

OS ESQUEMAS *RST* E A REPRESENTAÇÃO COMPUTACIONAL DA ORGANIZAÇÃO RETÓRICA MACROESTRUTURAL DE TEXTOS

Elizabet Beatriz FOLLMANN Oliveira
(Universidade do Vale do Rio dos Sinos)
elizabetfollmann@gmail.com

RESUMO: Neste trabalho, são apresentados os esquemas *RST*, conforme Mann e Thompson (1988), observados na análise macroestrutural de dois artigos de opinião autoral e dois artigos de divulgação científica, e é discutida a representação computacional dessas análises macroestruturais, utilizando-se como interface gráfica a ferramenta *RSTTool* (O'Donnell, 1997). São apresentadas as facilidades proporcionadas pelo uso da ferramenta e suas limitações, visto que as relações que se estabelecem entre núcleos e satélites no *corpus* analisado não se limitam às restrições indicadas para a aplicação de esquemas *RST*.

PALAVRAS-CHAVE: Análise retórica macroestrutural; Esquemas *RST*; Ferramenta *RSTTool*; Artigo de opinião autoral; Artigo de divulgação científica.

ABSTRACT: *In this paper the RST schemas are presented, according to Mann and Thompson (1988). Those schemas are observed in macro-structural analyses of two authorial opinion articles and two science journalism texts, and the computing representation of those analyses are discussed, utilizing RSTTool (O'Donnell, 1997) as a graphic interface. The facilitation provided by the use of the tool and its limitations are presented, since the relations which are established between the nuclei and the satellites in the analysed corpus are not limited to the restrictions indicated in the application of the RST schemas.*

KEYWORDS: *Macro-structural rhetoric analyses; RST schemas; RSTTool; Authorial opinion article; Science journalism text.*

0. Introdução

Ao considerar o texto uma unidade comunicativa formada por outras unidades menores, nas quais determinadas estratégias são utilizadas para atingir os objetivos parciais correspondentes, e levando

em conta as circunstâncias contextuais e o objetivo da comunicação, Bernárdez (1995) propõe a adoção do modelo *RST* (*Rhetorical Structure Theory*), de Mann e Thompson (op.cit.), para o estudo do texto.

Bernárdez vincula a *RST* à idéia de que a organização textual pode ser entendida com uma série de vias ou opções de continuidade etiquetadas com relações da *RST*, pois o produtor opta por determinada organização, estrategicamente escolhida, visando ao cumprimento do fim comunicativo de seu texto. A noção de estratégia, considerando o processamento de textos, é definida pelo autor (1995: 163) como “uma sucessão de ações que conduzem a um objetivo específico, em dependência das condições contextuais”.

A *Rhetorical Structure Theory*, oriunda de estudos sobre a geração automática de textos, concentra-se na descrição da organização textual, atribuindo um papel e uma intenção a cada unidade de informação do texto, a partir da concepção de que todas as partes contribuem para formar um todo coerente.

As intenções ou os efeitos pretendidos com determinadas escolhas lingüísticas do produtor textual são expressos por meio de relações retóricas que se estabelecem entre unidades informacionais nucleares ou satélites de um texto. As unidades nucleares são consideradas as partes mais importantes do texto, essenciais para o cumprimento dos propósitos do produtor. As unidades satélites, por sua vez, oferecem informações adicionais sobre o núcleo, sendo consideradas informações secundárias, de acordo com Mann e Taboada (2006).

Os esquemas *RST* são padrões abstratos que especificam as condições para aplicação das relações retóricas entre os segmentos textuais, indicando os tipos de relações que ocorrem entre segmentos nucleares e satélites e as relações estabelecidas entre os núcleos e o texto como um todo, determinando as possíveis estruturas *RST* de textos.

Uma análise estrutural de um texto, como a proposta, é, então, um conjunto de aplicações de esquemas para o qual há as seguintes restrições:

- (a) a aplicação de um esquema que contém um conjunto de segmentos textuais constitui o texto todo;
- (b) exceto para o texto inteiro como um segmento textual, cada segmento textual na análise é ou uma unidade mínima, ou constituinte de outro esquema da análise;
- (c) cada esquema consiste de um conjunto diferente de segmentos textuais e, dentro de um esquema multi-relação, cada relação aplica-se a um conjunto diferente de segmentos;

- (d) os segmentos textuais de cada esquema constituem-se de segmentos contíguos de texto.

