


Projeto COMET
avanços e carências

Stella E. O. Tagnin – FFLCH / DLM
 1º. Workshop de Linguística Computacional
 17/11/2009



Principal
 O Projeto
 Equipe
 CorTec
 CoMAprend
 CorTrad
 Artigos, etc.
 Links
 Informativo
 Contato
 Site FFLCH
 Site USP

PROJETO COrpus
Multilíngüe para
Ensino e
Tradução

Última atualização: outubro de 2009 (veja aqui)
 Webmasters: Elisa Duarte Teixeira
 Sabrina Matuda
 © Projeto Comet - FFLCH / USP

Auxílios do CNPq

- **COMET - CorTec: Coleta, organização e disponibilização *online* dos Corpora Técnicos do Corpus Multilíngüe para Ensino e Tradução (COMET)**
 Processo: 403120/03-9
- **COMET - *Corpus* Multilíngüe para Ensino e Tradução: disponibilização pública e gratuita de *corpora* e ferramentas de exploração (fase 2) -**
 Processo: 400988/2006-2



Principal
 O Projeto
 Equipe
 CorTec
 CoMAprend
 CorTrad
 Artigos, etc.
 Links
 Informativo
 Contato
 Site FFLCH
 Site USP

PROJETO COrpus
Multilíngüe para
Ensino e
Tradução

Última atualização: outubro de 2009 (veja aqui)
 Webmasters: Elisa Duarte Teixeira
 Sabrina Matuda
 © Projeto Comet - FFLCH / USP

COMET

- **CorTec** – corpus técnico, COMPARÁVEL, textos originais em ambas as línguas (inglês e português)
- **CoMAprend** – corpus de aprendizes, redações, 5 línguas: alemão, espanhol, francês, inglês, italiano.
- **CorTrad** – corpus PARALELO, originais e respectivas traduções: jornalístico, técnico-científico, literário.



Principal
 O Projeto
 Equipe
 CorTec
 CoMAprend
 CorTrad
 Artigos, etc.
 Links
 Informativo
 Contato
 Site FFLCH
 Site USP

CORTEC

Busca: O que é:
 Passo 1/3: Escolhendo os Corpora.
 Abaixo estão listados os corpora que compõem o CorTec.
 Selecione os corpora que deseja pesquisar:

Lista de Corpora

- Culinária
- Ecoturismo
- Hipertensão
- Informática
- Instrumentos Contratuais
- Astronomia
- Insuficiência Renal
- Linguística
- Medidores Eletromagnéticos de Vazão
- Suplementos Nutricionais
- Computação
- Futebol
- Café
- Turismo Cultural
- Esportes 2

14 corpora construídos para pesquisas de pós ou em cursos de tradução

COMET
Corpus Multilingue para Ensino e Tradução

Principal
O Projeto
Equipe
Co/Tec
Co/Trad
MÁprend
Co/Trad
Rigos, etc.
Links
Formativo
Contato
Re FFLCH
Site USP

Os Corpora Técnicos

1 - **Corpus de Culinária - Receitas**

O corpus de **Culinária** é um corpus comparável composto de 1.535 receitas originalmente escritas em português brasileiro e 2.076 em inglês britânico, coletadas da Internet nos últimos quatro anos. Todas as categorias culinárias estão representadas nesse corpus, sejam receitas doces ou salgadas.

Culinária	Inglês	Português
Ocorrências/tokens	368.227	252.149
Formas/types	7.281	7.150
T/T ratio	1,98	2,84

Este corpus foi compilado por **Elisa Duarte Teixeira**.

2 - **Corpus de Meio Ambiente - Ecoturismo**

O corpus de **Meio Ambiente** é composto por textos originais de conteúdo comparável em inglês e português sobre **Ecoturismo**, retirados de sites do governo, de entidades ambientalistas e de agências de turismo.

Ecoturismo	Inglês	Português
Ocorrências/tokens	201.826	200.887
Formas/types	10.009	17.938
T/T ratio	4,96	8,93

Este corpus foi compilado por **Josimeire Martins**.

COMET
Corpus Multilingue para Ensino e Tradução

Principal
O Projeto
Equipe
Co/Tec
Co/Trad
MÁprend
Co/Trad
Rigos, etc.
Links
Formativo
Contato
Re FFLCH
Site USP

Abaixo estão listados os corpora que compõem o Co/Tec. Seleccione os corpora que deseja pesquisar:

Lista de Corpora

- Culinária
- Ecoturismo
- Hipertensão
- Informática
- Instrumentos Contratuais
- Astronomia
- Insuficiência Renal
- Linguística
- Medidores Eletromagnéticos de Vazio
- Suplementos Nutricionais
- Computação
- Futebol
- Café
- Turismo Cultural
- Culinária 2

Lingua

- Português
- Inglês

Co/Tec - **Corpus Técnico Científico** >> Próximo Passo

COMET
Corpus Multilingue para Ensino e Tradução

Principal
O Projeto
Equipe
Co/Tec
Co/Trad
MÁprend
Co/Trad
Rigos, etc.
Links
Formativo
Contato
Re FFLCH
Site USP

português / inglês

Busca: O que é:

Passo 2/3: Escolhendo a Ferramenta.

Selecione abaixo a ferramenta a ser aplicada:

Lista de Ferramentas

- Concordanciador
- Gerador de Lista de Palavras
- Gerador de N-Gramas

Co/Tec - **Corpus Técnico Científico** >> Próximo Passo

COMET
Corpus Multilingue para Ensino e Tradução

Principal
O Projeto
Equipe
Co/Tec
Co/Trad
MÁprend
Co/Trad
Rigos, etc.
Links
Formativo
Contato
Re FFLCH
Site USP

bem vindo ao site do Proj

Busca: O que é:

Passo 3/3: Aplicando a Ferramenta sobre o Corpus.

