Rhetalho: um corpus de referência anotado retoricamente

Thiago Alexandre Salgueiro Pardo (ICMC-USP)
Eloize Rossi Marques Seno (ICMC-USP)

Apresenta-se, neste resumo, a descrição do Rhetalho, um corpus anotado segundo a *Rhetorical Structure Theory* – RST, uma das teorias discursivas mais utilizadas nas áreas de Lingüística e Lingüística Computacional.

Segundo a RST, o conteúdo textual pode ser representado por uma estrutura em que as proposições (isto é, unidades mínimas de significado) expressas em um texto são relacionadas por relações retóricas, diferenciando-se a importância destas proposições para a satisfação do objetivo do texto. Na RST, são definidas 24 relações que foram estendidas para 32, dadas as especificidades do corpus.

O Rhetalho é composto por 50 textos: 20 seções de introdução e 10 seções de conclusão de artigos científicos do domínio da Computação; 20 textos do jornal *on-line* Folha de São Paulo, mais especificamente, 7 textos da Seção Cotidiano, 7 da Seção Mundo e 6 da Seção Ciência.

Este trabalho visou à produção de um corpus de referência, isto é, que pudesse ser utilizado pela comunidade de pesquisa. Para isso, a anotação do corpus foi realizada por 2 humanos especialistas em RST, buscando produzir análises discursivas em que houvesse plena concordância por parte dos anotadores, garantindo, desta forma, a consistência da anotação. A anotação foi feita com o auxílio da ferramenta de edição RSTTool, utilizando-se um protocolo de anotação desenvolvido especialmente para este fim.

Na Lingüística, algumas possíveis aplicações para o Rhetalho são o estudo de fenômenos lingüísticos no âmbito do discurso, como anáforas e elipses, e a investigação do uso de marcadores superficiais do discurso. Na Lingüística Computacional, pode-se utilizar o Rhetalho para o desenvolvimento de sistemas que manipulam conhecimento discursivo, como analisadores discursivos e sumarizadores automáticos.

O Rhetalho foi desenvolvido no NILC (Núcleo Interinstitucional de Lingüística Computacional) e encontra-se disponível no endereço http://www.nilc.icmc.usp.br/~thiago/rhetalho.html