Conforme Mann e Thompson (1988), aplicadas as restrições, a análise *RST* resultante é representada por uma estrutura arbórea. “Estruturas *RST* são tipicamente representadas como árvores, com uma relação de nível superior abrangendo outras relações em níveis menores”, reforçam Mann e Taboada (2006: 6)¹.

Entretanto, Mann e Thompson (1988) observam que, dependendo do tipo de texto, é necessário que os critérios sejam abrandados para que sejam obtidas análises melhores.

1. Corpus

Os textos do *corpus* deste trabalho foram selecionados dentre os artigos que compõem os *corpora* dos projetos Organização Retórica de Textos de Opinião (O.R.T.O.) e Organização Retórica de Textos de Divulgação Científica (O.R.T.D.C.), desenvolvidos no âmbito do Programa de Pós-Graduação em Linguística Aplicada da Universidade do Vale do Rio dos Sinos (UNISINOS). Dois textos de cada projeto foram escolhidos tendo em vista a apresentação das dificuldades enfrentadas na representação computacional da organização retórica macroestrutural.

Trata-se de uma análise qualitativa, a partir da apresentação dos diagramas criados na *RSTTool*, tendo como base as planilhas de análise dos projetos O.R.T.O. e O.R.T.D.C. para cada um dos textos.

Os artigos estudados são os seguintes:

- a) *Bactérias transformam isopor em plástico biodegradável*, publicado na revista *Scientific American Brasil*, e *Onde está a raiz do problema?*, publicado na revista *Ciência Hoje*, retirados do projeto O.R.T.D.C.;
- b) *Alhos e bugalhos* e *Os fins não justificam os meios*, veiculados no jornal *Folha de São Paulo* - retirados do projeto O.R.T.O.

¹ Todas as citações presentes neste artigo foram traduzidas da língua em que foram originalmente escritas (inglês e espanhol) pela autora.

2. A RSTTool

A procura por ferramentas que oferecessem a possibilidade de representar computacionalmente essas análises macroestruturais levou à ferramenta *RSTTool* (O'Donnell, 1997), uma interface gráfica que auxilia na construção da estrutura retórica de textos, já utilizada por pesquisadores que realizam análises microestruturais de textos (Pardo, 2005)².

Para representar essas análises, ao conjunto das relações utilizadas nos projetos O.R.T.O. e O.R.T.D.C., acrescentou-se a relação *Same-Unit*, proposta por Carlson e Marcu (2001), conforme Tabela 1. Os autores a definem como “uma pseudo-relação usada como um recurso para ligar dois fragmentos descontínuos do texto que formam, na verdade, uma única unidade elementar do discurso, mas que está separada por uma unidade encaixada” (p. 66).

As unidades elementares do discurso (*Elementary Discourse Units* ou EDUs, em inglês) a que Carlson e Marcu (op.cit.) referem-se, são as unidades mínimas consideradas por eles na análise de textos. As unidades encaixadas, *embedded units*, são, por exemplo, orações adjetivas explicativas, que seccionam a oração principal. Essas unidades encaixadas nem sempre são orações, mas comentários entre parênteses, apostos, etc.

A definição da unidade de análise é arbitrária na RST, podendo ser definida de acordo com o propósito do estudo realizado. Para os fins dos projetos O.R.T.O. e O.R.T.D.C. e deste trabalho, a unidade mínima de análise adotada foi, segundo Giering (2005^a: 140), “uma ou mais seqüências consecutivas (compostas de uma frase, de um parágrafo ou de um conjunto de parágrafos) que expressam uma macroproposição”.

² Pardo (op.cit.) usa uma versão expandida da *RSTTool*, disponível em: <<http://www.isi.edu/~marcu/discourse/AnnotationSoftware.html>>. Acesso em: 05 set. 2006.

TABELA 1 - Conjunto de relações utilizadas.

