

Project CoMET

A multilingual corpus for teaching and translation

Stella E. O. Tagnin
University of São Paulo

Simon Fraser University
June 30, 2013

Outline

Project CoMET

- CorTec – technical corpus
- CorTrad – translation corpus
- CoMAprend – learner corpus
- Illustrated with possible queries



Corpus Multilíngue para Ensino e Tradução

USP

o Comet!

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

PROJETO **CO**rpus

Multilíngüe para

Ensino e

Tradução

Última atualização: **outubro de 2009** ([veja aqui](#))

Webmasters: **Elisa Duarte Teixeira**

Sabrina Matuda

© Projeto Comet - FFLCH / USP

www.fflch.usp.br/dlm/comet

The COMET project

Beginning: 1998

- First “corpora”: 1999-2005
 - Students in the Translation course built small corpora and compiled glossaries
- Officially launched online: September 2005 (CNPq grant):
 - CorTec (Technical Corpus)
 - http://www.fflch.usp.br/dlm/comet/consulta_cortec.html
 - CoMAprend (Learner Corpus)
 - <http://www.fflch.usp.br/dlm/comet/comaprend.html>

CorTec 2005

- 5 **COMPARABLE** corpora:
 - Cooking – recipes
 - Environment - Ecotourism
 - Computing - General
 - Cardiology – Hypertension
 - Law – agreements
- English - Portuguese
- +/- 200,000 words each

CorTec 2005

- Tools
 - Frequency Counter
 - Concordancer
 - Same as
 - Starting with
 - Ending in
 - Containing
 - N-grams



CORTEC

Português

Search What It Is

Step 1/3: Corpus selection.

Select which one(s) you want to work on.

List of Corpora

- ☐ Culinária
- ☐ Ecoturismo
- ☒ Hipertensão
- ☐ Informática
- ☐ Instrumentos Contratuais
- ☐ Astronomia
- ☐ Insuficiência Renal
- ☐ Linguística
- ☐ Medidores Eletromagnéticos de Vazão
- ☐ Suplementos Nutricionais
- ☐ Computação
- ☐ Futebol
- ☐ Café
- ☐ Turismo Cultural
- ☐ Culinária 2
- ☐ Prostodontia
- ☐ Fotografia
- ☐ Autoclaves
- ☐ Moda
- ☐ Turismo - Hotelaria

Language

- ☐ Portuguese
- ☒ English

>> Next Step

CorTec – Technical Scientific Corpus

Login



Corpus Multilíngue para Ensino e Tradução

USP

o o CorTrad - nosso novo corpus

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

CORTEC

Português

Search

What It Is

Step 2/3: Select your tool.

List of Tools

- ☒ Concordancer
- ☐ Frequency Counters
- ☐ N-Gram Generator

>> Next Step

CorTec – Technical Scientific *Corpus*

Login



Corpus Multilíngue para Ensino e Tradução

USP

novas publicações no link Artigos

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

CORTEC

Português

Search What It Is

Step 3/3: Applying Tool to *Corpus*.

Select the tool according to the options below:

Concordancer

Expression or word: Same as

Case sensitive: ☐

- Same as
- Starting with
- Ending in
- Containing

Size of reduced context (in characters):

>> Go

Select another tool <<

CorTec – Technical Scientific *Corpus*

Login



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

CORTEC

Português

Search

What It Is

Step 3/3: Applying Tool to *Corpus*.

Select the tool according to the options below:

Concordancer

Expression or word: Starting with

Case sensitive: ☐

Same as
Starting with
Ending in
Containing

Size of reduced context (in characters):

>> Go

Select another tool <<



Search What Is CorTec

Foram encontradas 449 ocorrências!! DOWNLOAD hits.

Click on search word to obtain expanded context (150 characters).

1 e for percutaneous use in the cardiac catheterization laboratory
2 is is often made on a single cardiac cycle and on a single echo p
3 rring in the presence of epi cardiac coronary stenosis and stress
4 any provocative test for myo cardiac ischemiaScan offer. The syst
5 the coronary sinus (or great cardiac vein) and cau ses the infus
6 ged 28 mmHg, whereas subendo cardiac and LV enddiastolic pressure
7 in the arrested heart during cardiac surgery 27 . Another indirec
8 canine 14 , while increasing cardiac output in the presence of a
9 res, systolic blood pressure, cardiac output and heart rate in a
10 oplasty (PTCA), SRP increased cardiac output and LV stroke work i
11 trograde cardioplegia during cardiac surgery, and b) drug retroi
12 plegia solution to rightsided cardiac structures. Shaper and cold
13 opsies in 42 patients during cardiac surgery and compared the ul
14 seems to be a consensus among cardiac surgeons that in patients w
15 eled metoprolol in the great cardiac vein and unlabeled metoprol
16 times higher during great cardiac vein compared to right atri
17 me amount given in the great cardiac vein 69 . In another study 7
18 a in dogs treated with great cardiac vein (GCV) retroinfusion of
19 of right atrial versus great cardiac vein infusions of the oxyge
20 y 15 times higher with great cardiac vein versus systemic intrav
21 n. In patients with previous cardiac surgery it may also be more
22 d be positioned in the great cardiac vein, in the vicinity of the
23 ure, right heart pressures or cardiac output, and more importantl
24 ion fraction, stroke work and cardiac output were significantly b
25 myocardial protection during cardiac surgery need to be further i
26 heter positioned in the great cardiac vein. In at least two patie



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

Ocorrência 18

nd during
the initial phase of reperfusion. In this case the
Fig. 9\$Infarct size (IS) expressed as percent of risk area in dogs treated
with great **cardiac** vein (GCV) retroinfusion of the oxygen free radical
scavenger superoxide dismutase and catalase (SOD & CAT) given 15
minutes prior to reperfusion

Ocorrência 19

atients with acute myocardial
infarction undergoing thrombolytic therapy. Our
experimental studies, comparing the use of right
atrial versus great **cardiac** vein infusions of the
oxygen free radical scavenger superoxide dismutase
and catalase prior to reperfusion in dogs with LAD
occlusion 75 , have s

