

Da sintaxe ao discurso: Evidências introdutórias da correlação entre *Universal Dependencies* e *Rhetorical Structure Theory*

Jackson Wilke da Cruz Souza¹, Paula Christina Figueira Cardoso²

¹Programa de Pós-Graduação em Língua e Cultura
Universidade Federal da Bahia - Salvador/BA

²Instituto de ciências exatas e naturais - Faculdade de Computação
Universidade Federal do Pará – Belém/PA

{jacksoncruz@ufba.br; pcardoso@ufpa.br}

Abstract: *The automatic identification of discourse relations constitutes a central challenge in Natural Language Processing, particularly for Brazilian Portuguese, due to the scarcity of annotated resources and the variability of discourse signaling mechanisms. In this preliminary study, we investigate the correlation between discourse relations from Rhetorical Structure Theory (RST) and syntactic dependency relations within the Universal Dependencies (UD) framework in journalistic texts. We analyze five RST relations that are frequent in this genre, combining quantitative and qualitative analyses. The results indicate that each discourse relation exhibits distinct syntactic-discursive characteristics.*

Resumo: *A identificação automática de relações discursivas constitui um desafio central no Processamento de Línguas Naturais, especialmente para o português brasileiro, devido à escassez de recursos anotados e à variabilidade dos mecanismos de sinalização discursiva. Neste trabalho introdutório investigamos a correlação entre relações discursivas da Rhetorical Structure Theory (RST) e relações de dependência sintática do modelo Universal Dependencies (UD) em textos jornalísticos. Analisamos cinco relações RST frequentes no gênero, combinando análises quantitativas e qualitativas. Os resultados indicam que cada relação discursiva apresenta características sintático-discursivas distintas.*

1. Introdução

O nível discursivo de análise linguística impõe uma série de desafios a pesquisadores da área de Processamento de Línguas Naturais (PLN), já que muitos fenômenos exigem mais poder de processamento e/ou descrição visando a implementação computacional. Além disso, há escassez de recursos linguístico-computacionais robustos capazes de lidar adequadamente com esse nível de análise, sobretudo no contexto do português do Brasil (PB).

A *Rhetorical Structure Theory* (RST) [Mann e Thompson, 1988] tem sido bastante utilizada em tarefas de PLN, como sumarização automática, detecção de notícias falsas e avaliação automática de redações [Souza, Cardoso e Rodrigues, 2023], bem como auxiliando

propriamente na análise discursiva (Liu *et al.*, 2021). Tais pesquisas, de maneira geral, utilizam marcadores discursivos (MDs), tipicamente conjunções e preposições, para rotular possíveis relações entre as *Elementary Discourse Units* (EDUs), tidas como unidades mínimas de análise.

Em consonância a trabalhos de destaque internacional, Rodrigues, Souza e Cardoso (2023) e Cardoso *et al.* (2024) identificaram que relações retóricas em português também podem ocorrer entre EDUs sem a necessidade de marcadores discursivos tradicionalmente descritos. Esses estudos adotaram a noção de Sinalizadores Discursivos (SDs), em que diferentes categorias de pistas linguístico-estruturais podem contribuir para a identificação de relações discursivas. Neste estudo, entretanto, restringimo-nos a sinalizadores de natureza sintática, operacionalizados por meio das relações de dependência propostas pela *Universal Dependencies* (UD) [Nivre, 2015], por constituírem o foco da investigação aqui desenvolvida. Essa delimitação permite investigar, de forma mais sistemática, possíveis regularidades na interface entre sintaxe e discurso.

Na UD, as relações sintáticas são definidas a partir da dependência entre elementos que compõem a sentença, permitindo a extração de informações gramaticais no nível morfossintático. No cenário do PLN do Português, o modelo UD vem sendo aplicado em diferentes gêneros textuais, além de estar agregado a uma gama de recursos (como *corpora* e léxicos) e ferramentas (como analisadores e modelos pré-treinados de anotação automática) [Di-Felippo *et al.*, 2021].

