

PROJETO CONTEMPLADO DO EDITAL SEAD- UFRGS 09/2008

Veja proposta em

<http://www6.ufrgs.br/sead1/edital09/listagens/ver.php?id=19>

Etapa em que se encontra o Projeto, em relação ao cronograma proposto

Emissão deste reatório: em agosto de 2008

Antes de entrarmos propriamente na etapa em que o projeto se encontra, acreditamos importante acrescentar algumas informações sobre as duas linhas de pesquisa:

Catálogo de Expressões Recorrentes:

A partir de um corpus composto por 283 textos da revista bilíngüe português-inglês *Jornal de Pediatria*, já fizemos o reconhecimento dos itens/expressões mais recorrentes que farão parte do Catálogo. Em um primeiro momento do trabalho, foram consideradas recorrentes todas as combinatórias de duas, três ou quatro palavras que ocorressem juntas no corpus em português com uma frequência igual ou maior do que cinco ocorrências.

Assim, expressões do tipo bigrama (de duas palavras) como **VACINA BCG**, do tipo trigrama (de três palavras) como **PREVALÊNCIA DE HEPATITE**, ou ainda do tipo quadrigrama (de quatro palavras) como **RECÉM-NASCIDOS PRÉ-TERMO** foram consideradas candidatas a entradas no Catálogo.

A listagem de expressões potencialmente catalogáveis, obtida nessa primeira etapa, foi muito extensa. Foi preciso foi preciso filtrar os **36.471** itens obtidos, pois o conjunto inicial incluiu uma série de associações que não interessariam ao usuário, tais como **COM A** [925 ocorrências], **PACIENTES COM** [756 ocorrências] ou **EM CRIANÇAS** [845 ocorrências].

Essa primeira filtragem, que visava eliminar itens indesejáveis, que apresentavam um determinado padrão composicional, foi feita com o uso de programas de computação elaborados pelo bolsista de Ciências da Computação, Adriano Zanette. A partir dessa primeira filtragem, restaram então **7321** expressões. Foram feitas verificações manuais para se chegar a uma lista de **3645** expressões cuja inserção no Catálogo foi avaliada manual e subjetivamente em função da pertinência da expressão em meio à terminologia de *Pediatria* e da pertinência da associação de palavras para o aprendiz de tradução. Em nossa última reciclagem, **2180** elementos foram excluídos manualmente.

Cada item ou expressão incluída no Catálogo estará acompanhado da indicação do seu percentual de ocorrência/frequência no corpus de *Pediatria*, por contextos de uso em português, em inglês e por notas de uso em português e notas de tradução. Para qualificar a seleção dos melhores contextos a oferecer para o usuário, considerando que tratamos com expressões com uso muito freqüente, o bolsista Adriano também criou uma ferramenta que é um GERADOR DE CONTEXTOS. Essa ferramenta classifica os contextos de ocorrência de modo a podermos avaliar os melhores e mais completos, tanto em termos nocionais quanto de emprego da expressão.

Contando agora a pré-seleção de construções recorrentes e com a lista dos melhores contextos dessas expressões geradas automaticamente, estamos implementando as construções recorrentes selecionadas.

Além desse trabalho, estamos finalizando o Dicionário de *Pediatria*, que ficou incompleto no edital/SEAD-07. Esse dicionário está *on-line* e possui 1006

verbetes, sendo que o trabalho de coleta de definições é o maior desafio a vencer. Vale notar que vários itens do Catálogo, itens que figuram no interior de alguma expressão recorrente, ao constituírem terminologia, estarão associados ao Dicionário. Assim, haverá uma estrutura de *links* entre o Catálogo de Expressões Recorrentes e o Dicionário de Pediatria. O dicionário enfatiza a informação conceitual/nocional; o catálogo a combinatória de palavras e os diferentes usos das expressões recorrentes em meio ao corpus.

Por fim, cabe dizer que os bolsistas Adriano Zanette e Gustavo Feix apresentarão trabalho no XX Salão de Iniciação Científica da UFRGS, em final de outubro de 2008, mostrando a concepção do Catálogo, seu embasamento teórico e sua pertinência para a formação de tradutores. Integram-se assim a pesquisa em EAD e a pesquisa de IC.

Disciplina Moodle Sintaxe do Texto:

Quanto à disciplina, ressaltamos que se encontra em pleno funcionamento na plataforma Moodle, visto que está sendo oferecida em 2008/2. Assim, não é exagero dizer que sua configuração é testada e ajustada em tempo real, simultâneo ao uso pelos alunos.

Estão disponíveis no site da disciplina: Programa da Disciplina, Cronograma das Atividades, Auto-Avaliação, Biblioteca Virtual, que dispõe dos textos utilizados em aula e complementares, Índice da Biblioteca Virtual e Monitoria – contamos com o auxílio da MONITORA PROGRAD Carina Roesler Arsego atuando constantemente, sobretudo na parte de uma interação com os alunos e na de atendimento sobre eventuais dúvidas. O site da disciplina na plataforma Moodle, no seu conjunto, oferece um recurso de complementação para as atividades de ensino presenciais.

Resta ainda a inserção de atividades de aula específicas com o Catálogo de Expressões Recorrentes, entre a sucessão de aulas programadas, já que a associação recorrente de expressões em textos técnicos e científicos integra um dos assuntos da disciplina Sintaxe do Texto em meio ao curso de Letras Bacharelado em Tradução. Essa inserção, acompanhada de uma apresentação presencial do trabalho de Iniciação Científica, antes citado sobre o Catálogo de Expressões Recorrentes em Pediatria, será feita com o auxílio do bolsista Gustavo Feix.