Ocorrência 20

a compared to intravenous ad
ministration 66 . In that study, myocardial
procainamide concentrations were approximately 15
times higher with great **cardiac** vein versus
systemic intravenous infusions.
Coronary venous retroperfusion in man
Intracoronary interventional therapy for
obstructive coronary

Ocorrência 21

CorTec 2008

14 corpora (CNPq grant)

- Cooking – recipes
 - Environment - Ecotourism
 - Computing - General
 - Cardiology – Hypertension
 - Law – agreements
- Astronomy
 - Urology - Kidney failure
 - Linguistics
 - Flowmeters
 - Nutritional supplements
 - Football
 - Coffe
 - Cultural Tourism
 - Cooking 2

CorTec 2012

New additions - Total: 20 corpora

- Odontology – Prostodontics
- Photography
- Autoclaves
- Fashion
- Tourism – hotels

... and

- Football has been updated
- Cooking 1 and 2 conflated

Translation equivalents

- Portuguese: “contrato”
- English: *contract*?
- Portuguese corpus:

15

contrato

1678



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

FREQUENCY TABLE

Pos.	Word	Freq.
1	the	15398
2	of	9445
3	to	7055
4	or	6841
5	and	6778
6	in	4026
7	any	4025
8	shall	2760
9	be	2455
10	by	2426
11	a	2319
12	such	2220
13	as	1973
14	for	1628
15	this	1619
16	with	1558
17	other	1344
18	parte	1305
19	agreement	1245
20	not	1216
21	is	1194
22	that	1119
23	all	1101
24	on	1035
25	infosig	979
26	its	947
27	which	933
28	company	862



Corpus Multilíngue para Ensino e Tradução

Principal	152	expenses	145
O Projeto	153	borrower	145
Equipe	154	purchase	144
CorTec	155	equipment	144
CoMAprend	156	lender	143
CorTrad	157	entitled	143
Artigos, etc.	158	and/or	142
Links	159	insurance	142
Informativo	160	iii	141
Contato	161	claim	141
Site FFLCH	162	legal	141
Site USP	163	act	140
	164	costs	138
	165	contemplated	138
	166	contract	138
	167	full	138
	168	between	137
	169	assin	136
	170	either	135
	171	day	133
	172	nor	133
	173	delivered	132
	174	delivery	132
	175	performance	132
	176	obligation	131
	177	writing	131
	178	limitation	130
	179	corporation	130
	180	request	129
	181	result	129
	182	sale	129



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

9 PROFITS OR REVENUE, WHETHER ARISING IN CONTRACT OR TORT, EVEN IF CENTER HAS BEEN ADVISE
10 /parte> 's liability to you, whether in contract, tort (including negligence), or otherw
11 cument necessary to enter into any such contract and/or agreement, including but not li
12 the ownership of voting securities, by contract, or otherwise. "Business Day" shall me
13 Y WITHOUT REGARD TO FAULT, OR BREACH OF CONTRACT OR WARRANTY, FOR ANY CONSEQUENTIAL, SPE
14 A. ASSIGNMENT OF RELINQUISHED PROPERTY CONTRACT 1. Owner agrees to transfer (as perm
15 hed Property, pursuant to a real estate contract dated _____ (the "Rel
16 _____ (the "Relinquished Property Contract"). The Relinquished Property Contract
17 Contract"). The Relinquished Property Contract is attached hereto as Exhibit B and is
18 igrations) in the Relinquished Property Contract, and Exchangor agrees to accept assignm
19 ligations) in the Relinquished Property Contract. B. NOTICE TO PURCHASER On or before
20 r's rights in the Relinquished Property Contract have been assigned to Exchangor. ARTI
21 D. ASSIGNMENT OF REPLACEMENT PROPERTY CONTRACT 1. Owner will enter into any contract
22 CONTRACT 1. Owner will enter into any contract for the acquisition of the Replacement
23 nt Property (the "Replacement Property Contract") and Owner will assign its rights (but
24 ions) under each Replacement Property Contract to Exchangor. It is understood and agre
25 on to execute or take assignment of any contract, or to do any other act or thing contem
26 is Agreement, the Relinquished Property Contract or the Replacement Property Contract wi
27 ty Contract or the Replacement Property Contract without, in each case, receiving a writ
28 seller under said Replacement Property Contract, that Owner's rights in said Replacemen
29 r's rights in said Replacement Property Contract have been assigned to Exchangor. E.
30 TED DAMAGES Every Replacement Property Contract shall provide that (i) Exchangor's liab
31 ability under the Replacement Property Contract shall be limited to the down payment th
32 closing under the Replacement Property Contract, except for representations and warrant
33 ement. If any such Replacement Property Contract requires Exchangor to execute a purchas
34 l property and the Replacement Property Contract for the lease being assigned thereunder
35 assignment. If any Replacement Property Contract for related transaction) requires Excha
36 d pursuant to the Replacement Property Contract. The term "mortgage" as used in this Ex
37 r pursuant to the Relinquished Property Contract. C. REPLACEMENT PROPERTY CLOSING DAT
38 ty pursuant to the Replacement Property Contract relating thereto. E. CLOSING ADJUSTMEN
39 Agreement and the Replacement Property Contract(s). Owner acknowledges that in connecti
40 ge Agreement, the Relinquished Property Contract or the Replacement Property Contract, o
41 ty Contract or the Replacement Property Contract, or shall incur any expenses by reason
42 ge Agreement, the Relinquished Property Contract or the Replacement Property Contract, w
43 ty Contract or the Replacement Property Contract, whether as a tax, or for breach of con
44 act, whether as a tax, or for breach of contract, injury to person or property, or fines

USING COMET AS A MONOLINGUAL CORPUS

Cooking corpus

- How frequent are adverbs in *-ly*?
- Which are the most frequent?
- Which are their collocates?



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

CORTEC

Português

Search What It Is

Step 3/3: Applying Tool to *Corpus*.

Select the tool according to the options below:

Concordancer

Expression or word: Ending in

Case sensitive: ☐

Size of reduced context (in characters):

>> Go

Select another tool <<

CorTec – Technical Scientific *Corpus*

Login



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

CORTEC

Search What Is CorTec

Foram encontradas 6195 ocorrências!! **DOWNLOAD** hits.

Click on search word to obtain expanded context (150 characters).