Partindo do princípio de que há uma correlação entre Discurso e Sintaxe, nosso objetivo neste trabalho é investigar possíveis associações entre padrões sintáticos da UD e relações discursivas da RST, buscando identificar regularidades recorrentes. Para tanto, realizamos dois tipos de análises. Na primeira, de base qualitativa, investigamos como e quando relações dos dois modelos estão correlacionadas; na segunda, de base quantitativa, buscando verificar a recorrência de relações UD e como poderia resultar em padrões associáveis a relações RST. Este trabalho está organizado em 4 seções, além desta Introdução. Na Seção 2, apresentamos os trabalhos relacionados ao tema desta pesquisa. Na Seção 3 delimitamos os procedimentos metodológicos aqui empreendidos. Na Seção 4 demonstramos os resultados. Por fim, na Seção 5, tecemos nossas considerações finais.

2. Trabalhos relacionados

Conforme Cardoso, Souza e Rodrigues (2024), o nível discursivo pode ser compreendido como uma confluência de outros níveis de análise linguística, o que implica a ocorrência de fenômenos situados em zonas de fronteira entre esses níveis. No âmbito do PLN, a relação entre sintaxe e discurso tem sido amplamente investigada, destacando-se estudos que articulam esses domínios a partir de modelos teóricos como a RST e a UD.

Nesse contexto, Zeldes (2024) desenvolveu a ferramenta *rst2dep*, voltada à conversão entre representações discursivas da RST e estruturas sintáticas da UD. Um dos principais desafios enfrentados nessa proposta diz respeito ao pareamento das unidades de análise, uma vez que a RST opera sobre EDUs, enquanto a UD se baseia em tokens. Para lidar com essa assimetria, a ferramenta parte do pressuposto de que o discurso também apresenta organização estrutural, propondo estratégias para a identificação de EDUs nucleares, inclusive em construções multinucleares. Para tanto, o autor parte do princípio de que uma das soluções relevantes seja a transposição entre os modelos,

considerando que a UD constrói sua representação arbórea a partir de um nó raiz (root), do qual derivam todas as relações sintáticas.

Outra linha de investigação profícua em PLN que articula os níveis discursivo e sintático concentra-se no estudo de conectivos e MDs. A centralidade desses elementos decorre de seu papel como sinalizadores de relações discursivas, aliado à sua ampla diversidade morfosintática. Em diversos casos, tais sinalizadores exploram relações sintáticas, sobretudo intersentenciais, para evidenciar relações discursivas. À título de exemplo tem-se o que ocorre com o conector *mas*, que, ao mesmo tempo pode expressar contraposição sintática, mas também pode sinalizar contraste discursivo entre EDUs.

De acordo com Zufferey e Degand (2024), abordagens fundamentadas em modelos como a RST e representações de dependência evidenciam a presença de conectivos, embora estes não constituam condição necessária para a identificação de relações discursivas. Muitas relações podem ser inferidas implicitamente a partir de informações estruturais, semânticas e pragmáticas, o que reforça a distinção entre marcação explícita e coerência discursiva subjacente [Das e Taboada, 2014; Rodrigues, Souza e Cardoso, 2023].

Do ponto de vista computacional, essas considerações têm implicações diretas para o desenvolvimento de *parsers* discursivos e sistemas automáticos de análise do discurso. A ambiguidade funcional de conectivos e sua dependência do contexto indicam a necessidade de modelos integrados, capazes de articular informações sintáticas e discursivas de forma consistente. Assim, o estudo de conectivos constitui um eixo central para a compreensão e a modelagem da interface entre sintaxe e discurso em PLN.

A escolha do modelo sintático adotado neste trabalho justifica-se pelo fato de a UD oferecer um conjunto unificado de padrões para a representação sintática em diferentes línguas, o que possibilita comparações interlinguísticas [Freitas, 2025]. Ademais, para o português, há ferramentas treinadas e amplamente disponíveis para a anotação e análise automática de estruturas de dependência, tanto para textos formais [Lopes e Pardo, 2024] quanto para *corpora* multigênero [Di-Felippo *et al.*, 2024].

