



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

PROPOSTAS E DESAFIOS DOS NOVOS MEIOS DAS ANTIGAS FONTES: A PRESERVAÇÃO DA MEMÓRIA PELA LINGUÍSTICA DE CORPUS

Cristiane Namiuti-Temponi*
(UESB)

Jorge Viana Santos**
(UESB)

Cândida Mara Brito Leite***
(UESB)

RESUMO

O presente artigo apresenta as propostas e desafios lançados pelo projeto Corpora Digitais Para a História do Português Brasileiro - região Sudoeste da Bahia: Aliança PHPB-Tycho Brahe, que visa: (1) a localização e preservação de documentos antigos provenientes da região Sudoeste do Estado da Bahia, e, (2) o tratamento das fontes coletadas para estudos de natureza linguística.

PALAVRAS-CHAVE: Língua portuguesa, Português histórico, Linguística de Corpus

INTRODUÇÃO

A integração entre o tratamento filológico e o computacional na elaboração de corpus para o estudo do português brasileiro é especialmente importante para

*Doutora em Linguística pela UNICAMP. UESB. PRFPML/GPEL. UESB/CNPq/FAPESB. E-mail: cristianenamiuti@pq.cnpq.br

**Doutor em Linguística pela UNICAMP. UESB. GPEL. UESB/CNPq/FAPESB. E-mail: viana.jorge.viana@gmail.com

***Doutora em Linguística pela UNICAMP. UESB. GPEL. UESB/CNPq/FAPESB. E-mail: candidamara@gmail.com



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

a preservação e divulgação do patrimônio histórico-linguístico do sudoeste baiano, da Bahia e do Brasil.

O trabalho que vem sendo desenvolvido na Universidade Estadual do Sudoeste da Bahia, no âmbito do projeto Corpora Digitais Para a História do Português Brasileiro - região Sudoeste da Bahia: Aliança PHPB-Tycho Brahe (FAPESB Ação Referência) visa à localização e preservação de documentos antigos provenientes da região sudoeste do estado da Bahia. Neste artigo refletiremos sobre alguns desafios atuais da pesquisa em Linguística de Corpus, na sua vertente dedicada à História das Línguas – focalizando, particularmente, a experiência do trabalho com textos históricos da língua portuguesa em meio eletrônico.

Sustentaremos alguns caminhos que já se mostram promissores na exploração da fronteira da pesquisa representada pela união das esferas da **edição de textos, lingüística de corpus, e lingüística diacrônica**. Discutiremos a interação entre essas três práticas a partir das propostas e desafios lançadas no já referido projeto, que envolve, para além de textos impressos de longa data, um desafio extra: o trabalho com fontes primárias manuscritas em estados de degeneração de seu meio original – o papel.

Neste artigo abordaremos a importância da lingüística de corpus e dos Estudos Diacrônicos para a preservação da memória e apresentaremos os desafios metodológicos impostos pela ciência da lingüística, aliada às vertentes tecnológica (fotografia) e computacional (banco de dados, edição e anotação de textos).

A Linguística de Corpus e os Estudos Diacrônicos

A perspectiva histórica sobre a Língua Portuguesa vem alcançando importância central nas últimas décadas, com a retomada do interesse pelo olhar diacrônico e a renovação da relevância teórica dos estudos da mudança lingüística



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

em diferentes quadros [MATTOS SILVA 1988, KATO; ROBERTS 1993, CASTILHO 1998] – um processo que trouxe, como consequência, a intensificação do trabalho com textos antigos no Brasil [MEGALE; CAMBRAIA, 1999]. Para algumas pesquisas realizadas a partir da década de 1990, a junção dos estudos diacrônicos com a prática de edição de textos passa a conferir centralidade para um terceiro campo: a Lingüística de Corpus, compreendida como o trabalho com o dado de língua em meio eletrônico.

Segundo Paixão de Sousa (2006a/b) a questão central que se coloca para este trabalho com textos antigos como fundamentos para estudos lingüísticos no meio eletrônico é a busca por uma abordagem global do texto, em termos conceituais e tecnológicos, que se reflita numa integração entre diferentes planos de análise. De fato, os estudos históricos realizados com base em textos antigos dependem, antes de tudo, da garantia da fidelidade às formas originais dos textos – sendo este o pilar de sustentação que qualquer estudo lingüístico, em qualquer quadro teórico, deve pressupor. No caso dos corpora eletrônicos, esse pressuposto fundamental precisa ser integrado com requerimentos impostos pela vertente computacional e lingüística dos estudos – tais sejam: o arquivo virtual/digital, a confiabilidade e durabilidade do código, a necessidade de quantidade, agilidade e automação no trabalho de organização e seleção de dados.