Configure a ferramenta usando as opções abaixo:

Gerador de Lista de Palavras (Contador padrão)

Ordenado por:

- Ordem de frequência (decrecente)
- Ordem Alfabética

>> Aplicar Gerador de Lista de Palavras

Gerador de Lista de Palavras (Contador por palavra)

Número de palavras acima e abaixo da palavra escolhida:

Digite a palavra de busca:

>> Aplicar Gerador de Lista de Palavras

Escolher outra Ferramenta <<

COMET
Corpus Multilingue para Ensino e Tradução

Principal
O Projeto
Equipe
Co/Tec
Co/Trad
MÁprend
Co/Trad
Rigos, etc.
Links
Formativo
Contato
Re FFLCH
Site USP

Seja bem vindo ao site

Fazer DOWNLOAD do Resultado.

Ocorrências (tokens), considerando a Lexia Complexa de Nomes Próprios

Total de Ocorrências: 110824
Total de Ocorrências que aparecem uma vez: 6360
Total de Ocorrências que aparecem mais de uma vez: 104464

Palavras únicas (formas (types), considerando a Lexia Complexa de Nomes Próprios

Total de Palavras: 11185
Total de Palavras que aparecem uma vez: 5631
Total de Palavras que aparecem mais de uma vez: 5554

Índice Vocabular (tokens/type ratio)

Token/Type: 0.4966

TABELA DE FREQUÊNCIA

Pos.	Palavra	Freq.
1	o	4860
2	a	4312
3	de	3885
4	e	2810
5	que	2467

COMET
Corpus Multilingue para Ensino e Tradução

Principal
O Projeto
Equipe
Co/Tec
Co/Trad
MÁprend
Co/Trad
Rigos, etc.
Links
Formativo
Contato
Re FFLCH
Site USP

Seja bem vindo

Busca: O que é:

Passo 3/3: Aplicando a Ferramenta sobre o Corpus.

Configure a ferramenta usando as opções abaixo:

Gerador de Lista de Palavras (Contador padrão)

Ordenado por:

- Ordem de frequência (decrecente)
- Ordem Alfabética

>> Aplicar Gerador de Lista de Palavras

Gerador de Lista de Palavras (Contador por palavra)

Número de palavras acima e abaixo da palavra escolhida:

Digite a palavra de busca:

>> Aplicar Gerador de Lista de Palavras

Escolher outra Ferramenta <<

Principal
O Projeto
Equipe
CorTec
CMAprend
CorTrad
Rigos, etc.
Links
Formativo
Contato
Site FFLCH
Site USP

COMET
Copa Multilingue para Ensino e Tradução

USP

Confira

CORTEC

Busca: O que é

Fazer DOWNLOAD do Resultado.

árbitro	432
jogador	429
ou	411
dos	409
partida	406
->gol	382
equipe	379
ser	368
como	368
contra	352
time	347

>> Fazer Nova Busca

CorTec - Corpus Técnico-Científico

Principal
O Projeto
Equipe
CorTec
CMAprend
CorTrad
Rigos, etc.
Links
Formativo
Contato
Site FFLCH
Site USP

COMET
Copa Multilingue para Ensino e Tradução

Português / Inglês

CORTEC

Busca: O que é

Passo 2/3: Escolhendo a Ferramenta.

Selecione abaixo a ferramenta a ser aplicada:

Lista de Ferramentas

- Concordanciador
- Gerador de Lista de Palavras
- Gerador de N-Gramas

>> Próximo Passo

CorTec - Corpus Técnico-Científico

Login

Principal
O Projeto
Equipe
CorTec
CMAprend
CorTrad
Rigos, etc.
Links
Formativo
Contato
Site FFLCH
Site USP

COMET
Copa Multilingue para Ensino e Tradução

Ses no link Artigos

CORTEC

Busca: O que é

Passo 3/3: Aplicando a Ferramenta sobre o Corpus.

Configure a ferramenta usando as opções abaixo:

Concordanciador

Expressão ou Palavra: igual a gol

Diferenciar maiúsculas: Condição com Terminado com Contendo

Tamanho do contexto reduzido (caracteres): 30

>> Aplicar Concordanciador

Escolher outra Ferramenta <<

CorTec - Corpus Técnico-Científico

Login

Principal
O Projeto
Equipe
CorTec
CMAprend
CorTrad
Rigos, etc.
Links
Formativo
Contato
Site FFLCH
Site USP

COMET
Copa Multilingue para Ensino e Tradução

ando ao site do Projeto Co

CORTEC

Busca: O que é

Foram encontradas 308 ocorrências!! Fazer DOWNLOAD do resultado.

Clique nas palavras de busca para obter um contexto expandido com 150 caracteres.

1 jogo e fora de jogo 34 10 0 gol marcado 35 11 Impedimento 36
 2 a partida deve incluir se um gol é ou não marcado e o resultado
 3 go da partida - depois de um gol ter sido marcado - no começo
 4 opção, quando for o caso Um gol pode ser marcado diretamente
 5 pois que uma equipe marcou um gol, o tiro de saída é executado
 6 go de jogo 36 REGRA 10 O GOL MARCADO Um gol é marcado
 7 36 REGRA 10 O GOL MARCADO Um gol é marcado que
 8 GOL MARCADO Gol marcado Um gol é marcado quando a bola ultra
 9 ate pela equipe que marcou o gol. Equipe vencedora A equipe
 10 número de gols, ou se nenhum gol é marcado, a partida termina
 11 er outra pessoa 4, impede um gol ou frustra uma oportunidade c
 12 tra uma oportunidade clara de gol da equipe adversária imediat
 13 tra uma oportunidade clara de gol de um adversário, que se mov
 14 tira a camisa ao comemorar um gol deve ser advertido (cartão a
 15 a adversária, e concedido um gol - se um tiro livre direto a
 16 go. A bola entra na meta Um gol pode ser marcado somente se a
 17 esta a bola está em jogo. Um gol pode ser marcado diretamente
 18 ja executado ou repetido, um gol é concedido se, antes de pass
 19 entra na meta, é concedido um gol - se a bola não entra na met
 20 entra na meta, é concedido um gol - se a bola não entra na met

Principal
O Projeto
Equipe
CorTec
CMAprend
CorTrad
Rigos, etc.
Links
Formativo
Contato
Site FFLCH
Site USP

