades funcionais (o lado do significado), e a ideia de que esses esquemas se organizam numa rede (e mesmo numa rede ‘dinâmica’, segundo o modelo de Diessel), construída com base em diferentes tipos de relações construcionais. Essa é uma visão complementar que o modelo construcional tem para oferecer ao modelo funcional. Além disso, como já foi dito, considero que o conceito de significado deve ser definido de forma ampla e, como tal, a ‘definição’ de uma construção ou de qualquer fenômeno linguístico pode se beneficiar muito da descrição de nuances relacionadas ao contexto, reveladas por meio de análises funcionais detalhadas. A análise detalhada de aspectos discursivos e a modelagem quantitativa dos padrões observados nos “grandes dados” devem andar de mãos dadas.

Creio que ainda há trabalho a fazer para tornar a aproximação entre essas abordagens mais explícita na literatura, de modo a colocar mais explicitamente quais são os principais pontos de convergência, mas também os desafios (metodológicos e teóricos) na sua integração. Assim, a sua fusão de forma mais transparente pode constituir um objetivo futuro fulcral para os grupos de investigação (brasileiros) que trabalham nesses domínios.

(EBB e FSC): Que precauções devemos ter em conta quando trabalhamos com estes dois domínios teóricos em conjunto?

(RE): Esta é uma questão não muito fácil de responder. De um modo geral, considero que os modelos e as teorias linguísticas existem como instrumentos que permitem ao investigador organizar e explicar os padrões que observa nos dados, e não o contrário. Isso significaria que as observações são “forçadas” a caber num modelo

pré-estabelecido. Por isso, devemos ter cuidado para não nos deixarmos prender indevidamente a um determinado quadro ou abordagem. Penso que as precauções com esses campos teóricos poderiam ser alargadas à investigação linguística em geral. Aqui estão alguns desafios adicionais que precisam ser abordados.

Em primeiro lugar, é importante começar com uma terminologia clara. Como discutimos anteriormente, ambas as teorias têm terminologias e premissas conceituais específicas, mas também dentro de cada campo há pontos de vista divergentes (por exemplo, na Gramática de Construções, quando a noção de ‘construção’ se aplica realmente a um padrão observado e quando não). Um pesquisador deve trabalhar com um léxico transparente e com termos claramente definidos desde o início. Isso garante que os modelos teóricos não sejam aplicados prematuramente para interpretar os dados. Assim, eu encorajaria uma aplicação cuidadosa e baseada em evidências de ambas as teorias.

Em segundo lugar, tal como é importante para a investigação (linguística) em geral, os modelos têm de ser aplicados com suficiente rigor metodológico. Em particular, em função da ênfase de ambos os campos nos dados empíricos, penso que é realmente importante selecionar cuidadosamente as fontes de dados (ou seja, os *corpora*), os métodos e as técnicas de validação, dependendo da questão de investigação a que se pretende responder. Por exemplo, aprendi no passado que, se quisermos trabalhar com medidas de produtividade, um tópico que abordaremos mais tarde, mas que é amplamente utilizado na abordagem construcional, é crucial uma amostragem cuidadosa dos dados. Isso é especialmente importante se quisermos comparar contextos comunicativos, por exemplo, numa perspectiva histórica ou sociológica. Se você não prestar atenção suficiente a esse aspecto metodológico, corre o risco de suas conclusões empíricas serem enviesadas. De modo geral, a transparência é importante em nossa metodologia e em nossas análises, para que possam ser reproduzidas por outros. A Universidade de Gent implementou uma política, a partir de 2023, que exige que todos os conjuntos de dados utilizados em nossa investigação sejam publicados em formatos de acesso aberto, como o repositório Trolling. Essa iniciativa reforça a importância de práticas de investigação acessíveis e transparentes que facilitem a validação e a exploração posterior por nossos colegas.

