

Representação lingüístico-computacional dos adjetivos valenciais para o processamento automático do português. Ariani Di Felippo (B), Prof. Dr. Bento Carlos Dias da Silva (O). PPG- Lingüística e Língua Portuguesa / UNESP – Araraquara.

Justificativa

Do ponto de vista lingüístico, o léxico é o conjunto de unidades de uma determinada língua natural que, aparentemente, é caótico, mas que, quando examinado em maior profundidade, revela uma estrutura própria e uma intrincada rede de relações entre seus constituintes (Mel'cuk, 1988). Com o objetivo de investigar as propriedades léxico-semânticas que caracterizam a complexidade do léxico, a Semântica Lexical vem buscando estudar vários aspectos que revelam facetas diferentes e complementares do léxico, por exemplo: 1. estrutura argumental, 2. papéis temáticos, etc.

Tais investigações têm sido, por sua vez, objeto de pelo menos três domínios: Lingüística, Ciência Cognitiva e Processamento Automático das Línguas Naturais (PLN).

Especificamente para o domínio do PLN, os dois componentes centrais de um sistema são o analisador sintático (parser) e o "léxico", estrutura de dados lexicais organizada em termos do par [ATRIBUTO valor] como, por exemplo: [PRED 'Pedro']; [GEN +m]; (Dias-da-Silva, 1996). Aquele responsável pelo processamento sintático de frases e orações e este encarregado de fornecer as informações léxico-semânticas para o analisador. Diante do fato de que variados sistemas de PLN, construídos para "funcionar" em aplicações e situações reais, requererem "léxicos" robustos, tanto do ponto de vista da quantidade de formas quanto do ponto de vista da qualidade e precisão das informações a elas associadas, os lingüistas computacionais se depararam com a necessidade de superarem as limitações de performance atribuídas à inadequação dos léxicos.

No domínio do PLN, constatamos que o estudo léxico-gramatical das classes gramaticais abertas dos verbos e substantivos tem merecido uma atenção bem maior, por parte dos investigadores, que classes como adjetivos e advérbios. Recorte justificável ao considerarmos a centralidade dessas classes para a constituição de frases e orações simples, conforme definição dada por Quirk & Greenbaum (1985). Desse modo, neste trabalho, nosso recorte é diferente. Privilegiamos o estudo da classe dos adjetivos. A escolha dessa classe justifica-se pelo motivo acima apontado e, sobretudo, pela sua importância no âmbito do sistema e do uso lingüísticos.

Objetivo

Dentro da classe sintática escolhida para investigação, este trabalho tem como objeto a questão da representação lingüístico-computacional da subclasse dos adjetivos valenciais num léxico computacional para PLN.

Os adjetivos valências são aqueles que, em um nível abstrato de representação, correspondem a um predicado (Borba, 1996). Essa subclasse de adjetivos caracteriza-se por apresentar uma estrutura de argumentos semelhante àquela prevista para os verbos. O adjetivo transportável, na expressão uma carga transportável do navio para os armazéns por guindastes (Borba, 1996), pode projetar quatro argumentos, associados a quatro papéis temáticos: por guindastes (Arg 1 = Agente), uma carga (Arg 2 = Tema), do navio (Arg 3 = Origem) e para os armazéns (Arg 4 = Instrumental).

Baseando-se em tal objeto de estudo e na metodologia adotada, o trabalho aqui apresentado.

concentra-se em três objetivos complementares, que são:

1. Compilação adjetivos valenciais e análise de suas principais propriedades morfossintáticas e semânticas relevantes para a estruturação de um léxico computacional;
2. Formalização dessas propriedades com o objetivo de projetar um modelo de representação computacionalmente tratável para essa subclasse de adjetivos, ou seja, um modelo de estruturação e de representação do léxico adequado para ser incorporado a sistemas de PLN;
3. Construção de um editor, isto é, uma interface computacional para uso do lingüista, que apresente estrutura e funcionalidades para a operacionalização das análises e resultados alcançados a partir dos estudos realizados nas atividades anteriores.

Revisão bibliográfica

A pesquisa desenvolvida neste trabalho investiga três domínios complementares, o Lingüístico, o Representacional e o Implementacional/Computacional. No que concerne ao domínio Lingüístico, especificamente aos adjetivos valenciais, vale ressaltar os trabalhos realizados por Bresnan (1981, 1982), Cruse (1986), Chierchia & McConnell (1990), Borba (1996), Neves (1999) e Pustejovsky (1995), que fornecem os subsídios atuais e essenciais para pesquisas lingüísticas dessa natureza. No que diz respeito ao domínio Representacional ressaltamos como fonte de pesquisa deste trabalho os modelos propostos por Bresnan (1982), Miller & Felbaum (1991), Pustejovsky (1995) e Handke (1995).

