

CorTrad – a bi-directional English-Portuguese corpus to enhance teaching and translation

Stella E. O. Tagnin (DLM-USP)

with the cooperation of

Diana Santos (Linguatca & Univ. of Oslo)

Elisa Duarte Teixeira (Projeto CoMET)

ABRAPUI – May 6-9, 2012

Table of contents

- Short introduction to CorTrad
- Distinctive features
- Some contrastive issues

Parallel corpora and translation

- ✓ Isabelle (1992), McEnery & Wilson (1993) and Somers (1993) – potential for **machine translation**
- ✓ Johansson & Hofland (1994), Santos (1995) and Tagnin (2004) – potential for **contrastive studies**
- ✓ Malmkjaer (1998), Frankenberg-Garcia (1999) and Peters *et al.* (2000) – potential for **translation teaching**

CorTrad in a nutshell

- ✓ Annotated, multiversion pt-en-pt parallel corpus
 - ✓ POS tagged
 - ✓ Semantically tagged: color & clothing
- ✓ Started in May 2008
- ✓ Partnership:
 - ✓ Linguateca (design, development & implementation of computational framework)
 - ✓ CoMET Project (design & text collection and edition)
 - ✓ NILC - Inter Institutional Center for Computational Linguistics (web hosting)

CorTrad in a nutshell

- ✓ **Innovations** compared to other parallel corpora:
 - ✓ **Multiversion** format allows for
 - ✓ comparison of different translation stages
 - ✓ study of revision process
 - ✓ detection of students' difficulties (Literary subcorpus)
 - ✓ **Refined search system** – tailored especially for each genre and text type
 - ✓ **Semantic information** – added and human-revised

Search and annotation system

DISPARA (Santos 2002) – system to make parallel corpora available on the Web

✓ **Corpus processing system**

→ IMS-CWB (Christ *et al.* 1999), now **Open CWB** (Evert 2010)

✓ **Underlying parser and tagger**

→ Portuguese: **PALAVRAS** (Bick, 2000)

<http://visl.hum.sdu.dk/visl/pt/>

→ English: **CLAWS** (Rayson & Garside 1998)

<http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/>

→ Semantic annotation: **corte-e-costura** (Santos & Mota 2010)

✓ **Interface** (graphic design by Patricia Tagnin)

CorTrad's Parallel Subcorpora

Journalistic

Scientific
(pt → en)
1,076 texts

Technical-
Scientific

Cookbook
(pt → en)
130,000 words

USP PhD
abstracts
(pt → en)
Coming soon!

Literary

Australian
Short Stories
(en → pt)
28 texts

Canadian
Short Stories
(en → pt)
Coming soon!

Alice in
wonderland
(en → pt)
Coming soon!

Legal

Mercosul
Agreements
(pt → ← en)
Coming soon!

Journalistic (Science): Revista FAPESP



Original
(Brazilian Portuguese)



Published translation
(online publication)





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

Pesquisar no corpus ?

Original	<input checked="" type="radio"/> principal [sema="roupa.*"]	<input checked="" type="checkbox"/> ver
Tradução publicada	<input type="radio"/> principal	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

- | | |
|--|---|
| <input type="radio"/> Concordância | <input type="radio"/> Distribuição das formas |
| <input checked="" type="radio"/> Distribuição dos lemas | <input type="radio"/> Distribuição da categoria gramatical (PoS) |
| <input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal | <input type="radio"/> Distribuição de pessoa e/ou número |
| <input type="radio"/> Distribuição do gênero morfológico | <input type="radio"/> Distribuição da função sintática |
| <input type="radio"/> Distribuição por documento | <input type="radio"/> Distribuição por data de publicação |
| <input type="radio"/> Distribuição por gênero de texto | <input type="radio"/> Distribuição por tema |
| <input type="radio"/> Distribuição por campo semântico | <input type="radio"/> Distribuição por grupo (de cor, de vestuário, etc.) |

Opções

- Resultados por ordem alfabética (só distribuições)
- Resultados numerados (só concordância)

Pesquisar

Technical-Scientific: Cookbook



Original
(Brazilian
Portuguese)



**Translators'
first version**
(English)



**Revised
text**
(by native
speaker)



**Published
translation**
(not yet avail.
online)





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

CorTrad técnico-científico culinária

O CorTrad é um corpus aberto, sujeito a alterações. Clique em [atualizações](#) para saber mais.

Pesquisar no corpus ?

Original	<input checked="" type="radio"/> principal	<input cor.*"]"="" type="text" value="[sema="/>	<input checked="" type="checkbox"/> ver
Primeira tradução	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver
Tradução revisada	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

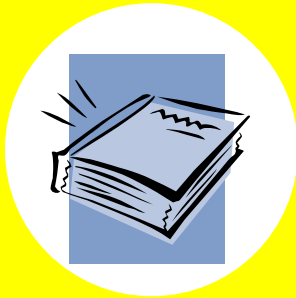
- Concordância
- Distribuição dos lemas
- Distribuição do tempo verbal e/ou do caso pronominal
- Distribuição do gênero morfológico
- Distribuição por parte de obra
- Distribuição por receita
- Distribuição por grupo (de cor, de vestuário, etc.)
- Distribuição das formas
- Distribuição da categoria gramatical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

Opções

Literary: Australian short stories (*learner corpus)



Original
(Australian
English)



**Student's
translation**
(Brazilian
Portuguese)



**Revised
draft** (after
teacher's
suggestions)



**Published
translation**





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

CorTrad literário contos

O CorTrad é um corpus aberto, sujeito a alterações. Veja [dados quantitativos](#) para informações atualizadas sobre o conteúdo do corpus.

