Corpus Linguistics at USP

Stella E. O. Tagnin
University of São Paulo

Encontro Acadêmico Brasil-Itália: entre Léxico e Corpora, aplicações práticas e teóricas
USP - August 2, 2013
Outline

Project CoMET
- CorTec – technical corpus
- CorTrad – translation corpus
- CoMAprend – learner corpus

Illustrated with possible queries
PROJETO CORpus Multílingüe 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)
  - CoMAprened (Learner Corpus)
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
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
Step 2/3: Select your tool.

List of Tools

- Concordancer
- Frequency Counters
- N-Gram Generator

CorTec – Technical Scientific Corpus
Step 3/3: Applying Tool to Corpus.

Select the tool according to the options below:

**Concordancer**

Expression or word:  
- Same as  
- cardia  

Case sensitive:  
- Starting with  
- Ending in  
- Containing  

Size of reduced context (in characters): 20

>> Go

Select another tool <<

CorTec – Technical Scientific Corpus
Step 3/3: Applying Tool to Corpus.

Select the tool according to the options below:

Concordancer

Expression or word: Starting with cardia
Case sensitive: □
Size of reduced context (in characters): 20

>> Go

Select another tool <<
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 ischemia Scan 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 indirect
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 (FTCA), SRP increased cardiac output and LV stroke work i
11 trograde cardioplegia during cardiac surgery, and b) drug retroi
12 n solution to right sided 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; owing great cardiac vein compared to right atria
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
Ocorrência 18

and 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

patients 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 86. 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 |
### FREQUENCY TABLE