Por último, a fim de evitar raciocínios circulares e, por vezes, interpretações demasiado intuitivas dos dados, penso que é importante procurar, tanto quanto possível, critérios de validação objetivos em ambos os domínios. Sobretudo quando se trata de fenômenos pragmáticos e semânticos, é imperativo não assumir interpretações funcionais sem uma análise rigorosa. Surgem questões como: por que um elemento específico é considerado um marcador de posicionamento? Que provas contextuais apoiam uma interpretação em detrimento de outra? Embora a investigação colocacional se tenha tornado um método predominante para desvendar as nuances do ‘significado’, essa abordagem nem sempre pode ser aplicável, ou pode apresentar

desafios, por exemplo, no estudo de marcadores pragmáticos. Nesses casos, continua a ser essencial que as análises funcionais se esforcem por identificar pistas mais objetivas. Essa abordagem garante que as interpretações não dependem apenas de nosso julgamento subjetivo, mas se baseiam em evidências observáveis e verificáveis.

(EBB e FSC): No que diz respeito à metodologia, como podemos definir uma amostra adequada (*corpus*) para uma investigação empírica de um fenômeno discursivo?

(RE): Esta é uma questão muito relevante que, de fato, se relaciona com um ponto que levantei anteriormente. Quando estudamos fenômenos relacionados ao discurso, é muito importante basear nossas análises numa amostra equilibrada e representativa. No entanto, apesar de todos os esforços para garantir a adequação da amostra, é crucial reconhecer que as nossas conclusões são específicas apenas para essa amostra e que devemos ser cautelosos antes de generalizar essas conclusões para diferentes gêneros ou contextos discursivos.

Assim, dada a natureza distinta dos fenômenos específicos do discurso, é importante ter acesso a dados orais representativos. No domínio das línguas românicas, o acesso a dados relevantes foi durante muito tempo limitado ou quase inexistente. Por exemplo, o primeiro *corpus* oral informal do espanhol peninsular só ficou disponível na década de 1970. Felizmente, os pesquisadores estão cada vez mais conscientes do fato de que os fenômenos relacionados ao discurso podem ser mais bem compreendidos através da análise de *corpora* falados, e que esses dados precisam ser recolhidos e preservados. Como resultado, para o espanhol, temos agora acesso a uma gama mais ampla de *corpora* falados que representam o discurso espontâneo do século XXI.

A nossa equipe de investigação deu um contributo significativo para esse campo ao gravar e transcrever recentemente três *corpora*: um *corpus* de espanhol coloquial chileno espontâneo, um *corpus* altamente comparável de espanhol falado em Madrid (conhecido como o *corpus* CORMA) e, finalmente, um *corpus* bilíngue espanhol-inglês de fala espontânea em El Paso, EUA. Apesar dessas contribuições, é importante notar que esses *corpora* representam variantes específicas da língua e são frequentemente limitados em termos de dimensão e escopo. De acordo com a nossa experiência, a recolha de dados autênticos e representativos é um processo extremamente dispendioso e moroso, um fato que é frequentemente ignorado pela comunidade acadêmica. No entanto, quando se trabalha com esse tipo de *corpus*, a validade ecológica da investigação é substancialmente assegurada. Por isso, defendo que a característica mais importante de uma amostra adequada é a sua autenticidade.

A seleção e as características de uma amostra devem, de fato, ser estreitamente adaptadas ao fenômeno investigado, o que indica que a amostra deve estar precisamente alinhada com as questões e os objetivos da investigação em causa. Por exemplo,

quando se estudam estruturas argumentais verbais, a inclusão dos dados da amostra pode não ser tão fortemente influenciada pela variedade de contextos discursivos em que são utilizadas construções ou verbos específicos. Por outro lado, se o foco for em fenômenos característicos da oralidade (como o uso de marcadores pragmáticos e estratégias de intensificação e atenuação), os contextos sociais e/ou comunicativos específicos em que esses são utilizados tornam-se significativamente mais relevantes. Essa distinção sugere a necessidade de diferentes abordagens à coleta de dados: para estudar estruturas argumentais verbais, a utilização de dados de uma fonte de dados ampla como o *Sketch Engine*, que agrega conteúdos de diversas fontes da internet, pode ser adequada devido a seu vasto conjunto de dados. No entanto, para a investigação de fenômenos discursivos em que o contexto social e comunicativo é crucial, é necessária uma abordagem mais controlada, com dados que incluam perfis sociológicos detalhados dos falantes e contextos discursivos específicos. Apesar do fascínio da extensa base de dados do *Sketch Engine*, a sua falta de perfis detalhados dos falantes torna-o menos ideal para o estudo de fenômenos relacionados com a fala, sublinhando a importância de selecionar fontes de dados que proporcionem uma compreensão abrangente dos antecedentes dos falantes e dos contextos de suas conversas.

