

---

# RELATÓRIO DE PÓS- DOUTORADO

**Proposta de um sistema  
de edições eletrônicas e  
seus aspectos  
tecnológicos:**

---

*a construção de um Piloto de  
Corpus Eletrônico a ser implantado  
na UEFS para o estudo da língua  
portuguesa no semi-árido baiano  
(séculos XVII-XXI)*

---

Zenaide de Oliveira Novais  
Carneiro

Supervisão/Responsável  
Institucional:

Prof<sup>a</sup> Dr<sup>a</sup> Charlotte Marie  
Chambelland Galves

---

Universidade Estadual de Campinas  
Instituto de Estudos da Linguagem

**PROPOSTA DE UM SISTEMA DE EDIÇÕES ELETRÔNICAS E SEUS ASPECTOS TECNOLÓGICOS: a construção de um *Piloto de Corpus Eletrônico* a ser implantado na UEFS para o estudo da língua portuguesa no semi-árido baiano (séculos XVII-XXI)**

**RELATÓRIO DE PÓS-DOCTORADO**

Zenaide de Oliveira Novais Carneiro Bolsista  
Bolsista FAPESB-Ba/1648/2009

**Supervisor/Responsável Institucional**  
Prof<sup>a</sup> Dr<sup>a</sup> Charlotte Marie Chambelland Galves

**Período**  
05 de março de 2009 a 31 de julho de 2010

## **Agradecimentos**

À Charlotte M. C. Galves, Antonio Almeida Carneiro, Pablo P.F de Faria, Cyntia Yano, Antonio Galves e aos colegas do Departamento de Letras e Artes da Universidade Estadual de Feira de Santana/UEFS e a equipe do Corpus Histórico do Português Tycho Brahe, IEL/Unicamp, que contribuíram de diferentes maneiras com esse estágio. E à Fundação de Amparo à Pesquisa da Bahia/Fapesb pelo financiamento da pesquisa.

## Sumário

Agradecimentos	03
Sumário	04
1. Apresentação	05
2. Objetivos gerais e específicos propostos no projeto	06
2.1 Gerais	06
2.2 Específicos	06
3. Resultados Obtidos	07
4. Infra-estrutura	09
5. Questões teórico-metodológicas	10
5.1 Da edição de documentos nos moldes filológicos tradicionais	10
5.2 Sobre as questões teóricas do conceito de mudança linguística e sobre a análise de fenômenos linguísticos	11
5.3 Da formação de bancos digitais eletrônicos	11
6. Produção Tecnológica	12
7. Serviços	12
8. Capacitação de Recursos Humanos	12
9. Difusão	13
10. Outros	13
11. Parceria Institucional	13
12. Impactos	14
12.1 Impacto Tecnológico	14
12.2 Impacto Econômico	15
12.3 Impacto Social	15
13 Produção bibliográfica gerada através do desenvolvimento da pesquisa	16
14. Comentários gerais finais e perspectivas	17

## 1. Apresentação

O projeto de pós-doutorado, realizado entre agosto de 2009 e agosto de 2010, junto ao *Corpus Histórico do Português Tycho Brahe* (CNPq - Edital Universal 2007 – Dez/2007 - Dez/2009) e do projeto interdisciplinar *Padrões Rítmicos, Fixação de Parâmetros e Mudança Lingüística - Fase II* (<http://www.tycho.iel.unicamp.br/~tycho/>), no Instituto de Linguagem da Unicamp, sob a supervisão da Prof<sup>a</sup> Dr<sup>a</sup> Charlotte C. Galves, teve como propósito fundamental conhecer novas metodologias para formação de grandes bancos de dados com processamento de texto para fins lingüísticos.

A partir da experiência adquirida, a idéia central, como o próprio nome do projeto sugere, consistiu em desenvolver um *Piloto de um Corpus Digital Eletrônico* nos moldes do *Corpus Histórico do Português Tycho Brahe*, a ser aplicado em um corpus constituído a partir de um extenso banco de textos feito durante mais de 10 anos no âmbito do projeto *Vozes do sertão em dados: história, povos e formação do português brasileiro* (CNPq 401433/2009-9), acessível em <http://www.uefs.br/nelp/zenaide-nelp/index.html>, subprojeto de um plano de trabalho iniciado, em 1997, por Ilza Maria de Oliveira Ribeiro/UEFS, no qual atuamos como vice-coordenadora (cf. histórico em [www.uefs.br/dohs/historico.html](http://www.uefs.br/dohs/historico.html))<sup>1</sup>.

Esse estágio buscou também consolidar uma parceria para troca de experiências, através da associação entre edição tradicional e edição computacional na construção desse piloto de *Corpus digital-eletrônico*. Também foi proposto a aplicação e uso da ferramenta integrada de anotação de corpus denominada *e-dictor* (<http://oncoto.dyndns.org:44880/projects/edictor>), desenvolvida como parte de desdobramentos do *Corpus Histórico do Português Tycho Brahe*, por Kepler e Paixão e Souza (2004) e implementada por Fábio Kepler (USP) e Pablo Faria (UNICAMP), a partir de 2009. Essa ferramenta permite não só a edição em XML de forma fácil e rápida como também a aplicação de outras ferramentas desenvolvidas pelo projeto, tais como: correção de etiquetagem automática e correção do *parser* automático, entre outras.

O projeto foi bem sucedido na medida em que atingiu as principais metas propostas, como explicitado a seguir:

- 1) Implementação do projeto Piloto na UEFS, com a aprovação do projeto CE-DOHS – Corpus Eletrônico de Documentos Históricos do Sertão (FAPESB 5566/2010), uma interface, em formato eletrônico, do banco digital DOHS – Documentos Históricos do Sertão ([www.uefs.br/dohs](http://www.uefs.br/dohs)), que faz parte do projeto Vozes do Sertão em Dados: história, povos e formação do português brasileiro (CNPq, 401433/2009-9) ([www.uefs.br/nelp](http://www.uefs.br/nelp)).

---

<sup>1</sup> O projeto *Vozes do Sertão em Dados* é filiado ao *Programa para a história do português/PROHPOR* (cf. [www.prohpor.ufba.br](http://www.prohpor.ufba.br)), sediado na UFBA, no arco temporal que compreende a história do português brasileiro, desenvolvendo agenda de pesquisa conjuntamente com o *Projeto Nacional Para a História do Português (PHPB)*. Ao projeto *Vozes do Sertão em Dados* coube/cabe fazer o levantamento, a compilação, a edição de documentação e os estudos lingüísticos de uma grande área da região semi-árida baiana. Atualmente, também faz parceria com o *Memória Conquistense: recuperação de documentos oitocentistas na implementação de um corpus digital*, coordenado por Cristiane Namiuti-Temponi, Jorge Santos e Cândida Mara, na UESB, submetido a FAPESB com o nome *Corpora digitais para a história do português brasileiro: Documentos históricos da região do Sudoeste da Bahia, aliança PHPB-TYCHO BRAHE* (FAPESB - 6171/ 2010), sob a coordenação de Jorge Viana Santos, e dos pesquisadores Cristiani Namiuti-Temponi e Cândida Brito, na UESB.