1 Abernethy Biscuits Surprisingly, Abernethy biscuits do not get their na
2 d rub in the butter until it is thoroughly mixed. Mix in the sugar and caraway see
3 ough with a knife. 2. Turn onto a lightly floured surface. Knead quickly until sm
4 o a lightly floured surface. Knead quickly until smooth. Roll out to about 1 cm (1
5 Almond Scones Scones are lovely served warm fresh from the oven. Makes
6 n the milk and almond essence. Mix lightly with a wooden spoon or fork until a sof
7 ugh on to a floured board and knead gently until smooth. Roll out to a thickness o
8 cutter. 4. Place the scones on a lightly greased baking sheet and brush the tops
9 e mixture forms a dough. 4. Knead quickly on a floured surface then roll out to 1
10 thick. Cut into small rounds approximately 5 cm (2 inches) in diameter. 5. Place
11 er and butter into a saucepan. Heat slowly until the butter melts, then bring to a
12 in the flour all at once. 3. Stir briskly until the mixture forms a soft ball and
13 4. Remove from the heat and cool slightly. Gradually add the eggs, a little at a
14 from the heat and cool slightly. Gradually add the eggs, a little at a time, beati
15 ed Unsalted peanuts - 75g (3 oz), roughly chopped Soft brown sugar - 75g (3 oz)
16 Milk - 250 ml (8 fl oz) Egg - 1, lightly beaten Walnut halves - to decorate ME
17 r, mix to a soft dough. Knead on a lightly floured surface until smooth. Place in
18 r until fluffy. Beat in the eggs gradually along with the flour. 2. Add the bakin
19 r and nuts. Fold well until thoroughly mixed. 3. Spread the mixture into a gr

The most common adverbs in -ly

- Freshly = 3117
- Finely = 3092
- Gently = 2345
- Lightly = 1524
- Thinly = 637
- Carefully = 635
- Immediately - 622
- Evenly = 327



Corpus Multilíngue para Ensino e Tradução

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

1 ugar - 4 tbsp Ground ginger - 1/2 tsp **Freshly** grated nutmeg Sultanas - 175g (6 oz)
2 - 1, zest only Caraway seeds - 4 tsp **Freshly** grated nutmeg - pinch Milk - 300 ml (1
3 for 20-25 minutes until golden. 7. Eat **freshly** baked and slightly warm.
4 Salt - 1 tsp Ground cinnamon - 2 tsp **Freshly** grated nutmeg - 1 tsp Ground cloves -
5 lted Tomatoes - 300g (11 oz) Salt and **freshly** ground pepper Oregano - 1 tsp Mushroo
6 the best and lightest flavour boil them **freshly**. If using leftovers, warm them in a con
7 gar - 110g (4 oz) Lime juice - 2 tbsp, **freshly** squeezed Galliano - 4 tsp Seedless gr
8 bsp Milk - 450 ml (3/4 pint) Egg - 1 **Freshly** grated nutmeg or ground cinnamon - a sp
9 anas, raisins or currants - 50g (2 oz) **Freshly** grated nutmeg - 1/4 tsp Double cream -
10 tablespoons marmalade and 3 tablespoons **freshly** squeezed orange juice. Serve hot as a s
11 dging with icing sugar and serving with **freshly** whipped cream. Serves 4-6 Milk - 600
12 Citrus Swing Milkshake Try to use the **freshly** squeezed juices - you will get a much b
13 l get a much better flavour. Serves 2 **Freshly** squeezed orange juice - 4 tbsp Freshly
14 Freshly squeezed orange juice - 4 tbsp **Freshly** squeezed lemon juice - 2 tbsp Lemon -
15 sugar cubes (or more or less to taste) **freshly** made strong coffee Double cream METHO
16 apot to the kettle and fill it with the **freshly** boiled water. 4. Stir the tea briefly,
17 e - 225 ml (8 fl oz) Parmesan cheese - **freshly** grated, to serve METHOD 1. Cook the 1
18 nguini to the anchovy mixture. Top with **freshly** grated Parmesan cheese. Clic
19 Tomato juice - 450 ml (3/4 pint) Salt **Freshly** ground black pepper METHOD 1. Put the
20 e 900g (2 lb) or 1 very large Salt and **freshly** ground pepper Sunflower oil - 1 tbsp
21 4 x 175g, 3 cm (11/4 inch) thick Salt **Freshly** ground black pepper Lemon juice - 4 ts
22 (1/2 oz) Plain flour - 2 tsp Salt and **freshly** ground black pepper Lemon juice - 1 ts
23 - 1, peeled and sliced thinly Salt and **freshly** ground pepper METHOD 1. Preheat a med
24 ack of frozen chopped spinach Salt and **freshly** ground black pepper Grated nutmeg - 1/
25 g (1 oz) Butter - 25g (1 oz) Salt and **freshly** ground black pepper Cheddar cheese - 7
26 ets Milk - 450 ml (3/4 pint) Salt and **freshly** ground pepper Butter - 65g (21/2 oz)
27 (3/4 oz) Hard boiled egg - 1, chopped **Freshly** ground black pepper Freshly made butte
28 , chopped Freshly ground black pepper **freshly** made buttered toast - 4 rounds, or frie
29 tatoes - 350g (12 oz), cooked Salt and **freshly** ground pepper Egg - 1, beaten Wholeme
30 110g (4 oz), rinded and diced Salt and **freshly** ground pepper Cream - 3 tbsp METHOD
31 ted Milk - 400 ml (14 fl oz) Salt and **freshly** ground black pepper Black peppercorns
32 hopped Eggs - 2 Watercress - 1 bunch **Freshly** ground pepper Whole prawns METHOD 1.