<table>
<thead>
<tr>
<th>Pos.</th>
<th>Word</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>the</td>
<td>15398</td>
</tr>
<tr>
<td>2</td>
<td>of</td>
<td>9445</td>
</tr>
<tr>
<td>3</td>
<td>to</td>
<td>7055</td>
</tr>
<tr>
<td>4</td>
<td>or</td>
<td>6841</td>
</tr>
<tr>
<td>5</td>
<td>and</td>
<td>6778</td>
</tr>
<tr>
<td>6</td>
<td>in</td>
<td>4026</td>
</tr>
<tr>
<td>7</td>
<td>any</td>
<td>4025</td>
</tr>
<tr>
<td>8</td>
<td>shall</td>
<td>2760</td>
</tr>
<tr>
<td>9</td>
<td>be</td>
<td>2455</td>
</tr>
<tr>
<td>10</td>
<td>by</td>
<td>2426</td>
</tr>
<tr>
<td>11</td>
<td>a</td>
<td>2319</td>
</tr>
<tr>
<td>12</td>
<td>such</td>
<td>2220</td>
</tr>
<tr>
<td>13</td>
<td>as</td>
<td>1973</td>
</tr>
<tr>
<td>14</td>
<td>for</td>
<td>1628</td>
</tr>
<tr>
<td>15</td>
<td>this</td>
<td>1619</td>
</tr>
<tr>
<td>16</td>
<td>with</td>
<td>1558</td>
</tr>
<tr>
<td>17</td>
<td>other</td>
<td>1344</td>
</tr>
<tr>
<td>18</td>
<td>parte</td>
<td>1305</td>
</tr>
<tr>
<td>19</td>
<td>agreement</td>
<td>1245</td>
</tr>
<tr>
<td>20</td>
<td>not</td>
<td>1216</td>
</tr>
<tr>
<td>21</td>
<td>is</td>
<td>1194</td>
</tr>
<tr>
<td>22</td>
<td>that</td>
<td>1119</td>
</tr>
<tr>
<td>23</td>
<td>all</td>
<td>1101</td>
</tr>
<tr>
<td>24</td>
<td>on</td>
<td>1035</td>
</tr>
<tr>
<td>25</td>
<td>infosig</td>
<td>979</td>
</tr>
<tr>
<td>26</td>
<td>its</td>
<td>947</td>
</tr>
<tr>
<td>27</td>
<td>which</td>
<td>933</td>
</tr>
<tr>
<td>28</td>
<td>company</td>
<td>863</td>
</tr>
<tr>
<td>Page</td>
<td>Word</td>
<td>Page</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>152</td>
<td>expenses</td>
<td>145</td>
</tr>
<tr>
<td>153</td>
<td>borrower</td>
<td>145</td>
</tr>
<tr>
<td>154</td>
<td>purchase</td>
<td>144</td>
</tr>
<tr>
<td>155</td>
<td>equipment</td>
<td>144</td>
</tr>
<tr>
<td>156</td>
<td>lender</td>
<td>143</td>
</tr>
<tr>
<td>157</td>
<td>entitled</td>
<td>143</td>
</tr>
<tr>
<td>158</td>
<td>and/or</td>
<td>142</td>
</tr>
<tr>
<td>159</td>
<td>insurance</td>
<td>142</td>
</tr>
<tr>
<td>160</td>
<td>iii</td>
<td>141</td>
</tr>
<tr>
<td>161</td>
<td>claim</td>
<td>141</td>
</tr>
<tr>
<td>162</td>
<td>legal</td>
<td>141</td>
</tr>
<tr>
<td>163</td>
<td>act</td>
<td>140</td>
</tr>
<tr>
<td>164</td>
<td>costs</td>
<td>138</td>
</tr>
<tr>
<td>165</td>
<td>contemplated</td>
<td>138</td>
</tr>
<tr>
<td>166</td>
<td>contract</td>
<td>138</td>
</tr>
<tr>
<td>167</td>
<td>full</td>
<td>138</td>
</tr>
<tr>
<td>168</td>
<td>between</td>
<td>137</td>
</tr>
<tr>
<td>169</td>
<td>assin</td>
<td>136</td>
</tr>
<tr>
<td>170</td>
<td>either</td>
<td>135</td>
</tr>
<tr>
<td>171</td>
<td>day</td>
<td>133</td>
</tr>
<tr>
<td>172</td>
<td>nor</td>
<td>133</td>
</tr>
<tr>
<td>173</td>
<td>delivered</td>
<td>132</td>
</tr>
<tr>
<td>174</td>
<td>delivery</td>
<td>132</td>
</tr>
<tr>
<td>175</td>
<td>performance</td>
<td>132</td>
</tr>
<tr>
<td>176</td>
<td>obligation</td>
<td>131</td>
</tr>
<tr>
<td>177</td>
<td>writing</td>
<td>131</td>
</tr>
<tr>
<td>178</td>
<td>limitation</td>
<td>130</td>
</tr>
<tr>
<td>179</td>
<td>corporation</td>
<td>130</td>
</tr>
<tr>
<td>180</td>
<td>request</td>
<td>129</td>
</tr>
<tr>
<td>181</td>
<td>result</td>
<td>129</td>
</tr>
<tr>
<td>182</td>
<td>sale</td>
<td>129</td>
</tr>
</tbody>
</table>
PROFITS OR REVENUE, WHETHER ARISING IN CONTRACT OR TORT, EVEN IF CENTER HAS BEEN ADVISED
of your liability to you, whether in contract, tort (including negligence), or otherw
ecessary to enter into any such contract and/or agreement, including but not li
the ownership of voting securities, by contract, or otherwise. "Business Day" shall me
Y WITHOUT REGARD TO FAULT, OR BREACH OF CONTRACT OR WARRANTY, FOR ANY CONSEQUENTIAL SPE
A. ASSIGNMENT OF RELINQUISHED PROPERTY CONTRACT
1. Owner agrees to transfer (as perm
hed Property, pursuant to a real estate contract dated __________ (the "Rel
_______ (the "Relinquished Property Contract"). The Relinquished Property Contract
17 Contract"). The Relinquished Property Contract is attached hereto as Exhibit B and is
ations) in the Relinquished Property Contract, and Exchanger agrees to accept assignm
ligations) in the Relinquished Property Contract. B. NOTICE TO PURCHASER On or before
r's rights in the Relinquished Property Contract have been assigned to Exchangor. ARTI
D. ASSIGNMENT OF REPLACEMENT PROPERTY CONTRACT
1. Owner will enter into any contract
"Replacement Property Contract") and Owner will assign its rights (but
ions) under each Replacement Property Contract to Exchangor. It is understood and agree
on to execute or take assignment of any contract, or to do any other act or thing conten
is Agreement, the Relinquished Property Contract or the Replacement Property Contract wi
ty Contract or the Replacement Property Contract without, in each case, receiving a writ
seller under said Replacement Property Contract, it being understood that Owner's rights in said replacemen
r's rights in said Replacement Property Contract have been assigned to Exchangor. E.
TED DAMAGES Every Replacement Property Contract shall provide that (i) Exchangor's liab
ability under the Replacement Property Contract shall be limited to the down payment th
closing under the Replacement Property Contract, except for representations and warrant
ement. If any such Replacement Property Contract requires Exchangor to execute a purchas
l property and the Replacement Property Contract or the lease being assigned thereunder
assignment. If any Replacement Property Contract or related transaction) requires Excha
The term "mortgage" as used in this Ex
r pursuant to the Replacement Property Contract. C. REPLACEMENT PROPERTY CLOSING DAT
pursuant to the Replacement Property Contract relating thereto. E. CLOSING ADJUSTME
Agreement and the Replacement Property Contract(s). Owner acknowledges that in connecti
ge Agreement, the Relinquished Property Contract or the Replacement Property Contract, o
ry Contract or the Replacement Property Contract, or shall incur any expenses by reason
ge Agreement, the Relinquished Property Contract or the Replacement Property Contract, w
ry Contract or the Replacement Property Contract, whether as a tax, or for breach of con
act, whether as a tax, or for breach of contract, injury to person or property, or fines
USING CORTEC AS A MONOLINGUAL CORPUS
Cooking corpus