COMET
Copa Multilingue para Ensino e Tradução

vo corpus paralelo multiver

Ocorrência 18

o o tempo de jogo foi prorrogado ao primeiro tempo ou ao final do tempo regulamentar para permitir que um tiro penal seja executado ou repetido, um gol é concedido se, antes de passar entre os postes e abaixo do travessão:

- a bola toca em um ou ambos os postes e ou no travessão e ou no gol

Ocorrência 19

uma infração

O árbitro infringe as Regras do Jogo:

- o árbitro permite que o tiro penal seja executado
- se a bola entra na meta, é concedido um gol
- se a bola não entra na meta, o tiro penal é repetido

Um componente do mecanismo do tiro infringe as Regras do Jogo:

- o árbitro permite que o o

Ocorrência 20

Um componente do mecanismo do tiro infringe as Regras do Jogo

- o árbitro permite que o tiro penal seja executado
- se a bola entra na meta, é concedido um gol
- se a bola não entra na meta, o tiro penal é repetido

Um jogador da equipe defensora e um da equipe atacante infringe as

Principal
O Projeto
Equipe
CorTec
CMAprend
CorTrad
Rigos, etc.
Links
Formativo
Contato
Site FFLCH
Site USP

COMET
Copa Multilingue para Ensino e Tradução

nas publicações no link Art

CORTEC

Busca: O que é

Passo 2/3: Escolhendo a Ferramenta.

Selecione abaixo a ferramenta a ser aplicada:

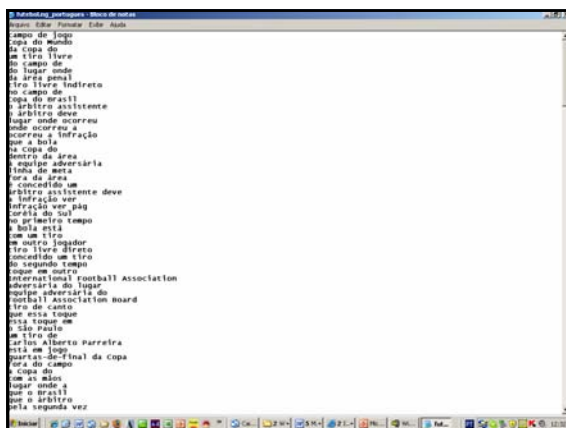
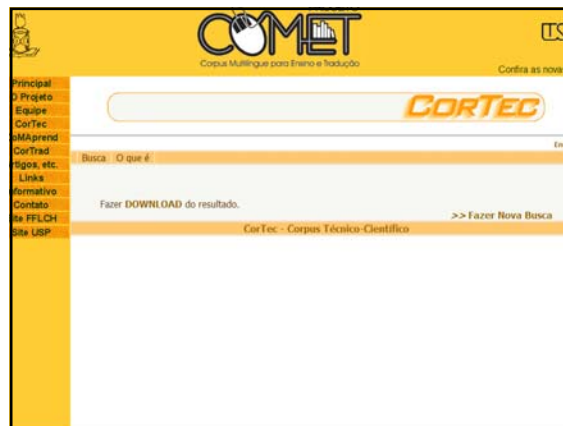
Lista de Ferramentas

- Concordanciador
- Gerador de Lista de Palavras
- Gerador de N-Gramas

>> Próximo Passo

CorTec - Corpus Técnico-Científico

Login



CorTec

Executora da 2ª. fase: Empresa Jr. do ICMC/USP – São Carlos

Avanços:

- Passamos de 5 para 14 *corpora*

Carências:

- Know-how para inclusão de novos *corpora*
- Dificuldade de diálogo com a empresa júnior
- Rotatividade dos envolvidos no projeto (empresa júnior)

