Using a parallel corpus to validate independent claims

Diana Santos and Signe Oksefjell

This paper examines the results from two corpus-based contrastive studies. Both studies offer cross-linguistic claims about the language pair English-Portuguese. We attempt to replicate the studies and check the findings against a different corpus, viz. the English-Portuguese part of the English-Norwegian Parallel Corpus, to see whether the regularities observed in the original corpora can be confirmed. After a brief presentation of each study, we describe how we gathered equivalent data, present our findings in the new corpus, and discuss some possible reasons for discrepancies in relation to the earlier studies. The topics investigated are boundary-crossing movement descriptions (after Slobin, 1997) and perception verbs (after Santos, 1998).

1. Introduction

This paper takes as its starting point hypotheses and claims put forward in two earlier contrastive studies. It discusses the possibility of validating results from corpus-based contrastive studies by checking the results againts a different corpus. The studies we will attempt to validate are:

- Slobin (1996, 1997b), which describe different expressions of movement
- Santos (1998), which discusses perception verbs

In both cases we will use a small English-Portuguese corpus – the English-Portuguese part of the English-Norwegian Parallel Corpus project (Johansson et al., 1999) – to test the hypotheses derived from these studies.[1]

Section 2 offers a background to some aspects of Slobin's studies of movement verbs in different languages, and in Section 3 we go on to validate his results. Section 4 comments on claims earlier made about perception verbs in English and Portuguese, while Section 5 offers a validation of these. We conclude by suggesting that claims about language should be sufficiently detailed to allow for truly successful validations. Finally, we stress that claims about differences between two languages and claims pertaining to translation have to be kept apart, since corpus-based contrastive studies mediated by translation have to consider both.

It is hoped that this study will serve as a useful contribution to issues such as validation methods and corpus comparison (see e.g. Kilgarriff and Salkie (1996) and Kilgarriff and Rose (1998) for monolingual studies).

2. Slobin's study of movement verbs

We set out to investigate some claims made by Dan Slobin about the translation of verb structures, more specifically the translation of boundary-crossing movement verbs (see e.g. Slobin 1996, 1997a, 1997b).

As a background for his studies, Slobin refers to Talmy (1985, 1991) who distinguishes two types of languages: satellite-framed and verb-framed. Such a distinction can be made "in terms of whether they [languages] prefer to express change of state in verbs or in elements associated with verbs" (Slobin 1997b:2).

Satellite-framed languages (S-languages) include English and German, along with most other Indo-European languages. Verb-framed languages (V-languages) include all Romance languages, and also Semitic and Turkic languages, Japanese, and Korean.

Examples of ways of expressing motion:

Satellite-framed: He *ran up* the stairs.

Verb-framed: *Subiu* as escadas. ('He ascended the stairs')

Talmy (1985:57) further assumes that "we can isolate elements separately within the domain of meaning and within the domain of surface expression". The semantic elements, or elements within the domain of meaning, relevant to the present paper are Motion, Path, and Manner, along with the surface elements verb and satellite. Table 1 presents Talmy's model of the two languages dealt with here: Portuguese and English.

Language	Language type	The particular components of a motion event characteristically represented in	
		Verb root	Satellite
Portuguese	V-framed	Motion + Path	Ø
English	S-framed	Motion + Cause/Manner	Path

Table 1: Typology of motion verbs and their satellites (cf. Talmy 1985: 114)

In his article on lexicalization patterns in narratives, Slobin (1997b) focuses on one typological difference between satellite-framed and verb-framed languages: the lexicalization of locative <u>transition</u> in satellite-framed languages vs. verb-framed languages. What characterizes the verb-framed languages in this respect is a main Motion verb which includes Path, e.g. *entrar* 'enter', *sair* 'exit'.[2] In contrast, the satellite-framed languages commonly give the path by means of a so-called satellite, i.e. a verb particle, e.g. *go in*, *go out*.

