

Using a parallel corpus to validate independent claims

Diana Santos
Signe Oksefjell

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

2

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?

3

Actual choices

- One influential set of claims, (partially) based on English-to-Spanish translation studies: Slobin 1997a, 1987, 1994, 1996, 1997b
- 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

4

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

5

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’)

6

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

7

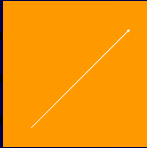
Movement in English



Finally he would sit up and wriggle *out of* his sheet. (AT1)
 Finalmente, ele se sentou e, *removendo-se, saiu do seu* lençol.
 He closed his eyes and without even trying, he dropped *off the edge into* sleep. (AT1)
 Fechava os olhos e sem nem mesmo tentar *mergulhava* no sono.
 ...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 *ran* to the door, sobbing violently, and *out of* the door, letting it crash. (DL2)
 E *correu* para a porta, soluçando com força, e *lá fora* chorou.

8

Movement in Portuguese



O escravo, tão suavemente como entrara, *saiu*.
 The slave *slipped out* as quietly as he had entered
 'The slave, as softly as he had entered, exited.'
 Depois, *foi* a uma mesa próxima
 He then *walked* to a nearby table
 'Then, he went to a nearby table'

9

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)
- 76% 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* OMISSION
 - I *went into the hall and through to the dining room*-> *entré en el hall y pasé al comedor* 2 MAIN CLAUSES
 - she *rusted out of the room* -> *salió del cuarto, acompañada del susurro ...* 2 CLAUSES

10

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

11

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

12

The translation of English Manner of Motion verbs + *into* or *out of* into Portuguese

Semantic elements in the Portuguese verb	No. of occur. (into)	No. of occur. (out of)	2 clauses (info)	2 clauses (out of)
Motion+Path	45	15	1	2
Motion+Manner	4	6	1	2
Motion+Manner +Path	34	10	2	
Path+Manner	7			
Path	15	1		
Manner	6			
rewrite	3	3		
Total	114	29	4	4

- ### 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-to-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 o 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 4 in Santos (1998:337):

could	VPinf 2	1 Perfeito VPinf, 1 gerúndio VPainf
	VPger 10	3 Imperfeito, 1 Imperfeito VPger, 1 Imperfeito VPainf, 1 Imperfeito VPpastpart, 1 Perfeito VPainf, 1 Perfeito VPinf, 1 gerúndio, 1 <i>pôde ver</i>
simple past	VPinf 16	7 Perfeito VPinf, 6 Perfeito, 2 Imperfeito, 1 Imperfeito VPinf
	VPger 12	3 Perfeito VPger, 3 Perfeito VPpastpart, 3 Perfeito, 1 Perfeito VPainf, 1 Imperfeito VPainf, 1 Mais que Perfeito
presperf	VPinf 3	1 Perfeito VPinf, 1 Perfeito VPainf, 1 Perfeito

Improved version, SD corpus

Preservation of VPinf and VPger form

English original	Portuguese translation	Total
VPinf 21	VPinf	10
	VPger,a+inf	2
	VPpartpass	0
	other	9
VPger 22	VPger,a+inf	8
	VPinf	1
	VPpartpass	4
	other	9
VPpastpart	VPinf	0
	VPa+inf	0

19

Translation of VP objects in EPC

English original	Portug. transl.	see	feel	hear	listen	Total
VPinf 39	VPinf	8	4	7	2	21
	VPger,a+inf	1	5	4	1	11
	VPpartpass		1			1
	other	5	1		1	7
VPger 33	VPger,a+inf	7	5	4		16
	VPinf	2	1	8		11
	VPpartpass	2				2
	other	3		1		4
VPpastpart 2	VPinf	1		1		2
	VPa+inf	1				1
Total 74		30	17	25	4	76

20

- ### 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
- 21

- ### 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
- 22

Translation of VP objects in EEPC

English original	Portug. transl.	see	feel	hear	listen	Total
VPinf 28	VPinf	5		7	2	14
	VPa+inf	1	4	2	1	8
	VPpartpass					
	other	4	1		1	6
VPger 22	VPa+inf	4	4	3		11
	VPinf	2	1	6		9
	VPpartpass					
	other	1		1		2
VPpastpart 1	VPinf					
	VPa+inf	1				1
Total 51		18	10	19	4	51

23

- ### 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 | 25 | 17 |
| - EEPC | 18 | 19 | 10 |
- 24

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?

25