Mais recentemente, *Large Language Models* (LLMs) também passaram a ser empregados em tarefas de segmentação e *parsing* discursivo, alcançando resultados robustos na identificação automática de relações retóricas por meio de estratégias de *fine-tuning* e *prompting* [Maekawa *et al.*, 2024; Thompson *et al.*, 2024]. Apesar de essas abordagens representem avanços significativos para a área de PLN, a análise discursiva permanece pouco interpretável, o que dificulta a compreensão de quais propriedades linguístico-estruturais efetivamente contribuem para a identificação das relações discursivas. Assim, ao passo que trabalhos recentes se concentram no aprimoramento do desempenho de *parsers* discursivos, este estudo busca compreender, ainda que sucintamente, quais possíveis regularidades sintáticas podem explicar a ocorrência de determinadas relações retóricas, oferecendo evidências linguísticas que podem, futuramente, subsidiar sistemas automáticos de análise discursiva.

3. Metodologia

Para o desenvolvimento deste trabalho, separamos uma amostra de sentenças anotadas previamente com o modelo RST provenientes do *corpus* CSTNews [Cardoso *et al.*, 2011], composto por textos jornalísticos em língua portuguesa. Por conta do gênero textual, construímos a amostra em função de relações que são mais características, a saber, *Sequence*, *Elaboration*, *Purpose*, *Attribution* e *Circumstance*.

O conjunto de sentenças foi submetido à ferramenta PortParserv2 [Lopes e Pardo, 2024], que realiza anotação UD automaticamente. A escolha por essa ferramenta se deu por conta de ela ter sido pré-treinada com textos jornalísticos. Essa dupla anotação (discursiva, via RST, e sintática, via UD) permite a investigação sistemática da interface entre sintaxe e discurso.

A partir das anotações UD, foram extraídos padrões sintáticos, definidos como sequências ordenadas de rótulos de dependência, respeitando a ordem dos tokens na sentença. Esses padrões representam abstrações da estrutura sintática e permitem a comparação entre sentenças de diferentes relações discursivas sem recorrer a informações lexicais específicas.

Para cada subconjunto correspondente a uma relação RST, foram computadas as frequências absolutas e relativas dos padrões UD observados. Em seguida, os padrões mais frequentes em cada relação foram selecionados, com o objetivo de identificar regularidades sintáticas potencialmente associadas a funções discursivas específicas.

Ainda, a abstração dos padrões consistiu na eliminação de modificadores opcionais da organização linear de constituintes e distinções exclusivamente morfológicas (como *amod* e *advmod*, quando não alteravam a configuração principal). Essa decisão baseou-se na ideia de preservar apenas as relações de dependência consideradas estruturantes da configuração sintática. O objetivo foi reduzir a esparsidade dos dados sem eliminar informações relevantes para caracterizar cada relação discursiva.

4. Resultados e discussão

Na Tabela 1, apresentamos a síntese dos resultados obtidos a partir da análise realizada. Nela reunimos os macro-padrões recorrentes nos trechos correspondentes para cada relação RST, organizados em termos de porcentagem (P). Os padrões de dependência foram extraídos a partir das anotações UD completas, agrupando-se sequências estruturalmente equivalentes por meio da eliminação de modificadores opcionais, variações de ordem linear e distinções morfológicas não centrais. Essa abstração permitiu reduzir a esparsidade dos dados e evidenciar regularidades sintático-discursivas associadas às relações RST analisadas.

Os números absolutos de exemplos para cada relação são: *Purpose* (32), *Attribution* (30), *Sequence* (25), *Elaboration* (20) e *Circumstance* (30).