CorTrad

- Began in May 2008
- pt-en-pt parallel corpus - bidirectional
- multiversion
- POS-tagged
- semantically annotated
- Joint project:
 - Linguateca (design, development & implementation of computational framework) – Diana Santos
 - 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
 - comparison of different translation stages
 - translation “learner corpus”
 - study of revision process
 - **Refined search system** – tailored especially for each genre and text type
 - **Semantic information** – added and human-revised

CorTrad Parallel Subcorpora

Journalistic

Scientific
(pt → en)
1,076 texts

Technical-
Scientific

Cookbook
(pt → en)
130,000 words

Literary

Australian
Short Stories
(en → pt)
28 texts

Canadian
Short Stories
(en → pt)
20 texts

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)



Technical-Scientific: Cookbook



Original
(Brazilian
Portuguese)



Translators' first version
(English)



Revised text
(by American
native speaker)



Published translation
(not yet available
online)

Literary: Australian short stories (*learner corpus)



Original
(Australian
English)



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



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



**Published
translation**



Literary: Canadian short stories (*learner corpus)



Original
(Canadian
English)



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



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



**Published
translation**



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)



Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Bem-vindo ao CorTrad

O **CorTrad** é o corpus paralelo de tradução (português-inglês) do **COMET**. Além das possibilidades de pesquisa normalmente presentes (i) a possibilidade de se comparar diferentes versões de um mesmo texto (original, versões revisadas e tradução publicada); (ii) mecanismos para pesquisar seções específicas dos diferentes tipos textuais.

O CorTrad é um corpus aberto e conta atualmente com três subcorpora:

- CorTrad **jornalístico** (por ora, divulgação científica)
- CorTrad **literário** (por ora, contos)
- CorTrad **técnico-científico** (por ora, culinária)

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/NILC/Linguateca, usando o sistema DISPARA.

Início

Jornalístico

Literário

Técnico-científico

Agradecimentos

Projeto conjunto

Como citar

Perguntas já respondidas

Publicações

Fale conosco

O CorTrad dispõe de pelo menos duas funcionalidades inovadoras: dados para cada gênero pesquisado - permitindo, por exemplo,

Comentários ou questões para a equipe do CorTrad





CorTrad jornalístico divulgação científica

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

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. Veja uma tabela pormenorizada por assunto e gênero. 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 type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver
Tradução publicada	<input checked="" type="radio"/> principal	<input type="text" value="home"/>	<input checked="" type="checkbox"/> ver
<input checked="" type="checkbox"/> Ignorar maiúsculas/minúsculas			
<input type="button" value="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 (de cor, de vestuário, etc.)

Opções

<input type="checkbox"/> Resultados por ordem alfabética (só distribuições)
<input type="checkbox"/> Resultados numerados (só concordância)

Exemplos de pesquisas

	Procurar:	Resultado:
a palavra <i>malha</i> no singular ou plural	{lema="malha"}	concordância
palavras seguindo <i>malha</i>	"malha" @[]	distribuição dos lemas
o termo <i>buraco negro</i>	"buraco" "negro"	concordância
palavras entre travessões	"--" @[] "--"	distribuição das formas
formas de cor	{sema="cor.*"}	concordância



Corpus Multilíngue para Ensino e Tradução

USP

Português / Inglês

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

☒ principal pesquisa

☒ ver

Tradução publicada

☐ principal

☒ ver

☐ Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

☒ Concordância

☐ Distribuição das formas

☐ Distribuição dos lemas

☐ Distribuição da categoria gramatical (PoS)

Internet

100%



Corpus Multilíngue para Ensino e Tradução

USP

ções no link Artigos

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

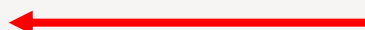
Literário

Técnico-científico

Ajuda

Se tem dúvidas sobre a sintaxe da pesquisa, pode ver por ora as seguintes páginas feitas para projetos relacionados:

- [Exemplos 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 a equipe do CorTrad



- ▶ [Home](#)
- ▶ [Search](#)
- ▶ [Help](#)

Search help
[Tutorial](#)

- ▶ [Texts in COMPARA](#)
- ▶ [General information](#)
- ▶ [Specific documentation](#)
- ▶ [Linguateca](#)

News [19 September 2008]
[Alignment constraints more visible in the concordances.](#)

- ▶ [Contact us](#)



Advanced Search Help

Help topics

- [1. Select search direction](#)
- [2. Enter query](#)
- [2.1 Word or expression query](#)
- [2.2 Further query options](#)
- [3. Use only a specific part of the corpus \(optional\)](#)
- [3.1 Choose specific varieties of Portuguese and English](#)
- [3.2 Select dates of publication](#)
- [3.3 Distinguish between source texts and translations](#)
- [3.4 Search only within specific texts](#)
- [3.5 Search within specific authors](#)
- [4. Choose other types of output](#)



1. Select search direction

Leave the "From Portuguese to English" option checked if you wish to enter a word or expression in Portuguese and see its English translation.

2. Enter query

2.1 Word or expression query

This is where you type in the word or expression you wish to research. It can be a word, a combination of words, a prefix, a suffix

As COMPARA contains only literary texts, the queries that work best are the ones that involve general vocabulary, the language of literature. It is very likely that you will find technical terms in COMPARA because our translation database does not, for now, include technical terminology.

We also advise you not to search for full sentences. To get good results from corpora, it is better to work with shorter chunks of text. It is more successful if you look up single words and conventional expressions such as *as soon as possible* than unique utterances such as *the king of the forest*.

What you want to search	Example of what to type in	What can be retrieved
<i>single word</i>	"this"	this
<i>two or more words</i>	"like" "this"	like this
<i>two words with any single word in between</i>	"a" ".*" "time"	a long time, a bad time, a short time, a good time, etc.
<i>two words with zero to three words in between</i>	"give" []* "up" within 3	give up, give it up, give them up, give it all up, etc.
<i>word with alternative spellings</i>	"reali[sz]e"	realise, realize
	"colo(u)?r"	colour, color
<i>alternative words</i>	"(big great)"	big, great
<i>words beginning with "dis"</i>	"dis.*"	dislike, disgusting, disappointing, disappear, disco, district
<i>words ending with "ly"</i>	".*ly"	finally, hopefully, beautifully, sadly, Sally, Billy, holy, etc.
<i>punctuation marks</i>	"\!"	!
	"\?"	?
	"\,"	,
<i>double quotation marks, single quotation marks, Portuguese "travessão"</i>	"(\« \»)"	«,»
	"(\' \')"	‘,’
	"\--"	--
<i>singular and plural</i>	[lema="boy"]	boy, boys
<i>verb inflections</i>	[lema="go"]	go, goes, going, gone, went