According to Slobin, these differences have consequences for the narrative style of the two languages, leading to differences in the attention to movement detail by the speakers. An example of the two kinds of discourse can be observed in example (1), where the English movement verb *wriggle* has been used in combination with the verb particle *out of*. In English, the verb tells us in what manner the movement was carried out and the result of the movement is expressed by the satellite *out of*. In the Portuguese rendering of the same sentence (2), however, two verbs are used: one expressing the crossing of a boundary (*saiu*), the other expressing an accompanying action (*remexendo-se*).

(1)   Finally he would sit up and *wriggle out of* his sheet. (AT1)
::
(2)   Finalmente, ele se sentou e, *remexendo-se*, *saiu* do seu lençol.
         'Finally, he sat up and, moving, exited from his sheet'

2.1 Slobin's main theses

Slobin claims that the description of movement is much more elaborate in S-languages (e.g. English) than in V-languages (e.g. Portuguese). Movement, in this context,

involves description of manner, number of ground elements per clause, number of clauses per journey, and the frequency of journey descriptions.[3] Further, Slobin states that a movement that crosses a boundary requires an independent predicate in a V-language, contrary to the case of an S-language: While English allows both (a) *He ran into the house* and (b) *He ran (up) to the house*, a verb-framed language would only allow the latter: *Correu até à casa*, where the Path is given in the PP (*até à casa*) and Motion+Manner in the main verb (*correu*). In the former, however, Motion+Path would be given in the main verb, whereas Manner would generally have to be expressed in a separate clause, or in a non-finite verb form: *Entrou na casa a correr* ('he entered the house running'); or perhaps more naturally, *Entrou na casa* ('he entered the house'), without mentioning in what manner.

The difference between (a) and (b) triggering the non-parallel solutions in a verb-framed language is claimed to be the boundary-crossing present in (a) but not in (b). The path in (a) crosses a boundary by means of the satellite *into*. According to Slobin, in a language such as Portuguese, "when a path crosses a boundary [...] it is no longer possible to accumulate a series of grounds to a single verb, because the state-change from one side of the boundary to the other will be expressed by a separate verb [...] This constraint is seen clearly in translations from a V-language to an S-language" (Slobin 1997b: 441). The boundary crossing present in our example (a), then, is the transition from being outside the house to being inside the house.

Slobin states that these typological differences have consequences for translation: loss of information or clause splitting may typically occur when translating from an S-language into a V-language. Let us restate Slobin's position as the following translational claim:

If a spatial boundary is crossed (in an intransitive Motion-Manner verb + Path particle) in English, (an adequately meaning-preserving) translation into Portuguese requires either a split into two clauses or further specification of manner by other means. Otherwise, there is loss of information.

2.2 Slobin's empirical studies

In his corpus-based investigation of these matters, Slobin (1996) makes use of four full novels for the language pair English-Spanish, randomly selecting twenty motion events from each.[4] He presents the following results (ibid:210):

As far as the description of trajectory is concerned, he finds that translation into Spanish only preserves it in 76% of the cases. In addition, the description of manner is omitted in 49% of the cases.

The kind of movement descriptions he looks at are also quantitatively described (ibid:207), in terms of ground elements in the English text, as 4% having none, 61% having one, 26% including two and 9% containing three or more ground elements.

3. Validation of Slobin's findings

Given Slobin's claim that Spanish and Portuguese are typologically similar as far as boundary-crossing movement is concerned, we will compare our English-Portuguese

material with his findings. Our corpus contains approximately 360,000 words from 16 novels in English (about 12,000 from each) and their translations into Portuguese.[5]
In order to restrict our study to a manageable amount of material we have chosen to concentrate on motion events which involve the verb particles *out of* and *into*.

We looked for all occurrences of *out of* and *into* in the corpus, and selected the cases involving movement (36 *out of*; 140 *into*), that is, where these particles combine with Motion-Manner verbs. In this connection, one should note that this proved to be a far from straightforward task, since the verbs were hard to classify. It is indeed hard to determine whether a verb contains only the features in question. For example, Slobin's contrastive description did not address the following questions: Is the same translational pattern to be expected

- if the verb describes accompanying manner (*slam*, *whisper*, *roar*, *struggle*, ...) instead of manner of motion (*swerve*, *slip*, *drive*, ...)?
- if we have a verb of position and not of movement (*rise*, *dangle*, *head*, *cram*, ...)?
- in metaphorical as well as concrete movement?
- if it is a group and not an individual that moves?

