

Padrões de Lexicalização e a Indexação da Base da Wordnet.Br

Ariani DI FELIPPO (UNESP/Ar.)

Bento Carlos DIAS-DA-SILVA (UNESP/Ar.)

A *WordNet*, desenvolvida para o inglês americano, é uma base de dados lexicais que se estrutura em conjuntos de sinônimos, os "synsets", que visam a representar o conceito lexicalizado pelas unidades que o compõem; cada synset constitui um nó e as ligações entre os diferentes nós, representada por meio de arcos rotulados, visam a exprimir a relação léxico-semântica de antonímia e as relações lógico-conceptuais de hiponímia, troponímia, meronímia, causa e acarretamento. Diante da relevância lingüística e da potencialidade tecnológica, o desenvolvimento de wordnets para diferentes línguas frutificou-se e sofisticou-se. As wordnets em construção para as línguas dos países da União Européia, por exemplo, estão aglutinadas em uma base multilíngüe, a *EuroWordNet*, em que os synsets das diferentes línguas que lexicalizam um conceito comum são interligados. Em uma das etapas futuras do desenvolvimento da wordnet brasileira, a Wordnet.Br, está prevista a interligação desta a wordnet americana, resultando, assim, em uma base bilíngüe que poderá ser utilizada em pesquisas lingüísticas e do PLN. Essa indexação, em especial, requer a identificação de *padrões de lexicalização* de conceitos (isto é, mapeamento ou associação regular entre conceitos e itens lexicais) do inglês e do português. Motivado por essa tarefa de indexação das wordnets brasileira e americana, apresenta-se neste trabalho a proposta de estudo do mapeamento dos padrões de lexicalização de conceitos codificados nos nomes concretos do inglês e do português. Além de subsidiar parte da futura indexação da Wordnet.Br a WordNet de Princeton, o trabalho ora proposto busca contribuir para o desenvolvimento de questões como o desenvolvimento de ontologias lingüísticas e conceituais e a especificação semântica dos itens lexicais. [CNPq]