On the way to the home in a cable tram, we agreed that the Jews we had met in the Communist Party were intelligent, warm-hearted people.	No caminho para o asilo, no bonde, concordamos que os judeus que havíamos encontrado no Partido Comunista eram inteligentes, pessoas de bom coração.	No caminho para o asilo, no bonde, concordamos que os judeus que havíamos encontrado no Partido Comunista eram inteligentes, pessoas de bom coração.	No caminho para o asilo, no bonde, concordamos que os judeus que havíamos encontrado no Partido Comunista eram inteligentes, pessoas de bom coração.
We finally decided he must have met Jews during his frequent travels away from home in his younger days.	Finalmente concluímos que ele provavelmente teria se encontrado com judeus durante as frequentes viagens que fazia na época em que era mais jovem.	Finalmente concluímos que ele provavelmente teria se encontrado com judeus durante as frequentes viagens que fazia na época em que era mais jovem.	Finalmente, concluímos que ele provavelmente teria se encontrado com judeus durante as frequentes viagens que fazia na época em que era mais jovem.
At the home hospital (a sort of hospice for the dying) we found he was sinking fast.	No hospital do asilo (uma espécie de abrigo para os moribundos) o encontramos definhando rapidamente.	No hospital do asilo (uma espécie de abrigo para os moribundos) o encontramos definhando rapidamente.	No hospital do asilo (uma espécie de abrigo para os moribundos) o encontramos definhando rapidamente.
When my father began to leave us for long periods for his work -- he sold things -- he left me with the responsibility of fuelling the home .	Quando meu pai começou a se ausentar por períodos longos -- ele vendia coisas -- passou-me a responsabilidade de não deixar a casa sem lenha.	Quando meu pai começou a se ausentar por períodos longos -- ele vendia coisas -- passou-me a responsabilidade de não deixar a casa sem lenha.	Quando o meu pai começou a ausentar-se por períodos longos -- ele vendia coisas --, passou-me a responsabilidade de não deixar a casa sem lenha.
Short, winter afternoons I spent up the back splitting pine for kindling, long, fragrant spines with neat grain, and I opened up the heads of mill ends and sawn blocks of she-oak my father brought home .	Passava as curtas tardes de inverno nos fundos rachando pinheiro com o intuito de fazer gravetos para o fogo, espinhos longos e aromáticos de fibra pura, e abria as cabeças do que saía dos moinhos e das toras serradas de carvalho que meu pai trazia para casa.	Passava as curtas tardes de inverno nos fundos rachando pinheiro com o intuito de fazer gravetos para o fogo, espinhos longos e aromáticos de fibra pura, e abria as cabeças do que saía dos moinhos e das toras serradas de carvalho que meu pai trazia para casa.	Eu passava as curtas tardes de inverno nos fundos, rachando pinheiro, sua longa e aromática espinha dorsal com textura clara, para fazer gravetos, e abria suas pontas em forma de cone e serrava as toras de carvalho que meu pai trazia para casa.
Then I met Elaine and we married and I left home .	Aí conheci Elaine, nos casamos e saí de casa.	Aí conheci Elaine, nos casamos e saí de casa.	Aí conheci Elaine, nós nos casamos e eu saí de casa.
Then my mother had her stroke and Jamie began demanding to be paid for woodchopping and Elaine went twice a week to cook and clean for them and I decided on the Home .	Em seguida minha mãe teve um derrame e Jamie começou a pedir dinheiro para cortar lenha; Elaine ia duas vezes por semana cozinhar e lavar para eles, então, decidi pelo asilo.	Em seguida minha mãe teve um derrame e Jamie começou a pedir dinheiro para cortar lenha; Elaine ia duas vezes por semana cozinhar e lavar para eles, então, decidi pelo asilo.	Em seguida, minha mãe teve um derrame e Jamie começou a pedir dinheiro para cortar lenha; e Elaine ia duas vezes por semana cozinhar e lavar para eles, então decidi-me pelo asilo.
I thought about getting the place converted to electricity but the Home was expensive and Elaine came to enjoy cooking on the old woodstove and it was worth paying Jamie a little to chop wood.	Pensei em mudar tudo para energia elétrica, mas o asilo era caro e Elaine acabou gostando de cozinhar no velho fogão à lenha, aí compensava pagar ao Jamie um pouquinho para cortar a lenha.	Pensei em mudar tudo para energia elétrica, mas o asilo era caro e Elaine acabou gostando de cozinhar no velho fogão à lenha, aí compensava pagar ao Jamie um pouquinho para cortar a lenha.	Pensei em mudar tudo para energia elétrica, mas o asilo era caro e Elaine acabou gostando de cozinhar no velho fogão à lenha, aí compensava pagar um pouquinho ao Jamie para rachar a lenha.
I strike it home and regain equilibrium.	Golpeio o ponto de origem e recupero o equilíbrio.	Golpeio o ponto de origem e recupero o equilíbrio.	Golpeio o alvo e recupero o equilíbrio.
I can think of nothing I've done to offend the neighbours -- I'm not a dog biter or anything -- though some of them grumbled about my putting my parents into the Home , as though it was any of their business.	Não consigo pensar em nada que eu pudesse ter feito para ofender os vizinhos -- não sou um cara que provoca cachorros ou algo parecido -- embora alguns resmungaram de eu ter colocado meus pais no asilo, como se isso fosse da conta deles.	Não consigo pensar em nada que eu pudesse ter feito para ofender os vizinhos -- não sou um cara que provoca cachorros ou algo parecido -- embora alguns resmungaram de eu ter colocado meus pais no asilo, como se isso fosse da conta deles.	Não consigo pensar em nada que eu pudesse ter feito para ofender os vizinhos -- não sou o tipo que provoca cachorros ou algo parecido -- embora alguns tenham resmungado por eu ter colocado meus pais no asilo, como se isso fosse da conta deles.
Carrying it home I have the feeling that I'm holding a stage property, not a tool; there are no signs of work on it and the head is so clean and smooth and shiny it does not seem intended for chopping.	Ao levá-lo para casa tenho a sensação de estar segurando um objeto para teatro, não uma ferramenta; não há marcas de trabalho nele e a cabeça está tão limpa, polida e brilhante que ele não parece ter sido feito para rachar lenha.	Ao levá-lo para casa tenho a sensação de estar segurando um objeto para teatro, não uma ferramenta; não há marcas de trabalho nele e a cabeça está tão limpa, polida e brilhante que ele não parece ter sido feito para rachar lenha.	Ao levá-lo para casa, tenho a sensação de estar segurando um objeto de teatro, não uma ferramenta; não há marcas de trabalho nele, e a cabeça está tão limpa, polida e brilhante que ele não parece ter sido feito para rachar lenha.
'The Home called', she says.	'Ligaram do asilo', ela diz.	'Ligaram do asilo', ela diz.	-- Ligaram do asilo -- ela diz.
She wandered all day looking for the way back home .	Vagou durante todo o dia em busca do caminho de volta para casa.	Vagou durante todo o dia em busca do caminho de volta para casa.	Vagou o dia todo tentando voltar para casa.
I've got some jeans at home , actually.	Na verdade, tenho jeans em casa.»	Na verdade, tenho jeans em casa.»	Na verdade, tenho jeans em casa.
But when they were going home for dinner he glanced disapprovingly at the propped-up posts.	Mas quando estavam indo para casa almoçar, lançou um olhar reprovador para as estacas escoradas.	Mas quando estavam indo para casa almoçar, lançou um olhar reprovador para as estacas escoradas.	Mas quando estavam indo para casa almoçar, lançou um olhar reprovador para as estacas escoradas.
Round about four o'clock he decided to go home .	Por volta das quatro horas decidi ir para casa.	Por volta das quatro horas decidi ir para casa.	Por volta das quatro horas, decidi ir para casa.
When Henry came home , rather late in the evening, and somewhat the worse for wine, he thought he had come to the wrong farm until Pietro came out of his room and carried his parcels for him.	Quando Henry voltou, já de noite, e mais pra lá do que pra cá de tanto vinho, pensou que tivesse entrado na fazenda errada até que Pietro saiu de seu quarto para carregar os pacotes para ele.	Quando Henry voltou, já de noite, e mais pra lá do que pra cá de tanto vinho, pensou que tivesse entrado na fazenda errada até que Pietro saiu de seu quarto para carregar os pacotes para ele.	Quando Henry voltou, já de noite, e mais pra lá do que pra cá de tanto vinho, pensou que tivesse entrado na fazenda errada, até que Pietro saiu do quarto para carregar os pacotes para ele.

