Using a parallel corpus to validate independent claims

**Motivation**
- Most corpus-based contrastive studies observe one corpus, refining and further testing is done on that same corpus
- How should one go about validating studies in other corpora?
- Verification of particular hypotheses in the English-Portuguese part of the ENPC: EPC

**Problems**
- How much is just intuitive speculation illustrated by a corpus?
- How much can actually be replicated by other researchers?
- How many English-Portuguese corpus-based contrastive studies are there that one would like to validate?

**Actual choices**
- One particular set of claims, based on a smaller “corpus”, which was Diana’s own interest to see whether it would escalate: Santos 1998, 1996

**Slobin’s main theses**
- Description of movement is much more elaborate in S-languages (English) than in V-languages (Portuguese)
  - description of manner
  - number of ground elements per clause
  - number of clauses per journey
  - frequency of journey descriptions
- Movement across a boundary requires an independent predicate in a V-language

**Two construction types to express motion**
- Manner of motion verb + Path particle (English)
  (S-languages) **Satellite-framed**:
  *He ran up the stairs.*
- Path of motion verb (+ Manner, expressed in a separate clause) (Portuguese)
  (V-languages) **Verb-framed**:
  *Subiu as escadas.* (‘He ascended the stairs’)

Signe Oksefjell & Diana Santos
Consequences for translation according to Slobin

Translation from an S- to a V-language results in loss of information or clause splitting
- Usual choice for translating Manner+Path motion is Path motion
- Manner verbs are less frequent in translations into V-languages
- Tendency to eliminate reference to ground when a manner verb is present
- The number of available manner verbs (types) is much smaller in V-languages

Movement in English

Finally he would sit up and wiggle out of his sheet. (AT1)

He closed his eyes and without even trying, he dropped of the bed. (AT1)

And she was pulled from the road into the shadow of the bushes. (PDJ3)

He closed his eyes and without even trying, he dropped out of the bed. (AT1)

Finally, ele se sentou e, remexendo-se, saiu do seu lençol.

Fechava os olhos e sem nem mesmo tentar mergulhava no sono.

E ela foi arrastada para as sombras dos arbustos.

… and she was pulled from the road into the shadow of the bushes. (PDJ3)

… e ela foi arrastada para as sombras dos arbustos.

… and she was pulled from the road into the shadow of the bushes. (PDJ3)

And she was pulled from the road into the shadow of the bushes. (PDJ3)

E ela foi arrastada para as sombras dos arbustos.

Slobin’s empirical studies

- 80 motion events from English to Spanish (corresponding to 165 verbs, 60 types)
- chosen at random in four different novels (20 from each)
- 78% preservation of trajectory information, 51% preservation of manner information
- presents 17 examples (3 of which with into / out of)
- she moved out into the sun and across the stony clearing -> la muchacha salió al claro rochoso
- I went into the hall and through to the dining room -> entré en el hall y pasé al comedor
- she rustled out of the room -> salió del cuarto, acompañada del susurro

Validation method

- Look for all occurrences of out of and into
- Select the cases involving movement (36 out of; 140 into)
- Analyze the translations
  - number of clauses
  - preservation of manner / ground elements
  - verb types

Interesting problems

- Manner of motion (swerve, slip, drive, ...) or accompanying manner (slam, whisper, roar, struggle, ...)
- Motion of the subject (fall, rise, ...) or of the object (pull, kick, shoot, ...)
- Motion or position (rise, dangle, head, cram)
- Kinds of movement
  - metaphorical / concrete
  - of an individual or of a group
The translation of English Manner of Motion verbs + into or out of into Portuguese

<table>
<thead>
<tr>
<th>Semantic elements in the Portuguese verb</th>
<th>No. of occur. (into)</th>
<th>No. of occur. (out of)</th>
<th>2 clauses (into)</th>
<th>2 clauses (out of)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion+Path</td>
<td>45</td>
<td>15</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Motion+Manner</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Motion+Manner + Path</td>
<td>34</td>
<td>10</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Path+Manner</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manner</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rewrite</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>114</strong></td>
<td><strong>29</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

Validation of Slobin’s claims

- Clear tendency to omit Manner altogether (more than half of the cases)
- Fewer (manner) verbs in Portuguese translations (20-31; 74-98)
- Clause splitting in conjoined cases (one English verb with two or more particles) was often observed

Perception of events (Santos 98)

- There is a close parallel [to gerund and bare infinitive in the VP object of an English perception verb] in Portuguese, namely infinitive and a-infinitive, which have roughly the same role and distribution — but no 1:1 mapping
- Both infinitive and gerund VPs suffer a drastic reduction in the Portuguese translation
- One category relevant for the description of the two languages is a VP object in the past participle