These are the sort of questions that can be left unanswered when one describes a phenomenon qualitatively, even when illustrated by authentic examples, but that have to be considered in a quantitative corpus study. Cf. Slobin's (1996: 206-207) description: "Simple appearances and disappearances from the scene were excluded, as were non-directional paths (e.g. *turning around*, *pacing up and down*). The only criterion was thus that the protagonist ended up in a different place within an uninterrupted stretch of narrative."[6]

We then analysed the Portuguese translations of the English sentences paying particular attention to the number of clauses, preservation of Manner / ground elements, and verb types. Tables 2 and 3 show the results.

Semantic elements in the Portuguese verb	No. of occurrences	2 clauses
Motion+Path (+Manner)	4	1
Motion+Manner (+Path)	6	1
Motion+Path	15	2
Path	1	
other	3	
Total	29	4

Table 2: The translation of English Motion+Manner verbs + *out of* into Portuguese

We see that in only four of the cases (an independent device conveying) Manner has been added to a translation involving a Motion+Path verb. An example is found in (3)-(4):

(3)   The picture showed the woman he had seen *driving north out of* London that
        morning with her husband, albeit taken a year earlier. (FF1)
::

(4)   Jim Rawlings examinou com uma lupa a fotografia recortada de uma revista, a
        qual mostrava a mulher que vira abandonar Londres, *de carro*, com o marido,
        naquela manhã, embora tivesse sido tirada há um ano.
        'JR examined with a magnifying glass the photograph taken from a magazine,
        which showed the woman he had seen leaving London, by car, with her
        husband, on that morning, although it had been taken a year ago'

A Motion+Manner solution, which Slobin does not mention as an option for translating
boundary-crossing movements,[7] is also fairly common in our material, with six
occurrences, as illustrated by examples (5) - (6).

(5)   Harris grumbled in the back of his throat, *jumped out of* bed and waddled over to
        the television set. (ST1)
::
(6)   Harris rosnou com o fundo da garganta, *saltou da cama* e passeou-se até ao
        televisor.
        'H. grumbled with the back of this throat, jumped from the bed and ...'

By far the most common means of translation is the Motion+Path solution, omitting
Manner entirely, as shown in (7)-(8).

(7)   He nodded, and in a moment had *slipped out of* the front door, and was gone.
        (DL2)
::
(8)   Philip anuiu e, pouco depois, *saía* a porta da rua.
        'P. agreed and, little after, exited the front door.'

Finally, a split into two clauses was only found in four cases, as in example (2) above.
        Slobin's hypothesis, then, that the translation of boundary-crossing events would
trigger either two clauses in the translation or further specification of manner does not
seem to be true of the material we have looked at: of the 29 examples with *out of*, only 8
follow that pattern (28%). Loss of manner specification is found in 66% of the cases.
        Turning now to the cases with *into*, summed up in Table 3, we find again that
only four of the 114 translations of Motion+Manner+*into* have been split into two
clauses. The picture is, however, somewhat different: Although the Motion+Path verbs
are still preferred, there is a strong tendency towards Motion+Manner verbs in the
translations as well.

Semantic elements in the Portuguese verb	Number of occurrences	Percentage of occurrence	2 clauses
Motion+Path	45	39%	1
Motion+Manner	4	4%	1
Motion+Manner (+Path)	34	30%	2
Path+Manner	7	6%	
Path	15	13%	

Manner	6	5%	
other	3	3%	
Total	114		4

Table 3: The translation of English Motion+Manner verbs + *into* into Portuguese

One conclusion to be drawn from Tables 2 and 3 is that the translator tends not to make use of two clauses in Portuguese even if a boundary is crossed in the English original. The tendency is rather to omit manner altogether both in the examples with *into* and, even more clearly so, in the examples with *out of*. Translated Portuguese does, however, show a much higher percentage of verbs expressing Manner than Slobin's and Talmy's original typology would foresee. Whether this reflects translationese or a genuine Portuguese feature is impossible to establish in the present study, since we have only looked at translated Portuguese.