Destaca-se que os padrões sintáticos apresentados na Tabela 1 não correspondem necessariamente a sequências fixas de palavras, mas sim a estruturas sintáticas representadas pelas relações de dependência da UD. Nesse caso, um mesmo padrão pode ser realizado por diferentes construções frásticas, desde que preserve a organização estrutural de seus constituintes. Nos trechos exemplificados, extraídos do *corpus* CSTNews, observa-se que relações como *case* + *det* + *obl* (+*nmod*) ou *mark* + *advcl* são instanciadas por diferentes realizações linguísticas, evidenciando que o foco da representação está sobre a função sintática dos constituintes e suas dependências, e não sobre combinações léxicas.

A relação *Sequence*, na RST, ocorre quando eventos ou estados são organizados em ordem temporal, procedural ou narrativamente, sem hierarquia entre as EDUs, sendo, portanto, uma relação multinuclear. Em textos jornalísticos, esse tipo de organização discursiva manifesta-se sobretudo por meio de coordenação de predicções, parataxe, marcas temporais e aspectuais e paralelismo sintático.

Relação RST	Padrão UD abstrato	P (%)	EXEMPLO
<i>Purpose</i>	mark + advcl	61	“quando Abadia se entregou à polícia colombiana”
	case + det + obl (+nmod)	23	“na manhã desta terça-feira”
	advmod + case + det + obl	16	“Desde julho de 2005 até agora”
<i>Attribution</i>	root (+nsubj) + punct	42	“O objetivo é prender 17.”
	case + det + obl 0(+flat:name)	38	“na Lapa”
	root + mark	20	“disse que”
<i>Sequence</i>	nsubj + root + conj (+obj / obl)	44	“o diretor assinou um termo de compromisso (...) e que Bruno será transferido para outra unidade”
	case + det + obl	32	“às 8h”
	root + cc + conj	24	“começou e ficou”
<i>Elaboration</i>	acl / acl:relcl	48	“que teria enviado toneladas de cocaína àquele país”
	case + det + nmod / obl	34	“na Lapa (zona oeste de São Paulo)”
	nsubj + cop + root	18	“a quadrilha de tráfico internacional de drogas e lavagem de dinheiro liderada por Abadia é uma das mais ricas do mundo”
<i>Circumstance</i>	case + det + obl (+nmod / amod)	46	“na manhã desta terça-feira”
	case + det + nmod / obl	34	“depois do motim”
	advmod + case + det + obl	31	“até agora”
	mark + advcl	23	“quando Abadia se entregou à polícia colombiana”

Tabela 1 – Distribuição de padrões sintáticos via UD.

A análise das relações UD presentes nas EDUs anotadas como *Sequence* evidencia: (i) coordenação explícita de eventos; (ii) presença de múltiplos núcleos predicativos; (iii) cláusulas adverbiais de progressão; (iv) marcadores discursivos que sinalizam ordenação, frequentemente realizados por advérbios; (v) paralelismo estrutural nominal-verbal; e (vi) complementos circunstanciais. Observa-se incidência quase categórica de coordenação (cc, conj), frequentemente associada a múltiplos núcleos predicativos, o que sinaliza a ausência de hierarquia retórica entre as EDUs e sustenta a organização do discurso como sucessão temporal ou procedimental. A recorrência de construções adverbiais e marcadores temporais ancora os eventos em uma progressão cronológica explícita, enquanto o paralelismo estrutural e a repetição de complementos circunstanciais contribuem para a coerência sequencial.

Já a relação *Elaboration* ocorre quando um segmento discursivo satélite acrescenta detalhes, especificações ou exemplificações a uma entidade ou evento descrito no núcleo, sem promover avanço na progressão temática principal. No gênero jornalístico, essa relação é frequentemente realizada por construções sintáticas que introduzem informação acessória de natureza explicativa ou descritiva.

Sob a perspectiva das UD, observam-se: (i) predominância de modificadores nominais e oblíquos, que viabilizam expansões referenciais de localização, quantificação ou caracterização; (ii) uso recorrente de orações relativas e outros modificadores clausais, responsáveis por acrescentar informação a referentes previamente introduzidos; (iii) estruturas adverbiais explicativas ou circunstanciais; (iv) construções apositivas e nomeações, associadas a estratégias jornalísticas de identificação nominal, cargos e nomes próprios; e (v) coordenação empregada como mecanismo de expansão informacional, acumulando atributos ou descrições sem alterar a macroestrutura discursiva.

