

Reflexões para a análise RST de textos

Juliana Thiesen Fuchs (juliana.tfs@gmail.com)
Maria Eduarda Giering (eduardag@unisinos.br)

Universidade do Vale do Rio dos Sinos

Neste trabalho, é discutido o uso da *Rhetorical Structure Theory* – RST (MANN; THOMPSON, 1988) para a análise de textos, sendo propostas algumas reflexões para o aprimoramento desse uso. As reflexões propostas neste trabalho têm origem nas pesquisas realizadas por Giering (2008a; 2008b; 2007)¹ e são desenvolvidas em Fuchs (2009). Essas reflexões se baseiam na hipótese de que o quadro teórico da RST é limitado, de forma que o analista, para ser capaz de analisar um texto a partir do modelo proposto pela RST, precisa conhecer previamente algumas concepções de como os textos funcionam.

Para a análise textual, a RST propõe um modelo de relações que se estabelecem, recursivamente, entre partes do texto consideradas pelo analista como núcleos e satélites. O primeiro passo para a análise do texto é a sua segmentação em unidades, que podem ser definidas de acordo com os critérios do analista (podem abranger tanto orações quanto parágrafos ou segmentos ainda maiores). Em seguida, o analista relaciona os segmentos; para escolher uma relação para determinados segmentos, o analista deve observar se a definição da relação plausivelmente se aplica ao trecho em questão. Ao relacionar os segmentos do texto, o analista constrói a sua estrutura hierárquica.

A RST está fundada na noção de coerência; por isso a teoria atribui um papel a cada parte do texto. Conforme Mann, Matthiessen e Thompson (1992), todo texto é produzido para atingir determinados fins, e cada parte do texto contribui para o alcance desses fins. Dessa forma, podem ser identificados, no texto, segmentos que parecem desempenhar um papel primordial para os fins do produtor; esses segmentos são considerados nucleares. Consequentemente, outros segmentos desempenham um papel secundário (satélite), contribuindo para a maneira como o leitor compreende ou aceita os segmentos nucleares. Assim, ao longo do texto, estabelecem-se diversas relações RST, cada uma delas ocorrendo entre um segmento considerado nuclear e um segmento considerado satélite (a não ser no caso de relações multinucleares, que se estabelecem entre dois ou mais núcleos). Cada relação pressupõe algum efeito causado no leitor.

Outra noção fundamental para a RST é a de hierarquia, que postula que os textos são organizados de forma que as partes são combinadas em partes maiores, que por sua vez são combinadas em partes maiores ainda, até formarem o texto inteiro. Por isso, é possível representar graficamente a estrutura de relações de um texto por meio de uma árvore, o que pode ser feito em ferramentas desenvolvidas para a representação arbórea de textos analisados com a RST, como a *RSTTool* (O'DONNELL, 2000).

As pesquisas de Giering (2008a; 2008b; 2007), dentro da área de investigação da Linguística Textual, utilizam a RST para a análise de segmentos macroestruturais de textos, procurando investigar se determinados gêneros textuais (no caso, artigos de opinião autorais e textos midiáticos de divulgação científica) apresentam configurações recorrentes de relações, o que aponta para uma organização macroestrutural prototípica dos textos desses gêneros. Nessas

¹ Trata-se dos projetos de pesquisa DCEROT (Divulgação Científica: Estratégias Retóricas e Organização Textual), ORTDC (Organização Textual de Textos de Divulgação Científica) e ORTO (Organização Retórica de Textos de Opinião), coordenados pela Profa. Dra. Maria Eduarda Giering e dos quais fez parte a primeira autora deste trabalho, Juliana Thiesen Fuchs.

pesquisas, portanto, a análise RST dos textos é guiada por uma série de concepções sobre como o texto funciona.

Uma dessas concepções é a de que o gênero a que o texto pertence determina a maneira como ele vai se organizar. Essa concepção não está presente nos postulados da RST; a teoria, apesar de enfatizar o ponto de vista comunicacional da linguagem, partindo do princípio de que os textos são produzidos para alcançar determinados fins comunicativos em relação a um leitor, não aborda as regularidades situacionais e culturais que fazem com que os textos funcionem a partir de determinados gêneros.