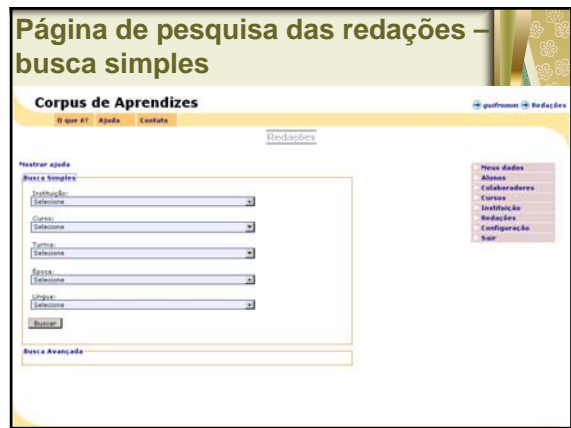
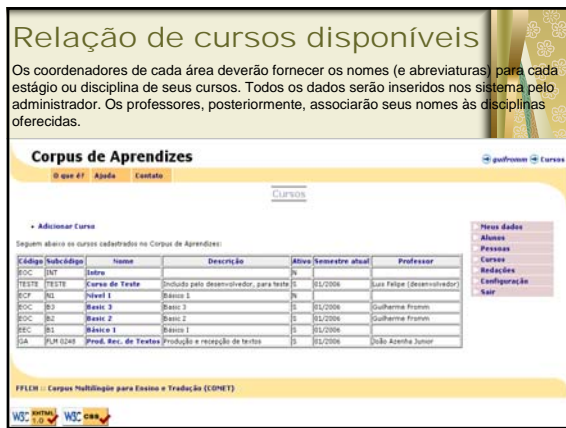
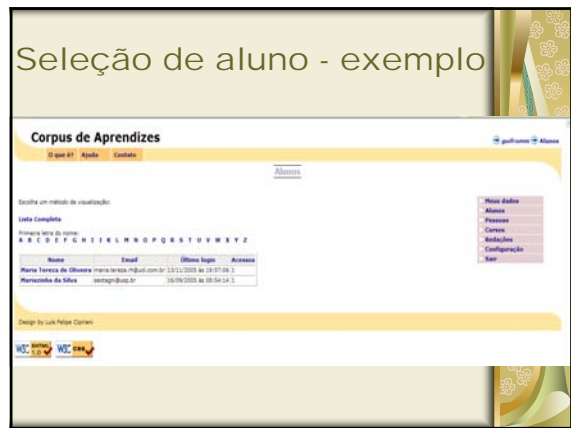
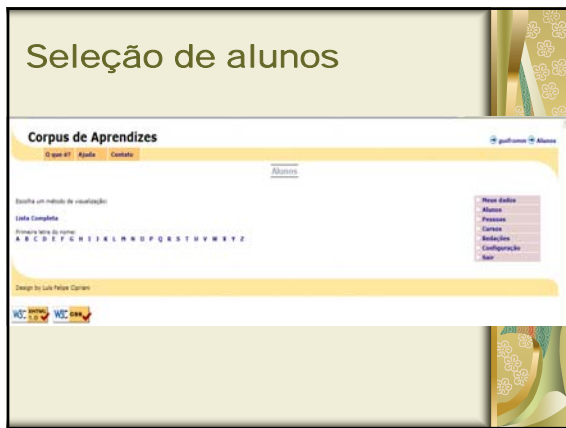
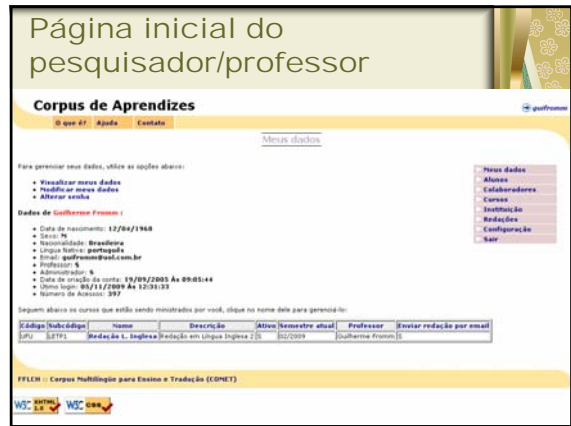
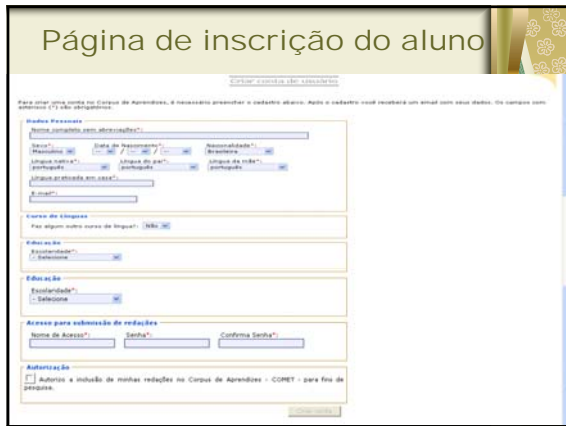
CoMAprend

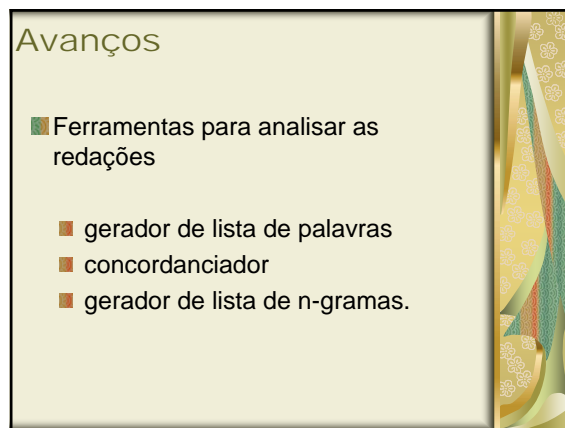
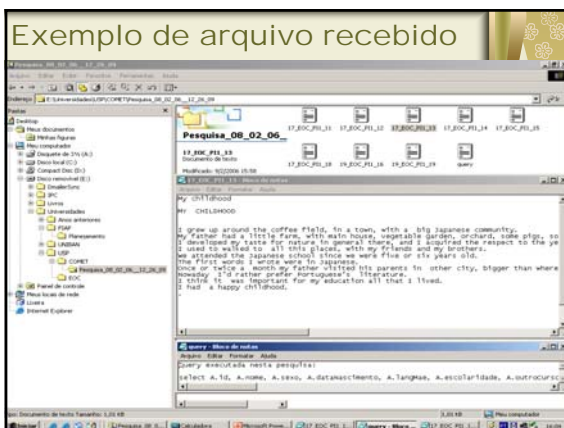
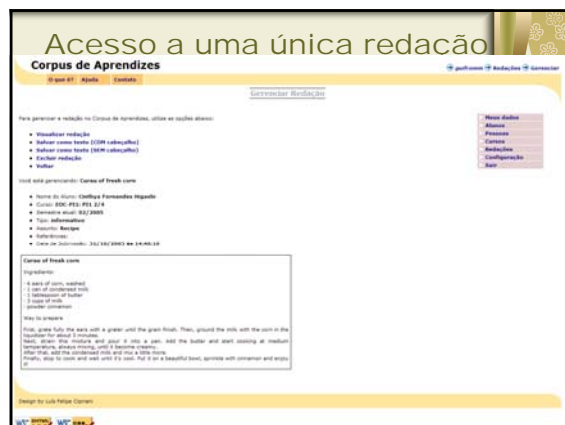
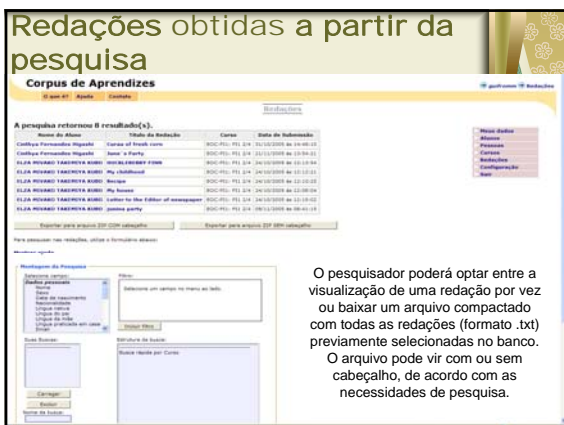
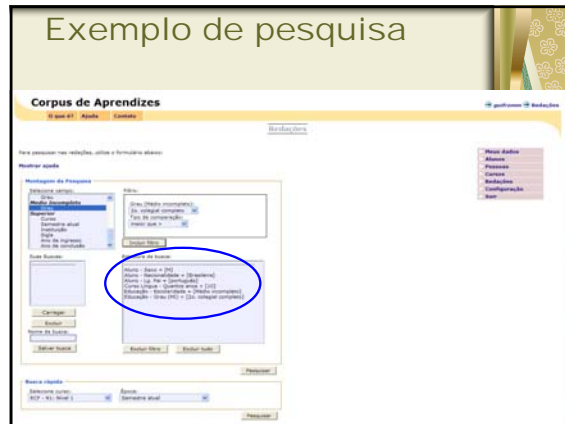
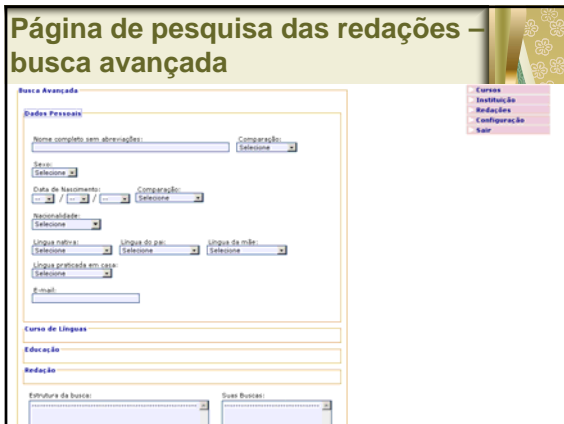
CORPUS DE APRENDIZES – COMET/USP