A **parte literária do CorTrad** consiste atualmente de 28 contos australianos traduzidos para o português por alunos do Curso de Especialização em Tradução *latu sensu* da Universidade de São Paulo, no 2º semestre de 2003. A tradução foi proposta como trabalho final da disciplina Tradução Literária, ministrada por Stella E. O. Tagnin. A disponibilização do CorTrad na rede é um projeto conjunto entre o COMET, a Linguateca e o NILC, usando o sistema [DISPARA](#).

Pesquisar no corpus



Original	<input checked="" type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver
Primeira tradução	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver
Tradução revisada	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver
Tradução publicada	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

<input checked="" type="radio"/> Concordância	<input type="radio"/> Distribuição das formas
<input type="radio"/> Distribuição dos lemas	<input type="radio"/> Distribuição da categoria gramatical (PoS)
<input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal	<input type="radio"/> Distribuição de pessoa e/ou número
<input type="radio"/> Distribuição do gênero morfológico	<input type="radio"/> Distribuição da função sintática
<input type="radio"/> Distribuição por obra	<input type="radio"/> Distribuição por autor
<input type="radio"/> Distribuição por campo semântico	<input type="radio"/> Distribuição por grupo (de cor, de vestuário, etc.)

Opções



Corpus Multilíngue para Ensino e Tradução

USP

omet!

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

Literário

Técnico-científico

Textos incluídos no corpus jornalístico

A **parte jornalística do CorTrad** conta atualmente com textos das edições de 2001, 2002 e 2003 da [Revista Pesquisa FAPESP](#), totalizando 20 números. As seções incluídas foram: Humanidades, Ciência, Tecnologia, Estratégias, Laboratório, Linha de Produção e Política de C&T.

Lista de revistas FAPESP incluídas:

- 2001: edições 62 (mar-01), 63 (abr-01), 64 (mai-01), 65 (jun-01), 67 (ago-01), 68 (set-01), 69 (out-01) e 70 (nov-01)
- 2002: edições 76 (jun-02), 77 (jul-02), 78 (ago-02), 79 (set-02), 80 (out-02), 81 (nov-02) e 82 (dez-02)
- 2003: edições 83 (jan-03), 84 (fev-03), 85 (mar-03), 86 (abr-03) e 87 (mai-03)

Atenção, esta página ainda é preliminar, em particular a classificação quanto ao gênero e assunto, segundo a [tabela](#) proposta pelo Lácio-Web, está sendo revista neste momento.

ID	Gênero	Assunto
ca-abr01v01	Carta	Direito
ca-abr01v02	Carta	Cultura
ca-abr01v03	Carta	Cultura



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

Literário

Técnico-científico

CorTrad jornalístico divulgação científica

O CorTrad é um corpus aberto, sujeito a alterações. Clique em [atualizações](#) para saber mais.

Pesquisar no corpus

Original	<input checked="" type="radio"/> principal <input type="text" value="pesquisa"/>	<input checked="" type="checkbox"/> ver
Tradução publicada	<input type="radio"/> principal <input type="text"/>	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

Concordância

Distribuição das formas

Distribuição dos lemas

Distribuição da categoria gramatical (PoS)



Corpus Multilíngue para Ensino e Tradução

; novas publicações no lir

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

Literário

Técnico-científico

CorTrad jornalístico divulgação científica

Expressão de busca: "pesquisa"

Resultado escolhido: concordância em contexto

Corpus pesquisado: originais

Apresenta-se uma amostra aleatória de 500 das 1322 ocorrências encontradas.



Voltar

Nova pesquisa

Se quiser mais de 500 resultados da sua pesquisa, contate-nos.

Original

A iniciativa privada tem flexibilidade e agilidade necessárias para levar avante um projeto como esse, já que as instituições públicas não têm cultura de mercado, diz Camargo. O CAT é um dos dez Centros de Pesquisa, Inovação e Difusão (Cepids) qualificados pela FAPESP e a parceria com a indústria farmacêutica para o desenvolvimento do novo medicamento atende a um dos principais quesitos do programa que é o de

Tradução publicada

says Dr. Camargo. The CAT is one of the ten Research Centers, Innovation and Diffusion (Cepids) certified by FAPESP and the partnership with the pharmaceutical industry for the development of the new medicine responds to one of the main requirements of the program which is that of putting together the academic activities of research with the market forces.



Corpus Multilíngue para Ensino e Tradução

Português / Inglês

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Original

A iniciativa privada tem flexibilidade e agilidade necessárias para levar avante um projeto como esse, já que as instituições públicas não têm cultura de mercado, diz Camargo. O CAT é um dos dez Centros de Pesquisa, Inovação e Difusão (Cepids) qualificados pela FAPESP e a parceria com a indústria farmacêutica para o desenvolvimento do novo medicamento atende a um dos principais quesitos do programa que é o de aproximar as atividades acadêmicas de **pesquisa** com o mercado.