Relações Mononucleares		
Antítese	Comentário	Preparação
Alternativa	Elaboração	Propósito
Avaliação-N	Evidência	Reformulação
Avaliação-S	Fundo	Resultado
Capacitação	Interpretação	Resumo-N
Causalidade	Justificativa	Resumo-S
Circunstância	Método	Solução
Concessão	Motivação	
Condição	Não-Condiciona	
Condição Inversa		
Relações Multinucleares		
Contraste	Reformulação	Seqüência
Lista	Same-unit	União
Esquemas		
Título	Intertítulo	Texto

3. Metodologia de análise do *corpus*

Conforme Carlson e Marcu (2001), podem ser assumidas diferentes estratégias de anotação retórica. Os analistas podem, conforme segmentam o texto, ir relacionando as proposições à estrutura retórica - análise chamada incremental, ou podem segmentar todas as unidades informacionais e, em seguida, estabelecer as relações retóricas que enlaçam essas unidades - análise chamada modular.

Para os fins deste trabalho, adotou-se a mesma metodologia de análise proposta pelos projetos O.R.T.O. e O.R.T.D.C. – uma análise modular. Primeiramente, os textos foram salvos em formato *plain text* para, então, serem importados para a ferramenta *RSTTool*. A seguir, foram automaticamente segmentados em frases. A segmentação automática foi conferida e, depois, foi especificado o conjunto de relações retóricas a ser utilizado.

Para estruturar o texto, adotou-se o seguinte procedimento metodológico: união das frases, por meio da relação *Same-Unit*, até o estabelecimento das unidades macroproposicionais indicadas pelos analistas dos projetos O.R.T.O. e O.R.T.D.C.; estabelecimento das relações mononucleares e/ou multinucleares entre as macroproposições; aplicação dos esquemas superestruturais e estabelecimento das relações retóricas que enlaçam esses esquemas.

4. Análise do *corpus*

Para o cumprimento do fim discursivo do texto *Bactérias transformam isopor em plástico biodegradável*, "divulgar pesquisa que revela que as bactérias da classe *Pseudomonas putida* transformam isopor em plástico biodegradável", a organização retórica macroestrutural compõe-se de quatro movimentos, representados pelas seguintes relações *RST*:

- (a) Resumo-N - que se estabelece entre o título do texto, unidade nuclear [1], e o corpo do texto, unidade satélite [2 a 17];
- (b) Preparação - que enlaça o segmento nuclear [4] e o segmento satélite [2 a 3];
- (c) Elaboração - que abarca o segmento satélite [5 a 11] e o segmento [4], que se estabelece como núcleo dessa relação
- (d) Comentário - que une o segmento nuclear [2 a 11] e o segmento satélite [12 a 17], conforme Figura 1.

As relações propostas pela *RST*, utilizadas pelos analistas dos projetos O.R.T.O. e O.R.T.D.C., explicitam processos estratégicos, processos de planejamento do discurso que resultam em uma organização textual coerente. Assim, pode-se explicar o título *Bactérias transformam isopor em plástico biodegradável*, por exemplo, aplicando-se a ele os critérios da relação de Resumo-N.

Carlson e Marcu (2001) sugerem que essa relação possa tanto apresentar o resumo no núcleo quanto no satélite, dividindo a relação em Resumo-N e Resumo-S, conforme a seguinte definição:

Em uma relação RESUMO-S, o satélite resume a informação apresentada no núcleo. A ênfase é na situação apresentada no núcleo. O tamanho do resumo (o satélite) é menor do que o tamanho do núcleo. Em uma relação RESUMO-N, o núcleo resume a informação apresentada no satélite. A ênfase é no resumo. O tamanho do resumo (o núcleo) é menor do que o tamanho do satélite (Carlson e Marcu, op.cit: 68).

