Resumo do Projeto

Status da Solicitação: Aprovado Parcialmente
Nº de Alunos Aprovados: 3

Matrícula:	292289
Orientador(a):	Juanito Ornelas De Avelar
E-mail do Orientador(a):	juanitoavelar@iel.unicamp.br
Ramal do Orientador(a):	11564
Título do Projeto:	**Edição Eletrônica de Corpora Orais para o Estudo de Variedades do Português**
Data do Cadastro:	28/10/2015 11:38:13
Status:	Aprovado Parcialmente
Área do Projeto:	Aprimoramento técnico – Humanas
Unidade de desenvolvimento do projeto:	IEL - Instituto De Estudos Da Linguagem
Campus do Projeto:	Campinas
Local específico de desenvolvimento do projeto:	Projeto Tycho Brahe - Pavilhão dos Docentes - IEL/UNICAMP
Período para a realização das atividades do bolsista:	Horário Flexível
O projeto está sendo apoiado ou financiado por entidade(s) ou órgão(s):	Sim, FAPESP
Número de bolsistas solicitados:	6
Área do curso:	Exatas, Humanas, Tecnológicas
Resumo do Projeto:	Este projeto visa à edição de corpora orais elaborados para o estudo de variedades do português faladas no Brasil e na África. Mais precisamente, o projeto tem por objetivo inserir transcrições de amostras orais do português na base de dados Tycho Brahe (http://www.tycho.iel.unicamp.br/corpus/index.html). O Corpus Tycho Brahe, desenvolvido no Instituto de Estudos da Linguagem da UNICAMP, constitui atualmente uma das maiores bases eletrônicas de textos para o estudo dos períodos clássico e moderno do português. A construção do Corpus teve início em 1998, sob a coordenação da Profa. Dra. Charlotte Galves (IEL/UNICAMP), e passou por diferentes

Descrição das atividades do Projeto, incluindo o cronograma:

fases que envolveram, entre outros aspectos, a etiquetagem morfológica e a anotação sintática de documentos escritos em diferentes estágios do português. Na sua fase mais recente, iniciada em 2012 por meio do projeto temático "A Língua Portuguesa no Tempo e no Espaço" (FAPESP 12/06078-9), o Corpus Tycho Brahe deu início a uma frente de trabalho para incluir em sua base amostras do português contemporâneo falado no Brasil e na África. Parte dessas amostras vem se constituindo com o apoio de pesquisadores da Universidade de Estocolmo (Suécia), por meio do grupo internacional de pesquisa "Estudos Linguísticos Afro-Latinos" (www.iel.unicamp.br/projetos/afrolatinos). Esse grupo tem se ocupado da constituição de corpora para a realização de estudos comparativos sobre variedades brasileiras e africanas do português, reunindo atualmente cerca de 130 horas de áudio com falantes do português afro-brasileiro, angolano, moçambicano e cabo-verdiano. Uma parte desse vasto material coletado pelo grupo será, a partir de 2016, inserido na base de dados do Tycho Brahe, o que exigirá a utilização de ferramentas computacionais voltadas à edição textual, etiquetagem morfológica e anotação sintática. Para executar essas etapas, o projeto espera contar com a atuação de bolsistas que realizarão atividades técnicas específicas, necessárias à organização eletrônica dos corpora.

O projeto irá se ocupar da edição de duas amostras de fala, representativas das seguintes variedades do português: (i) o português afro-brasileiro da comunidade do Cafundó (Sorocaba/SP) e (ii) o português falado na região de Cabinda (Angola). O material relativo ao Cafundó foi coletado no início da década de 80 por uma equipe liderada pelos pesquisadores Carlos Vogt e Peter Fry. Esse material compõe hoje o acervo do CEDAE-IEL/UNICAMP. Os áudios (em torno de 46h) foram inteiramente digitalizados em 2013 com recursos cedidos pela Universidade de Estocolmo. A amostra do português falado na região de Cabinda (Angola) também se encontra em fase de transcrição, que será concluída em fevereiro de 2016. Os áudios que compõem essa amostra foram gravados em 2014 por pesquisadores da Universidade de Estocolmo e perfazem cerca de 22h. A participação dos bolsistas se dará por meio das seguintes etapas (considerando

março de 2016 como o mês de início): (I) março e abril de 2016: treinamento para utilização das ferramentas de edição textual e etiquetagem morfológica do Tycho Brahe; (II) maio a dezembro de 2016: edição textual e etiquetagem morfológica das transcrições dos áudios de cada amostra na plataforma do Tycho Brahe; (III) novembro de 2016 a fevereiro de 2017: testagem e revisão dos materiais editados e etiquetados.

