

APRESENTAÇÃO

Gramática de Dependências: fundamentos e perspectivas

Adriana PAGANO 

Universidade Federal de Minas Gerais (UFMG)

Marta Deysiane Alves Faria SOUSA 

Universidade Federal de Sergipe (UFS)

Jorge BAPTISTA 

Universidade do Algarve - INESC-ID Lisboa

RESUMO

Dada a importância atual da gramática de dependências, este dossiê surgiu como iniciativa de reunir estudos enfocando a língua portuguesa, no sentido de alavancar discussões sobre o potencial dessa abordagem, tanto para a reflexão linguística teórica em geral, como também para a descrição e comparação de línguas em particular. Neste sentido, buscou promover uma visão de complementaridade e reciprocidade entre linguística e linguística computacional, com ganhos significativos para ambos os campos. A chamada realizada convocou estudos em que fossem discutidos temas como a utilização de estruturas de dependência na linguística teórica; a adoção da gramática de dependências nos estudos de tipologia, linguística comparada e estudos da tradução; a incorporação de estruturas de dependência na linguística de corpus; os avanços mais recentes em parsing de dependência e geração de textos. Foram também sugeridos estudos que indagassem as bases históricas e epistemológicas da gramática de dependências, bem como sua relação com outros formalismos; e também a relação entre a gramática de dependência e outros campos disciplinares, como a psicolinguística e os estudos da cognição. Temas correlatos, tratados tanto de um ponto de vista teórico quanto de um ponto de vista aplicado, também estavam no escopo da chamada. Assim, os seis artigos, que ora compõem este dossiê, tratam do tema da gramática de dependências trazendo reflexões nos eixos teórico, aplicado e avaliativo.



OPEN ACCESS

COMO CITAR

Pagano, A.; Souza, M. D. A. F.; Baptista, J. (2024). Gramática de Dependências: fundamentos e perspectivas. *Revista da Abralín*, v. 22, n. 2, p. 158-162, 2024.

Apresentação

A gramática de dependências constitui uma abordagem de longa tradição nos estudos da linguagem, associada a trabalhos que se remontam a tempos antes da Era Comum, como é o caso da obra de Panini (520 AEC. – 460 AEC.), e a séculos recentes, em particular, com a obra de Louis Hjelmslev (1899-1965) e Lucien Tesnière (1893-1954). Trata-se de uma proposta alternativa à gramática de constituência que, diferentemente desta última, privilegia as relações de interdependência entre as unidades linguísticas e representa a centralidade do verbo na sentença. Apesar de sua longa trajetória, a gramática de dependências teve impacto moderado nos círculos europeus, sendo sua repercussão escassa na América do Norte e na América Latina. No caso do Brasil, mais especificamente, pode-se afirmar que constitui uma abordagem praticamente, senão totalmente, ausente nos cursos de formação em linguística. Paradoxalmente, a gramática de dependências conta com maior impacto na ciência da computação, sobretudo na linguística computacional, onde vem ganhando cada vez mais destaque, diante de seu potencial como teoria que informa a anotação de *corpora* para o desenvolvimento de bancos de árvores sintáticas ou *treebanks*, orientados para a tarefa de análise sintática ou *parsing*, crucial nas chamadas *pipelines* de processamento de linguagem natural. Nesse sentido, enquanto a linguística teórica continua alheia ao crescimento da gramática de dependências e às demandas por estudos que subsidiem a formação de linguistas preparados para coordenar e participar em projetos de anotação de *corpora*, a linguística computacional vem preenchendo essa lacuna para resolver as necessidades do processamento de linguagem natural.

Ainda em decorrência de sua relevância no processamento de linguagem natural, a abordagem da gramática de dependências ganhou mais destaque, nesta última década, com a criação e consolidação do projeto Universal Dependencies (UD)¹. Este projeto visa desenvolver uma anotação consistente de *treebanks* em distintas línguas de forma tal que possa haver comparabilidade entre elas. Se bem esse desejo de comparabilidade entre línguas, capturado pela palavra "universal" no nome do projeto, foi motivado pelo objetivo de desenvolvimento de *parsers* multilíngues com aplicação no processamento de linguagem natural, a comparabilidade fez com que o projeto fosse tendo impacto cada vez mais crescente nos estudos linguísticos, sobretudo sob a perspectiva da tipologia linguística. De fato, o conjunto de categorias e etiquetas para a anotação constituem uma base universal no sentido de facilitar a anotação consistente de construções semelhantes em línguas distintas; todavia, o projeto UD contempla a possibilidade de adaptações específicas para cada língua, quando necessário.

