You don’t need to be a specialist to translate specialized language. Corpus Linguistics can come to your rescue.

Stella E. O. Tagnin
University of São Paulo

CATS Conference
Victoria, Canada
June 3-5, 2013
Outline

- Translator training
  - “deliberate practice” (Shreve 2006)
- Actual practice
  - Case studies
- The “real world”
  - From submission to publication
Translator training

- Translators cannot be specialists in all areas they translate.
- Adequate translator training can prepare students to cope with different specialized areas.
- Same goes for professional translators.
- Actual “real world” task:
  - Série 1001 Termos – what is of interest to the publishing house – market needs
  - Defining the domain: Chemistry
“Deliberate practice” (Shreve 2006)

- Motivating task: well designed and appropriately difficult
  - a “real world” task
  - best glossaries would be submitted for publication
- Feedback by teacher
- Chance to redo / correct task
Actual practice: the task

- To build a **corpus-driven** technical glossary for publication
  - 2005: various groups – various domains
  - 2008: various groups – one domain: Photography
Actual practice: Methodology - Corpus Linguistics

- Introduction to Corpus Linguistics
  - Compiling a specialized bilingual corpus
    - English – Portuguese
    - Comparable corpus: same topics, same genres etc
  - Balancing
    - Text types
    - Number of texts
    - Number of words
Actual practice: Methodology - Corpus Linguistics

- Corpus investigation tools
  - WordSmith Tools (Scott 2012)
  - AntConc (Laurence 2006)

- Term extraction
  - WordLists
  - KeyWord Lists
### WordList

<table>
<thead>
<tr>
<th>N</th>
<th>Word</th>
<th>Freq.</th>
<th>%</th>
<th>Texts</th>
<th>% Lemmas</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>THE</td>
<td>89,881</td>
<td>6.98</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>#</td>
<td>68,650</td>
<td>5.33</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TO</td>
<td>32,824</td>
<td>2.55</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>AND</td>
<td>25,422</td>
<td>1.97</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>24,471</td>
<td>1.90</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>OF</td>
<td>22,229</td>
<td>1.73</td>
<td>126</td>
<td>97.67</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>IN</td>
<td>20,314</td>
<td>1.58</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>IS</td>
<td>17,456</td>
<td>1.36</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>OR</td>
<td>14,885</td>
<td>1.16</td>
<td>126</td>
<td>97.67</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>YOU</td>
<td>12,614</td>
<td>0.98</td>
<td>114</td>
<td>88.37</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CAMERA</td>
<td>2,189</td>
<td>0.15</td>
<td>101</td>
<td>78.29</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>FOR</td>
<td>10,711</td>
<td>0.83</td>
<td>126</td>
<td>97.67</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>ON</td>
<td>10,506</td>
<td>0.82</td>
<td>127</td>
<td>98.45</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>IMAGE</td>
<td>9,946</td>
<td>0.77</td>
<td>111</td>
<td>86.05</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>WITH</td>
<td>9,498</td>
<td>0.74</td>
<td>128</td>
<td>99.22</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>BE</td>
<td>8,755</td>
<td>0.68</td>
<td>123</td>
<td>95.35</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>ARE</td>
<td>7,518</td>
<td>0.58</td>
<td>119</td>
<td>92.25</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>YOUR</td>
<td>7,478</td>
<td>0.58</td>
<td>111</td>
<td>86.05</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>CAN</td>
<td>7,388</td>
<td>0.57</td>
<td>121</td>
<td>93.80</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>THAT</td>
<td>7,009</td>
<td>0.54</td>
<td>120</td>
<td>93.02</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>IT</td>
<td>6,205</td>
<td>0.48</td>
<td>126</td>
<td>97.67</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>IMAGES</td>
<td>6,099</td>
<td>0.47</td>
<td>97</td>
<td>75.19</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>NOT</td>
<td>6,012</td>
<td>0.47</td>
<td>116</td>
<td>89.92</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>THIS</td>
<td>5,937</td>
<td>0.46</td>
<td>120</td>
<td>93.02</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>WHEN</td>
<td>5,785</td>
<td>0.45</td>
<td>120</td>
<td>93.02</td>
<td></td>
</tr>
</tbody>
</table>
Study corpus WL vs Reference/comparison corpus WL