- How frequent are adverbs in –ly?
- Which are the most frequent?
- Which are their collocates?
Step 3/3: Applying Tool to Corpus.

Select the tool according to the options below:

**Concordancer**

Expression or word: Ending **in** ly

Case sensitive: ☐

Size of reduced context (in characters): 20

>> Go

Select another tool <<
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 name rub in the butter until it is thoroughly mixed. Mix in the sugar and caraway seed with a knife. 2. Turn onto a lightly floured surface. Knead quickly until smooth. Roll out to about 1 cm (1/2 inch) thick. Cut into small rounds approximately 5 cm (2 inches) in diameter. 3. Place the scones on a lightly greased baking sheet and brush the tops with the milk and almond essence. Mix lightly with a wooden spoon or fork until a soft dough forms. Knead gently until smooth. Roll out to a thickness of 2cm (1 inch) thick. Cut into small rounds approximately 5 cm (2 inches) in diameter. 4. Place the scones on a lightly greased baking sheet and brush the tops with the milk and almond essence. Mix lightly with a wooden spoon or fork until a soft dough forms. Knead gently until smooth. Roll out to a thickness of 2cm (1 inch) thick. Cut into small rounds approximately 5 cm (2 inches) in diameter.

5. Place the scones on a lightly greased baking sheet and brush the tops with the milk and almond essence. Mix lightly with a wooden spoon or fork until a soft dough forms. Knead gently until smooth. Roll out to a thickness of 2cm (1 inch) thick. Cut into small rounds approximately 5 cm (2 inches) in diameter.

4. Remove from the heat and cool slightly. Gradually add the eggs, a little at a time, to the mixture. Beat until the eggs are fully incorporated. 5. Place the egg mixture into the bowl and fold in the flour mixture until just combined. 6. Fold gently until the mixture is smooth. Place in a preheated oven at 180°C (350°F) until the mixture is golden brown.

**Note:** The ingredients for the Scones include:
- 
145 g Unsalted peanuts - 75g (3 oz)
- 75g (3 oz)
- Soft brown sugar - 75g (3 oz)
- Milk - 250 ml (8 fl oz)
- Egg - 1, lightly beaten
- Walnut halves - to decorate
- Mix to a soft dough. Knead on a lightly floured surface until smooth. Place in a preheated oven at 180°C (350°F) until the mixture is golden brown.
The most common adverbs in -ly