Desenvolvimento:

Profa. Dra. Stella E. O Tagnin – FFLCH/USP
 Prof. Dr. Guilherme Fromm – (FFLCH/USP) ILEEL/UFU
 Luis Felipe Cipriani – ICMC USP/S.Carlos







Ferramentas de análise

2º passo: escolher as redações desejadas.

Lista de Corpora

- Pretty or unpretty?
- Mid-term test
- English: world's second language
- The "Black English" nowadays
- Language is just a tool for communication. There can only be growth and reflection in Literature
- Is literature a reflection of the reality or it has an own autonomy?
- How local identity can be influenced by external values
- Hope and Faith, even against all odds
- LITERATURE AND SOCIETY
- English as a foreign language: threaten or not?
- Prosa - Comarc, Escola 3
- A Restaurant Review
- Writing 1
- Frankenstein
- A letter to an advice columnist
- A letter to an old friend
- A Magazine Article
- Individualität muss bewährt werden
- Berndt
- Wie findest du Weihnachten?
- Unsiene Anstrengungen sollen auch gezeitigt werden
- Language is just a tool for communication: there can only be reflection and growth in literature
- Mid-term Test
- Prosa
- Reflection upon language
- The power of romantic creation through society
- A little about a friend
- The advantages and disadvantages of being a pharmacist
- Just some requests...

Ferramentas de análise

3º passo: aplicar uma das ferramentas desejadas.

Corpus de Aprendiz

COMPREND

Selecionar Todas | Desmarcar Todas | Voltar ao Passo 1

Passo 2/3: Escolhendo a Ferramenta.

Selecione abaixo a ferramenta a ser aplicada:

Lista de Ferramentas

- Concordanciador
- Gerador de Lista de Palavras
- Gerador de N-Gramas

>> Próximo Passo

FFLCH - Corpus Multilingua Ensino e Tradução (COMET)

Ferramentas de análise

4º passo: escolher a forma de busca.

Corpus de Aprendiz

COMPREND

Selecionar Todas | Desmarcar Todas | Voltar ao Passo 1

Passo 3/3: Aplicando a Ferramenta sobre o Corpus.

Configure a ferramenta usando as opções abaixo:

Concordanciador

Expressão ou Palavra: [igreja] [out]

Diferenciar maiúsculas e minúsculas:

Tamanho do contexto reduzido (caracteres): [15]

>> Aplicar Concordanciador

Escolher outra Ferramenta <<

Ferramentas de análise

5º passo: obter os resultados.

Corpus de Aprendiz

COMPREND

Foram encontradas 18 ocorrências!! Fazer DOWNLOAD do resultado.

Clique na palavra de busca para obter um contexto expandido com 150 caracteres

1 and personal dramas but they also have reve
 2 mily to bankruptcy, but he still refuses to
 3 d as a Manicheism, but the movie makes use
 4 nd our own identity, but it also helps us fo
 5 is a possible way, but no the only one.
 6 y exist separately, but the ideal would be
 7 without each other, but we can better appre
 8 through literature but to have a better un
 9 how to communicate, but on the essential is
 10 nds both languages, but also communicates a
 11 s not only talking, but thinking in differe
 12 English" (AAVE). But what exactly is the
 13 angerous Minds. "But I ain't never cro
 14 really hate to trip but I gotta loc (...)
 15 out" (to relax). But why is it that the
 16 o Standard English, but how it can help the
 17 evaluate arguments. But language can also b
 18 scriptive function, but an action function.

CoMAprend - carências

1. O Comaprend e o Cortec poderiam ter uma lista de exclusão (stoplist) do português e do inglês na Lista de Palavras.
2. O download de listas está nos trazendo listagens em .dat, ao invés de .xls ou .txt.
3. Não temos como trabalhar os Lemas nestas listagens.
4. Gerador de lista de palavras-chave
 - Lacio-Ref como corpus de referência para o português???
 - ... e para o inglês???
 - ... e as outras línguas???

COMET

Corpus Multilingua para Ensino e Tradução

Bem-vindo ao CorTrad

O CorTrad é o corpus paralelo de tradução (português-ingles) e de tradução (ingles-português) do mesmo texto original, versões revisadas e tradução por outros tradutores. Perguntas já respondidas

O CorTrad é um corpus aberto e conta atualmente com:

- CorTrad jornalístico (por área: divulgação científica)
- CorTrad literário (por obra: contos)
- CorTrad técnico-científico (por área: ciências)

A lista das obras incluídas, assim como os agradecimentos devidos, encontram-se em Agradecimentos.