Outra concepção que guia as análises RST dos textos nas pesquisas de Giering (2008a; 2008b; 2007) é a de macroestrutura, de van Dijk (1998). Para analisar os segmentos macroestruturais dos textos, as pesquisas de Giering (2008a; 2008b; 2007) se baseiam na ideia de van Dijk (1998) de que os textos estão formados primariamente por macroproposições. Ou seja, durante o processo de compreensão de um texto, o leitor percebe que o texto se compõe de grandes movimentos (tanto que, após a leitura, ele é capaz de resumir ou recontar o texto em poucas frases). A macroproposição, dessa forma, constitui um conjunto de proposições que têm uma unidade semântica e que podem, a partir de um processo cognitivo de redução da informação semântica, ser reduzidas a uma única proposição. Cada macroproposição é relativa às proposições do nível imediatamente inferior; assim, o texto está formado por vários níveis macroproposicionais, sendo o nível mais alto composto de uma única macroproposição que dá conta do tema do texto inteiro.

Para van Dijk (1998), o processo cognitivo de formação das macroproposições depende das funções pragmáticas do texto, ou seja, dos macroatos de fala: da mesma forma como os atos de fala regulam as condições de adequação das sequências de proposições (uma proposição se relaciona com outras em função do cumprimento de uma finalidade do produtor textual), há atos de fala mais globais, os macroatos de fala, que orientam as sequências de macroproposições. Assim, pode ser atribuído um macroato de fala às macroproposições gerais do texto, ou seja, é possível dizer que o texto inteiro foi produzido a partir de uma única finalidade. É por isso que, para van Dijk (1998), a coerência local depende da coerência global do texto, e só é possível compreender as proposições do texto a partir da sua funcionalidade em relação às macroproposições. Os macroatos de fala e as macroproposições do texto dependem, por sua vez, do gênero de texto em questão. Afinal, uma informação considerada macroproposicional em um texto de divulgação científica, que tem a finalidade de divulgar uma pesquisa, pode ser considerada microproposicional em um texto de opinião, por exemplo.

Essa visão macroestrutural do texto proposta por van Dijk (1998) se assemelha à visão hierárquica do texto proposta pela RST: é possível, a partir de uma árvore de relações RST construída na *RSTTool*, visualizar as macroproposições e as proposições do texto. Porém, não basta a noção de hierarquia da RST para se construir uma representação do texto conforme a sua organização macroestrutural. Ou seja, os elementos teóricos presentes na RST não permitem que o analista necessariamente represente as macroproposições e as proposições do texto na árvore construída na *RSTTool*. Por meio do modelo da RST, o analista pode, por exemplo, eleger como unidades de análise as orações e iniciar a análise do texto relacionando uma oração com outra, até formar segmentos maiores e, dessa forma, construir uma estrutura arbórea. Porém, essa estrutura arbórea, tendo sido construída a partir da microestrutura, dificilmente coincidirá com a organização macroestrutural do texto (que pressupõe que se observe o texto primeiramente a partir de suas macroproposições, em função de um macroato de fala regulado por um gênero).

Portanto, se para a RST o analista tem um papel fundamental (escolhendo quais serão suas unidades de análise e atribuindo a elas relações conforme sua interpretação das unidades e da definição das relações), para que ele possa representar a organização macroestrutural do texto

ele precisa primeiramente ser capaz de observar como o texto está organizado macroestruturalmente. E, como a RST não apresenta elementos teóricos que deem conta de como se dá a organização macroestrutural do texto, o analista precisa conhecer outras concepções de como o texto funciona para ser capaz de representar essa organização macroestrutural.

Parte-se, assim, da hipótese de que a RST não dá conta da representação da organização textual se não for acrescida de outras concepções de como o texto funciona. É por isso que as pesquisas de Giering (2008a; 2008b; 2007), ao pretenderem verificar se há uma organização prototípica macroestrutural dos textos analisados, utilizam o modelo da RST associado a uma série concepções de como o texto funciona (gênero textual e macroestrutura).

Essa hipótese foi formulada durante a observação de como a RST é utilizada em pesquisas com fins computacionais, como a de Rino et al. (2007)². Essa pesquisa usa o modelo da RST como um dos passos para a sumarização automática de textos: a partir da árvore RST do texto produzida na *RSTTool*, a sumarização, de caráter extrativo, consiste na supressão dos satélites, para que se mantenham, no sumário gerado, apenas as informações nucleares do texto. O *corpus* da pesquisa – intitulado *corpus* Summit – está formado por 50 textos retirados do caderno Ciência do jornal Folha de São Paulo. Observou-se que, na pesquisa de Rino et al. (2007), o método de análise dos textos (representação da estrutura hierárquica dos textos do *corpus* na *RSTTool* para posteriormente serem sumarizados) envolve apenas os pressupostos da própria RST (coerência, hierarquia, nuclearidade). Dessa forma, percebeu-se que as árvores RST construídas para os textos podem não representar a sua organização macroestrutural.