No entanto, o trabalho com manuscritos antigos traz uma complexidade muitas vezes ausente dos textos autorais impressos (na maioria das vezes de caráter literário, conhecidos e já trabalhados filologicamente), e, para garantir a representatividade e utilidade de documentos antigos recolhidos em arquivos públicos ou privados, de qualquer gênero, é necessário investigar a origem deste documento. Na seção que se segue trataremos desta importante questão.



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

A Origem dos Documentos e sua Representatividade para o Estudo Linguístico

A investigação diacrônica depende dos textos antigos. No entanto, apesar dessa necessidade premente, grande parte dos documentos históricos que sobreviveram ao tempo, seja em arquivos públicos, privados ou pessoais, não está acessível do ponto de vista científico (enquanto corpus manipulável) ao pesquisador, nem tampouco do ponto de vista material ao cidadão. Apesar da dificuldade para acessar o dado que interessa aos estudos da linguagem realizados sob uma perspectiva histórica – seja na diacronia ou na sincronia –, muito tem sido feito desde a proposta de pesquisa anunciada por Castilho (1998) e Mattos e Silva (1998).

Diante da tarefa de constituição de um corpus histórico com dados do Sudoeste baiano, cabe estabelecer diretrizes que guiarão a seleção e, posterior, recolha de dados. Em função dos interesses desse projeto, a seleção dos dados será feita, inicialmente, a partir de fontes judiciais localizadas no fórum de Vitória da Conquista (BA), bem como no Arquivo Municipal da cidade de Rio de Contas (BA). Uma vez delimitado os objetivos iniciais que guiarão a composição do corpus, o primeiro passo corresponde ao levantamento das fontes⁴⁴⁴. Em seguida, dá-se a preparação eletrônica dos documentos para a análise linguística e armazenamento dos dados – etapa que será descrita na seção **Do papel ao texto digital**.

Considerando os pressupostos e objetivos do nosso projeto, que visa a construção do corpora de Documentos de Vitória da Conquista e região (doravante DOVIC), bem como o interesse de estudos futuros relacionados à mudança gramatical, será dada atenção especial, entre as fontes judiciais, aos Inventários e

⁴⁴⁴ O projeto em desenvolvimento já realizou duas expedições de coleta de documentos nas regiões de Feira de Santana e Salvador, foram visitadas as seguintes instituições onde encontramos documentos de Vitória da Conquista e região: CEDOC-UEFS, APEB e IGHB.

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

Testamentos. Segundo Castilho (2009, p. 665), ao lidar com documentos de arquivo é importante considerar o gênero discursivo com que se lida e as partes constitutivas desses gêneros, pois “é desigual o ritmo com que textos e suas partes documentam a mudança linguística”. Em função de característica como essa, após a realização do levantamento das fontes, estas serão organizadas e catalogadas em função de uma cronologia e agrupadas em períodos (a cada 10 ou 25 anos, por exemplo). A definição quanto a essa organização/catalogação só será estabelecida em função do corpus que for possível selecionar e tratar, conforme metodologia adotada pelo DOVIC.

Após efetuada a coleta dos dados, será iniciada a etapa de preparação dos documentos para a análise linguística. Os dados selecionados serão transcritos – inicialmente ou posteriormente, a depender do estado de conservação do documento, bem como das condições de legibilidade – e submetidos à catalogação que, para o caso dos Inventários e Testamentos, será importante localizar informações tais como as seguintes, já observadas por Castilho (2009):

1. Dados possíveis sobre o inventariado (a): profissão, nome dos pais do morto, nome do cônjuge, nome dos filhos, curador dos órfãos, procuradores, local e situação da morte etc.
2. Dados possíveis sobre a redação dos documentos: autoria, nome do juiz/escrivão/tabelião/avaliadores etc.

Como segunda etapa de trabalho, será importante observar a estrutura discursiva dos Inventários e Testamentos com o intuito de identificar as partes constitutivas desses documentos e selecioná-las.