No momento presente, a língua portuguesa conta já com diversos *treebanks* anotados no padrão UD, resultados do esforço de distintos grupos de pesquisa, alguns dos quais estão representados nesta publicação. No caso do português brasileiro, diretrizes, manuais e softwares de anotação foram desenvolvidos e são atualizados continuamente de acordo com as discussões no fórum internacional do projeto UD².

¹ Informações detalhadas em <https://universaldependencies.org/>.

² Cf. <https://ud-brasil.gitbook.io/portal>

Nesse contexto e dada a importância atual da gramática de dependências, este presente dossiê surgiu como iniciativa de reunir estudos enfocando a língua portuguesa, no sentido de alavancar discussões sobre o potencial dessa abordagem, tanto para a reflexão linguística teórica em geral, como também para a descrição e comparação de línguas em particular. Neste sentido, buscou promover uma visão de complementaridade e reciprocidade entre linguística e linguística computacional, com ganhos significativos para ambos os campos.

A chamada realizada convocou estudos em que se discutiam temas como a utilização de estruturas de dependência na linguística teórica, seja para modelar relações sintáticas, fenômenos semânticos e pragmáticos relacionados à estrutura da informação, ou fenômenos discursivos que transcendem o nível da sentença; a adoção da gramática de dependência nos estudos de tipologia, linguística comparada e estudos da tradução; a incorporação de estruturas de dependência na linguística de *corpus*, com ênfase na anotação de *corpora* e desenvolvimento de *treebanks* e recursos linguísticos; os avanços mais recentes em *parsing* de dependência e geração de textos. Foram também sugeridos estudos que indagassem as bases históricas e epistemológicas da gramática de dependências, bem como sua relação com outros formalismos; e também a relação entre a gramática de dependência e outros campos disciplinares, como a psicolinguística e os estudos da cognição. Ainda, temas correlatos, tratados tanto de um ponto de vista teórico quanto de um ponto de vista aplicado, também estavam no escopo da chamada.

Como resultado, apresentamos os artigos que ora compõem este dossiê, os quais tratam do tema da gramática de dependências trazendo reflexões nos eixos teórico, aplicado e avaliativo.

Dentro do eixo teórico, Pagani, em “Duas noções de dependência”, como o próprio título destaca, discute duas noções de dependência a partir de Hjelmslev (1975) e Tesnière (1965) para o estabelecimento de critérios de análise para a descrição de fenômenos linguísticos. Para tanto, Pagani contextualiza o surgimento das gramáticas de dependências, apresenta o conceito de dependência tanto em Hjelmslev quanto em Tesnière, comparando-os de acordo com os critérios para a postulação da relação de dependência e a natureza das entidades que apresentam ou não a dependência. Destacando a dificuldade de classificação da posição dependencial da preposição, o autor aponta para a importância de o analista ter critérios explícitos para determinar quais entidades devem ter suas dependências averiguadas de forma se ter uma análise mais consistente.

Coneglian, em “O modelo de dependências universais: assentando bases teóricas e revisando diretrizes metodológicas”, em direção análoga à de Pagani, conduz sua argumentação para a necessidade de se explicitar os pressupostos teóricos sobre o quadro categorial geral das Dependências Universais (UDs) quando de sua implementação para a descrição e a análise linguísticas. Para o autor, os conjuntos de etiquetas de classe de palavras (POS) e de relações de dependência (*deprels*) podem ser analisados como conceitos comparativos, conceitos tais que podem ser transformados em categorias descritivas específicas em conformidade com a língua para a qual se pretende implementar a anotação. Ao visualizar as UD's como conceitos e categorias, o autor argumenta que a isonomia e equivalência teórico-metodológica são garantidas, aprimorando-se, assim, tarefas de descrição e análise linguísticas com o auxílio de modelos computacionais.