Para um país cuja academia manteve corporativamente protegida da sociedade, a criação da revista deu visibilidade a linhas de **pesquisa** que não teriam sobrevivido ao massacre intramuros, avalia Neves.

Os exemplos citados pelo pesquisador são das mais importantes obras encontradas na **pesquisa** do século 16.

Em pauta, os dramáticos efeitos da conjuntura econômica do país, particularmente da crise cambial, sobre o fomento à **pesquisa** científica em São Paulo.

Tradução publicada

says Dr. Camargo. The CAT is one of the ten Research Centers, Innovation and Diffusion (Cepids) certified by FAPESP and the partnership with the pharmaceutical industry for the development of the new medicine responds to one of the main requirements of the program which is that of putting together the academic activities of **research** with the market forces.

«For a country whose academic circles have remained as a body protected from society, the creation of the magazine brought visibility to the lines of **research** that would not have survived the intramural massacre», reckons Neves.

Other examples cited by the researcher are among the most important works found in the **research** of the 16th century.

The subject of discussion was the dramatic effects of the economic crisis in the country, particularly the **exchange** crisis, on the development of scientific **research** in São Paulo.



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Pesquisar no corpus

Original	<input checked="" type="radio"/> principal	<input type="text" value="pesquisa"/>	<input checked="" type="checkbox"/> ver
Tradução publicada	<input type="radio"/> principal	<input research\""="" type="text" value="!\"/>	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

<input checked="" type="radio"/> Concordância	<input type="radio"/> Distribuição das formas
<input type="radio"/> Distribuição dos lemas	<input type="radio"/> Distribuição da categoria gramatical (PoS)
<input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal	<input type="radio"/> Distribuição de pessoa e/ou número
<input type="radio"/> Distribuição do gênero morfológico	<input type="radio"/> Distribuição da função sintática
<input type="radio"/> Distribuição por documento	<input type="radio"/> Distribuição por data de publicação
<input type="radio"/> Distribuição por gênero de texto	<input type="radio"/> Distribuição por tema
<input type="radio"/> Distribuição por campo semântico	<input type="radio"/> Distribuição por grupo (do cor, do vestuário, etc.)



Corpus Multilíngue para Ensino e Tradução

JOS

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

barco a motor, refrigerador, balanças, estufa, câmara climática, pagamento dos catadores e um equipamento de análise de imagens por computador, usado na mensuração de mais de 3 mil caranguejos.

O fato é que a abrangência da **pesquisa** foi crescendo na medida em que o projeto avançava.

O guaxinim pode permanecer sentado, imóvel, em frente à abertura da toca, esperando o caranguejo sair, conta Gláucia. A **pesquisa** revelou, contudo, uma falta de conhecimento dos catadores sobre características importantes do uçá.

Grupos de **pesquisa** abrem perspectivas para prevenção e tratamento mais eficaz.

Em relação aos animais domésticos incluídos na **pesquisa**, o teste com a técnica da reação de cadeia de polimerase (PCR) mostrou que cerca de 20 % dos 61 cães e 14 % dos 58 eqüinos testados apresentaram infecção por Leishmania.

Na outra ponta, entre os comportamentos que podem reduzir a incidência desse problema cardiovascular, a **pesquisa** chegou a um resultado surpreendente: o consumo de álcool se mostrou a mais efetiva forma de proteção contra o ataque cardíaco.

boat, a refrigerator, scales, an oven, an air conditioned chamber, payment of the gatherers, and an item of equipment for the computer analysis of images, used in measuring more than 3,000 crabs.

In fact the coverage of the **survey** increased as the project went on.

The raccoon can remain seated, motionless, in front of the entrance to the burrow, **waiting** for the crab to come out», says Gláucia. The **survey** showed, however, some gaps in the gatherers' knowledge of certain important characteristics of the uçá.

Research groups open up prospects for a more efficient method of prevention and treatment

In relation to the domestic animals included in the **study**, the test with the technique of Polymerase Chain Reaction (PCR) showed that around 20% of the 61 dogs and 14% of the 58 equines tested showed infection by Leishmania.

At the other extreme, amongst the components that can reduce the **incidence** of this cardiovascular problem, the **study** arrived at a surprising result: the consumption of alcohol has proved to be the most effective form of protection against a heart attack.



Verbal
collocations



Corpus Multilíngue para Ensino e Tradução



Ir ao site do Projeto Cor

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

e da Globalização Econômica (Sobeet) **realizam pesquisa** semelhante, mas com caráter regional e setorializado.

Globalization (Sobeet) have **carried out** similar studies, but with a regional character and in one sector only.

Mas não há qualquer impedimento para que, até essa data, empresas **desenvolvam pesquisa** sobre o princípio ativo desses medicamentos.

But there is no impediment to companies **undertaking** research on the active ingredient before that date.

Ele acredita que tempos virão em que os editores serão obrigados a dizer: Nós não vamos admitir esse conflito de interesse e não vamos **publicar pesquisa** onde esse tipo de conflito exista.

He believes that «the time will come» when editors will be obliged to say, «No, we are not going to accept this conflict of interest and we are not going to publish research involving this type of conflict of interest».

Segurança redobrada nos Eua por causa do terror **afeta pesquisa** no Brasil.