Isso me leva, naturalmente, a outra questão pendente, que é a dimensão do *corpus*. Em geral, a dimensão foi definida como crucial para garantir a significância estatística. No entanto, mais uma vez, isso depende dos objetivos da investigação. Se quiser fazer uma análise aprofundada e multidimensional de uma estrutura, de uma construção ou de uma expressão linguística, pode ser mais útil tentar captar a natureza específica do fenômeno sob estudo, concentrando-se numa amostra mais limitada, mas acrescentando parâmetros de investigação adicionais. Um aspecto fundamental a ter em conta é que a amostra deve ser gerível dentro dos limites do tempo disponível e dos recursos analíticos. Encontrar um equilíbrio entre a amplitude dos dados e a profundidade da análise garante que os resultados da investigação sejam significativos e perspicazes.

Em termos gerais, penso que a melhor forma de lidar com a questão da representatividade das amostras é ser transparente quanto à sua compilação, composição e processamento. Ao serem abertos e claros sobre esses aspectos, os pesquisadores facilitam a reprodutibilidade de suas análises por outros que utilizem amostras diferentes. Essa transparência é essencial para o avanço do campo de investigação, uma vez que permite a validação e a comparação dos resultados da pesquisa em vários estudos. Assim, a melhor prática para abordar a representatividade das amostras reside na documentação exaustiva e na abertura, proporcionando uma base para o diálogo e a exploração acadêmicos contínuos.

(EBB e FSC): Como conciliar e/ou equilibrar fatores quantitativos e qualitativos na investigação baseada no uso? Pode dar-nos um exemplo?

(RE): Essa questão extremamente importante foi parcialmente abordada anteriormente. Como já foi dito, é crucial complementar a investigação quantitativa com a qualitativa e vice-versa. Mesmo no caso da estatística multivariada, que leva em conta simultaneamente o impacto de diferentes parâmetros num mesmo resultado linguístico, a modelagem quantitativa tem de ser concebida e alimentada por informações provenientes de um exame qualitativo meticuloso dos dados. Essa etapa garante que a análise quantitativa se baseia numa compreensão criteriosa dos fenômenos linguísticos em estudo. Os pesquisadores com uma forte proficiência em métodos quantitativos podem não ter uma compreensão intuitiva das implicações pragmáticas sutis transmitidas pelo uso de estruturas específicas em contextos comunicativos particulares. Entretanto, os pesquisadores que se concentram predominantemente em análises qualitativas correm o risco de generalizar excessivamente a partir de um conjunto limitado de observações. Podem construir teorias abrangentes com base em exames detalhados de algumas instâncias de um fenômeno, sem verificar a significância estatística ou a generalização dos padrões e correlações observados.

Por isso, penso que uma primeira fase de qualquer projeto de investigação sobre um novo tópico deve consistir na leitura atenta de *corpora* representativos em que o fenômeno será estudado. Tal como salientado pelo exemplo do estudo das estratégias de intensificação no espanhol contemporâneo, um tópico que está sendo explorado por nossa aluna de doutoramento Linde Roels, essa abordagem é inestimável. Permite aos pesquisadores mergulharem no uso real da língua, examinando transcrições, identificando metodicamente uma série de variantes concorrentes e compreendendo sua função e nuances dentro da língua.

Essa análise qualitativa é crucial para descobrir como essas características linguísticas são utilizadas em várias estratégias comunicativas, incluindo as que visam à manutenção de polidez e ao gerenciamento de preservação de face entre os falantes. Essa análise permite compreender a natureza dinâmica da língua e seu uso na interação social, revelando a complexa interação entre as escolhas linguísticas e as intenções comunicativas. Só depois de essas percepções qualitativas terem sido estabelecidas é que os pesquisadores podem proceder à definição precisa das variáveis relevantes. Essa etapa é essencial para o desenvolvimento subsequente de modelos quantitativos mais sofisticados.

(EBB e FSC): Como as ferramentas/recursos computacionais têm ajudado nas pesquisas por você desenvolvidas?