- Freshly = 3117
- Finely = 3092
- Gently = 2345
- Lightly = 1524
- Thinly = 637
- Carefully = 635
- Immediately = 622
- Evenly = 327
1 ugar - 4 tbsp Ground ginger - 1/2 tsp Freshly grated nutmeg Sultanas - 175g (6 oz) 4. Salt - 1 tsp Ground cinnamon - 2 tsp Freshly grated nutmeg - 1 tsp Ground cloves - 5 1pted Tomatoes - 300g (11 oz) Salt and freshly ground pepper Oregano - 1 tsp Mushroom Freshly grated orange juice boil them until golden. If using leftovers, warm them in a con freshly squeezed Galliano - 4 tsp Seedless gr freshly squeezed orange juice. Serve hot as a side. 11 freshly whipped cream. Serves 4-6 Milk - 600 freshly squeezed juices - you will get a much b 14 Freshly squeezed orange juice - 4 tbsp Lemon - 15 sugar cubes (or more or less to taste) 16 apot to the kettle and fill it with the freshly boiled water. 4. Stir the tea briefly, E - 225 ml (8 fl oz) Parmesan cheese - Freshly grated, to serve METHOD 1. Cook the 1 18 niqui to the anchovy mixture. Top with 19 Tomato juice - 450 ml (3/4 pint) Salt 20 e 900g (2 lb) or 1 very large Salt and freshly ground pepper Sunflower oil - 1 tbsp Freshly ground black pepper Lemon juice - 4 ts Freshly ground black pepper Lemon juice - 1 ts Freshly ground black pepper METHOD 1. Preheat a med 24 chock of frozen chopped spinach Salt and freshly ground black pepper METHOD 1. Cook 25 g (1 oz) Butter - 25g (1 oz) Salt and freshly ground black pepper Butter - 65g (21/2 oz) 26 ets Milk - 450 ml (3/4 pint) Salt and freshly ground black pepper 27 (3/4 oz) Hard boiled egg - 1, chopped Freshly made buttered toast - 4 rounds, or frie 28 , chopped Freshly ground black pepper Freshly ground black pepper Egg - 1, beaten Wholeme 29 tatoes - 350g (12 oz), cooked Salt and Freshly ground black pepper Cream - 3 tbsp METHOD 30 110g (4 oz), rinded and diced Salt and freshly ground black pepper Black peppercorns 31 ted Milk - 400 ml (14 fl oz) Salt and freshely ground black 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)
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)
**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/](http://visl.hum.sdu.dk/visl/pt/)
  - English: **CLAWS** (Rayson & Garside 1998)
    - [http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/](http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/)
  - Semantic annotation: **corte-e-costura** (Santos & Mota 2010)

- **Interface** (graphic design by Patricia Tagnin)
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) mecanismo de pesquisar seções específicas dos diferentes tipos textuais.

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

- 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/Linguatexa, usando o sistema DISPARA.
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 da edição de 2001, 2002 e 2003 da Revista Pesquisa FAPESP, totalizando 29 números. As seções incluídas foram: Humanidades, Ciência, Tecnologia, Estratégia, 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 Língua e o NILC, usando o sistema DISPARA.

Pesquisar no corpus

<table>
<thead>
<tr>
<th>Original</th>
<th>principal</th>
<th>ver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tradução publicada</td>
<td>principal</td>
<td>home</td>
</tr>
</tbody>
</table>

Exemplos de pesquisas

<table>
<thead>
<tr>
<th>Procurar</th>
<th>Resultado</th>
</tr>
</thead>
<tbody>
<tr>
<td>[lema=&quot;malha&quot;]</td>
<td>concordância</td>
</tr>
<tr>
<td>&quot;malha&quot; @[]</td>
<td>distribuição dos lemas</td>
</tr>
<tr>
<td>&quot;burocrático&quot; &quot;negro&quot;</td>
<td>concordância</td>
</tr>
<tr>
<td>&quot;--&quot; @[] &quot;--&quot;</td>
<td>distribuição das formas</td>
</tr>
<tr>
<td>[cor=&quot;cor&quot;]</td>
<td>concordância</td>
</tr>
</tbody>
</table>
CorTrad jornalístico divulgação científica

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

Pesquisar no corpus

<table>
<thead>
<tr>
<th>Original</th>
<th>principal</th>
<th>pesquisa</th>
<th>ver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tradução publicada</td>
<td>principal</td>
<td></td>
<td>ver</td>
</tr>
</tbody>
</table>

Ignorar maiúsculas/minúsculas

Resultado

- Concordância
- Distribuição das formas
- Distribuição dos lemas
- Distribuição da categoria gramatical (PoS)
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.
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. To go the other way around, check the "From English to Portuguese" option.

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, or a number. As COMPARA contains only literary texts, the queries that work best are the ones that involve general vocabulary, the language that is present in books and periodicals. It is very likely that you will find technical terms in COMPARA because our translation database does not, for now, include technical texts.

