

Apresentação

A Lingüística Computacional e suas linhas de investigação

Susana Q. de Creus
Professora da PUCRS,
Doutorado em Lingüística Aplicada



A expansão das novas tecnologias, o crescente processo de globalização das comunicações e o plurilingüismo abriram uma via enriquecedora no campo do estudo das línguas e propiciaram a implementação e o desenvolvimento de uma nova área vinculada à Lingüística e à Informática: a Lingüística Computacional.

Nela se distinguem três linhas principais de investigação: a lingüística computacional teórica; a lingüística computacional aplicada; e a informática aplicada à lingüística, linhas que envolvem, entre outros, estudos sobre o complexo domínio da cognição, os problemas de representação do conhecimento e o processamento da linguagem.

A elaboração de uma ampla gama de modelos lingüísticos computacionais no nível da morfologia, da sintaxe, do léxico, da semântica e da pragmática são a prova do interesse despertado por esta disciplina relativamente recente.

Considerou-se oportuno reunir neste número da revista *Letras de Hoje* estudos que, com diversos enfoques, oferecem uma visão panorâmica sobre conceitos, relevância, aplicações e projeção de futuro da Lingüística Computacional.

A obra, em seu conjunto, além de propor uma reflexão sobre o tema central abordado, pretende oferecer material de referência a todos aqueles que estejam desenvolvendo trabalhos de pesquisa nessa área.

Entre os ensaios aqui reunidos encontra-se o dos catedráticos da Universidade de Barcelona (UB) Eugenio Martínez Celdrán e Ana M. Fernández Planas que apresentam o AMPER (Atlas Multimedia de la Prosodia del Espacio Románico), projeto internacio-

nal dedicado a investigar os aspectos temporais, de intensidade e entonativos das variedades geoprosódicas de frases enunciativas e interrogativas de todas as línguas românicas. O foco essencial do projeto é a fonética, com vínculos à fonologia, dialetologia e sociolinguística. Atualmente o projeto conta com um grupo no Brasil, um par de grupos no Chile e outro em Cuba.

Os pesquisadores Antonio Pamies Bertrán e José Manuel Pazos Breña, ambos professores da Universidade de Granada, Espanha, estudam a detecção automatizada de unidades fraseológicas em um corpus eletrônico com fins lexicográficos, mostrando como essa metodologia, que combina de forma complementar ferramentas estatísticas e linguísticas de grande simplicidade, pode ser aplicada com eficaz rendimento em um corpus de grande magnitude.

No seu trabalho, o professor Jorge Antonio Leoni de León da Universidade de Genebra, Suíça, apresenta o esboço de uma metodologia baseada em Princípios e Parâmetros (P&P) para o reconhecimento automático de unidades fraseológicas, especificamente as que o autor denomina como *lexemas plurimembres*.

Os professores Paulo Quaresma e Irene Rodrigues, da Universidade de Évora, Portugal; Carlos A. Prolo, da PUCRS, Brasil; e Renata Vieira, da UNISINOS, Brasil, descrevem a metodologia seguida para a construção de um sistema de pergunta-resposta sobre uma base de documentos em Português. O sistema utiliza técnicas de processamento de Língua Natural para criar uma base de conhecimento com a informação veiculada pelos diversos documentos.

Um abordagem sobre termos, relacionamentos e representatividade na indexação de texto para recuperação de informação é introduzida pelos professores Vera L. S. de Lima, Marco Gonzalez da PUCRS, Brasil, e José V. de Lima, da UFRGS, Brasil. Os autores mostram uma visão geral das estratégias de representação de textos que consideram dependência de termos.

Dina Wonsever, Serrana Caviglia, Aiala Rosá e Javier Couto, professores da Universidad de la República, Uruguay, explicam um sistema para a segmentação em proposições de textos em espanhol que opera a partir dos resultados de um etiquetador probabilístico e foi concebido para oferecer um serviço de base nas tarefas de extração de informação de texto e de análise sintático.

O professor Bento C. Dias da Silva, da Universidade Estadual Paulista, reflete sobre o estudo linguístico-computacional da linguagem, propondo um equacionamento metodológico global para o desenvolvimento da pesquisa na área do Processamento Automático de Línguas Naturais (PLN) e mostra recursos teórico-metodológicos para o estudo do PLN.