No eixo aplicado, por sua vez, há três artigos que contemplam a adoção das UD's para tarefas de anotação de *corpora*. Dois dos três artigos tecem considerações sobre a anotação em *corpora* especializados. *Corpora* especializados oferecem desafios importantes para o processamento de linguagem natural, principalmente pelo uso não-convencional da língua. Os desafios impostos na anotação de numerais em narrativas clínicas é tema do artigo de Perini e Guimarães “Anotação de expressões com numerais em narrativas clínicas”. Nesse estudo, os autores fazem um levantamento de dez classes mais frequentes de *tokens* de numerais em um *corpus* composto por 1.000 narrativas clínicas. Ao longo do texto, os autores apontam os desafios encontrados para se anotar esse *corpus*, oferecendo também orientações para a anotação de numerais utilizando as UD's.

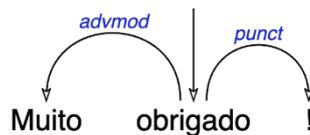
Assim como o estudo de Perini e Guimarães, Di Felippo *et al.*, em “The DANTEStocks Corpus: an analysis of the distribution of Universal Dependencies-based Part of Speech tags”, trabalham com um *corpus* especializado, o DANTEStocks, composto por *tweets* do mercado de ações anotados com informações morfosintáticas baseadas nas UD's. Além do desafio de se analisar etiquetagem em um *corpus* especializado, há também o fato de se tratar de textos categorizados como Conteúdos Gerados pelos Usuários (CGUs). Os autores fazem uma análise da distribuição das etiquetas de classe de palavra (POS) no intuito de suprir a falta de estudos sobre classes gramaticais em CGUs, contribuindo também como um ponto de partida para futuras investigações dentro deste escopo e também para a comparação com outros *corpora* da mesma natureza.

O artigo intitulado “Construções comparativas em português e sua anotação usando a sintaxe de dependências”, de Duran e colaboradores, também apresenta uma proposta de anotação baseada no modelo das UD's, mas, no caso, para construções comparativas em um *corpus* de língua geral. Essa proposta foi sistematizada a partir dos desafios encontrados pelos autores ao anotar sintaticamente construções comparativas no *corpus* Portinari-base, composto por sentenças avulsas do *corpus* Folha-Kaggle. Cumpre destacar, neste artigo, a apresentação de uma revisão de literatura crítica sobre análise sintática das construções comparativas no português brasileiro e a análise dos elementos constitutivos dessas construções com ilustrações feitas por meio de árvores de dependência extraídas do *corpus* Portinari-base, usado para a anotação. Os resultados do trabalho apresentam a descrição e os desafios de se anotar construções comparativas, além de ser um referencial para estudos que venham a abordar esse tipo de construção.

Por fim, no eixo avaliativo, De Souza e Freitas em “Avaliação da anotação automática de dependências sintáticas”, fazem uma avaliação do *treebank* PetroGold v2, anotado morfosintaticamente utilizando as diretrizes das UD's. Para tal empreendimento, os autores fizeram uma avaliação intrínseca da anotação feita no *corpus*, descrevendo, quantificando e analisando os tipos de erros encontrados para cada classe. Conforme apontam os autores, seu estudo possibilita a replicação do modelo para avaliação de *treebanks*, além de oferecer avanços significativos do ponto de vista linguístico na anotação morfosintática. Também demonstrou que as análises têm sido mais consistentes e, em termos de quantidade, relativamente, suficiente, havendo pouco espaço para melhorias linguísticas em *parsers* automáticos.

Com estes trabalhos, esperamos que outras iniciativas deem continuidade à contribuição inicial deste dossiê para expandir os estudos de gramática de dependências em língua portuguesa.

Inspirados pela publicação original de Lucien Tesnière, na qual o linguista faz a dedicatória da sua obra na forma de estema ou árvore de representação da sintaxe, concluímos esta apresentação com um estema do nosso agradecimento às contribuições recebidas.



Os organizadores

REFERÊNCIAS

HJELMSLEV, L. **Prolegômenos a uma teoria da linguagem**. Trad. J. Teixeira Coelho Netto. Perspectiva: 1975.

TESNIÈRE, L. **Eléments de syntaxe structurale**. Paris: Klincksieck, 1959.

TESNIÈRE, L. **Elements of structural syntax**. Trad. Timothy Osborne e Sylvain Kahane. Amsterdã, Filadélfia: John Benjamins Publishing Company, 2015.