(RE): Acho que todos concordamos que as várias ferramentas computacionais que temos hoje à nossa disposição constituem um recurso de enorme valor para a nossa pesquisa. Sem elas, a “virada empírica” e, mais especificamente, a “virada quantitativa” na linguística não teria sido possível. Comecei a minha investigação de doutoramento numa área em que ficávamos felizes por ter à nossa disposição os pacotes normais do Windows. Já não havia necessidade de recolher e transcrever manualmente exemplos relevantes de *corpus* – imagine fazer isso hoje em dia –, mas podíamos reunir todos esses dados em bases de dados, como o Access, hoje em dia não tão frequentemente utilizado, e o Excel, ainda indispensável na prática diária de pesquisadores na área da linguística. Pode parecer um recurso básico, mas alguns conhecimentos mais avançados da ferramenta permitem efetuar muitas operações com os dados: extrair amostras aleatórias, anotar cuidadosamente (e por vezes automaticamente) os dados, obter algumas ideias rápidas sobre padrões nos dados através de tabelas cruzadas, visualizar os dados etc. Em minha opinião, essa ferramenta continua a ser indispensável no domínio da linguística de *corpus*.

Utilizamos outras ferramentas para realizar atividades de investigação específicas: *Praat*, *EXMARaLDA* e *ELAN*, para a transcrição de gravações de dados conversacionais, por exemplo; a integração de ferramentas online em plataformas como o *Sketch Engine* facilita a pesquisa de *corpus* com etiquetagem PoS e baseados em *lemmas*, simplificando o processo de identificação de padrões linguísticos em grandes conjuntos de dados.

Para o processamento e a pesquisa concreta de padrões nos dados e, consequentemente, para a modelagem estatística, aprendi inicialmente a trabalhar com o programa SPSS. No entanto, rapidamente se tornou claro que o padrão de fato para a análise de dados era o ambiente de software livre e de código aberto *R*. O *R* fornece uma vasta gama de técnicas estatísticas e gráficas e é altamente extensível através de novos pacotes. O único inconveniente é a sua curva de aprendizagem acentuada: é necessário bastante esforço para se familiarizar com a linguagem de programação, mas, uma vez familiarizado com os fundamentos, é possível obter um vasto conjunto de habilidades, abrangendo manipulação de dados, cálculo estatístico e apresentação gráfica avançada.

Neste momento, estou explorando as possibilidades das aplicações GAI, como o *ChatGPT*, para a análise de dados de *corpus*. A minha primeira experiência é que, quando se consegue introduzir os dados num formato legível pelo *Chatbot*, ele pode tornar-se uma ferramenta muito útil (e mais acessível) para analisar padrões subjacentes nos dados. No entanto, como essas ferramentas ainda são relativamente novas,

os resultados devem ser tratados com cuidado e comparados com os resultados de outros recursos. Estou ansiosa para testemunhar a evolução futura desse campo e seu potencial para se tornar uma ferramenta integral no trabalho cotidiano de cada linguista (de *corpus*).

(EBB e FSC): Relativamente à produtividade, que aspecto(s) considera mais relevante(s) para medir a produtividade de uma construção? Por quê?

(RE): O fenômeno da produtividade linguística, enquanto propriedade fundamental das regras de nossa gramática mental, constitui, na minha opinião, um dos tópicos de investigação mais fascinantes. É interessante observar que as diferentes construções que compõem a nossa gramática têm o seu próprio domínio de aplicação, ou seja, a sua potencial aplicabilidade para incorporar novas palavras lexicais ou novos membros. Além disso, esse nível de produtividade interage com características sociais e cognitivas dos falantes que as produzem e se desenvolve ao longo do tempo. Além disso, a relação entre essa propriedade natural do sistema gramatical e o fenômeno da criatividade, definido como a capacidade do falante de usar a língua de formas novas e inovadoras, mostrando a imensa flexibilidade e adaptabilidade da linguagem humana, apresenta uma área rica para investigações futuras.