Por fim, salientamos também que os projetos WordNet (Base de dados lexicais on-line na qual nomes, adjetivos, verbos e advérbios do inglês estão organizados em conjuntos de sinônimos), SIMPLE (Semantic Information for Multifunctional Plurilingual Lexica - modelo que visa adição de informação semântica a uma parte das unidades lexicais codificadas morfológica e sintaticamente no projeto PAROLE) e TeP (Projeto desenvolvido pelo NILC que visa a compilação de um thesaurus eletrônico para o Português do Brasil) representam o que há de mais atual e relevante nas pesquisas no PLN. Tais projetos fornecem informações relevantes para os três domínios complementares que compõem este trabalho.

Metodologia

Na busca de equacionamento para esta pesquisa, tomou-se por base Handke (1995), Saint-Dizier & Viegas (1995) e Dias-da-Silva (1996), que fornecem os subsídios metodológicos essenciais para o desenvolvimento de projetos interdisciplinares no domínio do PLN.

Como decorrência desse recorte teórico-metodológico, este trabalho decompõe-se em domínios complementares: o Lingüístico, o Representacional e o Implementacional/Computacional (Dias-da-Silva, 1996). No domínio Lingüístico, as atividades concentram-se no aprofundamento dos estudos sobre os adjetivos valenciais do ponto de vista lingüístico. Para a compilação dos adjetivos, utiliza-se a Base do TeP e os corpora do Centro de Estudos Lexicográficos da FCL - UNESP/ Araraquara e do NILC. No Representacional, os conhecimentos lingüísticos reunidos são "projetados" em um sistema formal de representação. Por último, no domínio Implementacional, as atividades ficam concentradas nas questões pertinentes à codificação das representações elaboradas durante a fase anterior, em termos de linguagens de programação. É nesta fase que se pretende implementar o editor para a montagem das entradas dos adjetivos valenciais.

Resultados

Os resultados deste trabalho contribuem, de um lado, para a proposição de modelos lingüísticos mais completos, explícitos e operacionais e, conseqüentemente, mais apropriados para receberem um tratamento computacional, e, de outro, para a construção de sistemas de PLN lingüisticamente motivados e cada vez mais capazes de simular aspectos da competência e do comportamento lingüísticos humanos (Dias-da-Silva, 1996). Do ponto de vista prático, este trabalho gera informações acerca do léxico que contribuem para o aperfeiçoamento da Base de Dados Lexicais para o Português (BDL), que está sendo desenvolvida pelos pesquisadores do NILC.

Discussões e conclusões

Os resultados até agora obtidos reforçam a importância de trabalhos realizados na interface interdisciplinar entre lingüística e computação, domínio em que ambos os estudos, lingüístico e do PLN, têm se beneficiado, de fato, com o trabalho colaborativo.

Referências

- Borba, F. S. **Uma gramática de valências para o português**. Editora Ática, São Paulo, 1996.
- Bresnan J. (ed.) **The mental representation of grammatical relations**. The MIT Press, Cambridge, MA, 1982.
- Chierchia, G., McConnell-Ginet, S. **Meaning and grammar**. MIT Press, Cambridge, MA, 1990.
- Cruse, D. A. **Lexical semantics**. Cambridge University Press, Cambridge, MA, 1986.
- Dias-da-Silva, B. C. **A face tecnológica dos estudos da linguagem: o processamento automático dos línguas naturais**. Dissertação de Doutorado - Universidade Estadual Paulista, Araraquara, SP, 276p., 1996.
- Handke, J. **The structure of the Lexicon: human versus machine**. Mouton de Gruyter, Berlin, 1995.
- Mel'cuk, I. **Dependency Syntax : Theory and Practice**. The SUNY Press, Albany, N.Y, 428p. 1988.
- Miller, G. A., Fellbaum, C. **Semantic networks of English**. Cognition, Vol. 41, n.1-3, p.197-229, 1991.
- Neves, M.H.M. **Gramática de usos do português**. Editora UNESP, São Paulo, 2000.
- Pustelovsky. J. **The Generative lexicon**. Massachusetts Institute of Technology, Cambridge, 1995.
- Quirk, R, Greenbaum, S. **A comprehensive grammar of english**. Longman, London, 1985.
- Saint-Dizier, P., Viegas, E. **Computational lexical semantics**. Cambridge University Press, Cambridge, 1995.

Financiamento: Agência: CAPES, Mestrado.

Palavras-chave: adjetivos valenciais; léxico; PLN; semântica lexical; formalismos.