- 2) Convênio Guarda-Chuva entre a UEFS e a Unicamp e o Termo Aditivo de Transferência mútua de tecnologia, através do Corpus Histórico do Português Tycho Brahe ([www.tycho.iel.unicamp.br](http://www.tycho.iel.unicamp.br)), em tramitação na Unicamp.
- 3) Co-cooperação com projetos regionais. Isso tem sido feito através da parceria com o projeto *Corpora digitais para a história do português brasileiro: Documentos históricos da região do Sudoeste da Bahia, aliança PHPB-TYCHO BRAHE* (FAPESB 6171/ 2010), sob a coordenação de Jorge Viana Santos, da UESB.
- 4) Aplicação e uso do corpus na graduação e pós-graduação da UEFS, na área da Linguística Histórica.

## 2. Objetivos gerais e específicos propostos no projeto

### 2.1 Gerais

1. Implementação de um *corpus* em formato de texto computacionalmente manipulável, em linguagem XML, com o objetivo de permitir a recuperação de informações gráficas de cunho filológico em documentos originais e possibilitar o uso de mecanismos para geração de diferentes versões, acessíveis em processamentos de buscas automáticas, atingindo e satisfazendo públicos especializados e usuários diversos;
2. Permitir a geração de bases anotadas, morfológica e sintaticamente, para análise linguística;
3. Após o desenvolvimento de um Piloto, o objetivo seria implantar o banco com textos do “Projeto Vozes do Sertão em Dados”, em versão eletrônica manipulável, para uso na pós-graduação e, também, para ser disponibilizado na rede mundial de computadores;
4. Como sub-objetivo ligado ao objetivo 2, gerar léxicos das edições em XML, catálogos dos documentos, edições variadas mencionadas (conservadoras, modernizadas, etc.), além de outros subprodutos gerados pela transformação XML, nos moldes do projeto Corpus Histórico do Português Tycho Brahe.

### 2.2 Específicos

Esses objetivos estão ligados as metas e aos posteriores resultados do projeto, a saber:

1. Acompanhar o início da alimentação do *Corpus Histórico do Português Tycho Brahe*, no que se refere ao material editado para análise do volume 2, da minha tese de doutorado intitulada *Cartas brasileiras (1809-1904): um estudo linguístico-filológico* de forma a que venha servir de material para ser pensado um *corpus* piloto (cf. itens 45-48 <http://www.tycho.iel.unicamp.br/~tycho/corpus/index.htm>);
2. Estabelecimento de um *corpus* inicial formado por cartas escritas por brasileiros cultos nascidos entre fins do século XVIII e por brasileiros cultos nascidos até o terceiro quartel do século XIX, a ser editado em XML;
3. Codificação XML inicial de 185 cartas a partir de um *corpus* editado em Word, de acordo com as normas da tradição filológica para uso linguístico histórico;
4. Uso do e-dictor para controle do texto base bruto e transformação XML;
5. Testes da codificação XML;

6. Anotação morfológica através da equipe técnica do *Corpus Histórico do Português Tycho Brahe*;
7. Estabelecimento do formato da base de dados e do sítio-web do projeto piloto final;
8. Teste do projeto Piloto de forma que venha ser possível a alimentação de textos até atingir milhões de palavras em projeto contínuo na UEFS com metodologia do *Corpus Histórico do Português Tycho Brahe/Unicamp*.

### 3. Resultados Obtidos

Os resultados previstos nos objetivos gerais (1, 2, 3 e 4) e os específicos (1-7) foram alcançados. Esse Piloto que está no sítio no *Corpus Histórico do Português Tycho Brahe*, onde desenvolvemos o estágio de pós-doutoramento, pode ser acessado no site da UEFS no seguinte endereço: [www.uefs.br/dohs](http://www.uefs.br/dohs), do DOHS – *Documentos Históricos do Sertão*, onde há uma entrada intitulada, “Corpus (XML) - Acesso ao subconjunto de textos do corpus codificados e editados em XML, através do aplicativo E-Dictor - FAPESB/1658/2009”.

Dessa forma, transformamos o *Piloto de Corpus Digital Eletrônico* em um projeto de longo prazo, o que efetivamente já começamos a fazer. Esse projeto que em si já é um resultado do pós-doutorado, apresentado à FAPESB e contemplado no edital Referência/2010. O projeto tem o nome de *CE- DOHS – Corpus Eletrônico de Documentos Históricos do Sertão*. De forma resumida, os resultados obtidos a partir do que foi originalmente proposto – “Iniciar um plano de *Corpus Digital Eletrônico* permanente na UEFS, a partir da metodologia de edições eletrônicas em XML do *Corpus Histórico do Português Tycho Brahe*”, no que resultou no projeto CE-DOHS, se encontra em franco desenvolvimento. Ou seja, o que vamos fazer é transformar o banco de textos DOHS<sup>2</sup>, subprojeto Vozes do Sertão (CNPq), já em processo de publicação em formato digital, em um corpus eletrônico.

O CE-DOHS, portanto, é uma interface eletrônica do banco DOHS nos moldes do que nos propomos a fazer no Piloto do projeto de pós-doutoramento. O CE-DOHS vai inicialmente editar em XML (2010-2012) os documentos de *Cartas brasileiras (1808-2000): coletânea de fontes para o estudo do português brasileiro*<sup>3</sup>, composta com a edição em cd (no prelo), FAPESB -1493/2010, a saber:

---

<sup>2</sup> O portal “Dohs” – Documentos históricos do Sertão ([www.uefs.br/dohs](http://www.uefs.br/dohs)), sob a minha responsabilidade, se configura como um espaço de acesso aos acervos em dois formatos, *pdf (adobe)* e em XML, e dos projetos desenvolvidos no “Núcleo de Estudos de Língua Portuguesa/NELP/UEFS” ([www.uefs.br/nelp](http://www.uefs.br/nelp)), do qual participo como co-coordenadora e como pesquisadora, um núcleo que vem se firmando como centro interdisciplinar de acervos lingüístico-históricos raros e de pesquisas sobre o português brasileiro. Também nesse núcleo é possível acessar o projeto “Vozes do Sertão em dados (CNPq. 401433/2009-9) ([www.uefs.br/nelp](http://www.uefs.br/nelp)). Esse projeto, como vimos, teve como objetivo fazer um balanço de um amplo conjunto de textos inéditos compilado em toda grande área do semi-árido baiano e compor séries temáticas, como é o caso do *corpus* publicado como *Cartas brasileiras (1808-2008): coletânea de fontes para o estudo do português brasileiro*, aprovada pela FAPESB -1493/2010 no Edital Publicação de 2010.