No artigo *Computational Approaches to Pronoun Resolution* o professor Carlos Prolo analisa três abordagens principais que tem norteado a pesquisa de algoritmos para resolução de anáfora: saliência, *centering* e posição sintática do antecedente em relação ao pronome anafórico. No artigo são comparadas as propostas concretas originais de algoritmos baseados nestas diferentes abordagens, mostrando pontos fortes e fracos, avaliação de resultados publicados, limitações e possibilidades de longo prazo.

Vilson J. Leffa, professor da Universidade Católica de Pelotas, Brasil, descreve um tipo específico de Ferramenta de Autoria – a Ferramenta de Autoria do Professor (FAP) – voltada à produção de materiais de ensino. Trata-se de uma tecnologia computacional ainda pouco explorada em toda sua potencialidade, que visa instrumentalizar o professor para o ensino de línguas na rede mundial de computadores.

No seu trabalho, Costas Gabrielatos, pesquisador assistente da Universidade de Lancaster (Lancaster University), Reino Unido, e Simone Sarmento, professora da PUCRS e doutoranda da UFRGS (Lexicografia e Terminologia), analisam a frequência e distribuição dos verbos modais e as estruturas onde eles ocorrem em manuais de aviação. O estudo seguiu a metodologia de Lingüística de Corpus e utilizou o software Wordsmith Tools.

O professor Fábio Alves, da Universidade Federal de Minas Gerais, Brasil, e Augusto Buchweitz, doutorando da Universidade Federal de Santa Catarina, Brasil, mostram, através dos resultados de uma análise quantitativa, que a tradução da primeira língua para a segunda consome mais tempo e aumenta a segmentação da tradução, concluindo que a recursividade pode ser utilizada como uma medida de adaptação à dificuldade da tarefa de traduzir.

Uma investigação sobre as propriedades lógico-lingüísticas do conetivo 'E' em sua função inferencial, considerando as relações entre argumentos formais e informais, na interface Semântica-das-Condições-de-Verdade / Lógica de Predicados / Cálculo Proposicional (LPCP), é a proposta do professor Jorge Campos da Costa da PUCRS, Brasil.

O processo desenvolvido e os resultados obtidos através do projeto: *Preditibilidade: procedimentos utilizados por crianças de séries iniciais e escores de compreensão leitora*, – projeto situado em espaço de conexão entre Psicolingüística e Informática – é o tema abordado pela professora Vera Wannmacher Pereira, da PUCRS, Brasil e Maurício da Silveira Piccini, Bacharel em Ciências da Computação da UFRGS, Brasil.

Rove L. de Oliveira Chishman, professora da UNISINOS, Brasil e o Mestrando Anderson Bertoldi desenvolvem um estudo preliminar sobre a semântica dos adjetivos com vistas à sua inclusão em uma base de dados lexicais de um sistema de busca e extração de informação na *web*. O objetivo do trabalho é mostrar como uma aplicação computacional pode se beneficiar de uma descrição lingüística teoricamente bem fundamentada.

Considerações sobre conceitos gerais, aspectos e aplicações da Lingüística Computacional, área híbrida que envolve pesquisadores da Lingüística e da Informática, são apresentadas por Gabriel de Ávila Othero, doutorando em Lingüística Aplicada, da PUCRS.

Gabriela B. Hinrichs Conteratto, doutoranda da PUCRS, Brasil, introduz um recorte de uma pesquisa mais ampla sobre as construções com predicado secundário, mostrando que os estudos semânticos são essenciais para a eficiência de sistemas de Processamento de Linguagem Natural (PLN).

O volume de trabalhos incluídos nesse exemplar, em seus diversos enfoques, reflete os avanços, importância e projeção alcançadas pelas linhas de pesquisa que interligam a Lingüística e a Informática. Destacamos portanto a valiosa contribuição, neste número da revista *Letras de Hoje*, de professores e pesquisadores da Espanha, de Portugal, da Suíça, do Reino Unido, do Uruguai, de nossa Universidade e de outras do Brasil. Agradecemos a todos por ter-nos prestigiado com sua participação, bem como à Editora da PUCRS, na pessoa do Ir. Elvo Clemente, pela oportunidade que nos brindou de publicar este conjunto de estudos dedicados à Lingüística Computacional e suas linhas de investigação.