- **LEXICOLOGIA/LEXICOGRAFIA II**
  Dia 29 - Local: CL 13 - 14:00-18:00
  Coordenador(a): *Ariani di Felippo*

A SINONÍMIA NOS TEXTOS ESPECIALIZADOS: UM ESTUDO EM UM CORPUS DE ECONOMIA

Mariângela de Araújo (USP)

Embora, por muito tempo, os estudiosos da Terminologia tenham considerado a sinonímia e a polissemia empecilhos para a exatidão na comunicação especializada, atualmente há uma consciência científica de que tais fenômenos são inerentes aos discursos especializados, já que estes fazem parte das línguas naturais e, sendo assim, são passíveis de todos os fenômenos inerentes a elas.

Essa aceitação dos fenômenos da polissemia e da sinonímia pelos atuais estudiosos da Terminologia ocorreu devido aos numerosos estudos empíricos realizados e a uma nova vertente dos estudos terminológicos, denominada Socioterminologia.

Com essa perspectiva teórica, apresentamos nesta ocasião um estudo que temos desenvolvido com relação à sinonímia em um corpus constituído por manuais, teses e dissertações utilizados por alunos de graduação em Economia. Essa área do conhecimento tem demonstrado uma grande variedade denominativa (por exemplo, barreiras à entrada / barreiras à entrada de novas firmas / barreiras ao acesso de novas firmas, caminho de expansão / trajetória de expansão, curva de igual custo / isocusto), que, mais uma vez, comprova a tese de que as terminologias fazem parte das línguas naturais.

ALTERNÂNCIAS SINTÁTICAS, ESTRUTURA DE ARGUMENTOS, SYNSETS A REDE WORDNET.BR

Bento Carlos dias da Silva (UNESP), Maria Carolina Ávila

Os objetivos deste trabalho são (i) apresentar a metodologia de análise léxico-gramatical em que a hipótese de trabalho é considerar que a realização sintática da estrutura de argumentos do verbo é norteada por suas propriedades semânticas e (ii) demonstrar como esse tipo de análise, que opera na interface entre a sintaxe e a semântica, pode auxiliar na tarefa de refinamento da base lexical da rede Wordnet.Br. Trata-se de uma base relacional de unidades lexicais que se estrutura sob a forma de synsets (synonym sets, i.e. conjunto de sinônimos). Nessa rede, o sentido de uma unidade lexical emerge tanto das relações léxico-semânticas que essa unidade mantém com as demais unidades do synset de que é membro e quanto das relações lógico-conceituais que o synset que a contém mantém com outros synsets da rede. Em fase de construção, a base da rede Wordnet.Br consiste na ampliação qualitativa de uma outra base de dados lexicais, construída em projeto anterior, que é também estruturada em termos de conjuntos de sinônimos: a base do thesaurus eletrônico para o do português do Brasil. A base desses thesaurus contém cerca de 44 mil unidades lexicais, distribuídas entre verbos, adjetivos, substantivos e advérbios. A análise da consistência léxico-semântica de uma amostra de mais de 50% dos 4 mil synsets de verbos, revelou que parcela considerável das unidades verbais colocadas em um mesmo conjunto não compartilhava, com os demais membros, o mesmo comportamento sintático. Se a hipótese inicial estiver correta, esse fato sugere que os synsets que apresentam essa peculiaridade devem apontar para mais de um sentido. Decorre, então, que a aplicação do método de análise acima referido deverá contribuir não só para o desenvolvimento da rede Wordnet.Br, com a discriminação mais precisa do sentido subjacente ao synset, como também para a descrição sintático-semântica dessa complexa categoria lexical.

MODELO LINGÜÍSTICO-COMPUTACIONAL DA ESTRUTURA VALENCIAL DE ADJETIVOS DO PORTUGUÊS DO BRASIL

Ariani di Felippo (UNESP), Bento Carlos dias da Silva

No âmbito do Processamento Automático das Línguas Naturais (PLN), é premente a compilação de léxicos, monolíngües e multilíngües, que sejam computacionalmente tratáveis e lingüisticamente motivados. As atividades envolvidas nesse empreendimento podem ser fatoradas em três domínios: Domínio das Entidades e das Relações Lingüísticas, Domínio da Formalização e Domínio da Implementação. Focalizando, neste trabalho, parte das atividades pertencentes aos dois primeiros domínios, propõe-se uma representação lingüístico-computacional para os adjetivos valenciais do português. Uma das motivações que norteou a escolha da classe dos adjetivos para proceder à representação formal é o fato de que, por meio deles, codificam-se aspectos subjetivos do uso lingüístico. Decorre que um sistema de PLN, para melhor emular a análise e síntese dos enunciados lingüísticos, não pode prescindir da manipulação das unidades que integram essa classe. A escolha pelo subconjunto dos adjetivos valenciais, justifica-se pelo fato de que estes, assim como os verbos, projetam uma estrutura de argumentos, que tem sido amplamente empregada para representar parte das informações léxico-gramaticais, tanto nos