O presente trabalho procura refletir sobre como o uso da RST para a análise de textos pode ser aprimorado a partir da contribuição de outras concepções de organização textual, fora do escopo da RST. Para tanto, primeiramente, são apresentadas algumas concepções de organização textual.

As concepções de organização textual em foco neste trabalho são a de que a organização do texto depende do gênero de texto em questão e de que o texto se organiza partindo do nível macroestrutural até o nível microestrutural. Essas duas concepções têm em comum o fato de mostrarem que o texto se organiza de forma intimamente ligada ao contexto: o gênero textual em questão envolve o fato de que o texto é produzido em uma determinada situação de comunicação, a partir de uma determinada finalidade; esses elementos ligados ao gênero determinam como o texto vai se organizar, ou seja, como será construída a sua macroestrutura. Dessa forma, o quadro teórico apresentado aqui se subdivide em dois aspectos: a relação entre texto e contexto e a influência dessa relação no processo de produção do texto.

Para abordar a relação entre texto e contexto, são apresentados alguns aspectos das teorias de Bernárdez (1995) e de Halliday e Hasan (1989). A noção de texto como eleição de estratégias para a concretização de um objetivo dentro de um contexto prototípico faz com que a teoria de Bernárdez (1995) seja, essencialmente, funcional. Nesse aspecto, entre outros, essa teoria se aproxima da teoria de Halliday e Hasan (1989), que conceitua o texto a partir da sua íntima relação com o contexto. As duas teorias explicam o funcionamento do texto (linguagem em uso) como uma eleição contínua de estratégias em função de um objetivo dentro de um contexto.

Para abordar a forma como se dá a eleição de estratégias para a formação do texto, são apresentadas, neste trabalho, primeiramente, as considerações de Dik (1997) e Hengeveld

² Trata-se do projeto ProCaCoSA (Processamento de Cadeias de Correferência para a Sumarização Automática de Textos em Português), coordenado pela Profa. Dra. Lucia Helena Machado Rino. A primeira autora deste artigo, Juliana Thiesen Fuchs, participou do projeto na construção das estruturas hierárquicas na *RSTTool* dos textos do *corpus*.

(2004) sobre a Gramática Funcional Discursiva (GFD), modelo que mostra, a partir de conceitos adaptados do modelo da Gramática Funcional (que explica basicamente o funcionamento da oração), como o texto se organiza macroestruturalmente e como cada unidade menor – a oração, por exemplo – só faz sentido em função das unidades maiores. Esse modelo legitima a ideia de que, a partir da influência do contexto, o texto é o resultado de escolhas que operam de forma *top-down*. A abordagem da GFD é relevante para este trabalho porque explica como se dá o processo de escolhas do produtor na formação do texto. Ela mostra a possibilidade (e a urgência) de acomodar o discurso no interior das abordagens funcionais da linguística e o desafio de representar o discurso formalmente.

Outra teoria apresentada aqui para elucidar como se dá a eleição de estratégias para a formação do texto é a de van Dijk (2004). Sua teoria faz referência à GFD ao considerar que uma análise funcional do discurso é composta de dois aspectos: um mais linguístico, que se ocupa das relações funcionais entre estruturas gramaticais e entre estruturas ou estratégias textuais, e um mais interdisciplinar, que se ocupa das relações funcionais entre essas estruturas e as estruturas do contexto (cognição, interação etc.). O aspecto linguístico elucidava a forma como o texto funciona, porque apresenta categorias próprias ao nível textual. Algumas dessas categorias são as macroestruturas (estruturas globais do significado), as superestruturas (estruturas globais da forma) e os macroatos de fala (estruturas globais pragmáticas). Essas estruturas implicam unidades de análise dos textos: as proposições, as macroproposições e os atos de fala. Nesse nível linguístico da análise funcional do discurso proposta por van Dijk (2004), podem ser acomodadas as relações RST (que, conforme Dik, 1997, também podem ser contempladas no modelo da GFD). Afinal, essas relações seriam uma formalização da coerência funcional das proposições em um texto.

Todas as teorias apresentadas neste trabalho convergem para a ideia de que os aspectos contextuais (finalidade do produtor, configuração prototípica do texto em função de um contexto prototípico) são cruciais para a compreensão de como o texto se organiza: por meio de um processo contínuo *top-down* de eleição de estratégias que afetam primeiramente os níveis macroestruturais. Conforme argumentado neste trabalho, essa compreensão de como se dá a organização textual é fundamental para o analista que deseja representar o texto por meio de modelos como o da RST. Os princípios de organização abordados aqui (relação entre texto e contexto e processamento estratégico *top-down* do texto) influem nas escolhas do analista no momento de construir estruturas hierárquicas RST para os textos. A RST, como teoria, não contempla esses princípios, sem os quais o modelo pode gerar uma representação limitada da organização textual.