A disponibilização do CorTrad na rede é um projeto conjunto COMET/INCL/Inglês/UEA, usando o sistema DISPARA.

Comentários ou questões para equipe do CorTrad

COMET
Cópia Multilingue para Ensino e Tradução

Seja bem-vindo

Principal | Projeto | Equipe | CorTec | CoTéc | MMaprend | CorTrad | Rigos, etc. | Links | Formativo | Contato | FFLCH | Site USP

Projeto conjunto COMET/NILC/Linguoteca

As responsabilidades dos parceiros neste projeto conjunto são:

- COMET: planejamento do corpus, criação e fornecimento dos textos em formato eletrônico, negociação de autorizações de uso e apresentação do site.
- Responsáveis: Stella E. O. Taglio e Elise D. Tulecia
- Desenho/concepção gráfica do site do CorTrad: Patricia Taglio
- Linguoteca: aplicação do sistema DSPARA ao CorTrad; criação dos corpora em formato IMS (Open) CVID; desenvolvimento de novas funcionalidades para este projeto específico, em particular a procura em várias traduções e a adaptação da interface para textos técnicos; documentação e implementação para aumento posterior do corpus.
- Responsável: Diana Santos
- NILC: alinhamento do CorTrad em seu servidor; manutenção e aperfeiçoamento da parte computacional do serviço a partir da data de uma primeira versão estável.
- Responsável: Sandra Aluísio

De acordo com o art. 48, inciso VII da Lei nº 9.610, DE 19 DE FEVEREIRO DE 1998, não constitui infração aos direitos autorais "a reprodução, em qualquer obra, de pequenos trechos de obras preexistentes, de qualquer natureza, (...), sempre que a reprodução em si não seja o objetivo principal da obra nova e que não prejudique a exploração normal da obra reproduzida nem cause um prejuízo injustificado aos legítimos interesses dos autores".

Se você constatar a presença de textos no CorTrad que ferem os direitos autorais, por favor, entre em contato conosco e tornaremos as providências cabíveis.

Comentários ou questões para equipe do CorTrad

COMET
Cópia Multilingue para Ensino e Tradução

Conheça o CorTrad - nos

CorTrad técnico-científico culinária

O CorTrad é um corpus aberto, sujeito a alterações. Clique em atualizações para saber mais.

Principal | Projeto | Equipe | CorTec | CoTéc | MMaprend | CorTrad | Rigos, etc. | Links | Formativo | Contato | FFLCH | Site USP

CorTrad técnico-científico culinária

O CorTrad é um corpus aberto, sujeito a alterações. Clique em atualizações para saber mais.

Principal | Projeto | Equipe | CorTec | CoTéc | MMaprend | CorTrad | Rigos, etc. | Links | Formativo | Contato | FFLCH | Site USP

CorTrad técnico-científico culinária

O CorTrad é um corpus aberto, sujeito a alterações. Clique em atualizações para saber mais.

Principal | Projeto | Equipe | CorTec | CoTéc | MMaprend | CorTrad | Rigos, etc. | Links | Formativo | Contato | FFLCH | Site USP

Original: principal | ver

Primeira tradução: principal | ver

Tradução revisada: principal | ver

Ignorar maiúsculas/minúsculas

Resultado

- Concordância
- Distribuição dos lemas
- Distribuição do tempo verbal e/ou do caso pronominal
- Distribuição do gênero morfológico
- Distribuição por parte de obra
- Distribuição por receita
- Distribuição por grupo (de cor, de vestuário, etc.)
- Distribuição da categoria gramatical (Ps5)
- Distribuição de pessoa e/ou número
- Distribuição da função sintática
- Distribuição por tipo de texto
- Distribuição por campo semântico

Opções

COMET
Cópia Multilingue para Ensino e Tradução

multiversão portuguesa / inglês

CorTrad técnico-científico culinária

Expressão de busca "picadinho"
Resultado escolhido: concordância em contexto
Corpus pesquisado: original

60 ocorrências.

Original	Primeira tradução	Tradução revisada
1 ramo de alecrim picadinho	1 sprig rosemary	1 sprig rosemary
1 dente de alho picadinho	1 garlic clove, chopped	1 garlic clove, minced
1 dente de alho picadinho, já alveitado se quiser	2 tablespoons olive oil	1 clove garlic, chopped (previously boiled, if you want a milder flavor)
1 dente de alho picadinho	1 garlic clove, finely chopped	1 garlic clove, minced
1 dente de alho picadinho	1 garlic clove, finely chopped	1 garlic clove, minced
1 dente de alho picadinho	1 garlic clove, finely chopped	1 garlic clove, minced
1 dente de alho picadinho	1 garlic clove, chopped	1 garlic clove, minced
1 dente de alho picadinho	1 garlic clove, chopped	1 garlic clove, minced
1 colher (sopa) de endro (80) picadinho	1 tablespoon finely chopped dill weed	1 tablespoon minced dill weed
1 dente de alho picadinho	1 garlic clove, finely chopped	1 garlic clove, minced
1/2 colher (chá) de alho picadinho	SEM TRADUÇÃO	hour to dust
2 colheres (sopa) de endro (80) picadinho	2 garlic cloves, baked or boiled	2 heaping tablespoons of 'Ancienne Dijon mustard (white-gran old-fashioned mustard) 2 tablespoons tarragon
para a cobertura 200 g de chocolate ao leite picadinho	200 g (7 oz) finely chopped milk chocolate	hour to dust 200 g (7 oz) finely chopped milk chocolate

COMET
Cópia Multilingue para Ensino e Tradução

nosso novo corpus paralelo

CorTrad literário contos

O CorTrad é um corpus aberto, sujeito a alterações. Clique em atualizações para saber mais.

Principal | Projeto | Equipe | CorTec | CoTéc | MMaprend | CorTrad | Rigos, etc. | Links | Formativo | Contato | FFLCH | Site USP

CorTrad literário contos

O CorTrad é um corpus aberto, sujeito a alterações. Clique em atualizações para saber mais.