<sup>3</sup> A proposta de organização da coleção, nos termos colocados, foi pensada a partir de dois artigos de Rosa Virgínia Mattos e Silva, “Idéias para a história do português brasileiro: fragmentos para uma composição posterior” e “De fontes sócio-históricas para a história social lingüística do Brasil: em busca de indícios” apresentados, respectivamente, nos dois primeiros encontros do *Projeto Para a História do Português Brasileiro* (PHPB), em 1997 e 1998, publicados posteriormente nas Atas, que em última instância se propõe a construir a história do PB a partir de dois pólos: história do português culto e história do português popular.

- Livro 1.** *Cartas brasileiras (1808-1904): coletânea de fontes para o estudo do português brasileiro*<sup>4</sup> com 3 cds: cartas avulsas para vários destinatários; cartas para Severino Vieira e cartas para o barão de Jeremoabo.
- Livro 2.** *Cartas brasileiras (1902-2000): coletânea de fontes para o estudo do português brasileiro*, com 03 cds, cartas para Dantas Jr.; cartas de mulheres da família Costa Pinto e cartas pessoais de Valente, Bahia.
- Livro 3.** *Cartas brasileiras populares (1975-1992): coletânea de fontes para o estudo do português brasileiro*, em versão apenas impressa.

Dessa forma, buscaremos dar uma nova face ao banco de textos do DOHS. Esse banco eletrônico, o CE-DOHS, vai ser desenvolvido em parceria com a Unicamp, conforme será especificado abaixo no item que trata de “parcerias institucionais”. Esse banco tem perspectiva de crescimento, já que o DOHS tem um grande acervo<sup>5</sup> e vem agregando uma equipe formada por estudantes da pós-graduação (Mestrado em Estudos Lingüísticos, recentemente implantado, com uma linha em Lingüística Histórica, justamente a partir desses resultados conquistados nos últimos dez anos), assim como por estudantes da graduação (I.C.) e voluntários), entre outros, e por pesquisadores integrantes do projeto (cf. [www.uefs.br/dohs/equipe.html](http://www.uefs.br/dohs/equipe.html)).

Outro resultado importante desse projeto é o livro que está em fase de planejamento. Trata-se do volume composto por estudos sobre o português culto, como descrito no livro 1 da *Coleção Diacrônica (Volume 1: cartas de brasileiros cultos do século XIX)*. Esse material digital foi a base para o piloto do pós-doutorado (cf. <http://www.uefs.br/dohs/va004.html>) e já pode ser manipulado. Dessa forma, esperamos organizar o livro já usando as buscas no *Corpusearch* <http://www.uefs.br/dohs/va004.html>).

O livro tem o seguinte título provisório:

CARNEIRO, Zenaide O. N. (Org.); GALVES, Charlotte Marie C. (Org.); LOBO, T. C. F. (Org.); *Gramáticas em conflito: estudos sobre o português escrito por cultos no Brasil dos séculos XIX e XX*. (em preparação).

Cabe destacar que esse livro, a ser publicado em 2011, serve, portanto, como um duplo produto. De um lado, enquanto análise lingüística, esse livro é um produto do *Projeto Vozes do Sertão em Dados: história, povos e formação do português brasileiro* ([www.uefs.br/nelp](http://www.uefs.br/nelp)), financiado pelo CNPq/2009). De outro, como dito acima, é um livro já gerado a partir de um banco piloto em edições XML, já anotado tanto morfológica quanto sintaticamente, um dos primeiros a ser feito nesse formato no Brasil. As buscas automáticas, que substituem as manuais, para estudos lingüísticos, temas do livro, usam o produto deste projeto de pós-doutoramento, que por sua vez servirá de parâmetro para outro livro com materiais de populares, seguindo a premissa da *Coleção Diacrônica* (culto *versus* popular), fruto do projeto Vozes do Sertão em Dados. A configuração eletrônica desse material,

---

<sup>4</sup> A proposta de organização da coleção, nos termos colocados, foi pensada a partir de dois artigos de Rosa Virgínia Mattos e Silva, “Idéias para a história do português brasileiro: fragmentos para uma composição posterior” e “De fontes sócio-históricas para a história social lingüística do Brasil: em busca de indícios” apresentados, respectivamente, nos dois primeiros encontros do *Projeto Para a História do Português Brasileiro* (PHPB), em 1997 e 1998, publicados posteriormente nas Atas, que em última instância se propõe a construir a história do PB, a partir de dois pólos: história do português culto e história do português popular.

<sup>5</sup> Para uma visão do conjunto de documentos, cf. <http://www.uefs.br/dohs/historico.html>.

piloto, também se constitui no primeiro material do CE-DOHS, a ser feito em parceria com o *Corpus Histórico do Português Tycho Brahe*.

Sobre o resultado proposto “Colaborar na metodologia de aplicação do e-dictor (cf <http://oncoto.dyndns.org:44880/projects/edictor>) para troca mútua de experiências, a descrição de como se deram as discussões propostas estão detalhadas no relatório parcial (*em anexo*). De modo geral, podemos dizer que o E-dictor, pensado a partir de desdobramento da *I Oficina do Projeto Tycho Brahe*, da Unicamp, promovida na UFBA, sob a coordenação de Tânia Lobo, com a participação efetiva da equipe da UEFS, vem sendo constantemente atualizado, como comumente ocorre com programas computacionais. Parte da atualização resulta da parceria/interface para busca de adequações entre edições filológicas tradicionais e edições eletrônicas.

Sobre o proposto: “disponibilizar a transferência das metodologias para outros bancos regionais”, já estabelecemos parceria com um projeto semelhante, o que foi mencionado na nota 1, o projeto *Memória Conquistense: recuperação de documentos oitocentistas na implementação de um corpus digital*, coordenado por Cristiane Namiuti-Temponi e Jorge Santos, da UESB e no qual atuamos como consultora. Também, estamos com co-participação com a formação de um banco digital com UFBA, através do Prohpor, mas ainda em fase de elaboração por Klebson Oliveira e Tânia Lobo.

