

Relações de coerência do português brasileiro: um estudo bibliográfico-documental da RST e seus sinalizadores discursivos

Larissa Jesus Santa Bárbara¹, Roana Rodrigues², Jackson Wilke da Cruz Souza³

¹Universidade Federal de Sergipe (UFS) – São Cristóvão/SE

²Programa de Pós-Graduação em Letras – Universidade Federal de Sergipe (UFS) – São Cristóvão/SE

³Programa de Pós-Graduação em Língua e Cultura - Universidade Federal da Bahia (UFBA) – Salvador/BA

{larissajesus,roana}@academico.ufs.br; jackcruzsouza@gmail.com

Abstract. *This work tends to the development of a bibliographic study (Gil, 2002) in research in RST (Rhetorical Structure Theory), a theory that focuses on the organization of texts, in Brazilian Portuguese (BP) from searches on journals and databases. After the data collection process, the works that are part of this research were organized and described. It is expected to be possible to have a basis and an idea about the works in RST in BP.*

Resumo. *Este trabalho tende para o desenvolvimento de um estudo bibliográfico (Gil, 2002) nas pesquisas em RST (Rhetorical Structure Theory), teoria que tem como foco a organização, entre fragmentos de um texto, no português brasileiro (PB), a partir de buscas em sites de revistas da área e em base de dados. Depois do processo de coleta de dados, foram organizados e classificados os trabalhos que fazem parte desta pesquisa. Espera-se que seja possível ter uma base da condição em que a área se encontra, partindo dos trabalhos acessíveis aqui descritos em RST no PB.*

1. Introdução

A proposta *Rhetorical Structure Theory* – RST [Mann e Thompson 1987], é ser “uma teoria descritiva que tem por objeto o estudo da organização dos textos, caracterizando as relações que se estabelecem entre as partes do texto” [Antonio 2012, p.57]. Trata-se, portanto, de uma proposta de exame da organização de um texto, a partir da análise da coerência textual.

Das e Taboada (2017, p. 1) apontam que “as relações de coerência são frequentemente assinaladas por marcadores discursivos (MDs), que são geralmente considerados como os mais típicos (ou às vezes o único tipo de) sinais no discurso”. Segundo Taboada e Das (2018), os MDs são tidos frequentemente como conectivos, representados majoritariamente por preposições e conjunções. Os autores indicam que utilizar apenas os MDs como identificadores de uma relação RST pode causar duas situações complexas: (i) o número pequeno de relações que podem ser identificadas pelos MDs; e (ii) o uso de um MD pode resultar na confusão entre relações.

Além disso, Antonio (2017) aborda que no modelo RST algumas relações podem ser entendidas com ou sem a presença de um marcador discursivo. Diante disso,

salienta-se que os MDs não são apenas marcadores comuns de um discurso, mas sim um facilitador no momento de identificar e interpretar uma relação.

No contexto do português brasileiro (PB), Rodrigues, Souza e Cardoso (2023) abordaram a necessidade de estudo dos sinalizadores discursivos (SDs). Os SDs podem ser encontrados em diversos tipos de construções linguísticas, facilitando a identificação de outras relações não marcadas por MDs. Cabe pontuar que o que diferencia os MDs dos SDs é o fato de estes poderem características que apenas a de promover conexão, como sinais de natureza fonológica, semântica, sintática e pragmática.

Diante desse cenário, objetivou-se nesta pesquisa realizar um estudo bibliográfico-documental sobre a produção científica centrada na investigação de MDs e SDs em PB. Nossa intuito é contribuir com (i) o fomento de trabalhos centralizados na teoria RST no PB; e (ii) a ampliação da captura de novos sinalizadores que podem ser usados no PB.

2. Pesquisa bibliográfica-documental

Para a realização desta pesquisa bibliográfica-documental [Gil 2002], foram consultadas ferramentas e plataformas de buscas (como Google Acadêmico e Publish or Perish) e, para a realização da descrição aqui proposta, foram utilizados os seguintes critérios de seleção das obras: (i) a maneira em que a teoria RST estava sendo trabalhada; (ii) se a área do texto era focada na linguística/uso na linguagem; e (iii) se os trabalhos focaram no uso da RST no PB.

Nesta seção apresentam-se o trabalhos selecionados nesta pesquisa, bem como alguns dos SDs indicados pelos autores dos trabalhos.

