



PREPARAÇÃO DE CORPUS LINGUÍSTICO PARA INFORMATIZAÇÃO: O LÉXICO HISTÓRICO DO PORTUGUÊS BRASILEIRO

Autoria: Vanderci de Andrade Aguilera - - -

Resumo: O Léxico Histórico do Português Brasileiro – LhisPB – em andamento na Universidade Estadual de Londrina, integra o projeto nacional e interinstitucional Para uma História do Português Brasileiro, sob a coordenação geral do Dr. Ataliba de Castilho e dirigido regionalmente por pesquisadores da área da Linguística Histórica e suas vertentes. O LHisPB, inicialmente composto por docentes e discentes da área da Linguística, conta atualmente com a participação de uma equipe interdisciplinar e interinstitucional de pesquisadores da área da Informática. Para tanto, tornou-se necessário adequar a elaboração do Léxico Histórico às normas dessa disciplina, principalmente no que se refere à preparação dos manuscritos e dos verbetes para serem colocados em portal para acesso universal de usuários. Nesta comunicação, expomos as mudanças implementadas nesse processo de adequação de um corpus linguístico para sua apresentação informatizada. O portal – em desenvolvimento – possibilitará ao pesquisador consultar os verbetes, suas abonações, os fac-símiles dos documentos originais onde ocorrem estes verbetes e respectivas transcrições dos documentos. O banco de dados inicialmente contará com as informações organizadas no projeto LHisPAR – Léxico Histórico do Paraná e gradativamente todas as informações dos projetos dos outros Estados serão acrescentadas. O modo como o corpus está sendo preparado viabiliza a utilização de recursos de Processamento de Linguagem Natural, fato que tem colaborado para que as informações possam também ser acessadas por meio de um robô de conversação. Os testes preliminares com o banco de dados do LHisPAR demonstraram que o portal terá grande agilidade e rapidez na consulta, mesmo considerando os diferentes formatos de arquivos que comporta – textos, imagens e eventuais recursos gráficos.