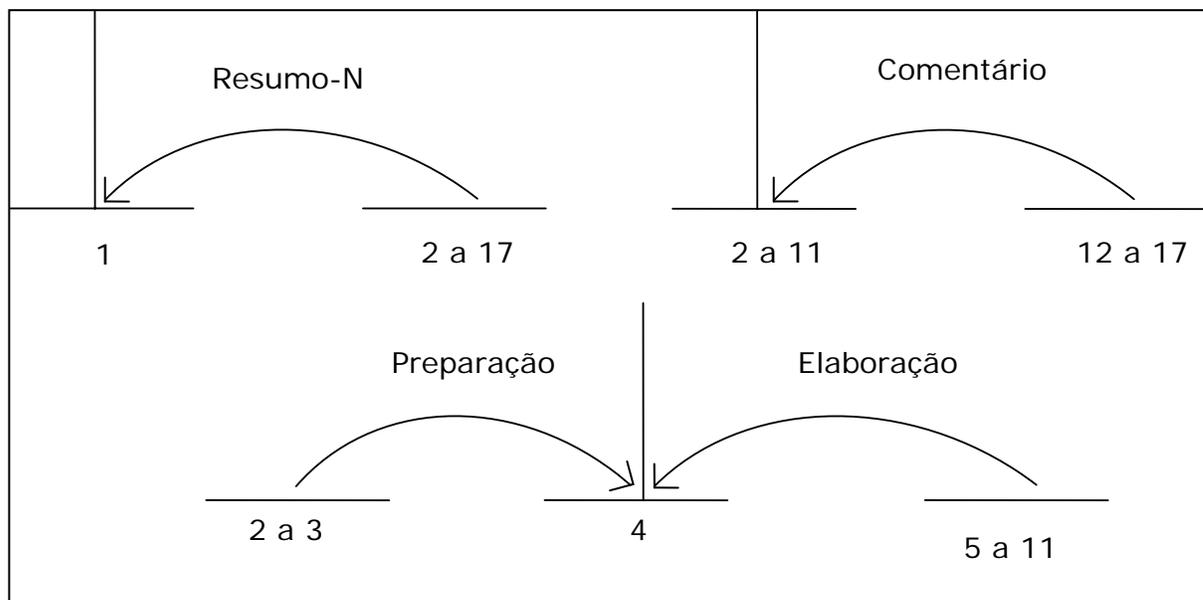


FIGURA 1 - Esquemas *RST* observados no texto *Bactérias transformam isopor em plástico biodegradável*.

No texto analisado, observa-se que o produtor textual orienta-se a celebrar as informações divulgadas pelos cientistas, utilizando como ponto de partida de seu discurso a novidade da pesquisa. Por isso, já no título, pareceu plausível aos analistas que o produtor apresentasse um breve resumo (*Bactérias transformam isopor em plástico biodegradável*), fazendo com que o leitor o reconhecesse como uma reformulação reduzida de todo o texto. Essa reformulação apresenta quais são as classes de bactérias que fazem essa transformação, como elas fazem isso e o benefício que a transformação representa para o meio ambiente.

A organização retórica macroestrutural do texto supracitado foi representada graficamente, utilizando-se a ferramenta *RSTTool*, conforme Figura 2. Os resultados, entretanto, apontam para a impossibilidade de representar graficamente todos os esquemas *RST* observados nos textos que formam o *corpus* deste trabalho.