Keywords = term candidates
Keywords - Photography

<table>
<thead>
<tr>
<th>N</th>
<th>Key word</th>
<th>Freq.</th>
<th>%</th>
<th>RC.</th>
<th>RC. %</th>
<th>t</th>
<th>P</th>
<th>Lemmas</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CAMERA</td>
<td>12,189</td>
<td>0.95</td>
<td>6</td>
<td>0.01</td>
<td>811.72</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>#</td>
<td>68,650</td>
<td>5.33</td>
<td>1,274</td>
<td>2.72</td>
<td>742.49</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IMAGE</td>
<td>9,946</td>
<td>0.77</td>
<td>12</td>
<td>0.03</td>
<td>608.32</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>YOU</td>
<td>12,614</td>
<td>0.98</td>
<td>105</td>
<td>0.22</td>
<td>390.87</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BUTTON</td>
<td>5,441</td>
<td>0.42</td>
<td>3</td>
<td>0.35</td>
<td>358.65</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>IMAGES</td>
<td>6,099</td>
<td>0.47</td>
<td>11</td>
<td>0.02</td>
<td>349.40</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>FLASH</td>
<td>4,396</td>
<td>0.34</td>
<td>0</td>
<td>0.314</td>
<td>314.64</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DIGITAL</td>
<td>4,503</td>
<td>0.35</td>
<td>1</td>
<td>0.31</td>
<td>310.18</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>MODE</td>
<td>4,865</td>
<td>0.38</td>
<td>5</td>
<td>0.01</td>
<td>302.92</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>OR</td>
<td>14,885</td>
<td>1.16</td>
<td>198</td>
<td>0.42</td>
<td>283.23</td>
<td>0.0000(C)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Actual practice: Methodology - Corpus Linguistics

- Identification of equivalents
  - Prima facie translations
  - Collocates
  - Context
2 TBSP
ADD
MINS
OIL
UNTIL
CHOPPED
PAN
OZ
BUTTER
INGREDIENTS
HEAT
SALT
SUGAR
PEPPER
COOKING
COOK
TSP
AND
METHOD
SERVES
SERVE
CREAM
SAUCE
TIME
PREPARATION
OVEN
FLOUR
GARLIC
OLIVE
STIR
INTO
MIXTURE
BOWL
Looking for equivalents