Sobre o proposto: “subsidiar o *Corpus Digital Eletrônico* de forma que acompanhe os avanços científicos, através de avaliações com outros grupos” foi já estabelecido. Uma nova forma de acesso ao DOHS, composto por ampla documentação dos séculos XVIII-XX, digitalizada, classificada e já com tratamento editorial filológico diplomático-interpretativo, começou a tomar forma, mostrando o resultado do projeto de pós-doutorado em associação com o “Corpus Histórico do Português Tycho Brahe” ([www.tycho.iel.unicamp.br/~tycho](http://www.tycho.iel.unicamp.br/~tycho)), desenvolvido na Unicamp, sob a coordenação da Prof<sup>a</sup> Dr<sup>a</sup> Charlotte C. Galves, iniciada com este projeto de pós-doutoramento, que gerou o CE-DOHS. Esse projeto se constitui em um meio para essa atualização. Portanto, o portal DOHS ([www.uefs.br/dohs](http://www.uefs.br/dohs)), pensado de forma a explorar o potencial do nosso banco de dados, vai assumir uma nova face, o CE-DOHS, em parceria com a Unicamp, via meios institucionais (cf. parcerias institucionais). O CE-DOHS será preparado para exploração automática, através de tecnologias de processamento, como dito. Assim, acreditamos que esse estágio abriu uma nova linha de pesquisa usando tecnologias para acesso a banco de dados. Essa parceria vem mostrando como a combinação de competências é crucial para a construção de um grande *corpus* de textos escritos em português no Brasil.

Dessa forma, os objetivos subseqüentes foram atendidos, como “disponibilizar formas de acesso a produção automática propiciadas pela transformação XML, tais como léxico das edições, catálogos, etc.; “produzir materiais através do uso direto do Corpus Digital Eletrônico na rede mundial de computadores para levantamento extensivo de dados de forma confiável e rápida para uso acadêmico. Sem dúvida, o livro já citado, assim como os projetos individuais de ex-bolsistas I.C, mostram a “elevação do número de artigos dos membros participantes. Isso, como dito, também irá “subsidiar, também, o programa de pós-graduação *stricto sensu* em Estudos Lingüísticos da UEFS, já citado.

#### 4. Infra-estrutura

Na UEFS, implantamos uma sala com recursos do CNPq disponibilizados para o projeto *Vozes do Sertão em Dados* que também vem servindo para a estrutura inicial para funcionamento do CE-DOHS. Os materiais são os seguintes:

01 computador de mesa com um tela de 29” que facilita a leitura de manuscritos;

01 laptop HP para trabalhos de campo;  
01 impressora HP laser  
01 HD externo com capacidade de 1 tera para *backup* do banco de dados;  
01 armário com divisórias e chaves para armazenamento dos documentos originais e digitalizados;  
01 máquina fotográfica semi-profissional;  
02 pontos de Internet  
02 cadeiras  
01 banco  
01 armário com livros e teses.

A UEFS disponibilizou um espaço físico localizado no MT 25 no NELP. Esse espaço será alargado a partir de recursos a serem solicitados ao FINEP/2011, com a construção de mais salas junto ao programa de pós-graduação.

Na parte de implementação do cronograma, estamos elaborando um Seminário Interdisciplinar e Interinstitucional (UEFS, UESB, UFBA e Unicamp) para projeção de agenda 2011 e realização de oficinas.

## 5. Questões teórico-metodológicas

A parte teórica está largamente especificada no projeto de pós-doutorado. No projeto, dividimos essa parte em quatro planos, com o objetivo de melhor delinear a formatação do piloto. Essa divisão foi feita da seguinte forma:

1. Discussão sobre a questão da edição de documentos nos moldes filológicos tradicionais;
2. Apresentação do conceito de mudança lingüística;
3. Discussão sobre a análise de fenômenos lingüísticos e
4. Discrção sobre a formação de bancos digitais eletrônicos nos moldes do *Corpus Histórico do Português Tycho Brahe*.

### 5.1 Da edição de documentos nos moldes filológicos tradicionais

Em resumo, essa parte foi centrada no projeto *Para a História do Português/Prohpor*, na UFBA ([www.prohpor.ufba.br](http://www.prohpor.ufba.br)) e no *Projeto Nacional Para a História do Português do Brasil/PHPB* do qual fazemos partes. A base empírica para a construção do nosso banco de texto é discutida e construída nas reuniões com esses projetos. São discussões que fomentam as bases teórico-metodológicas na construção de grandes *corpora*, de 1997 até os dias atuais<sup>6</sup>. Cabe salientar que participamos de todas as reuniões desse projeto e participamos ativamente na construção do corpus, especificamente na grande área do semi-árido. Portanto, são projetos que configuram as suas balizas na área da Lingüística Histórica em três campos de atuação: construção material/*corpus*, *base histórica que fundamenta a composição do material e análise da mudança lingüística*. As especificidades da Equipe da UEFS, que se volta ao estudo da língua portuguesa no semi-árido, bem como os produtos

---

<sup>6</sup> Na última reunião realizada entre 31 de maio a 4 de junho de 2010, no âmbito do “Projeto Para a História do Português Brasileiro”/PHPB, na Paráiba, foi discutida também já a viabilidade de um corpus eletrônico.

das fases de atuação, estão descritas no histórico do DOHS ([www.uefs.br/dohs](http://www.uefs.br/dohs)).

No caso dos estudos históricos do PB, alguns procedimentos têm orientado a formação dos *corpora*, a saber:

- a) Separação de documentos produzidos por indivíduos que têm o português como primeira língua (L1) ou como segunda (L2) e, nesse caso, tanto em situações “regulares” quanto naquelas que resultante de transmissão lingüística irregular;
- b) Separação de documentos escritos por indivíduos sem contato prolongado com a escola (populares) ou com relativo contato (semi-popular e semi-cultos) daqueles com muito contato (cultos).

## 5.2 Sobre as questões teóricas do conceito de mudança lingüística e sobre a análise de fenômenos lingüísticos

A orientação que sedimenta o olhar sobre o *corpus* é baseada na concepção de que a mudança ocorre durante o processo de aquisição de linguagem, nos moldes do é proposto no modelo de Princípios e Parâmetros (Chomsky, 1981). Para isso, assumimos a distinção entre língua-I e língua-E, proposta pelo autor. Desse modo, embora os estudos sejam feitos com dados de língua-E, o interesse é observar fenômenos que caracterizam a língua-I. Dentro dessa visão, as mudanças são tratadas como alterações paramétricas (Lightfoot, 1999, Kroch, 2001, entre outros). Essas alterações no PB, com conseqüências importantes para a mudança paramétrica no português como apontado nos trabalhos de Galves (1987, 2001) na linha dos trabalhos de autores como Tarallo (1989) e dos trabalhos organizados por Roberts e Kato (1993), entre outros. No âmbito da lingüística histórica tem sido considerado fundamental o uso de dados quantificados que apontem as alterações nas freqüências e sejam interpretados como indicativos de mudança. Desse modo, fazer estudos de Lingüística Diacrônica nessa perspectiva requer o manuseio de um número bastante grande de dados.