When is
Portuguese “natural”
not translated as
natural in English?

natural vs !natural

When “natural” is NOT “natural”



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. Clique em [atualizações](#) para saber mais.

Pesquisar no corpus

Ajuda

Original

☒ principal natural

☒ ver

Primeira tradução

☐ principal !natural

☒ ver

Tradução revisada

☐ principal !natural

☒ ver

☐ Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

☒ Concordância

☐ Distribuição das formas

☐ Distribuição dos lemas

☐ Distribuição da categoria gramatical (PoS)

☐ Distribuição do tempo verbal e/ou do caso pronominal

☐ Distribuição de pessoa e/ou número

☐ Distribuição do gênero morfológico

☐ Distribuição da função sintática

☐ Distribuição por parte de obra

☐ Distribuição por tipo de texto

☐ Distribuição por receita

☐ Distribuição por campo semântico

☐ Distribuição por grupo (de cor, de vestuário, etc.)

Opções

Result: “natural” ≠ “*natural*”



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_TRAD1 !"natural" :CORTRAD_CULI_TRAD2 !"natural"

Resultado escolhido: concordância em contexto

Corpus pesquisado: originais

[ajuda]

14 ocorrências.

[voltar]

[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	½ cup whole milk plain yogurt	½ cup plain whole milk yogurt
3/4 de xícara de iogurte natural	1/3 cup sugar	? cup sugar ¼ cup white wine vinegar
1 1/2 xícara de iogurte natural	1 ½ cups whole milk plain yogurt	1½ cups plain whole milk yogurt
2 1/3 de xícaras de iogurte natural	2 1/3 cups whole milk plain yogurt	2? cups plain whole milk yogurt
1 xícara de iogurte natural	1 cup whole milk plain yogurt	1 cup plain whole milk yogurt
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 (3½ fl oz) plain whole milk yogurt
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
200 ml de iogurte natural (1 copinho)	200 ml (7 fl oz) whole milk plain yogurt	200 ml (7 fl oz) plain whole milk yogurt
200 ml de iogurte natural (1 copinho)	200 ml (7 fl oz) whole milk plain yogurt	200 ml (7 fl oz) plain whole milk yogurt
1 litro de suco de laranja natural	1 L (1.1 qt) freshly squeezed orange juice	1 L (1 qt) freshly squeezed orange juice
1 litro de suco de laranja natural	1 L (1.1 qt) freshly squeezed orange juice	1 L (1 qt) freshly squeezed orange juice

CorTrad's semantic annotation

Semantic annotation for **colour**
in English and Portuguese



For **clothes** – only in Portuguese
so far



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

Exemplos de pesquisas

	Procurar:	Resultado:
a palavra <i>down</i>	down	concordância
palavras começando por <i>viol</i>	"viol.*"	concordância
a expressão <i>at least</i>	"at" "least"	concordância
palavras antecedidas por dois pontos	":" @[]	distribuição das formas
formas do verbo <i>see</i> em contexto	[lema="see"]	concordância
palavras modificadas por <i>rather</i>	"rather" @[]	distribuição de formas
acsos em que a tradução não tem <i>vermelho</i>	! [lema="vermelho"]	concordância

Clique em atualizações para saber do andamento do projeto.

Semantic information: colour

	Cooking	Scientific news	Short stories	Totals
Pure colour	574	372	344	1290
Conventional	310	153	2	465
Race	0	45	13	58
Human	0	7	39	46
Absence	7	21	22	50
Wine	87	1	6	94
Totals	985	599	428	
Word count	134,093	776,284	121,253	

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

Result type: semantic field

10 most recurring colour terms

Short stories	Scientific news	Cooking
white	black	brown
black	color	black
blue	white	white
red	green	red
grey	red	green
brown	yellow	color
green	blue	golden
yellow	yellowing	yellow
colour	greenhouse	purple
pink	gray	brown

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

Result type: lemma distribution

Concordances

Yellowing:

- The final result will probably take the form of a vaccine against the **yellowing disease**.
- It was another important victory in the fight against the **yellowing disease**.

