

068

CORREÇÃO DE ERROS NA PREPARAÇÃO DO CORPUS PARA PROCESSAMENTO COMPUTACIONAL. *Clarissa Pinheiro, Roberta Rossoni, Renata Vieira, Rove Chishman* (Projeto ANACORT – Anotação Automática de Co-Referência Textual, Centro de Ciências Exatas e Tecnológicas / Centro de Ciências da Comunicação – UNISINOS)

O Projeto ANACORT, coordenado pela Prof. Renata Vieira, da UNISINOS, é uma pesquisa da área de lingüística computacional que visa realizar o processamento computacional de linguagem natural em Língua Portuguesa e que está unindo pesquisadores dos cursos de Informática e Letras. Os pesquisadores da área de Letras contribuem em diferentes etapas desta pesquisa, cujo objetivo é processar computacionalmente as expressões definidas de um corpus composto de 15 artigos do jornal Correio do Povo. Dentre as etapas de preparação do corpus para a execução de trabalhos mais detalhados, submeteram-se os textos a uma análise sintática fornecida pelo software interativo do projeto VISL (Visual Interactive Syntax Learning) da University of Southern Dinamarque. Nossa contribuição nesta fase do projeto consistiu em revisar manualmente as listas geradas por esse software. Na ocorrência de erros, realizamos sua correção e devida adaptação para um melhor processamento das informações. Além da identificação das principais falhas – representação das expressões coordenadas, expressões numéricas, apositivas e uso de vírgulas -, pudemos iniciar o processo de classificação dos diferentes usos das descrições definidas, o que envolve estudar três possibilidades de usos das descrições definidas dentro dos textos: diretamente anafóricas, associativas e novas no discurso. (FAPERGS / PIBIC - UNISINOS)