O passo seguinte deste trabalho é apresentar uma amostra da análise qualitativa realizada com o *corpus* Summit³ (a análise é exemplificada neste trabalho a partir da representação arbórea na *RSTTool* de um texto do *corpus*). Durante essa análise, foram observados padrões, nas estruturas arbóreas RST dos textos do *corpus* Summit, que conduzissem a generalizações sobre o *corpus*. Primeiramente, foram observados padrões macroestruturais nos textos analisados, procurando-se verificar a incidência de determinadas relações RST nesse nível. Em seguida, foi observado se as relações RST que ocorrem no nível macroestrutural dos textos se organizam internamente por meio de sequências recorrentes de determinadas relações. Então, esses padrões de relações foram relacionados com elementos contextuais, que restringem as possibilidades de organização de relações. Assim, pôde-se evidenciar a influência de concepções de organização textual fora do escopo da RST na representação dos textos, demonstrando-se que a estrutura relacional retórica é apenas um aspecto da organização textual e que é possível – e imprescindível – relacionar o modelo da RST com outros princípios de organização textual.

³ Essa análise foi realizada em Fuchs (2009).

Por fim, neste trabalho, são apresentadas algumas sugestões de como o analista pode lançar mão de concepções de organização textual fora do escopo da RST ao analisar um texto com o modelo da RST.

REFERÊNCIAS

BERNÁRDEZ, Enrique. *Teoría y epistemología del texto*. Madrid: Cátedra, 1995.

DIK, Simon. *The theory of functional grammar: part II: complex and derived constructions*. Berlin: Mouton de Gruyter, 1997.

FUCHS, Juliana T. *Rhetorical Structure Theory: limites e possibilidades de representação da organização textual*. 2009. 164 f. Dissertação (Mestrado em Lingüística Aplicada) – Programa de Pós-Graduação em Lingüística Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, 2009.

GIERING, M. E. Contexto midiático, fim discursivo e organização retórica de artigo de divulgação científica. In: MAGALHÃES, José Sueli de; TRAVAGLIA, Luiz Carlos (Orgs.). *Múltiplas perspectivas em Lingüística*. 1 ed. Uberlândia: Edufu, 2008a. v. 1, p. 2057-2067.

_____. Gênero de discurso artigo de divulgação científica para crianças: estratégias retóricas e estrutura composicional. *Investigações*, Recife, v. 21, p. 241-260, 2008b.

_____. Organização retórica do artigo de opinião autoral: configuração prototípica. *Círculo de Lingüística Aplicada a la Comunicación - CLAC*, v. 29, p. 1-19, 2007.

HALLIDAY, M.A.K.; HASAN, R. *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford University Press, 1989.

HENGEVELD, K. The architecture of a functional discourse grammar. In: MACKENZIE, J. L.; GÓMEZ-GONZÁLEZ, M. Á. (Ed.). *A new architecture for functional grammar*. Berlin: Mouton de Gruyter, 2004, p. 1-21. (Functional Grammar Series, 24).

MANN, W.C.; THOMPSON, S.A. Rhetorical Structure Theory: toward a functional theory of text organization. *Text*, v. 8, n. 3, p. 243-281, 1988.

MANN, W.C.; MATTHIESSEN, C.M.I.M.; THOMPSON, S.A. Rhetorical Structure Theory and Text Analysis. In: MANN, W.C.; THOMPSON, S.A. *Discourse description: diverse linguistic analyses of a fund-raising text*. Amsterdam: John Benjamins, 1992. p. 39-78.

O'DONNELL, Michael. RSTTool 2.4: A Markup Tool for Rhetorical Structure Theory. *Proceedings of the International Natural Language Generation Conference (INLG'2000)*, 13-16 Jun. 2000, Mitzpe Ramon, Israel. 253-256.

RINO, L.H.M.; ABREU, S.C.; CARBONEL, T.I.; FUCHS, J.T.; COELHO, J.C.B.; VIEIRA, R. Summ-it: um corpus anotado com informações discursivas visando à sumarização automática. In: QUENTAL, Violeta (Ed.). *Proceedings of the V Workshop on Information and Human Language Technology (TIL 2007)*. Rio de Janeiro, RJ: Instituto Militar de Engenharia, 2007. v. 1. p. 1605-1614.

VAN DIJK, T.A. *Cognição, discurso e interação*. 6. ed. São Paulo: Contexto, 2004.

_____. *Estructuras y funciones del discurso*. 12. ed. Madrid: Siglo Veintiuno Editores. 1998.