A relação *Purpose*, na RST, organiza o discurso de modo que um satélite explicita a finalidade, o objetivo ou a intenção de uma ação descrita no núcleo. Em textos jornalísticos, essa relação é comumente realizada por orações adverbiais finais, construções infinitivas, subordinação introduzida por marcadores de finalidade e verbos de ação orientados a complementos que expressam resultado.

A análise dos padrões UD associados à relação *Purpose* revela: (i) marcação explícita de finalidade; (ii) orações adverbiais finais; (iii) modificadores causais de valor final; (iv) complementação infinitiva; (v) objetos sintáticos explícitos da ação final; (vi) complementos circunstanciais associados à finalidade; e (vii) marcadores de intensidade ou delimitação. A elevada quantidade de objetos explícitos e de complementação infinitiva sugere que a finalidade é concebida, nos textos analisados, como uma ação orientada a um resultado claramente identificável. Ademais, a recorrência de modificadores causais (*acl*) indica processos de nominalização de objetivos. Destacam-se ainda combinações frequentes entre relações UD, como *mark+advcl* e *acl+obj*, que apontam para um alinhamento entre sintaxe e discurso nessa relação RST.

A relação *Attribution*, na RST, ocorre quando uma EDU expressa um conteúdo informacional cuja responsabilidade enunciativa é atribuída a uma fonte explícita. Em textos jornalísticos, essa relação é tipicamente caracterizada por verbos de dizer, marcadores de subordinação e menções diretas ou indiretas à fonte da informação.

No plano das UD, observam-se: (i) presença de núcleo predicativo explícito; (ii) marcação formal de atribuição, associada a construções de discurso reportado; (iii) identificação nominal da fonte, reforçando a ancoragem enunciativa; (iv) estruturas nominais que enquadram institucionalmente a fonte, configurando atribuição indireta; e (v) uso de pontuação delimitadora, como vírgulas ou aspas. Mais da metade dos padrões analisados apresenta núcleo predicativo explícito, frequentemente acompanhado de sujeito lexicalizado, o que evidencia a centralidade de atos de dizer, afirmar, declarar ou opinar. Também se observa recorrência de marcadores subordinativos e construções oblíquas introduzidas por preposição, típicas de expressões como “segundo o ministro” ou “de acordo com a assessoria”. Em conjunto, esses resultados apontam para forte convergência entre as relações UD, especialmente *root*, *nsubj*, *mark*, *obl*, *flat:name* e *appos*.

Por fim, a relação *Circumstance*, na RST, expressa informação contextual que enquadra o evento principal em termos de tempo, lugar, modo, condição, duração ou situação, sem promover progressão narrativa. No discurso jornalístico, essa relação é predominantemente realizada por complementos circunstanciais, advérbios e orações adverbiais não finais.

A análise das relações UD evidencia (i) predominância de complementos oblíquos circunstanciais; (ii) modificação adverbial explícita; (iii) orações adverbiais de valor circunstancial; (iv) marcadores subordinantes; e (v) baixa incidência de predicação nuclear. Os resultados indicam uso recorrente de complementos oblíquos e modificadores adverbiais, frequentemente combinados em construções preposicionais múltiplas, características da expressão de informação temporal, espacial ou situacional. A incidência reduzida de núcleos verbais explícitos reforça o caráter acessório dessa relação, que atua majoritariamente como satélite na RST. Além disso, a presença de orações adverbiais e de marcadores subordinativos de não-finalidade confirma que *Circumstance* desempenha a função de fornecer pano de fundo interpretativo ao leitor.