Tal como defendemos em investigações anteriores realizadas por nosso grupo de pesquisa interdisciplinar *Language Productivity@work* (ver o nosso sítio Web: <http://www.languageproductivity.ugent.be>), o principal problema do conceito de produtividade é que constitui uma caixa preta: é crucial para a linguagem, mas faz parte do conhecimento linguístico implícito dos falantes, de modo que não pode ser observado diretamente. Trata-se de um construto teórico que deve ser deduzido a partir de suas manifestações observáveis. Mas é aqui que surge outra dificuldade: há muitas manifestações diferentes: na produção linguística concreta, mas também na forma como a língua é processada tanto pelo falante quanto pelo ouvinte quando produzida, e isso para cada falante ou ouvinte individualmente. Por isso, se você quer alcançar uma imagem completa da produtividade de uma construção específica, diria que a melhor maneira de o fazer é adotar uma abordagem multimétodo, para ver como a sua produtividade é realizada na linguagem contemporânea e/ou passada, como os falantes avaliam instâncias específicas de uma construção (*offline*), ou o que acontece no cérebro (*online*) quando são confrontadas com instâncias novas (e possivelmente criativas) de um padrão e, finalmente, como todas essas medidas podem variar entre diferentes comunidades ou indivíduos.

Penso que a questão está sobretudo relacionada com a visão de um/uma linguista de *corpus* e com a forma como ele/ela observa a produtividade. De fato, como linguistas de *corpus*, olhamos para o ‘uso acumulado da língua’, para o resultado da

produtividade em jogo. Assim, efetuamos diferentes tipos de contagens em grandes bases de dados de texto. Estes são os mais importantes, os quais considero todos relevantes:

- i. a frequência *token*, ou frequência geral de ocorrência de uma construção;
- ii. a frequência *type*, ou diferentes tipos lexicais atestados numa determinada posição de uma construção. A ideia geral é que, quanto mais tipos diferentes forem registrados, mais produtiva é a construção. Isso nos levou a aferir a razão entre as frequências *type* e *token*;
- iii. outra medida importante é a proporção de ocorrências de *hapax*, que mede o número de novos tipos diferentes que aparecem numa construção. A ideia é que, quanto maior for o número de ‘*types* únicos’, ou *hapaxes* diferentes, mais a construção é potencialmente produtiva.

Medidas que são muito menos discutidas na literatura, mas que são pelo menos tão perspicazes, creio eu, são as medidas potenciais de “antiprodutividade”. Essas medidas indicam o que é possivelmente prejudicial à produtividade e incluem as contagens dos preenchimentos mais frequentes, a média e o desvio padrão dos três preenchimentos mais frequentes e a curva de distribuição Zipfiana ajustada (das contagens de frequência de *tokens* como uma medida de produtividade baseada numa curva). O princípio subjacente aqui é que alta frequência *token* e a convencionalização dos itens mais frequentes prejudicariam a produtividade. Para uma leitura mais aprofundada sobre esse tema, recomendo a tese de doutoramento de Niek Van Wettere (2021).

Voltando à sua pergunta: as diferentes medidas destacam diferentes aspectos da produtividade, e eu recomendaria o cálculo de todas elas para ter uma ideia mais precisa da produtividade de uma ou mais construções. Ainda assim, mostramos num artigo recente (Van Hulle, Lauwers, Enghels, 2024)¹ que as chamadas medidas de ‘antiprodutividade’, por um lado, e as medidas de produtividade, por outro, estão fortemente correlacionadas e constituem opostos ao longo da mesma dimensão de abertura lexical.

Isso me leva ao último ponto, nomeadamente a necessidade de integrar o “significado” nos estudos de produtividade. Em vez de contar apenas o número de instâncias, pode-se querer saber qual é a cobertura semântica de uma construção ou de microconstruções (concorrentes). Ou se as novas instâncias de uma construção aparecem dentro de um domínio semântico preferido, ou se estão por todo o espaço onomasiológico. Isso só é abordado marginalmente pelas medidas anteriormente mencionadas, por exemplo, na razão entre frequências *type/token*. No entanto, muito

¹ O artigo foi publicado em: <https://doi.org/10.1515/ling-2023-0087>

pode ser aprendido sobre o grau de ‘abertura semântica’ de uma construção por meio da análise de medidas adicionais, tais como a amplitude semântica (como a proporção de grupos semânticos cobertos por uma (micro)construção específica) e a dispersão semântica (como a relação semântica média dos *types* de uma (micro)construção).

(EBB e FSC): Você investigou os marcadores pragmáticos do espanhol a partir de uma perspectiva construcionista. Quais desafios enfrentou ao lidar com fenômenos discursivos a partir desse quadro teórico e o que fez para superá-los?