estudos lingüísticos como computacionais da linguagem humana. Para fundamentar a proposta e proceder à representação, partimos de uma breve revisão dos principais modelos formais de representação das informações léxico-gramaticais. Na seqüência, delineamos, do ponto de (psico)lingüístico, os tipos de informação lexical lingüisticamente relevantes para o PLN e descrevemos as principais propriedades dos adjetivos valenciais do português. Como conclusão, apresentamos o modelo formal da entrada lexical para os adjetivos analisados.

OS VERBOS-SUPORTE NAS CONSTRUÇÕES COM NOMES DEVERBAIS ABSTRATOS

Bruno Oliveira Maroneze (USP)

Os nomes deverbais abstratos, da mesma forma que seus verbos cognatos, têm a capacidade de predicar; porém, na língua portuguesa, tais predicações não podem constituir sentenças isoladamente; é necessário utilizar verbos sem conteúdo semântico lexical, que têm apenas a função de ser o suporte da predicação. São os chamados verbos-suporte.

Por exemplo, um deverbal como visita é utilizado com verbos do tipo fazer ou receber:

(1) João fez uma visita (= visitou) ao irmão.

(2) Carlos recebeu a visita (= foi visitado) do irmão.

Nos dicionários, os verbos-suporte são descritos como acepções diferentes no verbete do verbo. Isso faz com que, por um lado, se utilizem definições muito genéricas (por exemplo, fazer é definido como executar, realizar), e por outro, alguns verbos sejam descritos com acepções muito específicas (por exemplo, uma das acepções do verbo dar, no dicionário Aurélio, é permitir, consentir, acepção que só aparece em expressões como dar consentimento). Além disso, os verbetes dos nomes deverbais raramente indicam qual é o verbo-suporte utilizado.

O presente trabalho visa propor um modelo de verbete para os nomes deverbais que leve em consideração as construções com verbos-suporte, organizando um padrão para as remissivas entre o deverbal e o(s) verbo(s)-suporte utilizado(s). Tal modelo será aproveitado no Dicionário de neologismos do português brasileiro contemporâneo, em elaboração na Universidade de São Paulo.

PADRÕES LEXICAIS NO PROCESSO DE EXTRAÇÃO AUTOMÁTICA DE INFORMAÇÃO DE TEXTOS

Lucielen Porfirio (UNIOESTE), Jorge Bidarra

A extração automática de informação de textos é uma área de investigação relativamente nova, mas que vem atraindo um significativo numero de pesquisadores de diversas áreas do conhecimento, em particular os lingüistas computacionais. O processo de extração consiste fundamentalmente na identificação de aspectos lingüísticos relevantes, sejam de natureza lexical, sintática e/ ou semântico-conceitual, capazes de permitir não só a sumarização do texto, bem como a sua própria interpretação, não manualmente. A grande maioria dos sistemas de extração automática de informação textual presentes em qualquer texto escrito tendem a se basear fortemente em regras sintáticas. Nesse estudo, propomos uma abordagem um tanto diferente. O foco desse estudo está concentrado no léxico, ou mais especificamente nas relações léxico-semanticas estabelecidas, tanto intra quanto inter palavras. Para tanto, partimos de fenômenos lingüísticos, tais como prosódia semântica, colocação e coligação (Berber Sardinha, 2000) e (Lewis, 2000).

PRODUÇÃO DE ONTOLOGIAS ESPECÍFICAS: A MODELAGEM DA ONTO-ECO

Claudia Zavaglia (UNESP)

A partir da ontologia proposta por Zavaglia (2002) para o subdomínio da Ecologia, reestruturamos o conhecimento especializado desse subdomínio em classes e subclasses, segundo os preceitos de Gruber (1993), e descrevemos formalmente seus conceitos e as relações existentes entre eles. O uso de ontologias tem sido amplamente empregado em representações do conhecimento de domínios restritos, máxime para sistemas de busca de informação e indexação de documentos, em Processamento de Línguas Naturais - PLN.