Golden:

- Bake for about 20 minutes, until rolls are slightly **golden** on all sides and lose the appearance of raw dough.
- Lower oven temperature to 200C (400F / moderately hot / Gas 6) and bake for 25 minutes or so, until bread loaves are risen and **golden brown**.

“white” collocates in ≠ genres

Short stories	Scientific News	Cooking
man	dwarf	wine
hand	cube	chocolate
feather	house	rice
noodle	spot	part
crockery	blood	pith
fence	crab	pepper
Camellia	shrimp	sandwich
knuckle	fluid	button
handbag	stripe	bean
cockatoo	hair	hominy

Concordances

Dwarf

- A physicist from Rio Grande do Sul shows how to make use of the variations in the brightness from pulsating **white dwarf stars**.

Cube

- The concept of the «**white cube**» arose in 1939, at the inauguration of the then new building of the New York Museum of Modern Art (MoMA), in which the paintings are hung at the viewer's eye height in completely neutral surroundings.

Blood

- ...the benefic action of which consists of increasing the speed of recovery of the neutrophils, a kind of **white blood cell** specialized in ...
- In the lymphocytes, a kind of **white blood globule**, the rate of aneuploidy is 3%.

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 sobre **arroz integral**
- **arborígene**
- **pão de forma**
- **vinho tinto**
- **repolho roxo**

Some remarks on colour

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

Corpus-based TS are revealing because

- you discover challenges the translator had to cope with
- you discover vagueness in both languages
- you find problems you had never thought about before
- you find creative solutions

BUT...

CorTrad

can be used

as a monolingual corpus

Possible Queries

Which adjectives
precede “contribution”?

`[pos="JJ.*"] "contribution"`

→ possible adjectival collocations

Adjectival collocations of *contribution*

We have received issue n 66 of the Pesquisa FAPESP magazine, which will bring a **great contribution** towards the dissemination of actions of technological research and innovation in the ambit of mini and small enterprises in São Paulo.

They could weigh up to 5 tons and reach a length of 6 meters. The proof that these mammals from the past came to share, for some time, the same environment as the first American paleoindian populations is a **weighty contribution** of the prehistoric sites of Minas Gerais to archeology.

Another **important contribution** was to manipulate the LTR plasmid component - which sets off the expression of the DNA sequence that comes next, to prevent the viral vector expression from decaying.

What will be the **major contribution** of genomic studies to this new form of history?

Still on the MgB₂, the group led by the physicist Dr. Oscar de Lima, who is also part of the thematic project, made an **unprecedented contribution**, determining that the superconductivity properties of MgB₂ depend on the direction in which the external magnetic field is applied».

The other report, making use of the opportunity of the recent inauguration of the first stage of an industrial unit from the company Embraer in the town of Gavião Peixoto, São Paulo, shows on what footing one can find the research projects that make up part of the program of Partnership for Innovation in Aero Space Science and Technology (PICTA), supported by FAPESP, which represents a **fundamental contribution** to the transformation of that region of the state of São Paulo into a large center of aeronautical development.

Another **important contribution** will come from Sparc, an association of American and European publishers and bookstores specialized in the publication of university magazines on-line.

Very well: an **important contribution** to the efforts by specialists from several parts of the world to locate the field of communication is magnificently set out in a new book by a Brazilian: Antropológica do Espelho (The Mirror Anthropology), published by Vozes, launched at the end of last April.

Sacks, the author of the book Seeing Voices: a Journey through the World of the Deaf, emphasizes the importance of the deaf creating their own language and the **effective contribution** made by the libras

Native speaker “contributions”

- COCA
 - significant - 377
 - important - 295
 - major - 171

important	3
financial	2
good	2
weighty	1
major	1
unprecedented	1
significant	1
social	1
effective	1
big	1
fundamental	1
great	1
technological	1
scientific	1
possible	1
like	1
brazilian	1

Comparing text types

- How do verbs *think* and *believe* behave in different text types?
- [lema="think"] + Distribution by text genre
- [lema="believe"] + Distribution by text genre



CorTrad jornalístico divulgação científica

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

A **parte jornalística do CorTrad** conta atualmente com textos das edições de 2001, 2002 e 2003 da [Revista Pesquisa FAPESP](#), [Estratégias](#), [Laboratório](#), [Linha de Produção](#) e [Política de C & T](#). Veja [uma tabela pormenorizada por assunto e gênero](#). A disponibilidade de textos varia de acordo com o assunto, sendo que alguns textos não estão disponíveis usando o sistema [DISPARA](#).

Pesquisar no corpus



Original

☐ principal

☒ ver

Tradução publicada

☒ principal

[lema="think"]

☒ ver

☒ Ignorar maiúsculas/minúsculas

Resultado

☐ Concordância

☐ Distribuição

☐ Distribuição dos lemas

☐ Distribuição

☐ Distribuição do tempo verbal e/ou do caso pronominal

☐ Distribuição

☐ Distribuição do gênero morfológico

☐ Distribuição

☐ Distribuição por documento

☐ Distribuição

☒ Distribuição por gênero de texto

☐ Distribuição

Think vs. Believe

Think		Believe	
Story	95	Story	125
Interview	81	Article	76
Article	66	News	58
News	33	Interview	18
Editorial	6	Editorial	1
Letter	2		
Review	1		

AND LAST...
BUT NOT LEAST...



Corpus de Aprendizes

O que é? Ajuda Contato

Bem-vindo ao Corpus de Aprendizes



Aluno

Se você é aluno, entre no sistema preenchendo os campos abaixo:

Login:

Senha:

Esqueceu sua senha?

Caso deseje obter um login para utilizar o sistema, basta [criar uma conta](#) no sistema.

>> [Crie uma conta.](#)



Professor/Pesquisador

Se você é professor ou pesquisador, entre no sistema preenchendo os campos abaixo:

Login:

Senha:

Esqueceu sua senha?

Caso deseje obter um login para utilizar o sistema, entre em contato com o administrador do Corpus.