## 5.3 Da formação de bancos digitais eletrônicos

Como dito no projeto, as análises quantitativas demandam o manuseio de dados extensos, lugar em que as contribuições de tecnologias computacionais são decisivas e fundamentais, como a desenvolvida pelo *Projeto Corpus Histórico do Português Tycho Brahe*. Esse tipo de banco de dados vem se mostrando uma tendência mundial. Elencamos os projetos que já trabalham com essa técnica: o pioneiro *Penn Helsinki Parsed Corpus of Middle English*, (<http://www.ling.upenn.edu/hist-corpora>), coordenado por Anthony Kroch na Universidade da Pensilvânia e os seus afiliados, o *York Helsinki Parsed Corpus of Old English Poetry*, por Susan Pintzuk e Leendert Plug, o *York Toronto Helsinki Parsed Corpus of Old English Prose*, por Ann Taylor, Anthony Warner, Susan Pintzuk, Frank Beths, ambos na Universidade de York, o *Penn Helsinki Parsed Corpus of Early Middle English*, por Anthony Kroch e Beatriz Santorini na Universidade da Pensilvânia, e enfim o *Parsed Corpus of Early English Correspondence*, por Ann Taylor, Anthony Warner, Susan Pintzuk na Universidade de York, e por Terttu Nevalainen e Arja Nurmi na Universidade de Helsinki, além de outros como o projeto de *Corpus annoté syntaxiquement de textes de français (9è au 17è siècle)*, por F. Martineau e Paul Hirschbuhler na Universidade de Ottawa, e o projeto *Corpus Dialetal Sintático (CordialSin)*, por Ana Maria Martins, na Universidade de Lisboa.

Defendeu-se, o uso da linguagem XML pelas vantagens que apresenta. o pioneiro *Penn Helsinki Parsed Corpus of Middle English*, (<http://www.ling.upenn.edu/hist-corpora>), coordenado por Anthony Kroch na Universidade da Pensilvânia e os seus afiliados

Sobre a parte metodológica, bastante detalhada no projeto de pós-doutoramento em anexo, há imagens demonstrativas do passo a passo com diversas ilustrações dispostas no relatório parcial em anexo.

## 6. Produção Tecnológica

O projeto piloto buscou colaborar com metodologias já usadas no *Corpus Histórico do Português Tycho Brahe*. Os procedimentos com o aprimoramento de ferramentas para uso no *Corpus Eletrônico* são inúmeros. Uma descrição breve pode ser vista de 1-3 e no projeto:

1. Anotação da estrutura XML permitindo o controle de interferências do editor, com agilidade e sem ter que atuar sobre o texto base.
2. Uso de linguagem XSLT (eXtensible Stylesheet Language for Transformation) para geração de transformações diversas. Os comandos permitem ainda alguns procedimentos finais úteis para o usuário (geração de ficha catalográfica com informações, por exemplo, de nacionalidade, data de nascimento, entre diversas outras codificadas na pesquisa. Além de diversos outros formatos e manuseio e reajuste do XML.
3. Geração de uma versão técnica (modernizada) para procedimentos de aplicação de *Módulos de Análise Lingüística Automática*. Essa parte é feita pelo *Corpus Histórico do Português Tycho Brahe* via parcerias institucionais. Nesse caso, há garantias de preservação de direitos de cada instituição, a exemplo do que pode ser visto no projeto onde já constam textos do português brasileiro: Cartas Brasileiras (~1800), piloto deste projeto de pós-doutoramento (UEFS), além de outros materiais, como as Atas dos Brasileiros (~1830) e Atas dos Africanos (~1800) (UFBA), editadas por Klebson Oliveira. Além do Corpus do DOViC (UESB) e o de “mãos inábeis” de Rita Marquilhas, de Portugal.

## 7. Serviços

Não houve receita, porque o projeto tem como propósito a disponibilização gratuita dos documentos na rede mundial de computadores. O ganho pode se dá através dos avanços científico-tecnológicos.

Os trabalhos técnicos são os seguintes:

- 1) Criação de um portal – DOHS (Documentos Históricos do Sertão), acessível em [www.uefs.br/dohs](http://www.uefs.br/dohs) hospedado na UEFS;
- 2) Criação do CE-Dohs – Corpus Eletrônico de Documentos Históricos do Sertão ([www.uefs.br/dohs/ce-dohs](http://www.uefs.br/dohs/ce-dohs), em co-parceria com o projeto Corpus Histórico do Português Tycho Brahe ([www.tycho.iel.unicamp.br](http://www.tycho.iel.unicamp.br)) em preparação;
- 3) Co-colaboração no Projeto já citado na apresentação, coordenado na UESB por Jorge Viana Santos, Cristiane Namiuti-Temponi e Cândida Brito;
- 4) Elaboração de Oficinas de uso do E-dictor, em andamento;  
Elaboração de um manual para usuários, compreendendo, desde a parte filológica até a parte de edição eletrônica (em preparação).

O objetivo do CE-DOHS é transpor a edição tradicional na versão *Word* para a versão com uso da XML, base para geração de versões de edições e para o acesso as informações sobre os documentos. O uso da XML será feita através da ferramenta E-dictor, como já explicitado.

## 8. Capacitação de Recursos Humanos

O projeto objetiva especificamente executar ainda algumas tarefas cruciais, citadas abaixo, que resultarão nas metas desejadas:

1. A formação de corpo técnico capacitado para transcrição dos manuscritos e para a aplicação de tecnologias eletrônicas (em preparação). A realização de oficinas com a participação dos pesquisadores externos e dos alunos que contribuirão para o pleno funcionamento do projeto e subsidiarão as discussões e a busca de soluções.

2. Buscar soluções que possam conciliar as metas do DOHS e CE-DOHS, entre a manutenção da fidelidade do texto original e fac-similado e a agilidade de processamento de um banco eletrônico, respectivamente. A linguagem XML permite a compatibilização de ambos.

## 9. Difusão

1. Realização de *Workshop* para formulação de diretrizes metodológicas para a transcrição dos manuscritos e preparação de textos para a edição eletrônica, através da realização de oficinas com a participação dos pesquisadores externos (UESB, Unicamp e UFBA).