A Onto-Eco prevê três subdomínios da Ecologia, a saber: Ecologia de Ecossistemas; de Populações; de Comunidades que se revelaram altamente produtivos tanto no processo de categorização, quanto no processo de nomeação de termos. Para a extração dos termos, confeccionamos um córpus da Ecologia com textos teóricos, revistas e textos eletrônicos que está sendo alimentado periodicamente e armazenado em uma ferramenta computacional, para gerenciamento de bases textuais. Na etapa de extração do termos propriamente dita, partimos de uma identificação primeiramente manual, para em seguida procedermos à sua automatização. Nesse sentido, valemo-nos de um extrator de termos presente no Núcleo Interinstitucional de Lingüística Computacional - NILC-USP/São Carlos e utilizamos a técnica da freqüência para a captura dos mesmos. Os termos foram validados por meio de dicionários especializados e de um consultor especializado em ecologia. Em seguida, esses termos foram etiquetados com informações morfossintáticas e semânticas referentes à Estrutura Qualia do Léxico Gerativo de Pustejovsky (1995). Após a distribuição dos itens lexicais na estrutura ontológica delineada, definimos e mapeamos as relaçãoes semânticas existentes entre eles. Para o armazenamento de todas essas informações, utilizamos o editor de ontologias Protégé-2000, sistema esse que permite a interoperabilidade e o reuso de estruturas ontológicas por sistemas de representação do conhecimento, além de caracterizar a Onto-Eco com uma Base de Conhecimento, uma vez que permite a introdução de informação semântica por meio de frames.

PRODUCÃO DE UM VOCABULÁRIO TECNICO-CIENTÍFICO DA ÁREA DE ORTOPEDIA TÉCNICA

Fernanda Mello Demai (USP)

Os textos dicionarísticos são subsídios fundamentais para a (re)construção e para a transformação das axiologias que lhe são subjacentes, dentro de uma sociedade sócio-lingüístico-cultural; tais textos sistematizam conhecimentos e práticas, disponibilizando à comunidade que o produziu (e a outras, contemporâneas ou não, conterrâneas ou não) um instrumento de controle, de transformação e de manutenção da herança tradicional de conhecimentos humanos - os dicionários são testemunhos de sua civilização-mãe.

Não há ciência sem terminologia, pois a terminologia é a substância da formação científica - "a codificação do universo natural pelo homem não é outra coisa senão a visão particular que dele tem, como indivíduo ou como grupo, de forma que esse universo passa a existir para ele, segundo o modelo com que foi estruturado e não pela sua natureza intrínseca, física e fisiológica" (BARBOSA, Maria Aparecida 1986. "Lexicologia: aspectos estruturais e semântico- sintáticos" in PAIS.. Manual de lingüística. Global. São Paulo. p. 91).

É uma necessidade que constitui um problema social a grande porcentagem de pessoas (10% da população mundial, segundo a OMS) que são pessoas portadoras de deficiência - PPD, devido a acidentes de natureza social, os de trânsito e os com armas de fogo e de natureza individual, os acidentes vasculares cerebrais, as paralisias, as patologias, as formações congênitas e os traumas diversos: um grande número de pessoas necessita de equipamentos para correção de sua compleição fisiológica: próteses e órteses para membros superiores e para membros inferiores, calçados ortopédicos para pés com alterações patológicas, etc.

Não há treinamento oficial, formal, para os profissionais que trabalham na confecção de equipamentos ortopédicos e não há, para esta área de conhecimento, textos dicionarísticos, instrumentos genitores de axiologias e de ideologias, responsáveis pela manutenência e pela modificação de valores sociais.

REPRESENTAÇÃO LINGÜÍSTICO-COMPUTACIONAL DE INFORMAÇÃO METAFÓRICA

Ana Eliza Barbosa de Oliveira (UNESP)

O objetivo deste trabalho é discutir a questão da representação formal do sentido figurado em léxicos computacionais e, em particular, na rede Wordnet.Br. Como ponto de partida da discussão, avaliamos, do ponto de vista da Teoria da Metáfora Cognitiva de Lakoff & Johnson, as noções de 'sentido figurado', 'sentido metafórico' e a própria noção de 'sentido', e como essas noções poderão ser sistematizadas. Nessa tarefa, abordamos a metáfora em dois domínios: o cognitivo, que abarca o plano conceptual, ou pré-lingüístico, da metáfora, e o semântico, que envolve o plano lingüístico, ou às metáforas enquanto expressões lingüísticas. Segundo Lakoff & Johnson, a 'metáfora conceptual', ou metáfora cognitiva, na medida em que é responsável pela estruturação do pensamento, atua como constructo para a criação e apreensão das expressões metafóricas. Em outras palavras, a expressão lingüística da metáfora é uma manifestação de uma metáfora conceptual. Na seqüência do trabalho, baseando-nos em análises realizadas no âmbito de outras redes wordnets, discutimos as principais questões que envolvem a representação do sentido metafórico sob duas perspectivas: lingüística e computacional. Grande parte da argumentação contemplará a investigação de como o sentido metafórico poderia ser registrado neste tipo particular de léxico e como estudos recentes podem subsidiar a proposição de uma codificação computacional consistente, sistemática e lingüisticamente motivada.