Como bem salienta Tarallo (1990), as fontes escritas que se aproximam da linguagem corrente são as mais indicadas para a pesquisa histórica. Essas fontes,



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

entretanto, não são de fácil localização. À primeira vista, poder-se-ia pensar que as tipologias típicas das fontes judiciais em nada poderiam colaborar quando o que se quer é ter acesso à linguagem corrente de uma época – ao vernáculo – dada a normatização requerida para escritos dessa natureza e ao provável grau de instrução daqueles que foram os responsáveis por tais escritos. Contudo, tanto os estudos históricos quanto aqueles realizados por pesquisadores das equipes do PHPB (CASTILHO, 1998; CASTILHO, 2009) indicam que narrativas de cunho pessoal podem ser encontradas em peças judiciais. Sendo assim, ampliam-se as possibilidades de estudos futuros que poderão ser realizadas a partir de tais fontes, fato que justifica o trabalho que será realizado por essa equipe.

Importância da Investigação Sócio-Histórica para os Estudos Linguísticos

Um dos objetivos do DOVIC é contribuir para recuperar a memória histórica e linguística da região. Diante de tal objetivo, pode-se questionar: qual a vantagem em relacionar História e Linguística?

Alguns estudos linguísticos, especialmente aqueles que são realizados no âmbito da Sociolinguística, asseguram a importância de lidar com a dicotomia história interna/história externa. Na verdade, a correlação de fatores de ordem linguística, sociológica, demográfica, bem como da História do cotidiano, mostra-se necessária tanto para os estudos sociolinguísticos quanto para as pesquisas da Sintaxe, Fonologia, Morfologia, Semântica, entre outras. Evidências dessa afirmação podem ser obtidas ao se verificar as propostas de trabalho apresentadas por Kroch (1989), Massini-Cagliari (1995), Marquilhas (1996), além da proposta inaugurada por Weinreich, Labov e Herzog (1968), que embasa todos os estudos de variação e mudança linguística que partem do pressuposto de que a inter-

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

relação entre língua e sociedade pode (e deve) ocupar a pauta dos estudos linguísticos.

Ao tratar das perspectivas para a História Social do português brasileiro, Ramos (1998) afirma:

Se aceitarmos as variações/mudanças como resultantes de contato e se assumirmos que o contato tem como consequência a produção de enunciados representativos de gramáticas distintas, poderemos buscar, na variação sintática presente no corpus selecionado, evidências da presença das especificidades das gramáticas presentes na comunidade de fala.

A partir desse ponto de vista, torna-se relevante recompor o caminho de cada variante no contexto sócio-histórico, de modo a identificar em que momento e em que lugar geográfico e social a variante ocorreu, o que equivale a responder à seguinte questão: com que gramática se deu o contato, quando, onde, como? (RAMOS, 1998, p. 157 – grifo nosso).

A observação de Ramos (1998) é pertinente e, apesar da autora se referir a aspectos sintáticos, a adequação da proposta contida nessa afirmação pode se estender aos demais aspectos da gramática de uma língua.

Quanto ao DOVIC, interessa, principalmente nesse primeiro momento de constituição dos corpora, a tarefa de procurar conhecer como se escrevia o português em espaços comunicativos da região Sudoeste. Para ter acesso a esse tipo de conhecimento, como bem registra Simões e Kewitz (2009, p. 699), faz-se necessário lidar com a história social de uma comunidade e/ou região. Nesse trabalho, como salientam os autores, é necessário fazer, e tentar responder, perguntas como: “quem falou/escreveu esse texto?, onde viviam?, como viviam?, como era sua movimentação?, que relações sociais estabeleciam os falantes/escritores?”

Como já foi dito, as investigações da sócio-história do Português, bem como de qualquer outra língua, é necessária e devem se somar aos resultados de



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

pesquisas eminentemente sincrônicas para que as interpretações acerca de aspectos linguísticos do português do Brasil possam ser melhor esclarecidas. Essas são, portanto, as diretrizes que guiarão a condução das investigações desse projeto de pesquisa.

Voltemos agora para a vertente tecnológica do nosso desafio. Na seção que se segue trataremos do transporte (conversão) de documentos escritos (manuscritos ou impressos) para o meio digital, considerando, primeiramente, a digitalização de documentos escritos empregando a fotografia.

Do Papel ao Texto Digital

Seguindo Santos (2010a; 2010b), colocamos a seguinte questão: Qual a viabilidade do uso da fotografia para a captação de documentos para compor corpora digitais, visando estudos linguísticos?, pergunta essa que, por recorte metodológico, leva a outra: “Quais são as complexidades intervenientes no processo de digitalização de documentos físicos escritos?”.