2. Divulgação do corpus DOHS e CE-DOHS para a comunidade na Feira do semi-árido a ser realizada em dezembro de 2010 na UEFS.

3. Continuar a disponibilizar, a partir do DOHS e CE-DOHS, os *corpora* digitais e os manipuláveis eletronicamente. Isso já está sendo feito em [www.uefs.br/dohs](http://www.uefs.br/dohs).

## 10. Outros

Submissão ao Conselho de Ética ([www.uefs.br/cep](http://www.uefs.br/cep)) para obtenção do *Termo de Consentimento Livre e Esclarecido* dos acervos privados.

## 11. Parceria Institucional

Com o fim de garantir a troca mútua de tecnologia e o uso efetivo da metodologia do projeto base *Corpus Histórico do Português Tycho Brahe* /Unicamp está sendo firmado, através de convênio guarda-chuva da UEFS com a Unicamp, um Termo Aditivo, uma iniciativa da Prof<sup>a</sup> Dr<sup>a</sup> Charlotte C. Galves (Unicamp), supervisora deste projeto de pós-doutorado (cf. minuta em tramitação na Unicamp).

O objetivo é aumentar as competências do banco, com a criação de Corpus Anotado feito com base na manipulação de tecnologias consolidadas pelo Projeto Tycho Brahe, fazendo uso de instrumentos que permitem a exploração do acervo com base em extração automática de informações lingüísticas, através de meios eletrônicos, no qual o projeto Tycho Brahe é pioneiro, desenvolvendo trabalho de excelência, através da elaboração de ferramentas específicas de etiquetagem automática de textos, a exemplo do software de anotação E-dictor (cf. <http://oncoto.dyndns.org:44880/projects/edictor>), desenvolvido por pesquisadores do projeto Tycho Brahe, e de outras ferramentas adaptadas, como o *CorpusDraw* e o *CorpusSearch* ([corpussearch.sourceforge.net](http://corpussearch.sourceforge.net)), além de outras ferramentas descritas no relatório parcial de pós-doutorado, em anexo.

Essas ferramentas são fundamentais para a preparação de dados para busca automática de dados lingüísticos, amplamente aplicadas a dados do português europeu pelo Projeto Tycho Brahe. E, através dessa parceria, a inserção de textos do português brasileiro, como por exemplo, as cartas escritas por brasileiros, na qual se baseia o piloto desse projeto. É importante salientar que o banco de dados da UEFS, na forma como foi desenvolvido, apesar das enormes contribuições para o conhecimento sobre a língua portuguesa e sobre a formação lingüística no Sertão da Bahia, estava sendo subutilizado, devido a limitações de acesso, devido às suas peculiaridades.

Considerando esses aspectos, o convênio e o termo aditivo mencionados são fundamentais nesse processo. Também estamos nos integrando com a UESB e outras instituições como a UFBA, como já dito.

## 12. Impactos

Como dito no projeto de pós-doutorado e que julgamos procedente enfatizar, o estudo da mudança lingüística, com fortes repercussões na área da linguagem, constitui-se a gênese da própria lingüística como ciência, iniciada em fins do século XIX, e avança com suas metodologias de acordo com o desenvolvimento da própria ciência em geral, tanto do ponto de vista social, biológico cerebral quanto no plano tecnológico na área da inteligência artificial, beneficiados com a automação tecnológica.

Como se trata de um campo que cresce com a troca de tecnologias e essas demandam grandes investimentos financeiros e acadêmicos, diversas universidades brasileiras vêm fazendo parcerias em todas as áreas de conhecimento e principalmente na construção de grandes bancos de dados compartilhados mundialmente na rede de computadores, como um recurso imprescindível para fazer lingüística na atualidade. No campo da Lingüística Histórica, no Brasil, um dos pioneiros nesse ramo é o projeto *Corpus Histórico do Português Tycho Brahe* (<http://www.tycho.iel.unicamp.br/~tycho>), inserido no projeto interdisciplinar “Padrões Rítmicos, Fixação de Parâmetros & Mudança Lingüística”, financiado pela FAPESP e CNPq, iniciado em 1998 (cf. relatórios em <http://www.tycho.iel.unicamp.br/~tycho/prfpml/fase2/relats.html>), na qual me filiei, através da UEFS, desde 2005 (cf. <http://dgp.cnpq.br/buscaoperacional/detalhepesq.jsp?pesq=5992506414152580>). No caso desse projeto, em particular, vem fazendo parcerias importantes com diversas instituições internacionais. A colaboração científica nessa parceria que nos propomos é importante porque esse tipo de banco de dados encontra problemas de adequação entre materiais editados na tradição filológica e sua conversão para linguagens computacionais, como a XML. No atual estágio, grandes e importantes discussões têm sido feitas com centros de excelência em filologia como a UFBA, no qual faço parte, através do PROHPOR ([www.prohpor.ufba.br](http://www.prohpor.ufba.br)) com diversas instituições, com o projeto nacional PHPB (conforme relatório <http://www.mundoalfal.org/Ataliba>). Entrei nessa discussão através do meu projeto de doutoramento na Unicamp, e continuo desde o seu término em 2005. No momento atual, passamos a colaborar com material para alimentação do *Corpus Histórico do Português Tycho Brahe* (ver cartas brasileiras (UEFS) atas de africanos (UFBA), itens 45-48 <http://www.tycho.iel.unicamp.br/~tycho/corpus/index.htm>). Mas queremos dar um passo à frente e trocarmos tecnologias para implantação de um banco na UEFS em parceria com a UNICAMP.

Os ganhos científicos se darão, principalmente, através do aprimoramento de técnicas de conversão para a XML ou Linguagem de Marcação Extensível (*extensible markup language*, cf. W3C, 1997), sem prejuízo do rigor filológico, desenvolvido com excelência na UEFS e na UFBA e agora também em um projeto que se inicia na UESB, formando uma parceira importante nesse cenário, unindo tecnologia e tradição, como referido anteriormente.