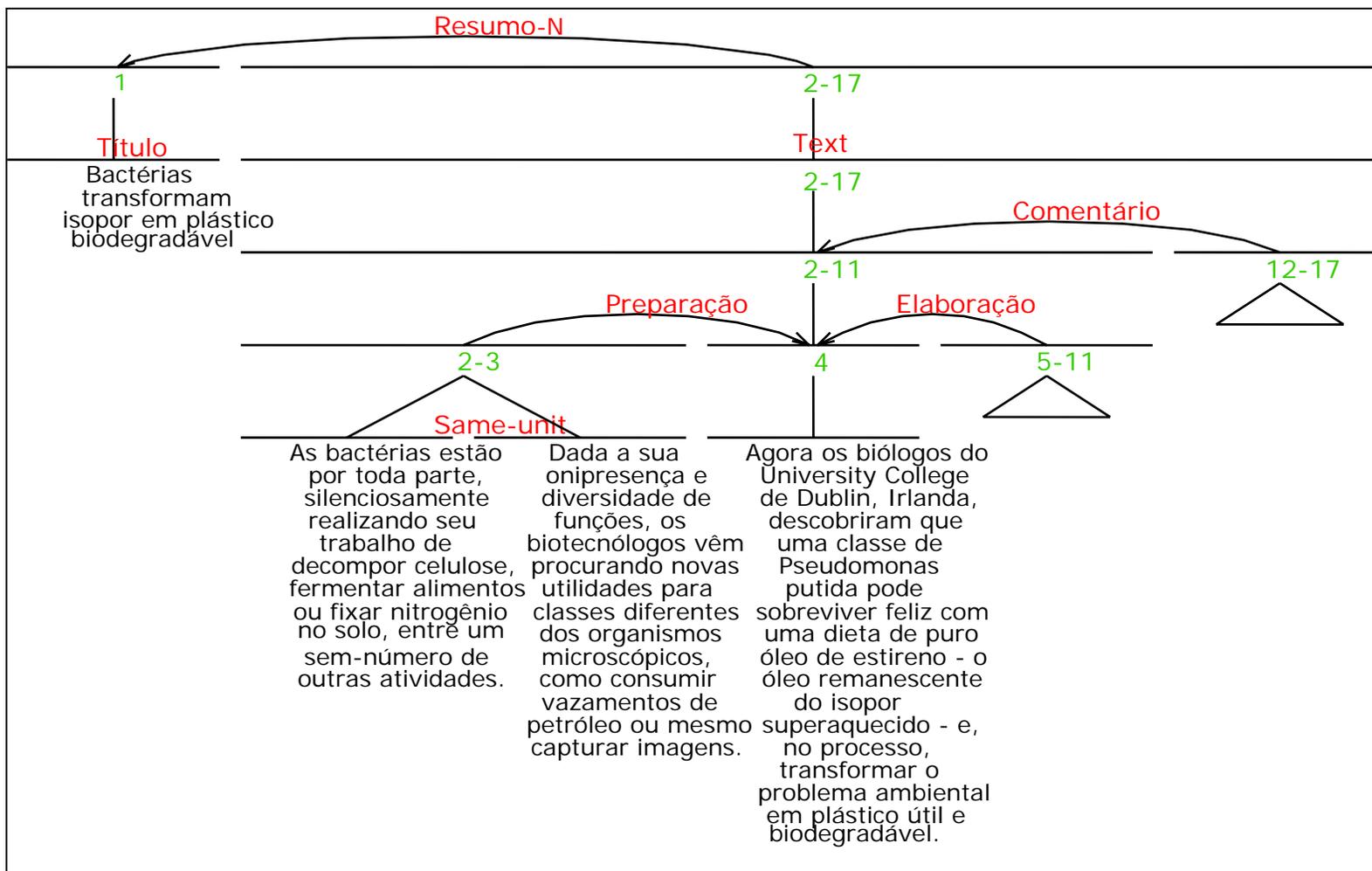


FIGURA 2 – Representação gráfica da organização retórica macroestrutural do texto *Bactérias transformam isopor em plástico biodegradável*.

Isso se deve à explicitação feita na análise macroestrutural de relações retóricas que se estabelecem entre segmentos não-contíguos de textos como, por exemplo, nas relações de Reformulação Multinuclear, segmentos [8] e [14 a 19], e Fundo, segmentos [2 a 5] e [8], observadas no texto *Os fins não justificam os meios*, conforme Figura 3.