Redoubled security in the USA because of terror affects research in Brazil

A Fundação segue apoiando os dez Centros de Pesquisa, Inovação e Difusão (Cepids), criados em 2000 para **desenvolver pesquisa** básica ou aplicada na fronteira do conhecimento.

The Foundation is carrying on supporting the ten Centers for Research, Innovation and Dissemination (Cepids), created in 2000 to **carry out** basic or applied research at the frontiers of knowledge.

Na visão de Esper Cavalheiro, atual presidente, o grande desafio do CNPq, agora, é recuperar o seu papel de fomentador do desenvolvimento científico e tecnológico e dar verbas para o pesquisador **fazer pesquisa**.

In the vision of its current president, Esper Cavalheiro, the CNPq's great challenge now is to recover its role in fostering scientific and technological development and to provide funds for the researcher to **do research**

Isso comprova que hoje seria impraticável **fazer pesquisa** sem uma rede de alta velocidade, afirma

«This proves that it would be impracticable **to do** research without a high speed network», remarks



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

Literário

Técnico-científico

CorTrad técnico-científico culinária

O CorTrad é um corpus aberto, sujeito a alterações. Veja [dados quantitativos](#) para informações atualizadas sobre o conteúdo do corpus.

A **parte culinária do CorTrad** conta atualmente com o conteúdo integral de um livro de 350 páginas em quatro versões: o texto original, escrito em português brasileiro, a tradução para o inglês, feita por duas tradutoras brasileiras, a versão revisada dessa tradução, feita por uma falante nativa do inglês, e a versão final, publicada -- esta, ainda em preparação. [Clique aqui para mais informações](#). A disponibilização do CorTrad na rede é um projeto conjunto entre o COMET, a Linguateca e o NILC, usando o sistema DISPARA.

Pesquisar no corpus

Original	<input checked="" type="radio"/> principal	<input type="text" value="natural"/>	<input checked="" type="checkbox"/> ver
Primeira tradução	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver
Tradução revisada	<input type="radio"/> principal	<input type="text" value="!natural"/>	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

- | | |
|--|--|
| <input checked="" type="radio"/> Concordância | <input type="radio"/> Distribuição das formas |
| <input type="radio"/> Distribuição dos lemas | <input type="radio"/> Distribuição da categoria gramatical (PoS) |
| <input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal | <input type="radio"/> Distribuição de pessoa e/ou número |
| <input type="radio"/> Distribuição do gênero morfológico | <input type="radio"/> Distribuição da função sintática |



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

Literário

Técnico-científico

CorTrad técnico-científico culinária

Expressão de busca: "natural" :CORTRAD_CULI_TRAD2 !"natural"

Resultado escolhido: **concordância em contexto**

Corpus pesquisado: **originais** (versão 1.1)



[Voltar](#)

16 ocorrências.

[Nova pesquisa](#)

Original	Primeira tradução	Tradução revisada
O mais natural é imaginar um almoço lá fora, mas por que não um jantar, com o frescor da noite?	When we think of an outdoor meal, it is usually lunch that comes to one's mind but why not dinner, enjoying the freshness of the night?	When we think of an outdoor meal, it is usually lunch that comes to one's mind but why not dinner, enjoying the freshness of the night?
1/2 xícara de iogurte natural	cup whole milk plain yogurt	cup plain whole milk yogurt
3/4 de xícara de iogurte natural	SEM TRADUÇÃO !!	1 eye of round roast (approximately 1 kg / 2 lb)
3/4 de xícara de iogurte natural	1/3 cup sugar	3 cups pineapple juice
1 1/2 xícara de iogurte natural	1 cups whole milk plain yogurt	1 onion, finely diced
2 1/3 de xícaras de iogurte natural	2 1/3 cups whole milk plain yogurt	cup honey
1 xícara de iogurte natural	1 cup whole milk plain yogurt	3 eggs
600 ml de iogurte natural (3 copinhos)	600 ml (21 fl oz) whole milk plain yogurt	600 ml (21 fl oz) plain whole milk yogurt
100 ml de iogurte natural	100 ml (3.5 fl oz) whole milk plain yogurt	100 ml (3 fl oz) plain whole milk yogurt
600 ml de iogurte natural (3 copinhos)	600 ml (21 fl oz) whole milk plain yogurt	3 cups milk
200 ml de iogurte natural (1 copinho)	200 ml (7 fl oz) whole milk plain yogurt	cup heavy cream
200 ml de iogurte natural (1 copinho)	200 ml (7 fl oz) whole milk plain yogurt	1 tablespoon lime juice
1 litro de suco de laranja natural	1 L (1.1 qt) freshly squeezed orange juice	? cup water

Comparing results

How do the verbs “acreditar” and “achar” compare across genres in the journalistic subcorpus?



CorTrad jornalístico divulgação científica

O CorTrad é um corpus aberto, sujeito a alterações. Clique em [atualizações](#) para saber mais.

Pesquisar no corpus ?

