

Proposta de uma Metodologia para a Identificação dos Argumentos dos Adjetivos de Valência 1 da Língua Portuguesa a partir de Córpus

Ariani Di Felippo (FCL-UNESP/Ar.)

Thiago Alexandre Salgueiro Pardo (ICMC-USP)

Sandra Maria Aluísio (ICMC-USP)

Propõe-se, neste trabalho, um método semi-supervisionado baseado em córpus para a identificação dos argumentos dos adjetivos qualificadores de valência 1. O estudo é feito para os adjetivos da língua portuguesa em posição adnominal (antepostos e pospostos). Nas pesquisas em Processamento de Línguas Naturais, a necessidade de léxicos enriquecidos semanticamente, com estruturas argumentais, por exemplo, faz-se evidente. Em geral, as pesquisas concentram-se no estudo dos verbos. Neste trabalho, uma investigação inicial para a classe dos adjetivos é conduzida.

Foram coletadas todas as ocorrências de dois adjetivos da língua – “bom” e “novo” – no córpus Mac-Morpho do projeto Lácio-Web. Para o adjetivo “bom” anteposto ao substantivo, foram encontradas 417 ocorrências, e, em posição posposta, 20; para o adjetivo “novo” anteposto ao substantivo, foram encontradas 1.063 ocorrências, e, em posição posposta, 101. Procedeu-se, então, a uma limpeza automática das ocorrências, mantendo-se somente os pares substantivo-adjetivo e adjetivo-substantivo. Utilizando-se o léxico do NILC, todas as palavras foram reduzidas às suas formas canônicas para uniformidade e maior eficácia do aprendizado dos argumentos.

A fim de possibilitar a generalização dos argumentos no aprendizado, mapearam-se estes em conceitos. Pelo fato de WordNet.Br não conter a relação de hiperonímia, utilizou-se a WordNet do inglês, e, com isso, fez-se necessário o uso de um dicionário bilíngüe para traduzir as palavras do português para o inglês. Consideraram-se, nessa tradução, todos os sentidos possíveis para as palavras. Em seguida, para cada palavra, buscou-se o hiperônimo mais imediato do primeiro sentido contido na WordNet.

Os pares substantivo-adjetivo e adjetivo-substantivo foram submetidos a um processo de cálculo de frequência para a determinação dos argumentos mais prováveis. Como esperado, verificaram-se, na lista de argumentos aprendidos, as seguintes propriedades dos adjetivos: quando antepostos, seus argumentos lexicalizam conceitos abstratos (por exemplo, o adjetivo “bom” e o argumento *feeling*); quando pospostos, seus argumentos lexicalizam conceitos concretos (por exemplo, o adjetivo “bom” e o argumento *object*).