### 12.1 Impacto Tecnológico

Como dito e defendido no projeto, o avanço científico de modo geral depende crucialmente da manipulação segura e rápida de um grande volume de dados com entrecruzamento com redes espalhadas por todo mundo, como uma necessidade global. Diversos projetos na área de lingüística para formação de grandes bancos de dados vêm avançando rapidamente. São exemplos desse o próprio projeto “*Corpus Histórico do Português Tycho Brahe* que é filiado ao *Penn-Helsinki Parsed Corpus of Middle English*, coordenado por Anthony Kroch, que vem se unindo em uma rede de colaboração mútua via conhecimento de experiências em outros grupos de pesquisa, no Brasil e no exterior, a exemplo do “Comportamento estocástico, Fenômenos Críticos e Identificação de Padrões Rítmicos nas Línguas Naturais (Pronex/Fapesp)”, e o Projeto “*Lácio-Web do Núcleo Interdisciplinar de Lingüística Computacional (NILC)*, além de contar com o suporte

computacional com a rede IME-USP. Sendo assim, o projeto de Corpus Digital Eletrônico, que estamos propondo, tem como impacto a sua inserção em uma rede internacional de construção de grandes *corpora* históricos anotados. A parceria, através da troca de experiência com o maior centro desse tipo no Brasil, é uma necessidade para a Bahia, a fim de que possa contribuir através da conversão de seus dados armazenados em projetos locais para linguagens mais adequadas. Os avanços se darão em várias frentes. No plano da pós-graduação (mestrado) na UEFS, em fase de implantação. Também será importante dar visibilidade a materiais já prontos, como será mostrado adiante e que estão sendo divulgados em ritmo lento por dificuldades no manuseio, dinamizando as pesquisas na UEFS, ao tempo que as insere nos contextos nacionais e internacionais.

## 12.2 Impacto Econômico

Também se mantém a asserção do projeto segundo a qual, esse tipo de banco de dados eletrônico, como o proposto, pode vir a contribuir para minimizar os custos futuros. O tipo de banco de dados digital em XML, embora com as dificuldades iniciais de implantação, motivadas principalmente pelas adequações em projetos pioneiros, pode sem dúvida ter um grande impacto econômico no sentido positivo, porque poderá gerar produtos com crescimento exponencial através de redes de cooperação e entrecruzamento de dados com custos cada vez menores. São materiais que podem se tornar disponíveis na rede mundial de computadores, com recuperação rápida e confiável de grandes quantidades de informação com custos inferiores àqueles manuseados por via impressa (livros digitais, e materiais de uso *on line*), etc. São inúmeras as contribuições com retornos cada vez maiores, seja em termos de produtos acadêmicos (teses, artigos, etc), seja através do seu uso mais direto pela comunidade (acessos *on line*), tanto de forma simétrica quanto assimétrica, através de ganhos com as facilidades tecnológicas no manuseio das informações.

Em resumo, de forma mais iminente, a cooperação entre projetos, como a que estamos propondo, permite a redução de custos através do compartilhamento de recursos tecnológicos e operacionais.

## 12.3 Impacto Social

Também na área social há impactos importantes. A aquisição de metodologia desenvolvida para a elaboração de grandes *corpora* anotados de língua vem se firmando como um recurso imprescindível para fazer estudos de mudança lingüística em uma das áreas mais importantes da lingüística hoje e que grandes contribuições vêm dando, inclusive na área educacional, através do conhecimento da língua exteriorizada, também, como instrumento legítimo e historicamente construído. E, é nesse sentido, esse projeto pode contribuir para o conhecimento do PB. Uma língua que assume papel de liderança em termos de usuários no chamado mundo lusófono. Conhecer a realidade lingüística do Brasil, do qual tem o grande papel a sociolingüística, desde o seu processo inicial de formação como país nos moldes de hoje, requer a manipulação de um grande volume de informações para legitimar todas as variedades lingüísticas dos povos que constituem as bases de sua formação. O Brasil enfrenta ainda hoje grandes problemas na área educacional, sobretudo com dificuldades na interpretação e leitura, o chamado analfabeto funcional.

Parte dessa dificuldade advém do desconhecimento do fosso que separa o Brasil no uso de uma língua padrão e suas legítimas variedades vernaculares que formam o português brasileiro, um contexto de diglossia lingüística com todas as implicações que isso possa ocasionar em termos educacionais. A lingüística histórica fornece a base empírica, ao lado de diversas outras áreas da lingüística, um suporte para uma adequada discussão sobre o problema educacional. Contudo, os avanços científicos na área da lingüística requerem uma base de dados eficiente e rápida que, acreditamos, possa ser dinamizada com projetos dessa natureza.

No que diz respeito diretamente ao projeto proposto, poderá permitir a diminuição da concentração regional da pesquisa, e propiciar a difusão do conhecimento entre instituições nacionais e internacionais, além de subsidiar outras do interior do Estado.

## 13 Produção bibliográfica gerada através do desenvolvimento da pesquisa

### Relatórios/ notas técnicas

1. Nota técnica relativa à sistematização catalográfica do material base para o corpus piloto (cf. [www.uefs.br/dohs](http://www.uefs.br/dohs) (acesso XML));
2. Composição de um *corpus* mínimo para alimentação do Corpus Anotado (Coleção Diacrônica – série cartas (no prelo));
3. Catálogo com os materiais digitalizados e editados em formato paleográfico no banco de texto do projeto “Vozes do Sertão em Dados”;
4. Elaboração do projeto “Vozes do sertão em Dados: *história, povos e formação do português brasileiro* (PB) - Período: julho de 2009/dezembro de 2010/Financiador: CNPq. Número do processo: 401433/2009-9/Edital/Chamada: Edital MCT/CNPq 02/2009 - Ciências Humanas, Sociais e Sociais Aplicadas, com o objetivo de fazer a seleção para alimentação do Corpus Anotado a ser instalado na UEFS, como produto do Piloto do Pós-doutorado.
5. Elaboração do projeto CE-DOHS – Corpus Eletrônico de Documentos Históricos do Sertão, aprovado no Edital Referência 2010 (FAPESB 5566/2010), uma interface, em formato eletrônico, do banco digital DOHS – Documentos Históricos do Sertão ([www.uefs.br/dohs](http://www.uefs.br/dohs)), que faz parte do projeto Vozes do Sertão em Dados: história, povos e formação do português brasileiro (CNPq. 401433/2009-9) ([www.uefs.br/nelp](http://www.uefs.br/nelp)).

### Anais

Três estudos: um sintático, com parte do material do “Vozes do Sertão em Dados, em (a) sobre avaliação sobre a escrita do material em (2) e uma avaliação sobre prospecção de material inédito relativo a escrita em contexto de aldeamentos nos século XVII-XVIII” em (c). Os trabalhos (b) e (c) definem-se como estudo de sistematização de material para aplicação no Corpus Anotado.