5. Considerações finais

Os resultados obtidos sugerem que os níveis discursivo e sintático, modelados respectivamente pela RST e pela UD, apresentam regularidades sistemáticas em sua interface. Enquanto a UD oferece uma representação estrutural da sintaxe, a RST possibilita observar a organização dessa estrutura em nível macrodiscursivo, evidenciando regularidades na interface entre sintaxe e discurso. Ainda que os achados ainda não permitam estabelecer associações estatisticamente generalizáveis, eles indicam a existência de regularidades promissoras para investigações futuras sobre a interface entre esses dois níveis de análise linguística

A análise comparativa das relações RST investigadas indica que cada uma apresenta uma assinatura sintático-discursiva própria. *Elaboration* associa-se a mecanismos de expansão descritiva; *Attribution* à ancoragem enunciativa explícita; *Sequence* ao encadeamento coordenado de eventos; *Purpose* à marcação formal de finalidade; e *Circumstance* ao fornecimento de enquadramento contextual periférico. Em conjunto, esses resultados corroboram a hipótese de convergência entre padrões de dependência UD e relações discursivas da RST.

Considerando o caráter exploratório deste estudo e o tamanho reduzido da amostra analisada, em nossos trabalhos futuros deveremos ampliar o conjunto de dados, contemplar outras relações RST e diferentes gêneros textuais, bem como incorporar análises estatísticas inferenciais para avaliar a significância das associações observadas entre relações sintáticas e discursivas. Ainda, será importante avaliar a especificidade dos padrões sintáticos identificados em relação às diferentes relações RST, de modo a avaliar a caracterização das relações discursivas, ou ainda quais padrões são compartilhados entre diferentes relações. Ademais, é relevante aprofundar a avaliação das anotações sintáticas geradas automaticamente, especialmente em casos de desalinhamento entre a atribuição da raiz sintática e o estatuto discursivo de satélite em relações RST mononucleares.

Cabe ainda destacar que, como as anotações UD foram produzidas automaticamente pelo PortParser v2, possíveis desvios de análise sintática podem ter influenciado parte dos padrões observados. Apesar de a ferramenta apresentar elevado desempenho para o português, sobretudo para o gênero discursivo analisado, em análises futuras deverá ser feitas avaliações quanto às saídas do *parser* com anotações revisadas manualmente.

Agradecimentos

Este projeto foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei N. 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Residência em TIC 13, DOU 01245.010222/2022-44. Além disso, o trabalho contou com suporte financeiro da FAPEB/CNPq da Chamada 004/2023 – Programa Primeiros Projetos.

Referências

- Ando, R. K., & Zhang, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6, 1817–1853. Disponível em: <https://www.jmlr.org/papers/volume6/ando05a/ando05a.pdf>.
- Cardoso, P. C. F., Maziero, E. G., Castro Jorge, M. L. R., Seno, E. M. R., Di Felippo, A., Rino, L. H. M., Nunes, M. G. V., & Pardo, T. A. S. (2011). CSTNews: A discourse-annotated corpus for single- and multi-document summarization of news texts in Brazilian Portuguese. In *Proceedings of the 3rd Brazilian RST Meeting* (pp. 88–105). Disponível em: <https://sites.icmc.usp.br/taspardo/rst2011-cardosoetal1.pdf>.
- Cardoso, P. C. F., Souza, J. W. da C., & Rodrigues, R. (2024). Modelos discursivos. In H. M. Caseli & M. G. V. Nunes (Eds.), *Processamento de linguagem natural: Conceitos, técnicas e aplicações em português* (3ª ed., Cap. 11). BPLN. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-discurso/cap-modelos-discursivos/cap-modelos-discursivos.html>.
- Cardoso, P., Souza, J., Rodrigues, R., Dantas, E., Santa Bárbara, L., Araújo, M., Gama, N., Almeida, T., & Cruz, G. (2024). A linguagem em foco: Anotação de sinalizadores discursivos em textos jornalísticos. In *Anais do XV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana* (pp. 247–256). Sociedade Brasileira de Computação. DOI: <https://doi.org/10.5753/stil.2024.245329>.
- Das, D., & Taboada, M. (2014). *RST signalling corpus annotation manual*. Disponível em: https://www.sfu.ca/~mtaboada/docs/publications/RST_Signalling_Corpus_Annotation_Manual.pdf.
- Di Felippo, A., et al. (2021). Descrição preliminar do Corpus DANTEStocks: Diretrizes de segmentação para anotação segundo Universal Dependencies. In *Proceedings of the XIII Brazilian Symposium in Information and Human Language Technology and Collocated Events* (p. 335). Disponível em: <https://aclanthology.org/2021.stil-1.36.pdf>.
- Di Felippo, A., et al. (2024). Genipapo: A multigenre dependency parser for Brazilian Portuguese. In *Proceedings of the Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2024)* (pp. 257–266). Sociedade Brasileira de Computação. DOI: <https://doi.org/10.5753/stil.2024.245415>.
- Freitas, I. S. (2025). Descrição da fragmentação estrutural de tweets do mercado financeiro via relação de "parataxis" do modelo Universal Dependencies [Trabalho de conclusão de curso, Universidade Federal de São Carlos]. Disponível em: <https://repositorio.ufscar.br/server/api/core/bitstreams/37ec0aca-bfea-473e-a29d-5b579989d69f/content>.