2.1 Proposta de Pardo (2005)

A tese de Pardo (2005) buscou investigar tanto abordagens simbólicas quanto estatísticas para a análise discursiva automática, demonstrando como diferentes tipos de conhecimentos podem contribuir para essa tarefa. Com isso, o autor utilizou o *Corpus TCC*, em que apenas o autor da tese realizou a anotação, sendo essa uma anotação retórica, para evitar qualquer variação nos resultados. Além disso, utilizou a ferramenta *RST Annotation Tool* [Carlson e Marcu 2001], em que é possível realizar a segmentação do texto, escolher as relações retóricas, além de indicar os núcleos e satélites.

Um dos objetivos da tese foi a criação de um analisador discursivo automático chamado DiZer. Além disso, o trabalho centrou nos marcadores discursivos, principais mecanismos linguísticos para a detecção de relações retóricas, bem como em palavras e frases indicativas específicas de gênero e domínio textual.

Ademais, no trabalho do autor ficou marcada a criação de 32 relações da teoria, focados no PB.

2.2 Proposta de Cardoso (2014)

Em sua tese, Cardoso (2014) trabalhou com a summarização automática de multidocumento, sendo este um de seus objetivos, intensificando, assim, a importância

dos sumários em diversas áreas do conhecimento, principalmente no momento de crescente produção de informações. Além disso, buscava simplificar as informações e ressaltar as que faziam grande impacto no texto.

A partir disso, a autora utilizou duas teorias: CST e RST, sendo esta segunda o foco deste artigo. O uso da teoria RST partiu devido à importância da teoria para os núcleos segmentados, já usada em monodocumentos. Então, Cardoso (2014) propôs realizar isso com uma sumarização de multidocumento, pois facilitaria a sumarização de informações pertinentes no texto. Para pôr em questão a função das teorias com a sumarização, a autora utilizou o *corpus* CSTNews [Cardoso *et al* 2011], que foi anotado em RST, permitindo a análise da estrutura retórica e a relevância da informação.

2.3 Proposta de Antonio (2017)

Essa pesquisa foi realizada para entender o uso e/ou a necessidade das relações RST quando ou não demarcadas por conectivos. Isso fez com que o autor trabalhasse com aspectos relacionados à intencionalidade, semântica, interpretação e entonação. Outrossim, o *corpus* trabalhado foi de língua oral e, por conta dessa condição, os textos passaram por uma retextualização (ou seja, foram transcritos), para que as marcações da conversação pudessem ser preservadas (como, entonação e pausas) na linguagem escrita. Em seguida, esses textos foram entregues para professores do ensino superior, que atuaram como analistas.

Como resultado, as relações foram identificadas por meio de outros sinalizadores, além de MDs. O autor fortaleceu o pressuposto da teoria RST, em que as relações retóricas são de *sentido* e não de *forma*, já que as relações estão presentes no discurso, mesmo na ausência de conectivos. Isso se justifica pelo fato de os informantes pontuarem algumas outras formas de facilitar a identificação das relações, como, por exemplo, tempo verbal, semântica e paralelismo sintático.

2.4 Proposta de Rodrigues, Souza e Cardoso (2023)

O trabalho realizado por Rodrigues, Souza e Cardoso (2023) teve como base as pesquisas de Antonio (2017) e Das e Taboada (2018). Os autores partiram do *corpus* jornalístico CSTNews [Cardoso *et al.* 2011] pré-anotado com o modelo RST e identificaram possíveis sinalizadores das relações retóricas. A escolha de se trabalhar com esse tipo de abordagem é ter uma visão de outras partes do texto que, eventualmente, não foram analisadas.

Os autores fizeram manualmente uma anotação em nove conjuntos de textos, identificando os SDs no texto maior de cada um dos conjuntos, chegando a um rol de sinalizadores para além dos MDs. Ainda, salientam que esses sinalizadores ocorreram de maneira isolada (ou *simples*) ou conjunta (ou *combinada*).

Abaixo, no Quadro 1, apresentam-se exemplos de sinalizadores em função dos trabalhos citados nesta pesquisa. Os trabalhos destacaram ora explícita ora implicitamente os SDs utilizados, com exceção de Cardoso (2014) que não os apresentou.

Quadro 1. Sinalizadores por trabalhos/documentos encontrados.