Defenderemos, a seguinte hipótese: Considerando que o processo de digitalização de arquivos de documentos físicos envolve pelo menos três complexidades: 1) a do objeto, 2) a da Fotografia, 3) a da Imagem digital fotográfica, defendemos que a Fotografia apresenta-se como forma altamente viável e produtora de digitalização, permitindo à Linguística acessar imagetivamente, de modo confiável, um objeto físico (o documento) não disponível no local da pesquisa.

Para tanto, o objetivo precípua aqui será apresentar propedeuticamente as três complexidades que envolvem a passagem de documentos físicos manuscritos ou impressos para o meio digital, com ênfase na complexidade 1, a do objeto.



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

A complexidade do objeto: o documento a se tornar imagem digital

A primeira complexidade que se apresenta a um pesquisador em Linguística em busca de um corpus documental é, sem dúvida, o arquivo físico que armazena e, em tese, estão organizados os documentos a serem coletados. Nesse sentido, há pelo menos três aspectos a considerar: a) o acesso; b) a forma; e c) a fragilidade e/ou raridade dos documentos.

O acesso

Por se tratar de arquivos normalmente institucionais, ou pessoais, nem sempre se tem **acesso** franqueado, sobretudo por longo tempo ou por repetidas vezes. Ou seja: quando se consegue acesso a este tipo de arquivo, o ideal será colher o maior número de informações documentais possível, pois o que pode não interessar imediatamente a um pesquisador, pode e muito ser até fundamental para outro (a exemplo da história que cerca o próprio documento arquivado). Então, é necessário se colocar na posição de um pesquisador formador de corpora (doravante PFC) e não apenas de um Pesquisador pragmático, aquele que se contenta apenas com seu objeto de pesquisa.

Além disso, não esqueçamos que o ideal é obter um documento digital que nos garanta a não necessidade de voltar ao original, pois além de um acesso à maioria dos arquivos envolver, por assim dizer, uma certa negociação política⁴⁴⁵, há no mais das vezes dificuldade do deslocamento geográfico do pesquisador.

⁴⁴⁵ Por exemplo, envio de ofícios explicativos sobre o motivo da consulta; agendamentos; compromissos formais sobre privacidade de informações, etc.

A forma

Um segundo fator de complexidade que um PFC se depara – e isto vai se refletir diretamente na complexidade 2 (Fotografia) – é a **forma** do objeto arquivado: Ele é físico, tridimensional, não padronizado e apresenta-se numa organização que, necessariamente, precisará ser desfeita (ou melhor: virtualizada) na etapa de digitalização fotográfica. Um livro, que tem páginas (às vezes sem identificação específica) todas presas e por isso automaticamente identificadas, será fotografado página por página, nem sempre na mesma ordem. Registrar essa ordem antes, durante a Fotografia e na própria imagem (fotografia) é, portanto, fundamental, sob pena de, depois, simplesmente tornar-se impossível a recuperação de dados organizacionais apenas pela imagem de um texto. Quer dizer: a digitalização pode ser improdente caso não seja acompanhada da devida identificação (recuperável) da organização original dos textos.

A fragilidade e/ou raridade

Enfim, um terceiro fator complexo envolvido no processo de digitalização, é a **fragilidade e/ou raridade dos documentos**. Isto impõe ao PFC muito cuidado no manuseio para evitar danos à fonte. Isto inclui o conhecimento e uso de certas técnicas e equipamentos⁴⁴⁶ que, ao mesmo tempo em que minimizam a possibilidade de dano, garante um registro fidedigno.

Diante dessas três complexidades do objeto, um pesquisador que queira passar à categoria de PFC, deve ter em conta que, sendo elas constitutivas de um tipo de objeto (o documento em papel), o fato de preparar-se para lidar com elas na formação de corpora⁴⁴⁷, trará, dentre outras, vantagens como: a) uma nova

⁴⁴⁶ Por limite metodológico, discutimos esse assunto em outros textos. Ver Santos (2010a; 2010b).

⁴⁴⁷ E, lembremos, com as outras duas complexidades: a da Fotografia, e a da imagem digital.