Examples

- Through the knife he *could feel* the muscle tighten hard
  *sentiu* na faca a resistência do manto
- Kino left his foot there and *watched* them move over it
  *Kino não se mexeu mais e* *ficou a observar a movimento das formigas.*
- In the pearl he *saw* Coyotito sitting at a little desk in a school
  *Na pérola,* *via* Coyotito na escola, *sentado à sua carteira*

Corpora involved

- SD 1 English text (26,000 words); 1 translation
- EPC 15 English texts (12,000 words); 16 Portuguese translations

Translation of VP objects in SD

<table>
<thead>
<tr>
<th>Tense</th>
<th>Portuguese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>could</td>
<td><em>Perfeito</em></td>
<td><em>perfeito</em></td>
</tr>
<tr>
<td>simple past</td>
<td><em>Perfeito</em></td>
<td><em>perfeito</em></td>
</tr>
<tr>
<td>present</td>
<td><em>Perfeito</em></td>
<td><em>perfeito</em></td>
</tr>
</tbody>
</table>

Table 4 in Santos (1998:337):
Using a parallel corpus to validate independent claims

31 October 1998

Signe Oksefjell & Diana Santos

19

Improved version, SD corpus

Preservation of VPinf and VPger form

<table>
<thead>
<tr>
<th>English original</th>
<th>Portuguese translation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPinf 21</td>
<td>VPinf</td>
<td>10</td>
</tr>
<tr>
<td>VPinf+VPger</td>
<td>VPinf+VPger</td>
<td>2</td>
</tr>
<tr>
<td>other</td>
<td>other</td>
<td>0</td>
</tr>
<tr>
<td>VPger</td>
<td>VPger</td>
<td>0</td>
</tr>
<tr>
<td>VPpastpart</td>
<td>VPpastpart</td>
<td>0</td>
</tr>
<tr>
<td>VPa+inf</td>
<td>VPa+inf</td>
<td>0</td>
</tr>
<tr>
<td>other</td>
<td>other</td>
<td>0</td>
</tr>
</tbody>
</table>

20

Translation of VP objects in EPC

<table>
<thead>
<tr>
<th>English original</th>
<th>Portuguese translation</th>
<th>VPinf</th>
<th>VPger,a+inf</th>
<th>VPger,a+inf</th>
<th>VPger,a+inf</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPinf</td>
<td>VPinf</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>VPger,a+inf</td>
<td>VPger,a+inf</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>VPger,a+inf</td>
<td>VPger,a+inf</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>VPger,a+inf</td>
<td>VPger,a+inf</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>other</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>VPpastpart</td>
<td>VPpastpart</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>VPa+inf</td>
<td>VPa+inf</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
<td>30</td>
<td>17</td>
<td>17</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

21

First validation of Santos’s claims

• there is a tendency to avoid complex VP objects in translation: 12 in 76 (SD 18 in 43)
• there is not a one-to-one mapping between VPinf and VPger (or VPa+inf): 54% VPinf matching, 48% VPger matching (SD 47%, 36%)

11/33 VPger→VPinf; 11/39 VPinf→VPger

• (almost) no trace of the use of could: 4% as opposed to 28% in SD

22

Corpora involved

• SD: 1 English text (26,000 words); 1 translation
• EPC: 15 English texts (12,000 words each); 16 Portuguese translations (10 European Portuguese, 6 Brazilian Portuguese translations)
• EEPC: 10 English texts (12,000 words each); 10 European Portuguese translations

23

Translation of VP objects in EEPC

<table>
<thead>
<tr>
<th>English original</th>
<th>Portuguese translation</th>
<th>VPinf</th>
<th>VPger,a+inf</th>
<th>VPger,a+inf</th>
<th>VPger,a+inf</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPinf</td>
<td>VPinf</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>VPa+inf</td>
<td>VPa+inf</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>VPger,a+inf</td>
<td>VPger,a+inf</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>VPger,a+inf</td>
<td>VPger,a+inf</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>other</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>VPpastpart</td>
<td>VPpastpart</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>VPa+inf</td>
<td>VPa+inf</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
<td>18</td>
<td>10</td>
<td>4</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

24

Comparison with the EEPC

• The original corpus, SD, is not at all deviant from the EEPC, except for too low a number of occurrences of hear

Proportion of see / hear / feel with VP objects

- SD 24 12 7
- EPC 30 28 17
- EEPC 18 19 10
Final remarks and doubts

• How convincing is this number manipulation?
• Is it necessarily interesting for translation studies that a translation conforms to average?
• Are the EPC and EEPC corpora big enough?
• How relevant is position inside a text?
• Is validation rewarding enough? Is distribution more important than the interesting examples themselves?
• Should validation be done in terms of corpora, or in terms of claims?