FFLCH :: Corpus Multilíngue para Ensino e Tradução (COMET)



Learner Corpus

Learner Corpus

Student written production

- English, French, German, Italian, Spanish

Automatic upload of compositions

Same tools: frequency list, concordancer

Search

- age group
- sex
- class
- language
- level
- Students fill out form with personal info
- Students grant permission for use of texts

Criar conta de usuário

Para criar uma conta no Corpus de Aprendizes, é necessário preencher o cadastro abaixo. Após o cadastro você receberá um email com seus dados. Os campos com asterisco (*) são obrigatórios.

Dados Pessoais

Nome completo sem abreviações*:

Sexo*:

Data de Nascimento*:

 / /

Nacionalidade*:

Língua do pai*:

Língua da mãe*:

Língua praticada em casa*:

E-mail*:

Curso de Línguas

Faz ou fez curso de língua?:

Educação

Escolaridade*:

Acesso para submissão de redações

Nome de Acesso*:

Senha*:

Confirma Senha*:

Semestre:

Disciplina:

Autorização

☐ Autorizo a inclusão de minhas redações no Corpus de Aprendizes - COMET - para fins de pesquisa.

Criar conta

Student
enrollment
page

Personal
information

Permission



Submitting a text

Corpus de Aprendizizes

 carolzupp

O que é ?

Ajuda

Contato

Personal data

Meus Dados

Seguem abaixo seus dados, utilize as opções do menu ao lado para gerenciá-los:

Dados de **Maria Carolina Zuppardo**:

- Data de nascimento: **11/08/1979**
- Sexo: **F**
- Nacionalidade: **Brasileira**
- Língua Nativa: **alemão**
- Língua Paterna: **português**
- Língua Materna: **português**
- Língua praticada em casa: **português**
- Email: **carolzupp@gmail.com**
- Faz outro curso de língua: **S**
 - Quais línguas: **inglês**
 - Anos de estudo: **10**
 - Já morou no país da língua estrangeira que está estudando: **S**
 - Hábitos com a língua: **internet, filmes, livros, revistas, jornais, outros**
- Escolaridade: **Superior Incompleto**
 - Curso: **Licenciatura em letras - inglês e português**
 - Semestre atual: **5o. semestre**
 - Instituição: **Universidade Paulista (UNIP)**
 - Ano de ingresso: **2008**
- Data de criação da conta: **22/10/2009 Às 15:34:13**
- Último login: **27/04/2010 Às 10:28:47**
- Número de Acessos: **5**

▶ Meus Dados

▶ Modificar Dados

▶ Alterar Senha

▶ Submeter Redação

▶ Sair

Teachers

- Can receive texts via e-mail
- Can compile their own “corpus”

Student production

Redações

A pesquisa retornou 7 resultado(s).

Nome do Aluno	Título da Redação	Curso	Data de Submissão
Renata Alabi	Application letter	LILC1-Lilc1: LingIngLingCont	29/04/2008 Às 14:19:30
Renata Alabi	Application form 2	LILC1-Lilc1: LingIngLingCont	29/04/2008 Às 14:20:19
Renata Alabi	My changing habits	LILC1-Lilc1: LingIngLingCont	28/05/2008 Às 14:04:35
Renata Alabi	To continue a story	LILC1-Lilc1: LingIngLingCont	13/06/2008 Às 12:10:08
Renata Alabi	A wedding reception - writing class	LILC1-Lilc1: LingIngLingCont	13/06/2008 Às 14:40:05
Renata Alabi	Article 1	LILC1-Lilc1: LingIngLingCont	19/06/2008 Às 22:52:25
Renata Alabi	Article 2	LILC1-Lilc1: LingIngLingCont	19/06/2008 Às 22:52:51

Exportar para arquivo ZIP COM cabeçalho

Exportar para arquivo ZIP SEM cabeçalho

479_LILC1_Lilc1_321.txt - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

Application form 2

250 Camargo Way
Flórida
32714
28th April 2008

Mr. Tombros
Lake Brantley High School
753 Winter Way

Dear Mr. Tombros

I would like to apply for the post of trainee teacher which was advertised in today's issue of the Altamont. I am 25 years-old and I will be graduating at the end of this academic year. I have, TELP, TKT and FCE Certificate. I have been involved in the production and carrying out of my University projects for the last three years and I have please find enclosed the name and address of one referee who can testify my conduct and character. I should be glad to attend an interview at any time which is convenient to you.

Yours sincerely,

Next steps

CorTec

- include new corpora

CorTrad

- revise alignment for new corpora
- revise semantic tagging
- include more parallel texts

CoMAprend

- correct “bugs”
- include new functionalities

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 (Linguateca)

References

- Bick, Eckhard. The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus University Press, 2000.
- Christ, Oliver, B. Schulze, A. Hofmann & E. Koenig (1999) "The IMS Corpus Workbench: Corpus QueryProcessor (CQP): User's Manual", Institute for Natural Language Processing, University of Stuttgart, March 8, 1999 (CQP V2.2)
- Evert. Stefan Evert and The OCWB Development Team. 2010a. The IMS Open Corpus Workbench (CWB). Corpus encoding tutorial. http://cwb.sourceforge.net/files/CWB_Encoding_Tutorial.pdf.
- Evert. Stefan Evert and The OCWB Development Team. 2010a. The IMS Open Corpus Workbench (CWB). Corpus encoding tutorial. http://cwb.sourceforge.net/files/CWB_Encoding_Tutorial.pdf.
- Santos, Diana. "DISPARA, a system for distributing parallel corpora on the Web". In Nuno Mamede & Elisabete Ranchhod (eds.), *Advances in Natural Language Processing (PorTAL 2002)* (Faro, Portugal, 23-26 de Junho de 2002), Berlin/Heidelberg : Springer-Verlag. Lecture Notes in Artificial Intelligence 2389, pp. 209-218.
- Santos, Diana & Cristina Mota. "Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora". In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)* (Valletta, Malta, 17-23 de Maio de 2010), European Language Resources Association, pp. 1437-1444.
- Rayson, P., and Garside, R. (1998). The CLAWS Web Tagger. *ICAME Journal*, no. 22. The HIT-centre - Norwegian Computing Centre for the Humanities, Bergen, pp. 121-123.

*Thank you
Stella*

seotagni@usp.br