

O PLN (Processamento de Línguas Naturais) é uma área da Inteligência Artificial que abrange modelos computacionais capazes de compreender e gerar textos em LN (Língua Natural). O desenvolvimento de um sistema de PLN pode envolver a utilização de uma série de recursos e etapas de processamento, como o lexical, sintático, semântico e pragmático-discursivo. Abordagens baseadas em conhecimento que procuram manipular o significado dos textos apresentam, normalmente, melhores resultados nas diversas aplicações do PLN, como a tradução e a sumarização automática, a manipulação de bases de dados, etc. Esses sistemas incluem, para tanto, um módulo responsável pela etapa de processamento semântico. Na compreensão da LN, foco deste trabalho, essa etapa é denominada “análise semântica” ou “geração conceitual”. Apesar de existirem vários modelos de geração conceitual na literatura, eles não contemplam características específicas da língua portuguesa. Além disso, esses modelos são, geralmente, baseados em recursos lingüísticos bastante complexos, inexistentes, ainda, para o português. Sua adaptação para o cenário em questão exigiria, portanto, a criação e instanciação de tais recursos com informações específicas, um processo que demandaria esforço e tempo consideráveis.

O objetivo do sistema ConPor (CONceitualização do PORTuguês) desenvolvido é a geração de estruturas conceituais, expressas na linguagem de representação conceitual UNL (*Universal Networking Language*), para sentenças do português do Brasil, limitadas a um vocabulário específico (horóscopo diário) e a um subconjunto das regras gramaticais dessa LN. Um objetivo secundário desse sistema é a sua independência de recursos e processos externos ao NILC (Núcleo Interinstitucional de Lingüística Computacional de São Carlos) e a possibilidade de utilização dos seus resultados em diversas aplicações de PLN, entre elas, a tradução e a sumarização automática.

A metodologia do ConPor é baseada em *corpus*, ou seja, em um conjunto de exemplos que delineiam a construção dos seus recursos e processos. Adicionalmente, é baseada na reutilização de recursos e processos disponíveis no NILC, além do desenvolvimento de recursos e processos específicos. A reutilização de um *parser* do português, em especial, permite uma arquitetura modular, segundo o método de compreensão sequencial (isto é, de análise semântica posterior à análise sintática), na qual o sistema explora os resultados desse *parser*, focalizando a etapa de geração conceitual. O modelo de geração conceitual do ConPor é genérico e flexível, definido segundo uma abordagem composicional e funcional, por meio de dois componentes principais: (a) uma gramática gerativo-transformacional de regras de projeção, que convertem os constituintes da sentença, de acordo com suas funções sintáticas, em conceitos da UNL associados às suas funções semânticas; e (b) um conjunto de *templates* de relacionamento, que distribuem esses conceitos em relações semânticas da UNL, gerando, com isso, as estruturas UNL.

Após sua implementação, o ConPor foi avaliado para verificar a qualidade das regras de mapeamento sintático-conceitual e o poder de expressividade das estruturas UNL resultantes, com relação ao significado das sentenças originais. A avaliação mostrou que o sistema apresenta um bom desempenho, na sua atual configuração e, além disso, que tem um bom potencial de geração conceitual considerando futuras expansões na sua abrangência. Essas expansões, por sua vez, segundo uma análise realizada, se mostraram factíveis e pouco custosas, uma vez que, como grande parte das regras pode ser reutilizada, o crescimento de regras para contemplar novas sentenças não é linear.

CONPOR: UM GERADOR CONCEITUAL PARA O PORTUGUÊS

SPECIA, LUCIA

RINO, LUCIA HELENA MACHADO

UFSCAR - UNIVERSIDADE FEDERAL DE SÃO CARLOS