- Gusfield, D. (1997). *Algorithms on strings, trees and sequences*. Cambridge University Press.
- Liu, Z., Shi, K., & Chen, N. (2021). DMRST: A joint framework for document-level multilingual RST discourse segmentation and parsing. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse* (pp. 154–164). DOI: [10.18653/v1/2021.codi-main.15](https://doi.org/10.18653/v1/2021.codi-main.15).
- Lopes, L., & Pardo, T. (2024). Towards PortParser: A highly accurate parsing system for Brazilian Portuguese following the Universal Dependencies framework. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese (Vol. 1)* (pp. 401–410). Disponível em: <https://aclanthology.org/2024.propor-1.41/>.
- Maekawa, A., Hirao, T., Kamigaito, H., & Okumura, M. (2024). Can we obtain significant success in RST discourse parsing by using Large Language Models? In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024)* (pp. 2803–2815). Association for Computational Linguistics. Disponível em: <https://doi.org/10.18653/v1/2024.eacl-long.171>.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243–281. Disponível em: https://www.sfu.ca/rst/05bibliographies/bibs/Mann_Thompson_1988.pdf.
- Nivre, J. (2015). Towards a universal grammar for natural language processing. In A. Gelbukh (Ed.), *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 3–16). Springer. DOI: https://doi.org/10.1007/978-3-319-18111-0_1.
- Rodrigues, R., Souza, J. W. da C., & Cardoso, P. C. F. (2023). Sinalizadores retórico-discursivos: Revisitando a anotação RST no corpus CSTNews. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)* (pp. 249–257). Sociedade Brasileira de Computação. DOI: <https://doi.org/10.5753/stil.2023.234120>.
- Rodrigues, R., Souza, J., & Cardoso, P. (2023). Sinalizadores retórico-discursivos: Revisitando a anotação RST no corpus CSTNews. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana* (pp. 249–257). Sociedade Brasileira de Computação. DOI: <https://doi.org/10.5753/stil.2023.234120>.
- Souza, J.W.C., Cardoso, P. & Rodrigues, R. Systematic Review of Studies on Rhetorical Structure Theory (RST) (2023). *Revista De Estudos Da Linguagem*, 31(3), 1643–1675. DOI: <https://doi.org/10.17851/2237-2083.31.3.1643-1675>.
- Thompson, K., Chaturvedi, A., Hunter, J., & Asher, N. (2024). Llamipa: An Incremental Discourse Parser. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 6418–6430). Association for Computational Linguistics. Disponível em: <https://doi.org/10.18653/v1/2024.findings-emnlp.373>.
- Zeldes, A. (2024). *rst2dep* [Computer software]. Disponível em: <https://github.com/amir-zeldes/rst2dep>. <https://pypi.org/project/rst2dep/>
- Zufferey, S., & Degand, L. (2024). *Connectives and discourse relations*. Cambridge University Press.