Original	<input checked="" type="radio"/> principal	<input acreditar"]"="" type="text" value="[lema="/>	<input checked="" type="checkbox"/> ver
Tradução publicada	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

- | | |
|--|---|
| <input type="radio"/> Concordância | <input type="radio"/> Distribuição das formas |
| <input type="radio"/> Distribuição dos lemas | <input type="radio"/> Distribuição da categoria gramatical (PoS) |
| <input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal | <input type="radio"/> Distribuição de pessoa e/ou número |
| <input type="radio"/> Distribuição do gênero morfológico | <input type="radio"/> Distribuição da função sintática |
| <input type="radio"/> Distribuição por documento | <input type="radio"/> Distribuição por data de publicação |
| <input checked="" type="radio"/> Distribuição por gênero de texto | <input type="radio"/> Distribuição por tema |
| <input type="radio"/> Distribuição por campo semântico | <input type="radio"/> Distribuição por grupo (de cor, de vestuário, etc.) |

Opções



CorTrad jornalístico divulgação científica

92 casos

Expressão de busca: [jema="acher"]
Resultado escolhido: distribuição de genero
Corpus pesquisado: originais

Houve 11 valores diferentes de genero.

Entrevista	67
Reportagem	58
artigo	24
Noticia	14
Artigo	8
Data	2
OU-Editorial	1
editorial	1
Editorial	1
Noticia-OU-section 1	1
section	1

Para informação sobre os códigos internos do atributo genero, consulte a página de

8-janeiro-2010. Esperamos que o CorTrad jornalístico divulgação científica lhe tenha sido útil!

CorTrad jornalístico divulgação científica

195 casos

Expressão de busca: [jema="acreditar"]
Resultado escolhido: distribuição de genero
Corpus pesquisado: originais

Houve 11 valores diferentes de genero.

Reportagem	92
Noticia	40
artigo	23
Artigo	21
Entrevista	11
Noticia-OU-section 3	3
section	2
opinion	1
artigo-OU-section 1	1
noticia-OU-section 1	1
Noticia -OU-Noticia 1	1

Para informação sobre os códigos internos do atributo genero, consulte a página do subcorpus

8-janeiro-2010. Esperamos que o CorTrad jornalístico divulgação científica lhe tenha sido útil!

Enhancing translation options

When is “then” not translated as
“então”?



- Principal
- O Projeto
- Equipe
- CorTec
- oMAprend
- CorTrad
- rtigos, etc.
- Links
- Informativo
- Contato
- Site FFLCH
- Site USP

CorTrad literário contos

Expressão de busca: "then" :CORTRAD_LIT_TRADE !"então"

Resultado escolhido: concordância em contexto

Corpus pesquisado: originais



[Voltar](#)

71 ocorrências.

[Nova pesquisa](#)

Original	Primeira tradução	Tradução revisada	Tradução publicada
She was by then a woman of forty-nine or fifty - small, straight, business-like , in a tailored suit and with her hair cut in a silver helmet.	Era, àquela altura, uma mulher de quarenta e nove ou cinquenta anos - pequena, ereta, metódica , trajando um conjunto feito sob encomenda e com o cabelo cortado em forma de elmo prateado.	Era, àquela altura, uma mulher de quarenta e nove ou cinquenta anos - pequena, ereta, estilo executivo , trajando um conjunto feito sob medida e com o cabelo cortado em forma de elmo prateado.	Àquela altura , ela era uma mulher de quarenta e nove ou cinquenta anos - pequena, ereta, com ar profissional , trajando um conjunto feito sob medida e com o cabelo cortado em forma de elmo prateado.
All the more alarming then , as I sat waiting on one of the cane-bottomed chairs in the corridor, while Ben Steinberg, Miss McIntyre's star pupil, played the Max Bruch, to hear the same voice oddly transmuted.	Ainda mais alarmante — quando eu estava sentado esperando em uma das cadeiras de assento de palha no corredor enquanto Ben Steinberg, o aluno modelo de Miss McIntyre, tocava Max Bruch — era 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 Steinberg, o aluno modelo de Miss McIntyre, tocava Max Bruch — era 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 Steinberg, o aluno modelo de Miss McIntyre, tocava Max Bruch — foi escutar a mesma voz estranhamente alterada.
We were alone for a little then .	SEM TRADUÇÃO	Depois disso, ficamos sozinhos por um tempo.	Depois disso , ficamos sozinhos por um tempo.
' Which one are you then ?	- Ora, ora, qual deles é você?	- Ora, ora, qual deles é você?	- Ora, ora , qual deles é você?
Just then my Aunt Ourania came into the room. At which my maid	Nesse momento, minha Tia Ourania entrou no sala, para a qual	Nesse momento, minha Tia Ourania entrou no sala para a qual	Nesse momento minha Tia



**Unexpected
revelations!!!**

CorTrad's Semantic annotation:

some findings on color and clothing

10 most recurring color terms

Scientific news (or)		Short stories (tr)		Cookbook (or)	
cor	86	branco	69	dourar	341
branco	57	azul	48	branco	144
vermelho	43	vermelho	37	vermelho	122
verde	42	cor	26	verde	113
negro	37	negro	25	preto	39
<i>buraco negro</i>	31	preto	22	cor	39
<i>ultravioleta</i>	28	verde	19	tinto	33
<i>infravermelho</i>	27	amarelo	18	dourado	27
azul	25	pálido	15	amarelo	16
<i>amarelinho</i>	22	cinza	13	roxo	14

Search expression: [sema="cor.*"]

Result type: lemma distribution

Color distribution by group

Search expression:
[sema="cor.*"]
Group distribution.