(RE): De fato, em alguns dos meus trabalhos sobre marcadores pragmáticos, defendi uma abordagem construcionista desses fenômenos no nível discursivo. Isso está em conformidade com um desenvolvimento recente no campo da Gramática de Construções, completamente de acordo com o seu princípio inclusivo e integrador, que evita distinguir padrões nucleares (morfofossintáticos) e periféricos (pragmáticos). Nessa perspectiva (defendida por Östman em 1999 e Östman & Fried em 2005, para citar apenas alguns), é crucial definir *frames* ou padrões discursivos que também se encontram em diferentes tipos de relações entre si, incluindo relações de herança dentro de uma rede construcional.

Desse ponto de vista, as expressões da língua falada que partilham funções e características formais semelhantes - tais como vários tipos de marcadores pragmáticos conversacionais - são consideradas exemplos de uma construção mais ampla. Essa construção abrangente atua como um modelo que não somente forma, mas também ajuda a interpretar novas expressões.

Depois de ter estudado diferentes marcadores pragmáticos individuais, encontrei muitos problemas, questões e dilemas pendentes, com os quais outros investigadores também se confrontaram, mas que aparentemente não puderam ser abordados de forma adequada. Essas questões incluem: a tensão entre micro e macrofunções, o desejo de trabalhar com critérios objetivos para definir a sua função, a necessidade de complementar a perspectiva semasiológica mais abrangente (partindo de formas concretas de marcadores pragmáticos) com uma perspectiva onomasiológica (partindo das necessidades funcionais do discurso que os marcadores pragmáticos preenchem). A certa altura, percebi que uma abordagem construcional poderia eventualmente dar resposta a algumas dessas questões. Por exemplo, ela permite dar conta de diferentes níveis de análise no comportamento funcional, referindo-se à sua polifuncionalidade como marcadores de polidez, marcadores modais e/ou elementos argumentativos e/ou metadiscursivos. Adotando uma perspectiva construcional, esses diversos níveis não são considerados como mutuamente exclusivos; em vez disso, enriquecem coletivamente o aspecto ‘significado’ do esquema. As características que constituem o lado formal da moeda incluem, então, padrões quanto a seu comportamento posicional ou à presença de elementos colocacionais específicos, que permitem discernir

subesquemas ou microconstruções do esquema pragmático global do marcador. Além disso, se partirmos da ideia de que os falantes desejam instanciar e exprimir padrões regulares com padrões específicos convencionalizados, alargamos naturalmente a nossa perspectiva. Essa abordagem transcende os limites de marcadores pragmáticos particulares categorizados pelas suas origens lexicais, tais como os derivados de verbos de movimento, verbos de percepção, e – creio eu – aproxima-se muito mais da forma como a linguagem funciona realmente no cérebro.

No entanto, uma vez que esse método é relativamente novo e não está tão difundido, a aplicação das ferramentas analíticas originalmente desenvolvidas para as construções de estrutura argumental a fenômenos discursivos de nível superior apresenta vários desafios (metodológicos) relacionados com a terminologia, com as categorias conceptuais e com as estratégias de representação. Por exemplo, tive dificuldade em definir a estrutura da própria rede: em que nível do *constructicon* operam as diferentes variáveis estruturais? Será que noções como microconstrução, mesoconstrução etc. se aplicam igualmente? O que é que conta como microconstrução ou subesquema do padrão global? Como definir os diferentes *slots* de uma construção no nível do discurso? E, à luz das medidas de produtividade que já discutimos, o que é que conta como um tipo diferente de construção no nível discursivo? Trata-se apenas da composição formal interna da construção ou engloba diferentes propriedades funcionais (incluindo pragmáticas)? Como incluir características de um padrão conversacional dialógico maior?

Assim, para além dessas questões remanescentes, um dos principais desafios, creio eu, é o de desvendar melhor o lado do ‘significado’ da construção, de modo a incluir não só características semânticas e pragmáticas dos padrões conversacionais, mas também o significado social indexical. Do ponto de vista formal, poderia ser feito mais trabalho no sentido de incluir a prosódia na definição das construções. Como outros critérios formais muitas vezes não conseguem ajudar e compreender o funcionamento dos marcadores pragmáticos, as características prosódicas podem constituir pistas adicionais.

Assim, essas questões, juntamente com muitas outras, realçam que este é um campo de investigação simultaneamente promissor e exigente, oferecendo muitas oportunidades para a realização de estudos de ponta num futuro próximo.