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 language. You are most likely to be successful if you look up single words and conventional expressions such as as soon as possible than unique utterances such as I don't know how to do it.
<table>
<thead>
<tr>
<th>What you want to search</th>
<th>Example of what to type in</th>
<th>What can be retrieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>single word</td>
<td>&quot;this&quot;</td>
<td>this</td>
</tr>
<tr>
<td>two or more words</td>
<td>&quot;like&quot; &quot;this&quot;</td>
<td>like this</td>
</tr>
<tr>
<td>two words with any single word in between</td>
<td>&quot;a&quot; &quot;.&quot; &quot;time&quot;</td>
<td>a long time, a bad time, a short time, a good time, etc.</td>
</tr>
<tr>
<td>two words with zero to three words in between</td>
<td>&quot;give&quot; []* &quot;up&quot; within 3</td>
<td>give up, give it up, give them up, give it all up, etc.</td>
</tr>
<tr>
<td>word with alternative spellings</td>
<td>&quot;reali[sz]e&quot;</td>
<td>realise, realize</td>
</tr>
<tr>
<td></td>
<td>&quot;colo(u)r&quot;</td>
<td>colour, color</td>
</tr>
<tr>
<td>alternative words</td>
<td>&quot;(big</td>
<td>great)&quot;</td>
</tr>
<tr>
<td>words beginning with &quot;dis&quot;</td>
<td>&quot;dis.&quot;&quot;</td>
<td>dislike, disgusting, disappointing, disappear, disco, distribution</td>
</tr>
<tr>
<td>words ending with &quot;ly&quot;</td>
<td>&quot;.&quot;.&quot;</td>
<td>finally, hopefully, beautifully, sadly, Sally, Billy, holy, etc.</td>
</tr>
<tr>
<td>punctuation marks</td>
<td>&quot;!&quot;</td>
<td>!</td>
</tr>
<tr>
<td></td>
<td>&quot;/?&quot;</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>&quot;,,&quot;</td>
<td>.</td>
</tr>
<tr>
<td>double quotation marks, single quotation marks, Portuguese &quot;travessão&quot;</td>
<td>&quot;( (&lt;</td>
<td>&gt;)&quot;)&quot;</td>
</tr>
<tr>
<td></td>
<td>&quot;(&quot; &quot;)&quot;</td>
<td>. ,</td>
</tr>
<tr>
<td>singular and plural</td>
<td>[lema=&quot;boy&quot;]</td>
<td>boy, boys</td>
</tr>
<tr>
<td>verb inflections</td>
<td>[lema=&quot;go&quot;]</td>
<td>go, goes, going, gone, went</td>
</tr>
</tbody>
</table>
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.

We finally decided not to meet Jews during his frequent travels away from home.

At the home hospital (a sort of hospice for the dying) we found he was sinking fast.

When my father began to leave for long periods for his work -- he sold things -- he left me with the responsibility of fuelling the home.

Short, winter afternoons I spent up the back splitting pine for kindling, long, fragrant spines with neat grain, and opened up the heads of pine and sawn blocks of she-oak my father brought home.

Then I met Elaine and we married and I left home.

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.

I thought about getting the place converted to electricity, but the Home was expensive and Elaine was cooking on gas so I was saving money paying Jamie a little for cutting wood.

I can think of nothing I have done to offend the neighbours -- I am not a dog barker or anything -- though some of them grumbled about my putting my parents into the Home, as though it was any of their business.

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 shed is so clean and smooth and shiny it doesn't seem intended for chopping.

"The Home called," she says.

She wandered all day looking for the way back home.

I've got some jeans at home, actually.

But when they were going home for dinner he glanced disapprovingly at the dropped-up posts.

Round about four o'clock he decided to go home.

When Henry came home, rather late in the evening, and somewhat 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.
When is Portuguese “natural” not translated as natural in English?

natural vs !natural
When "natural" is NOT "natural"
Result: “natural” ≠ “natural”

### CorTrad técnico-científico culinária

Expression of search: "natural":CORTRAD_CULL_TRAD1 "natural":CORTRAD_CULL_TRAD2 "natural"
Result selected: concordance in context
Corpus searched: original

14 occurrences.