4. Santos's study of perception verbs

We now turn to Santos (1998), whose subject is perception verbs in the two languages. This study is part of a larger investigation on the tense and aspect systems of English and Portuguese based on translation performance in the two directions (Santos, 1996).

Santos (1998) investigates the translation behaviour of original English sentences containing the verbs *see*, *hear* and *feel*, as well as original Portuguese sentences containing the corresponding verbs *ver*, *ouvir* and *sentir*. The study pays particular attention to tense, aspect, presence of the modal *could* and syntactic form of the direct object of the perception verb.

For the purpose of the present paper, we made a selection of the claims made in Santos (1998) which were backed by quantitative data and which concerned the English to Portuguese translation direction:[8]

1. Translation data do not show the expected one-to-one correspondence between gerundive and infinitive objects, although the constructions seem to have the same role in Portuguese. In the next example (9)-(10), we have a case of an infinitive translated by a gerundive VP, the opposite can be seen in (13)-(14) below:

(9)   Gracious, she was no prude, but she hated to hear *a woman <u>swear</u>*. (ST1)
::
(10)   Não é que fosse púdica, mas detestava ouvir *uma mulher <u>a falar</u> naquelas coisas*.

'Not that she were a prude, but she hated hearing a woman speaking about those things'

2. Both infinitive and gerundive verb phrases as objects of a perception verb suffer a drastic reduction in the translation from English into Portuguese, i.e., they are often translated by other, simpler, constructions (see example (11)-(12) below)

3. The modal *could* is often employed in English with verbs of perception, but its use seems to be controlled by considerations alien to Portuguese and does not show any reliable translation regularity[9]

Should these results be interpreted as reflecting general features of the two languages and of the translation between them, or are they due to a particular translator's style or to a source text which is special in that regard? Testing this against a different and larger material should provide some tentative answers.

5. Validation of Santos's findings

The material used in Santos (1998) consisted of one complete English novel (26,000 words) and its translation into Portuguese (henceforth the SD material).[10] Tables 4 and 5 show, respectively, the quantitative data presented in Santos (1998) and the corresponding distribution in the English-Portuguese part of the English-Norwegian Parallel Corpus (ENPC).[11]

English original	Portug. transl	Total
VPinf 21	VPinf	10
	VPger	2
	VPpastpart	0
	other	9
VPger 22	VPger	8
	VPinf	1
	VPpastpart	4
	other	9
VPpastpart	VPinf	0
	VPger	0
Total 43		43

Table 4: Distribution of the translation of VP objects in the SD material

English original	Portug. transl	*see*	*feel*	*hear*	*listen*	Total
VPinf 39	VPinf	8	4	7	2	21
	VPger	1	5	4	1	11
	VPpastpart		1			1
	other	5	1		1	7
VPger 33	VPger	7	5	4		16

	VPinf	2	1	8		11
	VPpastpart	2				2
	other	3		1		4
VPpastpart 2	VPinf	1		1		2
	VPger	1				1
Total 74 [12]		30	17	25	4	76

Table 5: Translation of VP objects in the English - Portuguese texts from the ENPC

Table 5 seems to present a moderate confirmation that there is a tendency to avoid complex VP objects of perception verbs when translating into Portuguese: there were 11 cases marked "other" in 76 examples (i.e., 11 cases where the translation is not a VP of some sort). Claim 2 seems thus to hold, but the tendency is far less striking than the 18 out of 43 instances of "other" in the SD material. Examples (11)-(12) illustrate the phenomenon in question.

