

138

ANACORT - EXTRAÇÃO DE SINTAGMAS NOMINAIS PARA O PROCESSAMENTO DE CO-REFERÊNCIA. *Daniela Rossi, Fabiano Gorziza* (Unisinos, São Leopoldo-RS)

A identificação de expressões que se referem a um mesmo elemento é importante em diversas aplicações de processamento de linguagem natural, como, por exemplo, extração e/ou recuperação de informação, geração de linguagem natural, geração automática de resumos e tradução automática. A área de processamento natural é rica em trabalhos voltados ao problema de resolução de anáforas pronominais, menor atenção tem sido dada a outros casos de co-referência. Nosso trabalho tem sido guiado pelo objetivo de tratar estes outros casos, mais especificamente as descrições definidas (sintagmas nominais iniciados por artigo definido). Trabalhos anteriores (Vieira, 1998; Poesio and Vieira, 1998; Vieira and Poesio, 1999) apresentam um estudo detalhado sobre o uso de descrições definidas na Língua Inglesa e propõem um sistema baseado em corpus para o processamento destas expressões. Esses trabalhos têm sido base para estudos sobre a resolução de descrições definidas da Língua Portuguesa. Para tal, foi preciso a construção de um corpus constituído por um conjunto de textos e as respectivas listas de sintagmas nominais. Nosso corpus, consiste das listas de sintagmas nominais extraídos de artigos do Jornal Correio do Povo, de Porto Alegre. A extração dos mesmos foi feita semi-automaticamente: de árvores sintáticas geradas pelo software interativo do projeto Visual Interactive Syntax Learning (VISL) foram extraídos automaticamente sintagmas nominais em forma de listas PROLOG. O resultado obtido passou por uma correção manual.