White	64	0	50
Black	47	Branco	44
Blue	46	Azul	39
Red	44	Vermelho	38
Grey	27	Preto	30
Green	23	Amarelo	19
Gold	22	Nãoespecificad a	19
Brown	21	Cinzeno	18
Unspecified	18	Verde	17
Yellow	18	Castanho	11
Naoespecificada	13	Rosa	10
Pink	12	Ausência	8
Purple	11	Creme	6
Beige	8	Dourado	4
Silver	7	Prateado	4
Orange	6	Roxo	3
Multiple	1	Vermelho_0	2
		Laranja	2

white ≠ “branc*”



site do Projeto Comet!

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

CorTrad literário contos

Expressão de busca: "white" %c :CORTRAD_LIT_TRADF !"branc.*" %c
Resultado escolhido: **concordância em contexto**
Corpus pesquisado: **originais** (versão 4.1)



5 ocorrências.



[Voltar](#)

[Nova pesquisa](#)

Original	Primeira tradução	Tradução revisada	Tradução publicada
She began to prowl between the desks, waving the white letter like a flag.	Então começou a rondar as carteiras, balançando a carta como se fosse uma bandeira.	Então começou a rondar as carteiras, balançando a carta como se fosse uma bandeira.	Então começou a rondar as carteiras, balançando a carta como se fosse uma bandeira.
We had never cared for Aunt Calliope, who had made us write essays and things, though her brothers loved her and would quarrel with her till the white hours over any boring political issue.	Nunca demos muita atenção a Tia Calliope, ela que nos fazia escrever ensaios e outras coisas, apesar de seus irmãos a amarem e serem capazes de discutir com ela até o amanhecer sobre qualquer tediosa questão política.	Nunca demos muita atenção a Tia Calliope, ela que nos fazia escrever ensaios e outras coisas, apesar de seus irmãos a amarem e serem capazes de discutir com ela até o amanhecer sobre qualquer tediosa questão política.	Nunca demos muita atenção a Tia Calliope, ela que nos fazia escrever ensaios e outras coisas, apesar de seus irmãos a amarem e serem capazes de discutir com ela até o amanhecer sobre qualquer tediosa questão política.
They eat margarine and ham on thin white sliced bread, and supermarket sponge rolls with jam and plastic cream, which is why they are all rather large.	Elas comem pão de forma com presunto e margarina, e pães esponjosos de supermercado com geléia e creme plástico, motivo pela qual elas estão um tanto gordas.	Elas comem pão de forma com presunto e margarina, e pães esponjosos de supermercado com geléia e creme plástico, motivo pelo qual elas estão um tanto gordas.	Elas comem pão de forma com presunto e margarina e pães esponjosos de supermercado com geléia e creme plástico, motivo pelo qual elas estão um tanto gordas.
He went straight to the shaft, let himself down carefully on to one of the old logs, and stooped to pick up the note, gleaming white in the sickly summer daylight.	Ele caminhou direto para o poço, desceu cuidadosamente até uma das velhas toras de madeira e abaixou-se para pegar o bilhete que brilhava naquele dia ensolarado.	Ele caminhou direto para o poço, desceu cuidadosamente até uma das velhas toras de madeira e abaixou-se para pegar o bilhete que brilhava naquele dia ensolarado.	Ele caminhou direto para o poço, desceu cuidadosamente até uma das velhas toras de madeira e abaixou-se para pegar o bilhete que brilhava na pálida luz daquela manhã de verão.
She sits on an upturned bucket, her hat pulled down hard on her head, her flat nose red with cold - queen of a spread of tins and jars and white and blue enamel bowls of uncut clams; my job is to prise the shells apart; she cuts and scoops them from the open fan.	Senta-se sobre um balde virado para baixo, seu chapéu enterrado na cabeça, seu nariz largo já vermelho de frio -- rainha de uma montoeira de latas e potes e bacias de mariscos inteiros; minha tarefa é arrancar as conchas; ela corta e tira o que sobrou na concha aberta.	Senta-se sobre um balde virado para baixo, seu chapéu enterrado na cabeça, seu nariz largo já vermelho de frio -- rainha de uma montoeira de latas e potes e bacias de mariscos inteiros; minha tarefa é arrancar as conchas; ela corta e tira o que sobrou na concha aberta.	Senta-se sobre um balde virado para baixo, o chapéu enterrado na cabeça, o nariz largo vermelho de frio -- rainha de uma montoeira de latas e potes, tigelas azuis esmaltadas cheias de mariscos inteiros. Minha tarefa é abrir as conchas; ela corta e retira o conteúdo da concha aberta.

Simple search: “dour.*” is not “golden.*”



Confira as novas

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

Literário

Técnico-científico

CorTrad técnico-científico culinária

O CorTrad é um corpus aberto, sujeito a alterações. Veja [dados quantitativos](#) para informações atualizadas sobre o conteúdo do corpus.

A **parte culinária do CorTrad** conta atualmente com o conteúdo integral de um livro de 350 páginas em quatro versões: o texto original, escrito em português brasileiro, a tradução para o inglês, feita por duas tradutoras brasileiras, a versão revisada dessa tradução, feita por uma falante nativa do inglês, e a versão final, publicada – esta, ainda em preparação. [Clique aqui para mais informações](#). A disponibilização do CorTrad na rede é um [projeto conjunto](#) entre o [COMET](#), a [Linguateca](#) e o [NILC](#), usando o sistema [DISPARA](#).