(11)   I don't like hearing *my own voice echo in there*, especially at night. (MA1)
::
(12)   Não gosto de ouvir *o eco da minha própria voz*, especialmente à noite
        'I don't like to hear the echo of my own voice, especially at night'

As to the first claim, we again note that there is not a one-to-one correspondence between VPinf and VPger across the two languages. While, in the original material, VPinf matches the corresponding form in 47% of the cases and VPger in 36%, in the ENPC VPinf and VPger match in 54% and 48% of the cases, respectively. In fact, for some of the mismatches the ENPC material even exceeds the original findings, translating 11 of the 33 English VPger by VPinf (as against one instance in 22 in the SD material), and 11 out of the English 39 VPinf by VPger (as against 2 instances in 21 in the SD material). Examples (13)-(14) show two instances of a gerundive VP translated by infinitive VPs:

(13)   Mind you, I heard *my da saying to my ma*, – when was the last time we heard *him howling at the moon*? (RDO1)
::
(14)   Repara – ouvi *o meu pai dizer à minha mãe* – quando foi a última vez que *o ouvimos uivar à lua*?
        'Notice: I heard my father say to my mother – when was the last time we heard him howl at the moon?'

The high frequency of the modal *could* before perception verbs attested in Claim 3 above was not matched in the ENPC material in the case of VP objects (although the more general tendency with other kinds of objects was confirmed[13]). In the SD material, 12 of the 43 instances (28% of the cases) were modified by *could*, whereas the ENPC contained only four examples in the whole set (three, if one notes that Table 4 concerns only perception verbs in the past tense), that is, a bare 4%.[14] While this shortage of

examples offers little to prove or disprove any (translation) tendency, it indicates that the SD material is somewhat special as far as this particular feature is concerned.[15]

Because there might be considerable differences between translation practice and/or language use in Brazilian Portuguese, we decided to consider the ten translations into European Portuguese separately, summed up in Table 6. No significantly different pattern emerged, though.

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

Table 6: Translation of VP objects in the European Portuguese texts

In order to give a better overview of all possible sub-corpora[16] involved and to check whether it was possible to detect deviance in this respect, we summed up several different corpus possibilities: In addition to the three corpora already presented, we combined the SD material and the European Portuguese part of the ENPC; see Table 7 below. It is not clear which, if any, can be considered most representative as far as perception is concerned.

English original	Portuguese translation	SD		EPC= EP[17] part of ENPC		EEPC= EEP[18] part of ENPC		SD+EPC		SD+EEPC	
Vpinf	Vpinf	10	23%	21	28%	14	27%	31	26%	24	26%
	VPger	2	5%	11	14%	8	16%	13	11%	10	11%
	VPpastpart			1	1%			1	1%		
	other	9	21%	7	9%	6	12%	16	13%	15	16%
VPger	VPger	8	19%	16	21%	1	22%	24	20%	19	20%
	Vpinf	1	2%	11	14%	9	18%	12	10%	10	11%
	VPpastpart	4	9%	2	3%			8	7%	4	4%
	other	9	21%	4	5%	2	4%	13	11%	11	12%
Vppastpart	VPinf			2	3%			2	2%		
	VPger			1	1%	1	2%	1	1%	1	1%
Total		43		76		51		119		94	

Table 7: Distribution of the translation of VP objects in several corpora

One possible conclusion is that none of the corpora is big enough for this kind of study. Given the shortage of English-Portuguese parallel corpora, we are in the same position that led to the present paper: we can only state that further research, making use of larger corpora, is necessary.

6. Concluding remarks

One thing that struck us during this investigation was that researchers may not provide enough information about the studies they conduct, which makes it difficult not only to replicate them, but also to compare them to other studies. For example, we took up several points which were not addressed in Slobin's description. Likewise, the presentation of data about perception verbs in Santos (1998) could have been more useful if classified by the particular perception verb (as we did in Tables 4 and 5), and/or by the particular chapter the examples came from. In this way, we might be able to identify special uses of a particular construction in a particular part of text, for example.