Dito isso, ressaltamos que o projeto Catálogo de Construções Recorrentes em Textos de Pediatria para Aprendizes de Tradução, como um todo, encontra-se na etapa de implementação das construções pré-selecionadas. Para essa implementação, ambos bolsistas estão escolhendo os melhores contextos para as expressões selecionadas, partindo da ferramenta criada. Estamos refinando e otimizando o desenho do Catálogo em função do que imaginamos ser importante para atender as necessidades do usuário principal, ou seja, um estudante de Tradução.

Até o final deste semestre do letivo, o Catálogo será apresentado para uso e testagem dos alunos matriculados em Sintaxe do Texto 2008/2. Essa testagem será muito importante para o aperfeiçoamento da ferramenta, que possui uma estrutura colaborativa. O Dicionário de Pediatria, que dialoga com os registros do Catálogo, já está em fase final de implementação e em breve estará disponível para a consulta no *site* Textquim (www.ufrgs.br/textquim).

Dificuldades encontradas: Algumas das construções selecionadas, especialmente as de baixa recorrência em português, ou seja, as que apresentaram menos de 10 contextos, não se evidenciaram construções recorrentes quando analisadas em inglês. Assim, o que é recorrente em português, língua tomada como referência para o Catálogo, não é necessariamente recorrente em inglês. Temos, então, uma dificuldade, visto que pretendemos mostrar a correspondência da recorrência na tradução para o inglês. Para ilustrar a dificuldade de pareamento entre as duas línguas, felizmente bastante centrada nos itens de frequência mínima, temos o exemplo da expressão **CONSUMO DE LEITE**. Essa expressão que possui cinco ocorrências no corpus em português possui QUATRO formas diferentes no corpus de tradução para o inglês: **(HIGH) INTAKE OF COW'S MILK**, **COW'S MILK WAS INGESTED**, **HUMAN MILK CONSUMPTION**, e **MILK CONSUMPTION**. Esse tipo de dado/fenômeno terá que ser informado ao nosso usuário na parte do Catálogo que contempla **notas de uso**, alertando-o que há diferença no grau de fixação/estabilidade da forma da expressão do português para o inglês.

Outras expressões, por sua vez, apresentaram diferenças em relação ao número de componentes do português para o inglês. Algumas construções bigramas no português tornam-se trigramas no inglês ou vice-versa etc. Como exemplo, o trigrama **ACONSELHAMENTO EM AMAMENTAÇÃO** corresponde ao bigrama **BREASTFEEDING COUNSELING**; o trigrama **ALEITAMENTO MATERNO EXCLUSIVO** torna-se **EXCLUSIVELY BREASTFED**. Assim, será importante fixar para o usuário que o fator número de componentes não é necessariamente homogêneo em muitos casos, fato inerente às diferenças entre uma língua românica como o português e uma germânica como o inglês. Por fim, cabe dizer que as diferenças inerentes às escolhas de tradução para o inglês, presentes em um corpus composto por textos publicados em revista conceituada, e as diferenças advindas do distinto padrão gramatical das línguas envolvidas terão que ser superadas em termos de organização da informação que o usuário receberá. As diferenças, enfim, são constitutivas, e devem ser espelhadas no Catálogo.

Quanto à disciplina, uma das dificuldades a ser superada na implantação no Moodle ainda é a baixa adesão ao ambiente por parte de muitos alunos. Vários deles sequer acessaram o *site*, e já estamos praticamente no meio do semestre. Sem essa interação, fica algumas vezes difícil acompanhar as necessidades de alteração das atividades propostas e avaliar como elas estão sendo recebidas. Como a disciplina inclui uma série de participações de convidados, que fazem palestras/relatos, criamos a opção de envios de arquivo com uma síntese das palestras assistidas. Assim, os alunos terão bons motivos para se familiarizar com a estrutura do site e com seu modo de operação. Outra dificuldade que prevemos será a inserção qualificada do tema das expressões recorrentes como atividade de AULA presencial e em EAD. Para tanto, abriremos um período para utilização e avaliação do Catálogo em combinação com o Dicionário de Pediatria no cronograma da disciplina.

Necessidades para a conclusão:

Os trabalhos estão bem encaminhados e, sobretudo graças ao auxílio da monitora PROGRAD, Carina Arsego, que atua junto à disciplina Sintaxe do Texto, temos podido interagir com os alunos a fim de aperfeiçoar o desenho das informações e das atividades em nosso site, na plataforma Moodle. Imaginamos que até novembro possamos fazer os primeiros testes e apresentações do funcionamento Catálogo para os alunos de Sintaxe do Texto. A estrutura do Catálogo já foi idealizada e já está sendo abastecida com as unidades selecionadas. Será preciso ainda reavaliar as inserções das expressões no Catálogo, ponderar suas pertinências para o usuário aprendiz de tradução, explicitar as diferenças de usos nas notas de uso e implantar o funcionamento das vinculações entre as expressões recorrentes e os itens do Dicionário com elas relacionados.

Bolsistas SEAD atuantes:

Gustavo de Azambuja Feix (Bacharelado em Letras, Tradução, UFRGS)

Adriano Zanette (Ciência da Computação, UFRGS).