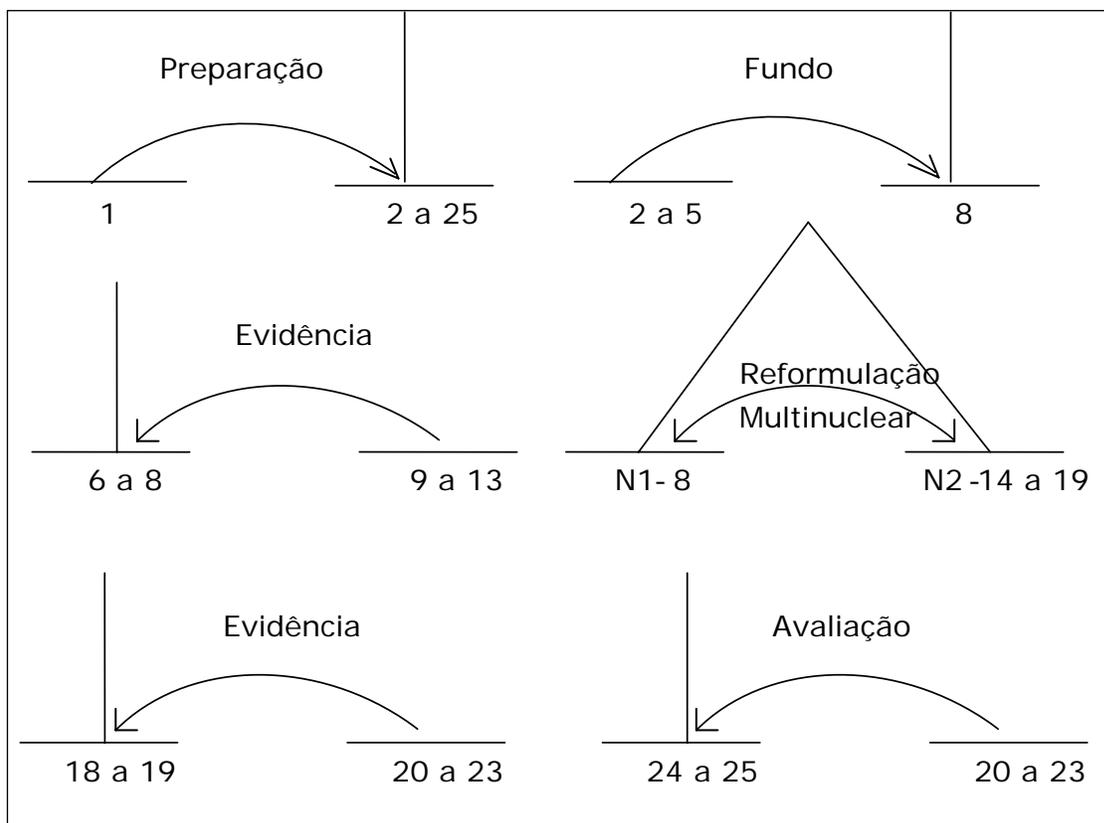


FIGURA 3 - Esquemas RST observados no texto *Os fins não justificam os meios*.

Também não é possível representar relações retóricas que ocorrem entre segmentos contíguos, mas que, por serem satélites de outros esquemas RST, não podem fazer parte de um novo esquema. No texto *Onde está a raiz do problema?*, o segmento [8 a 22] não pode ser representado como núcleo da relação de Interpretação, por já ser satélite da relação de Elaboração, conforme Figura 4.

As unidades textuais que formam o corpo do texto foram unidas por meio do esquema superestrutural Texto e, então, as relações de Preparação e Resumo-N foram aplicadas, de acordo com a Figura 5.

Porém, as relações de Interpretação e Comentário não foram representadas.

