

ANOTAÇÃO DE CORREFERÊNCIA TEXTUAL EM CORPUS LINGÜÍSTICO USANDO XML. *Karina de Moraes Crestani, Marcirio Silveira Chaves, Renata Vieira* (Centro de Ciências Exatas e Tecnológicas – Informática – UNISINOS).

A identificação de expressões que se referem a um mesmo elemento em textos da linguagem natural, também chamada resolução de correferência, é importante em diversas aplicações, por exemplo, extração e/ou recuperação de informação, geração automática de resumos e tradução automática. Entretanto, estabelecer essas relações entre elementos de um mesmo texto automaticamente é um problema ainda não solucionado. O desenvolvimento de sistemas para tratar esse problema pode se beneficiar de recursos lingüísticos, tais como, textos anotados, onde essas relações são explicitadas. Textos anotados dão suporte à elaboração e à avaliação desses sistemas. Este trabalho tem como objetivo a construção e a utilização de um modelo para anotação de correferência em textos da Língua Portuguesa, fazendo uso da linguagem de marcação XML - Linguagem de Marcação Extensível. Essa linguagem tem sido adotada como padrão para marcação lingüística de texto e de discurso, como por exemplo, no projeto MATE - Multilevel Annotation Tools Engineering que serviu de base para a construção do nosso modelo. No modelo proposto são marcados os sintagmas nominais do texto e as relações de correferência existentes entre as descrições definidas são indicadas (descrições definidas são sintagmas nominais iniciados por artigo definido *o, a, os, as*). Os resultados alcançados até agora constituem-se de um esquema para marcação em XML de um corpus que pode ser utilizado para pesquisas sobre correferência e descrições definidas da Língua Portuguesa. A próxima fase do trabalho é a exploração tanto de consultas sobre o texto marcado quanto da visualização do mesmo através do uso de XML e dos aplicativos associados a essa tecnologia. (UNIBIC).