- a) CARNEIRO, Z. O. N. ; GALVES, C. . A ênclise no português brasileiro: uma mudança fracassada. In: ROSAE - I Congresso Internacional em Lingüística Histórica, 2009, Salvador. ROSAE - Homenagem a Rosa Virgínia Mattos e Silva - I Congresso Internacional de Lingüística Histórica. Feira de Santana: Imprensa Universitária da UEFS, 2009. v. 1.
- b) CARNEIRO, Z. O. N. . Das normas em contraste, aproximações ou distanciamentos? Um estudo do português culto, semi-culto e popular no século XIX. In: II Simpósio Mundial de Estudos de Língua Portuguesa, 2009, Évora - Portugal. Panorâmica de Linguística, Literatura e Cultura do II Simpósio Mundial de Estudos de Língua Portuguesa. Évora : Universidade de Évora, 2009. v. 1. p. 64-64.
- c) CARNEIRO, Z. O. N. . Prática de escrita em aldeamentos no sertão baiano, século XVII e XVIII: novas fontes para estudo do português no âmbito do projeto Vozes do sertão em dados: história, povos e formação do português brasileiro. 2009. (Apresentação de Trabalho/Comunicação).

### Previstos:

#### Publicação da coleção de cartas brasileiras e do livro de análise lingüística.

*Cartas brasileiras (1808-2000): coletânea de fontes para o estudo do português brasileiro*, composta com a edição em cd (no prelo), FAPESB -1493/2010, a saber:

- Livro 1.** *Cartas brasileiras (1808-1904): coletânea de fontes para o estudo do português brasileiro com 3 cds: cartas avulsas para vários destinatários; cartas para Severino Vieira e cartas para o barão de Jeremoabo.*
- Livro 2.** *Cartas brasileiras (1902-2000): coletânea de fontes para o estudo do português brasileiro, com 03 cds, cartas para Dantas Jr.; cartas de mulheres da família Costa Pinto e cartas pessoais de Valente, Bahia.*
- Livro 3.** *Cartas brasileiras populares (1975-1992): coletânea de fontes para o estudo do português brasileiro, em versão apenas impressa.*

Análise de fenômenos lingüísticos: Produto do Piloto do Pós-doutorado.

CARNEIRO, Zenaide O. N. (Org.); GALVES, Charlotte Marie C. (Org.); LOBO, T. C. F. (Org.); Gramáticas em conflito: estudos sobre o português escrito por cultos no Brasil dos séculos XIX-XX. (em preparação).

## Eventos

ROSAE - Homenagem a Rosa Virgínia Mattos e Silva - I Congresso Internacional de Lingüística Histórica – Pestana – Salvador – FAPESB  
Data: 31/08/ 2009

II Simpósio Mundial de Estudos de Língua Portuguesa, 2009, Évora - Portugal. Panorâmica de Lingüística, Literatura e Cultura do II Simpósio Mundial de Estudos de Língua Portuguesa. Évora : Universidade de Évora, 2009.  
Data: 06/10/ 2009

- VIII SEMINÁRIO DO PHPB, realizado no período de 01 a 04 de junho de 2010 em João Pessoa. Oficina na UEFS em agosto /2010 para treinamento na UEFS para treinamento do E-dictor.

- Organização de balcão de pesquisa na Feira do Semi-Árido na UEFS com a participação de grupos da UFBA e da UESB.

- *International Conference on Pluricentric Languages*, 15-17 – September, Braga – Lisboa (Aceito) Oficina na UEFS em meados de Julho/2010 para treinamento na UEFS para treinamento do E-dictor.

- Organização de balcão de pesquisa na Feira do Semi-Árido na UEFS com a participação de grupos da UFBA e da UESB.

## 13. Comentários gerais finais e perspectivas

Implementar o CE-DOHS com a parceria com a Unicamp, via *Corpus Histórico do Português Tycho Brahe* ([www.tycho.unicamp.br](http://www.tycho.unicamp.br)), como já relatado, que vem a ser o projeto macro, através da metodologia subsidiada e desenvolvida nesse projeto com o sistema de edições eletrônicas. Ou seja, implementar o piloto do projeto do pós-doutorado, cuja meta foi a aplicação de tecnologias que possibilitem a convergência de edições tradicionais e limitadas a usos impressos ou mesmos digitais estáticos, em camadas de textos manipuláveis, via uso da linguagem XML, através do uso da ferramenta integrada de anotação de *corpus*, e-dictor, citada acima.

A iniciativa em projetos de mútua cooperação como esse que aqui se apresenta, entre a UEFS e o *Corpus Histórico do Português Tycho Brahe* é importante para fortalecer grupos de pesquisas em universidades que ainda se encontram em fases iniciais, sobretudo no que se refere ao

desenvolvimento de tecnologias novas e que necessitam de grandes investimento. Da parte dos resultados alcançados no projeto, pudemos contar com a qualidade da orientação e da infra-estrutura oferecida. A Prof<sup>a</sup> Dr<sup>a</sup> Charlotte C. Galves, pioneira no uso de corpus eletrônico no Brasil, vem desde 1998, fornecendo todo o instrumental de orientação e aperfeiçoamento de ferramentas para esse tipo de banco, como pode ser visto claramente no Projeto Tycho Brahe (<http://www.tycho.iel.unicamp.br/~tycho/>).

Também foi fundamental o apoio de participantes do projeto, como Cynthia Yano na preparação das edições e Mariana Pereira na etiquetagem morfológica. Foi fundamental o apoio de Pablo Faria, um profissional capacitado para dar todo o suporte técnico computacional e adequar as ferramentas às necessidades entre a edição tradicional paleográfica importante para os estudos lingüísticos e as edições técnicas computacionais, essencial para o manuseio de banco digitais eletrônicos. Inclusive, como dito, está prevista uma oficina na UEFS com a sua participação para efetiva formação de corpo técnico. Também está prevista a vinda da Prof<sup>a</sup> Dr<sup>a</sup> Charlotte C. Galves na UEFS, ocasião em que serão estreitadas as relações interinstitucionais. A professora Charlotte Galves ofereceu todo o aparato intelectual para o efetivo sucesso do projeto. Estamos agora em fase de integração com outras equipes regionais que têm os mesmos objetivos, o de formar bancos eletrônicos anotados, a exemplo da equipe da UESB, já citada, com o apoio da Profa. Dra. Charlotte Galves, que também participa na elaboração do livro com os resultados das análises lingüísticas com uso do piloto. A infra-estrutura física do projeto é adequada, oferecendo todas as condições para o efetivo sucesso do projeto. E para o efetivo sucesso da transformação do piloto em projeto na UEFS, estamos contando com o apoio institucional em tramitação na Unicamp para a troca de tecnologias e manutenção da parceria. Esse acordo está sendo intermediado pela administração da UEFS, através da AERI, como dito.