<table>
<thead>
<tr>
<th>Original</th>
<th>First translation</th>
<th>Revised translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>O mais natural é imaginar um almoço lá fora, mas por que não um jantar, com o frescor da noite?</td>
<td>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?</td>
<td>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?</td>
</tr>
<tr>
<td>1/2 xícara de iogurte natural</td>
<td>¼ cup whole milk plain yogurt</td>
<td>½ cup plain whole milk yogurt</td>
</tr>
<tr>
<td>3/4 da xícara de iogurte natural</td>
<td>1/3 cup sugar</td>
<td>1/3 cup sugar</td>
</tr>
<tr>
<td>1 1/2 xícara de iogurte natural</td>
<td>1 1/3 cups whole milk plain yogurt</td>
<td>1 1/3 cups whole milk plain yogurt</td>
</tr>
<tr>
<td>2 1/3 de xícaras de iogurte natural</td>
<td>2 1/3 cups whole milk plain yogurt</td>
<td>2 1/3 cups whole milk plain yogurt</td>
</tr>
<tr>
<td>1 xícara de iogurte natural</td>
<td>1 cup whole milk plain yogurt</td>
<td>1 cup whole milk plain yogurt</td>
</tr>
<tr>
<td>600 ml de iogurte natural (3 copinches)</td>
<td>600 ml (21 fl oz) whole milk plain yogurt</td>
<td>600 ml (21 fl oz) whole milk plain yogurt</td>
</tr>
<tr>
<td>100 ml de iogurte natural</td>
<td>100 ml (3.5 fl oz) whole milk plain yogurt</td>
<td>100 ml (3.5 fl oz) whole milk plain yogurt</td>
</tr>
<tr>
<td>100 ml de iogurte natural</td>
<td>100 ml (3.5 fl oz) whole milk plain yogurt</td>
<td>100 ml (3.5 fl oz) whole milk plain yogurt</td>
</tr>
<tr>
<td>200 ml de iogurte natural</td>
<td>200 ml (7 fl oz) whole milk plain yogurt</td>
<td>200 ml (7 fl oz) whole milk plain yogurt</td>
</tr>
<tr>
<td>200 ml de iogurte natural</td>
<td>200 ml (7 fl oz) whole milk plain yogurt</td>
<td>200 ml (7 fl oz) whole milk plain yogurt</td>
</tr>
<tr>
<td>1 litro de suco de laranja natural</td>
<td>1 L (1.1 qt) freshly squeezed orange juice</td>
<td>1 L (1.1 qt) freshly squeezed orange juice</td>
</tr>
<tr>
<td>1 litro de suco de laranja natural</td>
<td>1 L (1.1 qt) freshly squeezed orange juice</td>
<td>1 L (1.1 qt) freshly squeezed orange juice</td>
</tr>
</tbody>
</table>
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 localizado na 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. Tognin. A disponibilização do CorTrad na rede é um projeto conjunto entre o COMET, a Linguatexa e o NILC usando o sistema DISPARA.

### Pesquisar no corpus

<table>
<thead>
<tr>
<th>Tipo</th>
<th>Termo</th>
<th>Filtrar</th>
<th>Ver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>principal [semis=&quot;cor&quot;]</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Primeira tradução</td>
<td>principal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tradução revisada</td>
<td>principal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tradução publicada</td>
<td>principal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Ignorar maiúsculas/mínusculas**:

### Resultado

- **Concordância**
- **Distribuição dos lemas**
- **Distribuição do tempo verbal e/ou do caso pronominal**
- **Distribuição do gênero morfológico**
- **Distribuição por obra**
- **Distribuição por campo semântico**

### Opções

- **Resultados por ordem alfabética (apenas distribuições)**
- **Resultados numerados (apenas concordância)**

### Exemplos de pesquisas

<table>
<thead>
<tr>
<th></th>
<th>Precurso</th>
<th>Resultado</th>
</tr>
</thead>
<tbody>
<tr>
<td>down</td>
<td>down</td>
<td>concordância</td>
</tr>
<tr>
<td>inglês por inglês</td>
<td>&quot;ziol.*&quot;</td>
<td>concordância</td>
</tr>
<tr>
<td>inglês por inglês</td>
<td>&quot;at &quot; &quot;least&quot;</td>
<td>concordância</td>
</tr>
<tr>
<td>inglês por inglês</td>
<td>&quot;&quot;:&quot;&quot; &quot;leat&quot;&quot;</td>
<td>concordância</td>
</tr>
<tr>
<td>inglês por inglês</td>
<td>[ lemma=&quot;see&quot;]</td>
<td>concordância</td>
</tr>
<tr>
<td>inglês por inglês</td>
<td>&quot;rather&quot; [lemma]</td>
<td>distribuição das formas</td>
</tr>
<tr>
<td>inglês por inglês</td>
<td>[lemma=&quot;vermelho&quot;]</td>
<td>concordância</td>
</tr>
</tbody>
</table>