Principal | Projeto | Equipe | CorTec | CoTéc | MMaprend | CorTrad | Rigos, etc. | Links | Formativo | Contato | FFLCH | Site USP

Original: principal | ver

Primeira tradução: principal | ver

Tradução revisada: principal | ver

Tradução publicada: principal | ver

Ignorar maiúsculas/minúsculas

Resultado

- Concordância
- Distribuição por obra
- Resultados por ordem alfabética (só distribuições)
- Resultados numerados (só concordância)
- Distribuição das formas
- Distribuição por autor

Opções

COMET
Cópia Multilingue para Ensino e Tradução

Seja bem-vindo ao s

Principal | Projeto | Equipe | CorTec | CoTéc | MMaprend | CorTrad | Rigos, etc. | Links | Formativo | Contato | FFLCH | Site USP

Ajuda

Se tem dúvidas sobre o sistema de pesquisa, pode ver por ora as seguintes páginas feitas para projetos relacionados:

- Exemplo de pesquisa no AC/DC
- Ajuda à pesquisa no COMPARA

Em breve mais informação específica de ajuda ao próprio CorTrad será colocada aqui.

Comentários ou questões para equipe do CorTrad

COMET
Cópia Multilingue para Ensino e Tradução

off

Exemplos simples e diversificados: uma primeira visão

- a forma caso do verbo CASAR [casar<casar> & pos<V>,*']
- ocorrências do verbo DESEJAR [desear<desear>']
- adjetivos ligados de acesso ao masculino plural [pos<ADJ> & gen<M> & presnum<P>] [pos<M> & gen<M> & presnum<P>']
- os objectos do verbo MATAR no mínimo quatro palavras distintas [matar<matar>] [[10,3] [Euso<GADQ>']
- palavras identificadas como verbos pelo analisador através de heurísticas morfológicas [pos<V>,*' & desini<DEB,*>']
- formas do supletivo do conjuntivo [conjug<CONJ_F_203,*>']

Exemplos de consultas por atributos morfológicos

Linguoteca: AC/DC

Quer só tempos simples do passado?

Process: [tempo<+ (P) DEF>,*']
Resultado: Distribuição do tempo verbal

Note-se que, se pedir a concordância, esta não é apresentada por uma tal procura, de, em geral, um número excessivo de ocorrências. Experimente, por exemplo, restringir à terceira conjugação:

[lema<*,> & tempo<+ (P) DEF>,*']
Resultado: Concordância

Quer só o pretérito perfeito?

[tempo<+PS,*>']
Resultado: Distribuição das lemas

Mais sobre número excessivo. Se quiser ver apenas as primeiras 100 ocorrências, veja

COMET
Corpus Multilingue para Ensino e Tradução

Principal
Projeto
Equipe
Co/Tec
Co/Trad
SIGOS, etc.
Links
Formativo
Contato
FFLCH
Site USP

Resultado

Concordância Distribuição das formas

Distribuição por obra Distribuição por autor

Opções

Resultados por ordem alfabética (só distribuições)

Resultados numerados (só concordância)

Exemplos de pesquisas

	Procurar:	Resultado:
a palavra vovno	violino	concordância
palavras começando por eqic	"açúc.*"	concordância
palavras antecedidas por dois pontos	":* #[]	distribuição das formas
formas do verbo reunir em contexto	reunir.*	concordância
palavras modificadas por muito	"muito" #[]	distribuição de formas

Veja dados quantitativos para mais informações.

Última atualização do sistema: 03 de novembro de 2009
Última atualização do corpus: 12 de Fevereiro de 2009

[Comentários ou questões para equipe do Co/Tec](#)

COMET
Corpus Multilingue para Ensino e Tradução

Principal
Projeto
Equipe
Co/Tec
Co/Trad
SIGOS, etc.
Links
Formativo
Contato
FFLCH
Site USP

Início | **Jornalística** | Literário | Técnico-científica

CorTrad literário contos

48 casos

Expressão de busca: "walk"
Resultado escolhido: **distribuição das formas**
Corpus pesquisado: **originais**

Houve 5 valores diferentes de forma.

walked	20
walk	12
walking	8
walks	6
walkabout	2

Esperamos que o CorTrad literário contos lhe tenha sido útil!

COMET
Corpus Multilingue para Ensino e Tradução

Principal
Projeto
Equipe
Co/Tec
Co/Trad
SIGOS, etc.
Links
Formativo
Contato
FFLCH
Site USP

CorTrad literário contos

O CorTrad é um corpus aberto, sujeito a alterações. Clique em atualizações para saber mais.

Pesquisar no corpus

Original principal although ser ver

Primeira tradução principal ser ver

Tradução revisada principal ser ver

Tradução publicada principal ser ver

Ignorar maiúsculas/minúsculas

Resultado

Concordância Distribuição das formas

Distribuição por obra Distribuição por autor

Opções

Resultados por ordem alfabética (só distribuições)

Resultados numerados (só concordância)

COMET
Corpus Multilingue para Ensino e Tradução

Principal
Projeto
Equipe
Co/Tec
Co/Trad
SIGOS, etc.
Links
Formativo
Contato
FFLCH
Site USP

Início | **Jornalística** | Literário | Técnico-científica

CorTrad literário contos

Expressão de busca: "although"
Resultado escolhido: **concordância em contexto**
Corpus pesquisado: **originais**

8 ocorrências.