- *finely* (2.987 occ.) = “finamente” (337 occ.)
- **Not equivalent**
  (finely ≠ finamente)
- Indirect method: *chop* = “picar”
- Collocates of “pique/picad*”
  in Portuguese corpus
<table>
<thead>
<tr>
<th>N</th>
<th>WORD</th>
<th>TOTAL</th>
<th>LEFT</th>
<th>RIGHT</th>
<th>L5</th>
<th>L4</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DE</td>
<td>11121</td>
<td>5940</td>
<td>5181</td>
<td>783</td>
<td>0</td>
<td>939</td>
</tr>
<tr>
<td>2</td>
<td>PICADA</td>
<td>3727</td>
<td>239</td>
<td>230</td>
<td>155</td>
<td>46</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>3253</td>
<td>1647</td>
<td>1606</td>
<td>226</td>
<td>230</td>
<td>432</td>
</tr>
<tr>
<td>4</td>
<td>SOPA</td>
<td>2661</td>
<td>1503</td>
<td>1158</td>
<td>135</td>
<td>231</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>PICADO</td>
<td>2261</td>
<td>172</td>
<td>162</td>
<td>75</td>
<td>58</td>
<td>28</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>2066</td>
<td>884</td>
<td>1182</td>
<td>201</td>
<td>111</td>
<td>138</td>
</tr>
<tr>
<td>7</td>
<td>CHÁ</td>
<td>1737</td>
<td>849</td>
<td>888</td>
<td>78</td>
<td>182</td>
<td>579</td>
</tr>
<tr>
<td>8</td>
<td>CEBOLA</td>
<td>1681</td>
<td>1403</td>
<td>277</td>
<td>47</td>
<td>68</td>
<td>69</td>
</tr>
<tr>
<td>9</td>
<td>XÍCARA</td>
<td>1590</td>
<td>831</td>
<td>759</td>
<td>181</td>
<td>407</td>
<td>227</td>
</tr>
<tr>
<td>10</td>
<td>COLHERES</td>
<td>1446</td>
<td>675</td>
<td>771</td>
<td>217</td>
<td>449</td>
<td>82</td>
</tr>
<tr>
<td>11</td>
<td>COLHER</td>
<td>1372</td>
<td>656</td>
<td>716</td>
<td>198</td>
<td>438</td>
<td>19</td>
</tr>
<tr>
<td>12</td>
<td>ALHO</td>
<td>1228</td>
<td>833</td>
<td>395</td>
<td>83</td>
<td>79</td>
<td>39</td>
</tr>
<tr>
<td>13</td>
<td>PICADOS</td>
<td>1215</td>
<td>107</td>
<td>130</td>
<td>35</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td>14</td>
<td>PICADAS</td>
<td>1205</td>
<td>95</td>
<td>69</td>
<td>44</td>
<td>33</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>EM</td>
<td>1062</td>
<td>270</td>
<td>792</td>
<td>79</td>
<td>86</td>
<td>72</td>
</tr>
<tr>
<td>16</td>
<td>SALSINHA</td>
<td>1036</td>
<td>892</td>
<td>144</td>
<td>19</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td>17</td>
<td>G</td>
<td>957</td>
<td>511</td>
<td>446</td>
<td>179</td>
<td>132</td>
<td>195</td>
</tr>
<tr>
<td>18</td>
<td>SAL</td>
<td>770</td>
<td>141</td>
<td>629</td>
<td>39</td>
<td>48</td>
<td>33</td>
</tr>
<tr>
<td>19</td>
<td>PIQUE</td>
<td>740</td>
<td>5</td>
<td>74</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>AS</td>
<td>707</td>
<td>378</td>
<td>329</td>
<td>78</td>
<td>68</td>
<td>73</td>
</tr>
<tr>
<td>21</td>
<td>BEM</td>
<td>669</td>
<td>413</td>
<td>256</td>
<td>32</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>22</td>
<td>GOSTO</td>
<td>646</td>
<td>146</td>
<td>500</td>
<td>41</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>COM</td>
<td>643</td>
<td>357</td>
<td>286</td>
<td>80</td>
<td>72</td>
<td>119</td>
</tr>
<tr>
<td>24</td>
<td>SEM</td>
<td>628</td>
<td>403</td>
<td>225</td>
<td>126</td>
<td>30</td>
<td>47</td>
</tr>
<tr>
<td>25</td>
<td>DENTES</td>
<td>551</td>
<td>316</td>
<td>235</td>
<td>20</td>
<td>30</td>
<td>266</td>
</tr>
<tr>
<td>26</td>
<td>MODO</td>
<td>550</td>
<td>93</td>
<td>457</td>
<td>3</td>
<td>4</td>
<td>86</td>
</tr>
<tr>
<td>27</td>
<td>OS</td>
<td>545</td>
<td>229</td>
<td>316</td>
<td>51</td>
<td>46</td>
<td>37</td>
</tr>
<tr>
<td>28</td>
<td>CEBOLINHA</td>
<td>536</td>
<td>427</td>
<td>109</td>
<td>8</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>29</td>
<td>PREPARRO</td>
<td>512</td>
<td>147</td>
<td>365</td>
<td>18</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>30</td>
<td>TOMATES</td>
<td>450</td>
<td>262</td>
<td>188</td>
<td>23</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>31</td>
<td>SALSA</td>
<td>448</td>
<td>404</td>
<td>44</td>
<td>5</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>32</td>
<td>COL</td>
<td>429</td>
<td>228</td>
<td>201</td>
<td>137</td>
<td>190</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>VERDE</td>
<td>381</td>
<td>306</td>
<td>75</td>
<td>18</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>34</td>
<td>MANTEIGA</td>
<td>376</td>
<td>148</td>
<td>59</td>
<td>81</td>
<td>46</td>
<td>37</td>
</tr>
<tr>
<td>35</td>
<td>FOLHAS</td>
<td>374</td>
<td>258</td>
<td>116</td>
<td>29</td>
<td>146</td>
<td>13</td>
</tr>
<tr>
<td>36</td>
<td>PIMENTA-DO-REINO</td>
<td>351</td>
<td>74</td>
<td>277</td>
<td>16</td>
<td>33</td>
<td>7</td>
</tr>
</tbody>
</table>

