

Tratamento Automático de Ambigüidades na Recuperação da Informação.

Marisa Brascher

Este estudo trata da ambigüidade morfológica e lexical, mais especificamente da policategorização, da homografia e da polissemia. A pesquisa tem como objetivo geral solucionar casos de ambigüidades em textos científicos e técnicos em língua portuguesa (variante brasileira), por meio de um sistema de tratamento automático da linguagem natural, com vistas a promover maior precisão na recuperação da informação. Procura-se responder a duas questões básicas: como solucionar as ambigüidades que ocorrem em textos científicos e técnicos em língua portuguesa e que interferem na recuperação da informação e como introduzir informações semânticas num sistema de tratamento automático da linguagem natural, de maneira a possibilitar a solução destas ambigüidades. Discutem-se questões relativas à extração de informações contidas em textos completos e posterior recuperação, por meio de métodos de tratamento automático da linguagem natural. Além da extração de palavras do texto, procura-se manter as relações de significado que estas palavras possuem no contexto do discurso em que ocorrem. Os sistemas de recuperação de informação em linguagem natural devem ser capazes de tratar determinados fenômenos lingüísticos que afetam a qualidade da recuperação e necessitam, portanto, da incorporação de um sistema completo de tratamento da linguagem, constituído de componentes morfológicos, lexicais, sintáticos, semânticos e pragmáticos. Os aspectos semânticos devem ser contemplados para solucionar problemas não resolvidos pela análise sintática, como, por exemplo, o da ambigüidade. A ambigüidade causa ruído na recuperação da informação, pois, sob um mesmo termo, o usuário encontrará informação relevante e irrelevante. Como referencial teórico-metodológico para efetuar a análise e organização sintático-semântica dos lemas incluídos num Dicionário automático, utilizam-se a Gramática de Valências de Borba e a teoria de Gráficos Conceituais de Sowa. Emprega-se um sistema de tratamento automático da linguagem natural – o Zstation – e um corpus constituído de documentos oficiais do Mercosul, para operacionalização de dois tipos de tratamento de ambigüidade: tratamento automático de formas policategoriais e tratamento sintático-semântico de homografias e polissemias. Os resultados obtidos nos testes efetuados confirmam que a análise do contexto possibilita resolver casos simples de policategorização, mesmo sem utilizar uma gramática formal muito complexa. Quanto ao tratamento de polissemias e homografias, conclui-se que um sistema de tratamento da linguagem natural pode solucionar determinados tipos de ambigüidades quando dispõe de informações relativas à valência sintático-semântica das unidades lexicais que compõem um enunciado. Os resultados obtidos demonstram ser possível introduzir procedimentos automáticos de solução de ambigüidades em sistemas de tratamento da linguagem natural.