Pesquisar no corpus ?

Original	<input checked="" type="radio"/> principal "dour.*"	<input checked="" type="checkbox"/> ver
Primeira tradução	<input type="radio"/> principal	<input checked="" type="checkbox"/> ver
Tradução revisada	<input type="radio"/> principal !"gold.*"	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

<input checked="" type="radio"/> Concordância	<input type="radio"/> Distribuição das formas
<input type="radio"/> Distribuição dos lemas	<input type="radio"/> Distribuição da categoria gramatical (PoS)
<input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal	<input type="radio"/> Distribuição de pessoa e/ou número
<input type="radio"/> Distribuição do gênero morfológico	<input type="radio"/> Distribuição da função sintática
<input type="radio"/> Distribuição por parte de obra	<input type="radio"/> Distribuição por tipo de texto
<input type="radio"/> Distribuição por receita	<input type="radio"/> Distribuição por campo semântico

Options

Result: “dour.*” ≠ “golden.*”



	Português	Inglês	Português
Principal	Enquanto isso, doure ligeiramente os pinoli numa frigideira grande e seca e reserve.	Meanwhile, lightly toast pine nuts in a dry large frying pan and set aside.	Meanwhile, <u>lightly toast</u> pine nuts in a large dry frying pan and set aside.
O Projeto			
Equipe	Leve ao forno por uns 25 minutos para aquecer e dourar (se quiser, monte na véspera e guarde na geladeira sem assar).	Bake for about 25 minutes to heat thoroughly and brown the top (if you want, you can assemble lasagna a day ahead and refrigerate, unbaked).	Bake for about 25 minutes to heat thoroughly and <u>brown</u> the top (if you want, you can assemble lasagna a day ahead and refrigerate, unbaked).
CorTec			
CoMAprend			
CorTrad	Regue com um fio de azeite e asse por uns 15 minutos, até que a massa esteja bem dourada e crescida nas bordas.	Drizzle with olive oil and bake for about 15 minutes, until dough is dark brown and risen on the edges.	Drizzle with olive oil and bake for about 15 minutes, until dough is <u>dark brown</u> and risen on the edges.
Artigos, etc.	Aqueça um fio de azeite numa panela grande que possa ir ao forno, junte metade da carne e mantenha o fogo forte para dourar e não juntar água.	Heat a drizzle of olive oil in a large ovenproof casserole dish, add half the meat and sear, over a high heat to prevent juices from escaping, until cubes are browned.	Heat a <u>drizzle</u> of olive oil in a large pot that can go into the oven, add half the meat and sear, over high heat to prevent juices from escaping, until cubes are <u>browned</u> .
Links			
Informativo	Transfira os pedaços dourados para um prato, aqueça mais um pouco de azeite, doure a carne restante e passe também para o prato.	Transfer browned chunks to a plate, heat another drizzle of olive oil, sear remaining meat and transfer to the same plate.	Transfer <u>browned</u> chunks to a plate, heat another drizzle of olive oil, sear <u>remaining</u> meat and transfer to the same plate.
Contato			
Site FFLCH			
Site USP	Transfira os pedaços dourados para um prato, aqueça mais um pouco de azeite, doure a carne restante e passe também para o prato.	Transfer browned chunks to a plate, heat another drizzle of olive oil, sear remaining meat and transfer to the same plate.	Transfer browned chunks to a plate, heat another drizzle of olive oil, sear remaining meat and transfer to the same plate.
	Aqueça mais um fio de azeite, junte o bacon e, quando começar a dourar , acrescente a cebola, a cenoura, o salsão e 1 pitada de sal.	Heat a little more olive oil, add bacon and, when it begins to brown, stir in diced onion, carrot, celery, and a pinch of salt.	Heat a little more olive oil, add bacon and, when it begins to <u>brown</u> , stir in diced onion, carrot, celery, and a pinch of salt.
	Pincele o rolo com a gema e asse por uns 25 minutos, até começar a dourar .	Brush roll with egg yolk and bake for 25 minutes or so, until it begins to brown.	Brush cylinder with egg yolk and bake for 25 minutes or so, until it begins to <u>brown</u> .
	Espalhe as fatias sobre a assadeira e volte ao forno por mais 20 minutos, virando na metade do tempo para dourar por igual.	Spread slices out on baking sheet and return to oven for another 20 minutes, turning once halfway through, so they brown evenly.	in) thick slices. Spread slices out on baking sheet and return to oven for another 20 minutes, turning once halfway through, so they <u>brown</u> evenly.
	Para começar, como o clima é de muita descontração, vêm uns camarões de médios para grandes, passados no gergelim, dourados no forno e servidos com palitinhos com um molhinho agri-doce de abacaxi.	To start with, as the atmosphere exudes relaxation, there are the medium to large-sized shrimps, coated with sesame seeds, browned in the oven, and served with the aid of toothpicks for dipping in a sweet-n-sour pineapple sauce.	To start with, as the atmosphere exudes relaxation, there are the medium- to large-sized shrimps, coated with sesame seeds, <u>browned</u> in the oven, and served with the aid of toothpicks for dipping in a sweet-n-sour pineapple sauce.
	Leve o salmão ao forno por uns 15 minutos, apenas o suficiente para terminar de dourar por fora e deixar o filé macio e rosado no centro.	Transfer salmon to preheated oven and bake for 15 minutes or so, just enough to crisp the outer crust and to tenderize the opaque pink flesh in the center.	Transfer salmon to preheated oven and bake for 15 minutes or so, just enough to <u>crisp</u> the outer crust and to tenderize the opaque pink flesh in the center.
	A segunda leva folhas verdes, manga e tirinhas douradas e picantes de peito de frango.	The second takes green salad leaves, mango and browned and spicy strips of chicken breasts.	The second takes salad greens, mango and <u>browned</u> and spicy strips of chicken breasts.
	Em seguida, aqueça um fio de óleo numa frigideira grande e	Then, discard chicken marinade, heat a drizzle of oil in a	Then, discard chicken marinade, heat a drizzle of oil in a large