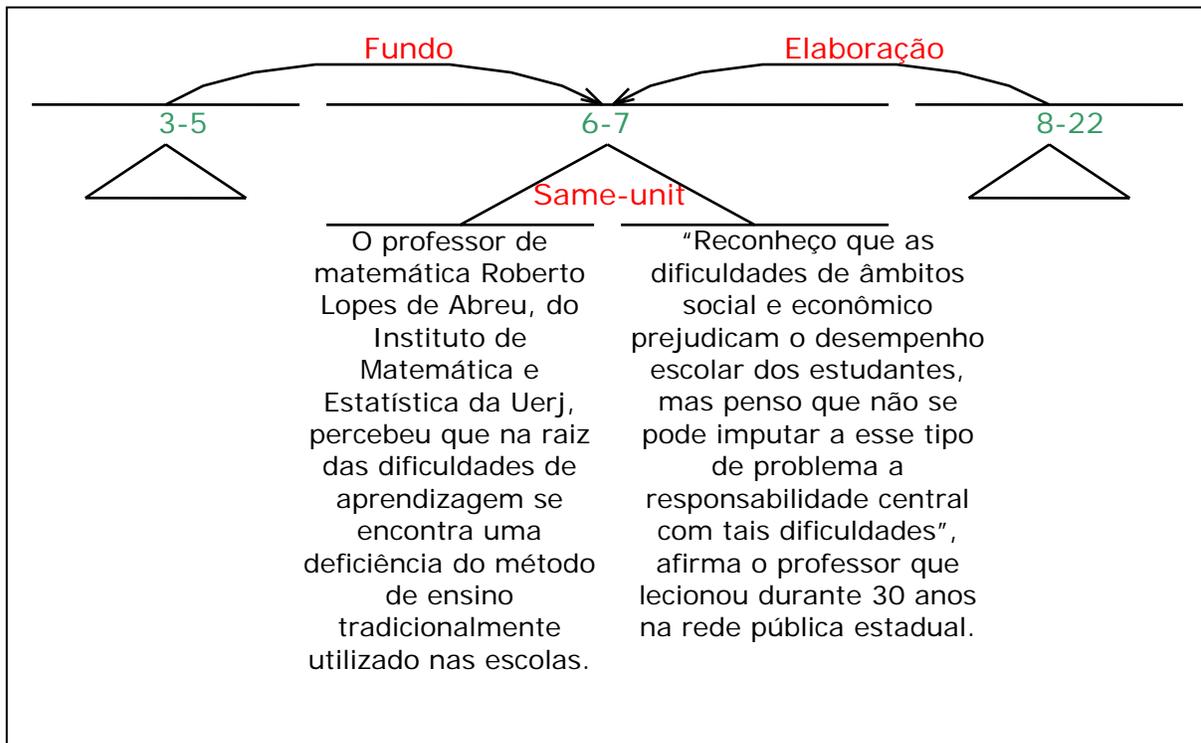


FIGURA 4 – Esquema multi-relação observado no texto *Onde está a raiz do problema?*

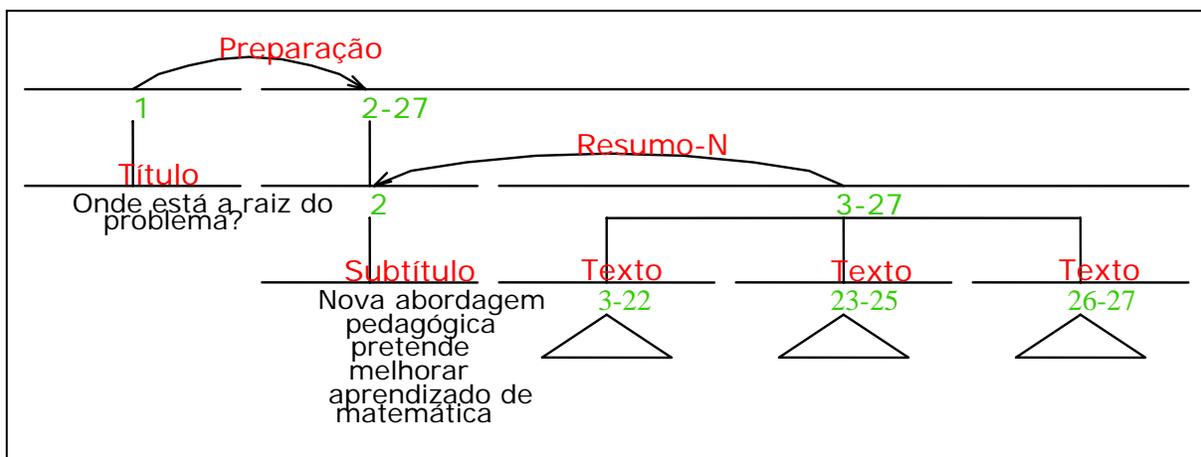


FIGURA 5 – Aplicação de esquemas superestruturais ao texto *Onde está a raiz do problema?*

A noção de que o texto é uma unidade comunicativa, composta por outras unidades menores, leva a considerações sobre o uso da relação *Same-Unit* neste trabalho. Por unir proposições redutíveis a uma macroproposição, a *Same-Unit* abarca relações que, em estudos que considerem a organização microestrutural, podem ser identificadas. Entretanto, o propósito das análises apresentadas é a identificação da organização retórica macroestrutural de artigos opinativos autorais e de artigos de divulgação científica.

A identificação do fim discursivo de um texto e das informações nucleares para o cumprimento desse fim discursivo são tarefas que ainda representam um desafio para os pesquisadores que buscam tratá-las automaticamente.

As dificuldades encontradas para representar os esquemas *RST* observados na análise macroestrutural vêm ao encontro de questionamentos de William Mann (2000)³, em "Areas por resolver em la RST". Indaga ele: "Quais são os diferentes esquemas que se encontram na realização das relações em um texto? Com que frequência se dá cada esquema? [...] Que esquemas observam-se na ordem das unidades nas análises retóricas?".