Clique em atualizações para saber do andamento do projeto.
### Semantic information: colour

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

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

Result type: semantic field
10 most recurrent colour terms

<table>
<thead>
<tr>
<th>Short stories</th>
<th>Scientific news</th>
<th>Cooking</th>
</tr>
</thead>
<tbody>
<tr>
<td>white</td>
<td>black</td>
<td>brown</td>
</tr>
<tr>
<td>black</td>
<td>color</td>
<td>black</td>
</tr>
<tr>
<td>blue</td>
<td>white</td>
<td>white</td>
</tr>
<tr>
<td>red</td>
<td>green</td>
<td>red</td>
</tr>
<tr>
<td>grey</td>
<td>red</td>
<td>green</td>
</tr>
<tr>
<td>brown</td>
<td>yellow</td>
<td>color</td>
</tr>
<tr>
<td>green</td>
<td>blue</td>
<td>golden</td>
</tr>
<tr>
<td>yellow</td>
<td>yellowing</td>
<td>yellow</td>
</tr>
<tr>
<td>colour</td>
<td>greenhouse</td>
<td>purple</td>
</tr>
<tr>
<td>pink</td>
<td>gray</td>
<td>brown</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Short stories</th>
<th>Scientific News</th>
<th>Cooking</th>
</tr>
</thead>
<tbody>
<tr>
<td>man</td>
<td>dwarf</td>
<td>wine</td>
</tr>
<tr>
<td>hand</td>
<td>cube</td>
<td>chocolate</td>
</tr>
<tr>
<td>feather</td>
<td>house</td>
<td>rice</td>
</tr>
<tr>
<td>noodle</td>
<td>spot</td>
<td>part</td>
</tr>
<tr>
<td>crockery</td>
<td>blood</td>
<td>pith</td>
</tr>
<tr>
<td>fence</td>
<td>crab</td>
<td>pepper</td>
</tr>
<tr>
<td>Camellia</td>
<td>shrimp</td>
<td>sandwich</td>
</tr>
<tr>
<td>knuckle</td>
<td>fluid</td>
<td>button</td>
</tr>
<tr>
<td>handbag</td>
<td>stripe</td>
<td>bean</td>
</tr>
<tr>
<td>cockatoo</td>
<td>hair</td>
<td>hominy</td>
</tr>
</tbody>
</table>
**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)
- Scientific news: a lot of (unexpected) colour in scientific terminology: names of diseases, stars, etc.
- Short stories: high correlation of clothing and colour
BUT...

CorTrad can be used as a translation learner corpus
Possible Queries

Which adjectives do students use with “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

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>important</td>
<td>3</td>
</tr>
<tr>
<td>financial</td>
<td>2</td>
</tr>
<tr>
<td>good</td>
<td>2</td>
</tr>
<tr>
<td>weighty</td>
<td>1</td>
</tr>
<tr>
<td>major</td>
<td>1</td>
</tr>
<tr>
<td>unprecedented</td>
<td>1</td>
</tr>
<tr>
<td>significant</td>
<td>1</td>
</tr>
<tr>
<td>social</td>
<td>1</td>
</tr>
<tr>
<td>effective</td>
<td>1</td>
</tr>
<tr>
<td>big</td>
<td>1</td>
</tr>
<tr>
<td>fundamental</td>
<td>1</td>
</tr>
<tr>
<td>great</td>
<td>1</td>
</tr>
<tr>
<td>technological</td>
<td>1</td>
</tr>
<tr>
<td>scientific</td>
<td>1</td>
</tr>
<tr>
<td>possible</td>
<td>1</td>
</tr>
<tr>
<td>like</td>
<td>1</td>
</tr>
<tr>
<td>brazilian</td>
<td>1</td>
</tr>
</tbody>
</table>
AND LAST...