Justificativa do solicitação e do número de bolsistas:

A edição eletrônica de grandes corpora tem conquistado um espaço cada vez maior nos estudos da linguagem, por meio da chamada Linguística de Corpus, campo que vem se firmando como uma das várias possibilidades de interface entre a linguística e a computação. As atividades desenvolvidas no âmbito do Tycho Brahe vêm contribuindo para colocar a UNICAMP na linha de frente das pesquisas em Linguística de Corpus realizados no Brasil, em particular, e na comunidade dos países de língua portuguesa, em geral. Ao investir na inclusão de amostras do português contemporâneo falado no Brasil e na África em sua base de dados, o acervo do Tycho Brahe permitirá realizar estudos online para estabelecer comparações entre novas variedades linguísticas do português, bem como elaborar novas ferramentas de etiquetagem, anotação e busca eletrônica de dados para futuros desenvolvimentos do projeto. Para tanto, é necessário contar, nesta fase, com um número amplo de bolsistas que possam se ocupar de atividades técnicas relacionadas à elaboração dos corpora. Estamos, por essa razão, solicitando 6 (seis) bolsistas para atuar na área de "aprimoramento técnico", por meio de atividades que envolverão desde a edição eletrônica das amostras à sua etiquetagem morfológica. Os bolsistas serão divididos em 2 grupos: um deles, formado por 4 membros, se ocupará do material relativo às 46 horas de transcrição da amostra do Português Afro-brasileiro, enquanto o outro grupo, formado por 2 membros, se ocupará das 22 horas de transcrição da amostra do Português de Cabinda/Angola.

Estimativa do número de pessoas beneficiadas pela presente solicitação, desrição do público-alvo e dos impactos esperados com a realização do projeto:

De modo direto, serão beneficiados todos os pesquisadores do grupo de pesquisa "Estudos Afro-Linguísticos", bem como os que atuam no projeto temático "A Língua Portuguesa no Tempo e no Espaço". A equipe dos "Estudos Linguísticos Afro-Latinos" reúne atualmente 13 docentes de universidades do Brasil e do exterior (Alemanha, Estados Unidos, Finlândia, Noruega, Suécia, Uruguai), que desenvolvem pesquisas sobre as dinâmicas de contato interlinguístico na África e na América Latina, em particular no que concerne à emergência de novas variedades de português e espanhol. O projeto "A Língua Portuguesa no Tempo e no Espaço", por sua vez, conta com 24 docentes de diferentes universidades brasileiras. Considerando os docentes integrados aos dois projetos, bem como os seus orientandos discentes, o número de beneficiados diretos com os resultados do presente projeto gira em torno de 130 pesquisadores. Ressalte-se, além disso, que os corpora serão disponibilizados na Web, à medida que a sua edição eletrônica for sendo concluída, o que significa que estará disponível, entre 2017 e 2018, a toda a comunidade acadêmica do Brasil e do exterior.

Quais os principais conhecimentos adquiridos pelo bolsista durante o cumprimento das atividades deste projeto?

A utilização de recursos computacionais para a edição de corpora tem sido uma tendência crescente não apenas nas áreas de Linguística e Letras, mas também em outros campos das Ciências Humanas e em áreas de investigação que requerem o trabalho com dados provenientes de grandes bases textuais. Ao participar de etapas essenciais à disponibilização eletrônica de corpora, o bolsista terá acesso a ferramentas importantes da tecnologia empregada em Linguística de Corpus, o que contribuirá para o desenvolvimento de competências e habilidades aplicadas à elaboração e disponibilização de grandes bases de dados em meios digitais. O bolsista terá, além disso, a oportunidade de ampliar seus conhecimentos em torno de aspectos linguísticos, sócio-históricos, culturais e políticos relacionados à difusão da língua portuguesa do mundo.

Alunos Vinculados ao Projeto

#	Nome	Curso	E-mail	Período	Carga Horária Total
1	Larissa Kaoane Firmino Ribeiro	Ciências Econômicas	117168 7@dac.unicamp.br	De 15/08/2016 a 01/03/2017	420 horas