More information regarding the material used in the studies would also be valuable. It is, however, likely that discrepancies between the description of a particular study and the requirements of further studies can never be totally avoided. In that light, the advantages of making the corpora used by the researchers generally available cannot be overemphasized.

Johansson (1975) discusses the need to evaluate thoroughly empirical claims in the realm of applied linguistics; this is pertinent, in our opinion, also for corpus-based claims. It is not the case, however, that the main contribution of corpus-based studies is merely quantitative. In fact, other factors such as the exposure to real text (and real translation), as well as the ability to perform possibly related tests when the need arises, can be equally (or even more) relevant in corpus-based translation studies, at least when one's main interest is linguistic differences between languages and not translation performance (as is the case in the two studies discussed in the present paper).

We conclude by raising some questions: Is validation rewarding enough? Should validation be done primarily with reference to corpora, or in terms of the claims themselves? Can one realistically expect to get an "average" corpus that can be used for evaluation of (claims based on) other corpora, no matter what claims are involved?

Our preliminary conclusion is that the same corpus can be close to standard in terms of movement description, but extremely deviant as far as the use of perception verbs is concerned, to take the examples of the two domains explored here. We therefore suggest that validation should be performed directly in terms of particular claims; and not in an indirect way in terms of the corpora used to establish them (by establishing a representative corpus on top of which claims are checked, and which is assumed "average" in every respect), even though we do not deny the interest of corpus comparison in itself.

References

Caenepeel, Mimo. 1989. Aspect, temporal ordering and perspective in narrative fiction. Phd dissertation, University of Edinburgh, Edinburgh, 1989.

Johansson, Stig. 1975. *Papers in Contrastive Linguistics and Language Testing.* Lund: CWK Gleerup.

Johansson, Stig, Jarle Ebeling, and Signe Oksefjell. 1999. English-Norwegian Parallel Corpus: Manual. Oslo: Department of British and American Studies, University of Oslo.

Kilgarriff, Adam and Raphael Salkie. 1996. "Corpus similarity and homogeneity via word frequency", in Martin Gellerstam et al. (eds.), *EURALEX '96: Papers submitted to the Seventh EURALEX International Congress on Lexicography* (Göteborg University, 1996), pp.121-130.

Kilgarriff, Adam and Tony Rose. 1998. "Measures for corpus similarity and homogeneity", *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing* (Granada, June 1998), pp. 46-52, also report ITRI-98-07, Information Technology Research Institute, University of Brighton.

Santos, Diana. 1994a. "Bilingual alignment and tense", *Proceedings of the Second Annual Workshop on Very Large Corpora* (Kyoto, 4 August 1994), pp.129-141.

Santos, Diana. 1994b. "Imperfeito: a broad-coverage study", *Actas do X Encontro da Associação Portuguesa de Linguística* (Évora, 6-8 October 1994), pp.525-539.

Santos, Diana. 1995. "L'Imperfeito portugais: étude systématique de ses fonctions et de comment en rendre compte en traduisant vers l'anglais", to appear in *Actes du XXIV Colloque sur la Linguistique des Langues Romanes* (Palermo, 18-24 September 1995).

Santos, Diana Maria de Sousa Marques Pinto dos. 1996. Tense and aspect in English and Portuguese: a contrastive semantical study. PhD dissertation, Instituto Superior Técnico, Lisboa, June 1996.

Santos, Diana. 1998. "Perception verbs in English and Portuguese", in Stig Johansson and Signe Oksefjell (eds.), *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Amsterdam: Rodopi, pp.319-342.

Slobin, Dan I. 1996. "Two ways to travel: Verbs of motion in English and Spanish", in M. Shibatani and S. A. Thompson (eds.), *Grammatical Constructions: Their Form and Meaning*, Oxford: Oxford University Press, pp.195-220.

Slobin, Dan I. 1997a. "Mind, code and text", in Joan Bybee, John Haiman, and Sandra A. Thompson (eds.), *Essays on Language Function and Language Type, Dedicated to T. Givón*, John Benjamins, pp.437-67.