Original	Primeira tradução	Tradução revisada	Tradução publicada
Father O'Shaughnessy could vouch for this, although he had not had the privilege, etcetera, etc.	Pai de O'Shaughnessy poderia responder por isso, embora ele não tivesse tido o privilégio de conhecê-lo, etc.	Pai de O'Shaughnessy podia garantir isso, embora ele não tivesse tido o privilégio de conhecê-lo, etc.	Pai de O'Shaughnessy podia garantir isso, embora ele não tivesse tido o privilégio de conhecê-lo, etc.
the dust shot out from under my shoes as I trudged along Stadium Street, for although I had intended to spend my holiday at Paton, Jack-Owren had at once suggested.	O pó ergueu-se sob meus pés conforme eu amarrota-me pela Rua Stadium, já que, embora eu tivesse planejado passar minhas férias em Paton, Tia Ouzena havia imediatamente insinuado: sua consciência vai permitir, com todos esses refugiados dormindo em colchões no corredor?	O pó ergueu-se sob meus pés conforme eu me amarrota pela Rua Stadium, já que, embora eu tivesse planejado passar as férias em Paton, Tia Ouzena imediatamente insinuou: sua consciência vai permitir, com todos esses refugiados dormindo em colchões no corredor?	O pó ergueu-se sob meus pés conforme eu me amarrota pela Rua Stadium, já que, embora eu tivesse planejado passar as férias em Paton, Tia Ouzena imediatamente insinuou: sua consciência vai permitir, com todos esses refugiados dormindo em colchões no corredor?
Mr. Trink did not care if she had broached or teased of them, although at the Christmas party, after the third or fourth glass, he had held her against him as they danced and told her she was a fine figure of a woman.	O Senhor Trink não se importava se ela havia tido um namorado ou vinte, embora na festa de Natal, após o terceiro ou quarto copo ele a tivesse segurado junto a si quando dançavam e lhe tivesse dito que tinha um belo físico de	O Senhor Trink não se importava se ela havia tido um namorado ou vinte, embora na festa de Natal, após o terceiro ou quarto copo ele a tivesse segurado junto a si quando dançavam e lhe tivesse dito que tinha um belo físico de	O Senhor Trink não se importava se ela havia tido um namorado ou vinte, embora na festa de Natal, após o terceiro ou quarto copo, ele a tivesse segurado junto a ela quando dançavam e lhe tivesse dito que ela tinha um belo físico.

COMET
Corpus Multilingue para Ensino e Tradução

Principal
Projeto
Equipe
Co/Tec
Co/Trad
SIGOS, etc.
Links
Formativo
Contato
FFLCH
Site USP

Início | **Jornalística** | Literário | Técnico-científica

CorTrad literário contos

O CorTrad é um corpus aberto, sujeito a alterações. Clique em atualizações para saber mais.

Pesquisar no corpus

Original principal Pan ser ver

Primeira tradução principal ser ver

Tradução revisada principal ser ver

Tradução publicada principal ser ver

Ignorar maiúsculas/minúsculas

Resultado

Concordância Distribuição das formas

Distribuição por obra Distribuição por autor

Opções

Resultados por ordem alfabética (só distribuições)

Resultados numerados (só concordância)

COMET
Corpus Multilingue para Ensino e Tradução

Principal
Projeto
Equipe
Co/Tec
Co/Trad
SIGOS, etc.
Links
Formativo
Contato
FFLCH
Site USP

Início | **Jornalística** | Literário | Técnico-científica

CorTrad literário contos

Expressão de busca: "then"
Resultado escolhido: **concordância em contexto**
Corpus pesquisado: **originais**

71 ocorrências.

Original	Primeira tradução	Tradução revisada	Tradução publicada
She was <u>then</u> a woman of forty, straight, <u>businesslike</u> in a tailored suit and with her hair cut in a silver helmet.	Era, aquela altura, uma mulher de quarenta e nove <u>profissional</u> e anos, jovem, ereta, <u>profissional</u> trajando um conjunto feito sob medida e com o cabelo cortado em forma de elmo prateado.	Era, aquela altura, uma mulher de quarenta e nove <u>profissional</u> e anos, jovem, ereta, <u>profissional</u> trajando um conjunto feito sob medida e com o cabelo cortado em forma de elmo prateado.	Aquela altura ela era uma mulher <u>profissional</u> e não se preocupava <u>profissional</u> trajando um conjunto.
All the more alarming then, as I sat eating on one of the cane-bottomed chairs in the corridor, while Ben Stenberg, Miss McInryne's star pupil, played the Man Struck, to hear the same waddy transmuted.	Ainda mais alarmante — quando eu estava sentado esperando em uma das cadeiras de assento de palha no corredor enquanto Ben Stenberg, o aluno modelo de Miss McInryne, tocava Man Struck — ao escutar a mesma voz estranhamente alterada.	Ainda mais alarmante — quando eu estava sentado esperando em uma das cadeiras de assento de palha no corredor enquanto Ben Stenberg, o aluno modelo de Miss McInryne, tocava Man Struck — ao escutar a mesma voz estranhamente alterada.	Mais alarmante ainda — um dia, em que eu esperava sentado numa das cadeiras de palha no corredor enquanto Ben Stenberg, o aluno modelo de Miss McInryne, tocava Man Struck — ao escutar a mesma voz estranhamente alterada.
We were alone for a little <u>then</u> .	SEM TRADUÇÃO	Depois disso, ficamos sozinhos por um tempo.	Depois disso, ficamos sozinhos por um tempo.
"Which one are you <u>then</u> ?"	— Ora, ora, qual deles é você?	— Ora, ora, qual deles é você?	— Ora, ora, qual deles é você?
Just then my Aunt Ouzena came.	Nesse momento, minha Tia	Nesse momento, minha Tia	Nesse momento, minha Tia

CorTrad

Avanços (em relação a outros corpora paralelos inglês-português)

- **Multiversão** – permite comparar várias etapas da tradução / revisão
- **Sistema de busca refinado** – permite buscas específicas para cada subcorpus

Sistema de busca e anotação

- **DISPARA** (Santos, 2002) - sistema parametrizável de disponibilização de corpora paralelos na rede
 - ✓ **Sistema de processamento de corpus**
 - IMS-CWB (Christ *et al.*, 1999), agora **Open CWB**
 - ✓ **Etiquetagem morfossintática**
 - Português: **PALAVRAS** (Bick, 2000)
<http://visl.hum.sdu.dk/visl/pt/>
 - Inglês: **CLAWS** (Rayson & Garside, 1998)
<http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/>
 - ✓ **Interface**
 - desenhada pela equipe e programada por Patricia Tagnin

Carências

- Por ora, nenhuma porque **Diana Santos** se encarrega de toda a parte computacional. Diálogo fácil!
- “mão de obra” para revisão de
 - alinhamento
 - análise sintática
 - análise semântica
 - preparação de novos textos