RELAÇÃO RST	TRABALHOS			
	Pardo (2005)	Cardoso (2014)	Antonio (2017)	Rodrigues, Souza e Cardoso (2023)
<i>Elaboration</i>	Locução verbal	-----	Pontuação (:)	Pontuação (,) + Pronome Relativo (que)
<i>Sequence</i>	MDs (a partir)	----	Tempo Verbal	MD (e) + tempo verbal

A partir do Quadro 1, notou-se que em EDUs anotadas com relações RST *Elaboration* e *Sequence*, por exemplo, ocorreram diferentes SDs e MDs de maneira isolada ou combinada. No trabalho de Antonio (2017), nas unidades anotadas com a relação *Elaboration* ocorreu Pontuação, enquanto, em Rodrigues, Souza e Cardoso (2023) a mesma relação foi identificada pela combinação de dois sinalizadores. Já na relação *Sequence* no trabalho de Pardo (2005) ocorreu uma Locução prepositiva, ao passo que no trabalho de Antonio (2017) o tempo verbal foi apontado como SD dessa relação.

Além do estudo sobre os sinalizadores, Pardo (2005) aplica a RST no desenvolvimento de um dos primeiros analisadores discursivos automáticos para o PB, o DiZer. Já no trabalho de Cardoso (2014), a RST foi aplicada na sumarização automática em PB. Antonio (2017) utiliza a teoria para averiguar, em um *corpus* oral, a possibilidade de existir e, consequentemente, identificar uma relação retórica com outros SDs. Por fim, Rodrigues, Souza e Cardoso (2023) buscaram a existência de outros SDs, realizando um estudo a partir de um *corpus* pré-anotado.

3. Conclusões

Nesta pesquisa, buscamos trabalhos científicos que abordassem a RST no PB, sobretudo com relação aos MDs e SDs destacados por eles. Foram encontrados diferentes estudos, entre eles os quatro em destaque. Foi possível perceber a abrangência de perspectivas com relação a esse tema, além da diversidade de áreas e aplicações com que a RST é inserida em português.

Importante destacar que houve limitações sobre o estudo por conta da falta de unificação das bases de dados, o que, eventualmente, sobreponha os resultados de busca. Assim, cabe, em trabalhos futuros, aplicar outros termos de busca, bem como verificar possíveis sombreamentos de resultados entre as bases bibliográficas. Além disso, pretende-se futuramente realizar um aprofundamento nas análises dos sinalizadores discursivos em PB, buscando outras formas de ocorrências e a frequência em diferentes gêneros textuais. Tais resultados poderão subsidiar pesquisas de identificação automática das relações RST, bem como fenômenos discursivos (como desinformação e argumentação) em gêneros textuais e domínios específicos.

Agradecimentos

Este trabalho foi realizado no âmbito do Centro de Inteligência Artificial da Universidade de São Paulo (C4AI -<http://c4ai.inova.usp.br/>), com o apoio da

Fundação de Amparo à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM. Este projeto também 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.

Referências

- Antonio, J. D. (2012) Relações retóricas estabelecidas por orações gerundiais adverbiais. *Alfa: Revista de Linguística*, v. 56, n. 1, p. 55-79.
- Antonio, J. D. (2017) Mecanismos utilizados pelos destinatários do discurso para identificação de relações de coerência não sinalizadas por conectores. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, v. 33, n. 1, p. 79–108.
- Carlson, L.; Marcu, D. (2001) Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, v. 54, n. 2001, p. 56.
- Cardoso, P.C.F. *et al.* (2011) CSTNews -A Discourse-Annotated Corpus for Single and Multi-Document Summarization of News Texts in Brazilian Portuguese. [s.l: s.n.]. Disponível em: <<https://sites.icmc.usp.br/taspardo/rst2011-cardosoetal1.pdf>> . Acesso em: 5 out. 2023.
- Cardoso, P.C.F. (2014) Exploração de métodos de sumarização automática multidocumento com base em conhecimento semântico-discursivo. 2014. 182p. Tese (Doutorado). Universidade de São Paulo. Instituto de Ciências Matemáticas e de Computação/Universidade de São Paulo.
- Gil, A. C. (2002) *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Editora Atlas.
- Pardo, T. A. S. (2005) *Métodos para análise discursiva automática*. 2005. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.
- Rodrigues, R; Souza, J. W. C; & Cardoso, P. Sinalizadores retórico-discursivos: revisitando a anotação RST no córpus CSTNews. In: *Simpósio Brasileiro De Tecnologia da Informação e da Linguagem Humana (STIL)*, 14., 2023, Belo Horizonte/MG. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 249-257.
- Taboada, M.; Das, D. (2013) Annotation upon Annotation: Adding Signalling Information to a Corpus of Discourse Relations. *Dialogue & Discourse*, v. 3, n. 2, p. 249–281.