
- **LINGÜÍSTICA DE CORPUS I**

Coordenador(a): *Denise Delegá Lúcio*

A CONSTRUÇÃO DE UM CORPUS PARALELO E ALINHADO PORTUGUÊS-ITALIANO-PORTUGUÊS PARA O DOMÍNIO LITERÁRIO

Monique Lopes Ferraresi (UNESP), Claudia Zavaglia (UNESP)

Nesta comunicação, temos por objetivo apresentar para a comunidade científica o projeto de elaboração de um corpus paralelo português-italiano-português do domínio literário e a sua importância. Definimos corpus como sendo um conjunto de textos produzidos em determinada língua e sistematizado de acordo com critérios específicos, com a finalidade de produzir resultados proveitosos para a descrição e análise de dados lingüísticos. Nos últimos anos, houve uma grande tendência na utilização de corpora para a elaboração de dicionários, no ensino e na prática tradutória, no aprendizado de idiomas, na Tradução Automática e nas aplicações em

Processamento de Línguas Naturais (PLN). No entanto, é incipiente, ainda, a produção desse tipo de material lingüístico para o português-italiano e vice-versa, em qualquer domínio, do literário ao técnico-científico. É com o intuito de auxiliar esse tipo de trabalho que propomos a elaboração de um corpus paralelo, isto é, composto por textos originais e suas respectivas traduções nas línguas citadas anteriormente. Além de serem ordenados paralelamente, os textos que compõem esse corpus serão alinhados sentencialmente, ou seja, possuirão marcas que identificam os pontos de correspondências entre o texto original e sua tradução, baseando-se nas sentenças dos textos originais. Cada texto e cada sentença alinhada deverá conter etiquetas que indicarão, respectivamente, a obra de que se trata, bem como o início e o fim da sentença original e aquela da tradução. Para esse protótipo, serão utilizadas as obras literárias mais representativas a partir da década de 70 até os dias de hoje.

A RELEXICALIZAÇÃO ADJETIVAL NAS REDAÇÕES DE ALUNOS BRASILEIROS DE INGLÊS - UM ESTUDO BASEADO EM CORPUS

Denise Delegá Lúcio (PUC-SP)

Este trabalho tem como objetivo sugerir relexicalizações para adjetivos sobreusados por alunos brasileiros de inglês, contribuindo, dessa maneira para maior coerência do texto por eles escrito. Para tal serão examinados dois corpora: o do BrICLE, que é constituído por redações em inglês de brasileiros e o do BNC, que é um corpus nativo com 100 milhões de palavras. Primeiramente verificarei quais são os adjetivos sobreusados e seu padrão de uso no BrICLE. A seguir verificarei o padrão de uso daqueles adjetivos no BNC e, observando seus colocados principais, encontrar relexicalizações para tais adjetivos.

ANÁLISE DE FERRAMENTAS PARA A EXTRAÇÃO DE TERMOS

Luciana Latarini Ginezi (USP)

A atuação de profissionais na Interpretação Consecutiva, especialmente na área de acompanhamento, mostra a necessidade do conhecimento da língua falada em diversos níveis de registro e de dialeto. Devido a essa necessidade, resolveu-se elaborar um corpus falado onde constem as falas autênticas de profissionais em níveis de registro e dialeto diferenciados. Café é o tema do corpus por se tratar de uma especialidade da pesquisadora como Intérprete. Dentro do tema serão pesquisados o Processamento e a Colheita do Café, partindo da idéia de que nesses sub-temas haveria variação de registro e dialeto suficientes para o objetivo do estudo. Os profissionais inseridos na pesquisa são o trabalhador rural, o produtor de café e engenheiros agrônomos. Para o desenvolvimento geral de pesquisa serão utilizados os princípios fundamentais da Lingüística de Corpus, aplicados desde a elaboração até sua etapa final. Tão logo o corpus esteja completo, será possível extrair termos de cada nível de registro e analisá-los, comparando-os entre si de acordo com seu uso em situações específicas. A etapa final compreenderá a análise dessa lista terminológica nos diferentes níveis de registro e dialeto propostos, por meio de ferramentas computacionais utilizadas pela Lingüística de Corpus. O objetivo desta apresentação é mostrar algumas estratégias para a extração de termos de um Corpus por meio da análise de ferramentas computacionais disponíveis e apropriadas. Os caminhos traçados no desenvolver do trabalho mostram que não há apenas uma ferramenta apropriada para esta pesquisa, mas uma combinação delas para que se consiga obter o resultado esperado.

COESÃO E COERÊNCIA TEXTUAIS: UMA ABORDAGEM NO ENSINO FUNDAMENTAL

Luciana Mirabile (PUC-SP)

Nosso trabalho é direcionado aos estudiosos e professores de língua portuguesa e tem aplicação, para tais profissionais, no ensino da redação em língua materna nas escolas do Brasil.

Temos o intuito aqui de verificar como acontecem as manifestações da coerência e coesão textuais em textos escolares, utilizaremos, portanto, como corpus, as produções textuais de alunos de uma sexta série da rede pública de ensino.

RECURSOS DE PRESENÇA NAS CRÔNICAS DE MILLÔR FERNANDES

Marta Maria Pagadigorria (UNESP)

Tem esta comunicação o objetivo de descrever os recursos de presença estudados pela retórica, nas crônicas de Millôr Fernandes, a partir dos processos de projeção, que a atual lingüística cognitiva defende ser um dos mais férteis mecanismos cognitivos utilizados pelos seres humanos. Nossa hipótese é a de que a maneira como o autor utiliza esses processos para criar presenças constitui parte importante de seu estilo. Em uma de suas crônicas, por exemplo, Millôr diz, a respeito do escritor, que se ele não pegar pelos chifres a vontade de criar é bem possível que ela não se deixe agarrar em outra oportunidade. Trata-se da projeção da situação de capturar um animal esquivo (um boi, um bode?) sobre a situação de manter sob controle uma idéia criativa. Constituem recursos de presença obtidos por processos de projeção a parábola, a comparação a metáfora e a metonímia.