Slobin, Dan I. 1997b. "In a manner of going: influences of lexicalization patterns on descriptions of motion events", Paper presented at the second Meeting of the Association for Linguistic Typology (ALT II), University of Oregon, Eugene, OR, September 11-14.

Talmy, Leonard. 1985. "Lexicalization patterns: semantic structure in lexical forms", in Timothy Shopen (ed.), *Language Typology and Semantic Description, Vol.3: Grammatical Categories and the Lexicon,* Cambridge: Cambridge University Press, 1985, pp.57-149.

Talmy, Leonard. 1991. "Path to realization: a typology of event conflation", *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, pp.480-519.

[1] A list of the texts, including the abbreviated references used in the examples in this paper, is given in http://www.hf.uio.no/iba/prosjekt/. The English-Portuguese part of the ENPC is considerably more heterogeneous than the material on which the two studies analysed are based, involving 31writers as opposed to 8 in the case of Slobin and 2 in the case of Santos.

[2] It should be mentioned that, although there are English verbs that incorporate Path in the verb root, these tend to be of Romance origin: e.g. *exit*, *enter*, *descend*.

[3] "In describing real-wold or fictional events [...] a narrator may present a series of linked paths or a path with waystations. I will call a complex path a *journey*" (Slobin, 1996:202).

[4] In the present paper we only describe the studies conducted on English material translated into Spanish. Slobin (1996) amasses further evidence both by comparing English and Spanish originals and by using elicited narratives in the two languages.

[5] One of the English texts has been counted twice since two translations of it have been included in the corpus; one into European Portuguese and one into Brazilian Portuguese.

[6] In addition, an interesting question appears: why is it that the languages differ only in self-movement, but are similar if it is the direct object (*pull*, *kick*, *shoot*, ...) and not the subject (*fall*, *rise*, ...) that moves?

[7] Although he mentions that such patterns are possible in Spanish, warning that "we will see that Spanish does not always behave like Talmy's characterization of a verb-framed language" (1996:196).

[8] In order not to bias the investigation, we did not redo the studies nor the presentation of their results, although we have access to the raw data, as opposed to the case of Slobin's material.

[9] For example, the two following similar examples (i)-(ii) and (iii)-(iv) followed different translation strategies:

(i) I *can feel* my throat tightening, a pain along the jawline. (MA1)

::

(ii) *Sinto* a garganta a apertar-se, uma dor ao longo do maxilar.
'I feel the throat tightening, a pain along the jawline'

(iii) We *could feel* the boards hopping under us. (RDO1)

::

(iv) *Podíamos sentir* as tábuas a saltar debaixo de nós.
'We were able to feel the planks jumping under us'

[10] *The Pearl,* by John Steinbeck, Bantam Books, 1975 (first edition, 1945), translated into (European) Portuguese by Mário Dionísio as: *A pérola*, Publicações Europa-América, 1977.

[11] The label VPger stands for gerundive VP objects as in *I saw him running* or *Eu vi-o a correr* or *Eu vi-o correndo*; VPinf is used for infinitive VP objects like *I saw him run* or *Vi-o correr*; VPpastpart applies to cases where the object is a past participle, like *I saw him seated* or *Vi-o sentado*.

[12] The difference in the "Total" columns is due to two translations of the same text, which means that two instances in Table 5 have been translated and counted twice.

[13] As regards instances of *could see*, for instance, only 19 translations of the 51 English occurrences employed the verbs *poder* or *conseguir*; likewise, the tense distribution in translations was 56% of
Imperfeito and 31% of Perfeito.

[14] Interestingly, the four cases all concerned VPger cases in English: the two of *see* received an "other" translation, the two of *feel* received a standard VPger translation.

[15] Note that the ENPC texts are taken from the beginnings of novels. In order to check whether the distribution of perception events in the SD material was idiosyncratic, one would need to compare with texts containing final chapters as well.

[16] We use the term "corpus" somewhat loosely in the present paragraph, for ease of exposition.

[17] English-Portuguese.

[18] English-European Portuguese.