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

forma de acesso da imagem: a visual, que, ao mesmo tempo em que, sendo digital, permite sua veiculação e virtual, pode com isso democratizar a consulta de fontes que, sendo únicas (e em papel), só podem ser vistas/consultadas in locu; b) a manipulação visual eletrônica do texto⁴⁴⁸ (ampliação, contraste, cor...); e – muito importante – a reprodutibilidade: um documento adequadamente digitalizado com auxílio da Fotografia pode ser reproduzido eletronicamente, o que contribui ao mesmo tempo para a preservação do original quanto para sua consulta e análise por pesquisadores de áreas as mais diversas.

A Vertente Computacional

Na exploração da fronteira da pesquisa representada pela união das esferas (i) da edição de textos, (ii) da lingüística de corpus, e (iii) da lingüística diacrônica temos nossa vertente computacional. A interação entre essas três práticas está bem exemplificada em nossa experiência anterior com a construção do "Corpus Anotado do Português Histórico Tycho Brahe" (www.tycho.iel.unicamp.br/corpus). Conforme apresenta Namiuti (2010), a construção do corpus Tycho Brahe envolve três etapas computacionais centrais: (i) a codificação e edição dos textos ortograficamente transcritos em linguagem XML pelo processador de texto EDICTOR; (ii) a etiquetagem morfológica automática pelo programa TAGGER; (iii) anotação automática da sintaxe dos textos pelo PARSER.⁴⁴⁹

O processo atual de trabalho com os textos desse corpus e que almejamos trazer para o DOVIC fundamenta-se na aplicação de tecnologias de anotação

⁴⁴⁸ Pode-se, por exemplo, ampliá-lo, modificar o contraste fundo/letra, alterar-se a cor, e outros recursos disponíveis para tratamento de imagem digitais, não aplicáveis, claro, ao documento em papel.

⁴⁴⁹ Ferramentas desenvolvidas e aplicadas no âmbito do projeto temático Padrões Rítmicos Fixação de Parâmetros e Mudança Lingüística/FAPESP-UNICAMP (Galves 2004-2010), para maiores detalhes remeto ao site do projeto: <http://www.tycho.iel.unicamp.br>



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

eletrônica de textos (particularmente, a Linguagem de Marcação Extensível, ou XML (W3C, 1997; TRIPPEL; PAIXÃO DE SOUSA, 2006) e aos procedimentos tradicionalmente seguidos pelas edições especializadas. A conjunção destes dois campos representa uma experiência interessante para o desenvolvimento do meio eletrônico como instrumento para diferentes campos que tomam o texto como objeto de estudo. A equipe Tycho Brahe acredita que o trabalho de codificação das diferentes camadas de informação em um texto (de sua organização gráfica a sua estrutura lingüística) pode abrir novos caminhos para a abordagem do texto como um sistema de raízes que pode ser escavado nele próprio (CARVALHO, 2003) – contribuindo, desta forma, na exploração das fronteiras de pesquisa delineadas pela união entre a Lingüística de Corpus e o estudo Histórico das Línguas.

CONCLUSÕES

A superação dos desafios do projeto aqui noticiado, com a construção do corpus DOVIC, possibilitará, pioneiramente, em termos de Bahia, o princípio da possibilidade de utilização das edições eletronicamente manipuláveis de manuscritos históricos para fins de pesquisas acadêmico-científicas de diversas áreas Além da Linguística, os documentos digitalizados poderão, pela sua natureza, servir para historiadores, sociólogos, antropólogos, para citarmos apenas alguns. Eis um avanço que, sem dúvida, compensa a superação das complexidades da digitalização.



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

REFERÊNCIAS:

- CARVALHO, R. B. S. A Filologia e seu Objeto: Diferentes perspectivas de estudo. In: **Philologus** - Revista do Círculo Fluminense de Estudos Filológicos e Lingüísticos, ano 9, n. 26, Rio de Janeiro, 2003.
- CASTILHO, A. T. de. Projeto de História do português de São Paulo. In: _____. (Org.) **Para a História do Português Brasileiro**. São Paulo: Humanitas Publicações/FFLCH/USP, v. I: primeiras ideias, p. 61-76, 1998.
- CASTILHO, C. M. M. de. Estrutura discursiva dos inventários e testamentos de São Paulo (séculos XVI-XVII). In: CASTILHO, A. T. (Org.). **Para a História do Português Paulista**. Série Estudos, vol 1. Campinas: FAPESP/ Setor de Publicação do Instituto de Estudos da Linguagem da Unicamp, p. 665-693, 2009.
- KATO, M. A.; ROBERTS, I. (Org.) **Português brasileiro: uma viagem Diacrônica**. Campinas: Editora da Unicamp, 1993.
- KROCH, A. Reflexes of Grammar in patterns of language change. **Language Variation and Change** 1,1989. p. 199-244.
- MARQUILHAS, M. R. B. **A faculdade das letras: leitura e escrita em Portugal no século XVII**. (Tese de doutorado). Lisboa: Universidade de Lisboa, 1996.
- MASSINI-CAGLIARI, G. **Cantigas de amigo: do ritmo poético ao linguístico: um estudo do percurso histórico da acentuação em português**. (Tese de Doutorado em Linguística). Campinas: IEL/UNICAMP, 1995.
- MATTOS e SILVA, R. V. Fluxo e refluxo: uma retrospectiva da lingüística histórica no Brasil. **D.E.L.T.A.**, 4.1: 85-113. São Paulo, 1988.
- _____. Idéias para a História do português brasileiro: fragmentos para uma composição posterior. In: _____. (Org.). **Para a História do Português Brasileiro**. São Paulo: Humanitas Publicações/FFLCH/USP: FAPESP, v.II: primeiros estudos, p. 21-52, 1998.
- MEGALE, H.; CAMBRAIA, C. N. Filologia Portuguesa no Brasil. In: **D.E.L.T.A.**, vol. 15, número especial:1:22. São Paulo, 1999.
- NAMIUTI-TEMPONI, C. As possibilidades de processamento automático nos novos meios das antigas fontes: a anotação e a busca automática do Corpus Tycho Brahe. **Conferência ministrada na I Oficina de Lingüística de Corpus da Bahia (UEFS, UESB, UFBA)**. Feira de Santana: UEFS, 2010.
- PAIXÃO DE SOUSA, M. C. Memórias do Texto. In: **Revista Texto Digital**, ISSN 1807-9288, ano 2 n.1 2006a.<<http://www.textodigital.ufsc.br/num02/paixao.htm>>.
- _____. Hypertext: conceptual and methodological frontiers. **Comunicação ao Seminário Internacional Literaturas: del Texto al Hipertexto**. Faculdade de Filología, Universidade Complutense de Madrid. Madri, 22 de Setembro, 2006b.



ISSN: 2175-5493

IX COLÓQUIO DO MUSEU PEDAGÓGICO

5 a 7 de outubro de 2011

RAMOS, J. História Social do Português Brasileiro: perspectivas. In: MATTOS e SILVA, R. V. (Org.). **Para a História do Português Brasileiro**. São Paulo: Humanitas Publicações/FFLCH/USP: FAPESP, v.II: primeiros estudos, p. 153-167, 1998.

SANTOS, J. V. Apresentação de meios para o transporte: digitalização de documentos manuscritos e impressos. **Conferência ministrada na I Oficina de Lingüística de Corpus da Bahia (UEFS, UESB, UFBA)**. Feira de Santana: UEFS, 2010a.

_____. Técnicas de transporte do texto manuscrito para o meio digital. Conferência **ministrada na I Oficina de Lingüística de Corpus da Bahia (UEFS, UESB, UFBA)**. Feira de Santana: UEFS, 2010b.

SIMÕES, J. da S.; KEWITZ, V. Norma linguística, história social, contatos linguísticos e tradições discursivas: transformando encruzilhadas em novas caminhos para a constituição de corpora diacrônicos. In.: CASTILHO, A. T (Org.). **Para a História do Português Paulista**. Série Estudos, vol 1. Campinas: FAPESP/ Setor de Publicação do Instituto de Estudos da Linguagem da Unicamp, p. 699-720, 2009.

TARALLO, F. **A pesquisa sociolinguística**. 3 ed. São Paulo: Ática, 1990.

WEINREICH, U., W. LABOV; M. HERZOG. Empirical foundations for a theory of language change. In.: LEHMANN; MALKIEL: **Directions for Historical Linguistics**. University of Texas, Austin, 1968.

TRIPPEL, T.; PAIXÃO DE SOUSA, M. C. Metadata and XML standards at work: a corpus repository of Historical Portuguese texts. In: **Papers from the V International Conference on Language Resources and Evaluation**. LREC, 2006.

W3C. "Extensible Markup Language", 1997. Disponível em: <<http://www.w3.org/XML>>