- **AZEITE**: 338, 108, 230, 53, 39, 15, 1, 0, 10, 43, 13, 38, 126
- **PIMENTÃO**: 320, 194, 126, 9, 12, 41, 105, 27, 0, 68, 22, 15, 21
- **MÉDIA**: 319, 256, 63, 12, 6, 8, 10, 221, 0, 0, 38, 18, 7
- **GROSSEIRAMENTE**: 311, 74, 237, 37, 2, 4, 1, 60, 182, 29, 0, 0
- **GROSSEIRAMENTE**: 311, 74, 237, 37, 2, 4, 1, 60, 182, 29, 0, 0
- **PICADINHO**: 260, 54, 15, 7, 19, 11, 7, 191, 0, 1, 1, 0, 0, 3, 6
- **PÉLL**: 125, 259, 184, 75, 4, 106, 16, 21, 28, 0, 18, 20, 31
- **PICADINHA**: 244, 41, 12, 7, 18, 0, 2, 1, 0, 216, 0, 1, 0, 40, 12, 12
- **ACRESCENTE**: 236, 167, 61, 62, 18, 22, 127, 0, 1, 4, 9, 12, 23
- **FINAMENTE**: 236, 197, 4, 4, 2, 1, 28, 0, 149, 28, 12, 4, 4
- **PEQUENOS**: 233, 165, 76, 14, 10, 4, 6, 0, 9, 4, 7, 40, 0
- **COLOQUE**: 228, 55, 173, 15, 6, 33, 0, 2, 29, 26, 33, 29, 2, 1
- **PIQUE**: 202, 76, 126, 15, 7, 30, 8, 0, 13, 8, 60, 44
- **QUEIO**: 198, 110, 88, 21, 11, 20, 43, 15, 0, 44, 3, 19, 14
- **CENOURA**: 196, 114, 82, 7, 15, 30, 49, 0, 37, 12, 13, 20
- **AMARELO**: 196, 155, 41, 4, 1, 0, 1, 2, 13, 25, 52, 0
- **CHOCOLATE**: 186, 267, 69, 16, 9, 5, 0, 25, 5, 11, 25, 0
- **AZEITAS**: 188, 122, 66, 9, 16, 6, 50, 38, 0, 25, 5, 11, 25
- **RESERVA**: 188, 90, 98, 7, 9, 1, 0, 0, 22, 16, 22, 12, 5
- **UMA**: 188, 82, 106, 32, 25, 11, 12, 0, 5, 22, 20, 24, 35
- **CUBOS**: 184, 52, 132, 84, 23, 14, 24, 1, 69, 39, 21, 12, 9
- **XÍCARAS**: 184, 71, 112, 29, 13, 6, 11, 11, 20, 7, 10, 4, 11, 4
- **DO**: 181, 78, 103, 24, 11, 12, 0, 1, 2, 24, 48, 28
- **HORTELA**: 182, 149, 33, 5, 8, 10, 120, 0, 0, 23, 19, 8, 7
Results for picad*

Most frequent collocates:

- "bem" (669 occurrences): "bem picada", "bem picado", "bem picadinho" etc.
- "picadinh*" (504 occurrences, while 69 are of "bem picadinh*")

Best equivalents for finely chopped

"bem picad*" ou "picadinh*"
Actual practice: Building the glossary

- A translator-oriented glossary
  - Equivalents
  - Authentic examples
  - Collocates and phraseology in general
  - Cultural information where relevant
  - Suggestion of substitutions (cooking)
Actual practice: Building the glossary

- Hand-in glossary
- Review by teacher
- New submission (optional)
Photography

- Life is not a bowl of cherries...
- Domain subdivided:
  - Camera
  - Studio
  - History
  - Light
  - Preservation
  - Digital photography
- Not all glossaries met publishing quality
- New group to correct and improve material produced
The real world: from submission to publication

- Submission to publisher
- Review by specialist
- Back to authors for corrections and changes
- New submission
- Wait...
- Wait...
- Wait...
- Publication!!!
Publications

Vocabulário para Química
Ana Julia Perrotti-Garcia
Rozane Rodrigues Rebechi
(2007)
Vocabulário para Culinária
Elisa Duarte Teixeira
Stella E. O. Tagnin
(2008)
Vocabulário para Fotografia

Eliana C. R. Antonopoulous
Helena Akemi Misumi
Moir Martins de Andrade
Maria Angelica Royo
Veridiana Rocha Schwenck
(2013)
Final remarks

- Not all glossaries were publishable

- Nevertheless
  - students became familiar with Corpus Linguistics
  - learned how to build a glossary
  - and, by extension, learned how to mine for lexical information
Says one student

“\[I \text{ knew nothing about Chemistry so that it was difficult to select the texts for the corpus. Consequently, some had to be discarded later on.}\]”

\[= \text{difficulty of task}\]
“the process really worked and using the corpus made it possible to overcome the problems you experience when working with an area in which you are not a specialist.”

becoming familiar with the area
Says the reviser

- herself a translator and Chemical Engineer:

   “I would have never been able to come up with the terms as they did”

- how efficient CL is
- language actually used by that professional community
Most importantly

“the learner’s authentic experience allows him/her to apply the knowledge acquired to any other area in which he/she may come to work in”  (Alves & Tagnin, 2010)
Thank you

Stella

seotagni@usp.br
References