Translation of figurative expressions and terminology

- just declared itself **out to the blue**
- It was to be **black tie**.
- She never refused to go to Melbourne, but it was her hoodoo city, **a black jinx**.
- The dog knew they were coming, and barked **blue murder**.
- would quarrel with her till **the white hours**
- knowing about **brown rice**
- **blackfellow**
- **thin white sliced bread**
- **red wine**
- **red cabbage**

- surgira de repente, **do nada**
- O traje é **a rigor**.
- Nunca se negava a ir a Melbourne, mas era uma cidade de azar, **mau agouro**.
- O cachorro sabia que eles estavam vindo e latiu **desesperadamente**.
- de discutir com ela até **o amanhecer**
- Eu entendia de **arroz integral**
- **arborígene**
- **pão de forma**
- **vinho tinto**
- **repolho roxo**

Deemed unnecessary ... or different associations

- She rushed to the back of the house and hauled the drowsy **black** pup out of the kennel.
- She began to prowl between the desks, waving the **white** letter like a flag.
- When the torrent of **white** water subsided
- The **brown** smell of his cigar
- **white** and blue enamel bowls
- She had the **tinted** view of the Irish
- long **golden** summer

- Ela correu para os fundos da casa e puxou o sonolento filhote para fora da casinha.
- Então começou a rondar as carteiras, balançando a carta como se fosse uma bandeira.
- Quando a corrente de águas espumantes cessou
- O aroma característico do charuto
- tigelas azuis esmaltadas
- Ela tinha a visão **cor-de-rosa** dos irlandeses
- verão longo e ensolarado

Concluding remarks on color

- Totally different translation patterns for
 - Figurative language (most cases do not preserve color)
 - Skin/race/culture color (more differentiation in English)
 - Real colour
- Scientific news: a lot of (unexpected) color in scientific terminology: disease names, stars, etc.

10 most recurring clothing words

Scientific News (or)		Short Stories (tr)		Cookbook (or)	
Total oc.	58 (20)	Total oc.	380 (72)	Total oc.	21 (12)
roupa	17	roupa	37	roupa	7
sapato	9	chapéu	30	avental	2
luva	5	vestir	27	vestir	2
calçado	4	sapato	26	chapéu	2
camiseta	4	vestido	24	...	1
vestir	2	usar	22		
bota	2	camisa	13		
anel	2	meia	10		
vestido	2	calça	10		
...	1	capa	10		

Search expression: [sema="roupa.*"]

Result type: lemma distribution

Semantic information: clothing

	Cooking	Scientific news	Short stories	Totals
Pure clothing	21	50	378	449
Figurative	0	6	2	8
Totals	21	56	380	
Word count	134,093	776,284	121,253	

Search expression: [sema="roupa.*"]

Result type: semantic field

Figurative expressions and non-clothing-referring clothes

- The arrangement **suit**ed Henry extremely well.
- He says the government will set an example in **'belt-tightening'**.
- desafio de **se pôr nos sapatos** de um arqueólogo
- do tamanho de uma caixa de **sapatos**
- pense na **roupa**

- O esquema **caiu como uma luva** para Henry.
- Ele diz que o governo estabelecerá um exemplo de **'apertar os cintos'**.
- task of **wearing the hat** of an archeologist
- the size of a **shoebox**
- think of what you will wear

Concluding remarks on clothing

- Also different translation patterns for
 - figurative language
 - the issue of *wear* vs. *vestir/calçar* and *usar*
- Scientific news: expressions related to clothing but no clothes (shoebox)
- Short stories: how culture dependent are some descriptions, and what was the stance taken by the translators?

Corpus-based TS are revealing because

- you discover the challenges the translator had to face
- you find problems you had never thought about before
- you find creative solutions
- you can use findings – **and corpora** – to enhance translator training

Acknowledgements

- Thanks to Eckhard Bick and Paul Rayson, for the use of PALAVRAS and CLAWS, respectively.
- Thanks to Sandra Aluísio and Arnaldo Candido Júnior at NILC for hosting and corresponding technical support.
- Thanks to Research Computing Services at Univ. Oslo
- This work was partially funded by the Portuguese government, UMIC, FCCN and the European Union (FEDER and FSE), under grant POSC/339/1.3/C/NAC (Linguatca)