Uma análise aprofundada dos esquemas *RST*, observados em determinados tipos textuais, poderia, eventualmente, resultar na ampliação do conjunto de esquemas reconhecidos pela teoria. Também, o estudo desses esquemas, observados em um maior número de textos, poderia levar ao desenvolvimento de um aplicativo computacional para representar a organização retórica macroestrutural, já que a *RSTTool* é um aplicativo desenvolvido a partir de estudos da organização retórica microestrutural de textos.

REFERÊNCIAS BIBLIOGRÁFICAS

- BERNÁRDEZ, E. *Teoría y epistemología del texto*. Madrid: Cátedra, 1995.
- BETTENCOURT, E. Os fins não justificam os meios. *Folha de São Paulo*. São Paulo, 22 jun. 2002. Opinião, p. A2.
- BIELLO, D. Bactérias transformam isopor em plástico biodegradável. 2006. *Revista Scientific American Brasil*. Disponível em: <<http://www2.uol.com.br/sciam/>>. Acesso em: 24 mar. 2006.
- CARLSON, L.; MARCU, D. *Discourse tagging reference manual*. ISI Technical Report ISI-TR-545, 2001.

³ MANN, William. Areas por resolver en la RST. Mar. 2000. Disponível em: <<http://www.sfu.ca/rst/03research/index.html>>. Acesso em: 3 nov. 2006.

- CONY, C. H. Alhos e bugalhos. *Folha de São Paulo*. São Paulo, 20 jun. 2002. Opinião, p. A2.
- FILHO, M. C. Onde está a raiz do problema? 2006. *Revista Ciência Hoje*. Disponível em: <<http://cienciahoje.uol.com.br/3470>>. Acesso em: 04 mar. 2006.
- GIERING, M. E. O artigo de opinião autoral: as escolhas estratégicas do produtor para o fazer-creer. *Calidoscópio: Revista de Lingüística Aplicada da UNISINOS*, São Leopoldo, v. 3, n. 2, p.138-143, maio/agosto, 2005a. _____ . *Projeto Organização Retórica de Textos de Divulgação Científica*. (UNIPP CON001110). São Leopoldo: UNISINOS, 2005b.
- MANN, W. *Areas por resolver en la RST*. Mar. 2000. Disponível em: <<http://www.sfu.ca/rst/03research/index.html>>. Acesso em: 3 nov. 2006.
- MANN, W.C.; THOMPSON, S.A. *Rhetorical Structure Theory: toward a functional theory of text organization*. *Text*, 8 (3). 1988, p. 243-281.
- MARCU, D. *RST Annotation Tool*. Disponível em: <<http://www.isi.edu/~marcu/discourse/AnnotationSoftware.html>>. Acesso em: 05 set. 2006.
- O'DONNELL, M. RST-Tool: an RST analysis tool. In: *Proceedings of the 6th European Workshop on Natural Language Generation*. March 24 - 26. 1997. Gerhard-Mercator University, Duisburg, Germany, 1997.
- _____. RSTTool 2.4 -- A Markup Tool for Rhetorical Structure Theory. In: *Proceedings of the International Natural Language Generation Conference (INLG'2000)*, 13-16 June, Mitzpe Ramon, Israel, 2000, p. 253-256.
- _____. *RSTTool - an RST Markup Tool*. Version 3.0 Users Guide. Disponível em: <<http://www.wagsoft.com/RSTTool/>> Acesso em: 03 mar. 2006.
- PARDO, T.A.S. *Métodos para Análise Discursiva Automática*. 2005. 211 f. Tese (Doutorado) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2005.
- TABOADA, M.; MANN, W. C. *Rhetorical Structure Theory: looking back and moving ahead*. 24 jan. 2006. Disponível em: <[http://www.sfu.ca/~mtaboada/docs/Taboada Mann_RST_Part1.pdf](http://www.sfu.ca/~mtaboada/docs/Taboada_Mann_RST_Part1.pdf)>. Acesso em: 05 set. 2006.

Recebido em setembro de 2007
Aprovado em fevereiro 2008