BUT NOT LEAST...
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*:
  - Masculino
- Data de Nascimento*:
  - / / 
- Nacionalidade*:
  - Brasileira
- Língua do pai*:
  - alemão
- Língua da mãe*:
  - alemão
- Língua praticada em casa*:
  - alemão
- E-mail*:

**Curso de Línguas**
- Faz ou faz curso de língua?:
  - Selecione

**Educação**
- Escolaridade*:
  - Selecione

**Acesso para submissão de redações**
- Nome de Acesso*:
- Senha*:
- Confirma Senha*:

**Semestre**
- Disciplina:
  - Selecione

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

[Create account button]
Submitting a text

Corpus de Aprendizes

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
- Número de Acessos: 5
Teachers

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

<table>
<thead>
<tr>
<th>Nome do Aluno</th>
<th>Título da Redação</th>
<th>Curso</th>
<th>Data de Submissão</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renata Alabi</td>
<td>Application letter</td>
<td>LILC1-LLIC1: LingLingCont</td>
<td>29/04/2008 às 14:19:30</td>
</tr>
<tr>
<td>Renata Alabi</td>
<td>Application form 2</td>
<td>LILC1-LLIC1: LingLingCont</td>
<td>29/04/2008 às 14:20:19</td>
</tr>
<tr>
<td>Renata Alabi</td>
<td>My changing habits</td>
<td>LILC1-LLIC1: LingLingCont</td>
<td>28/05/2008 às 14:04:35</td>
</tr>
<tr>
<td>Renata Alabi</td>
<td>To continue a story</td>
<td>LILC1-LLIC1: LingLingCont</td>
<td>13/06/2008 às 12:10:08</td>
</tr>
<tr>
<td>Renata Alabi</td>
<td>A wedding reception - writing class</td>
<td>LILC1-LLIC1: LingLingCont</td>
<td>13/06/2008 às 14:40:05</td>
</tr>
<tr>
<td>Renata Alabi</td>
<td>Article 1</td>
<td>LILC1-LLIC1: LingLingCont</td>
<td>19/06/2008 às 22:52:25</td>
</tr>
<tr>
<td>Renata Alabi</td>
<td>Article 2</td>
<td>LILC1-LLIC1: LingLingCont</td>
<td>19/06/2008 às 22:52:51</td>
</tr>
</tbody>
</table>

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,
Our LC Production

Teaching

- Business English:
  - usual collocations - Adriane
  - adverbial collocations - Andréa
- Academic English – abstracts: Carmen
- Specialized corpora in ESP teaching: Danilo (IC)
- LC and Multiliteracy: Cristina
- Student difficulties with scientific writing: Marlene
Terminology

- Orthodontics – building a corpus: Roberto
- Cooking 1) translation of recipes, 2) proposal for dictionary: Elisa
- Binomials in Agreements/Contracts: Luciana C.
- Ecotourism: Josimeire
- VoTec – online vocabulary for translators: Guilherme
- Coffee – regional variants: Luciana
- CL in Interpretation – building a working glossary: Carla
- Football: Sabrina
- Hotel industry: Sandra
Terminological publications

Vocabulário de Química
Ana Julia Perrotti-Garcia
Rozane Rodrigues Rebechi
(SBS, 2007)

Vocabulário de Culinária
Elisa Duarte Teixeira
Stella E. O. Tagnin
(SBS, 2008)
Vocabulário para Fotografia
Angelica Royo
Eliana C.R. Antonopoulos
Helena Akemi Misumi
Moira Martins de Andrade
Veridiana Rocha Schwenck
(SBS, 2013)
Translation

- Adverbial collocations (general): Helmara
- *Dubliners*: Lourdes
- Naturalness in translation: Alvamar
- Chico Buarque in translation: Sergio
- Adverbial collocations in Cooking and Law: Helmara
Research in progress

1. Brazilian cooking ingredients and dishes
2. Football from a cultural perspective
3. Translation learner corpus
4. Consecutive or simultaneous interpretation first?
5. Verbal collocations in student writing
6. “Get”: a semantic analysis
7. Humor in translation
8. Discourse in VBAC statements
9. Aviation “Basic English”
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

Thank you
Stella

seotagni@usp.br