

# Understanding Portuguese Translations with the Help of Corpora

Ana Frankenberg-Garcia  
Centre for Translation Studies  
University of Surrey

## 1. Introduction

Judging by the way the way people in general criticize machine translation output or sometimes complain about screen translation, where what actors say out loud might differ quite substantially from what is written in the translated subtitles displayed on the screen, it appears to be fairly straightforward for individuals who can speak more than one language to detect translation mistakes, at least the most glaring ones. What is not as easy to pin down is the opposite, i.e., what constitutes a decent translation, or simply what translation is all about. With the gradual establishment of Translation Studies as an independent academic discipline over the past decades, there has been a growing concern with describing translation from a neutral, non-judgemental perspective. According to Frawley (1984), translated texts are different from the source texts that give rise to them and, at the same time, they are different from target language texts that are not translations. This *third code* (Frawley 1984:168) is therefore a type of language that deserves being studied in its own right.

If until not very long ago it was only feasible to examine translations one text at a time, as Baker (1993) predicted, with the advent of corpora it is now possible to raise descriptive approaches to translation to new levels. Comparable corpora of translated and non-translated texts enable us to analyse enormous quantities of translated and non-translated language and understand some of the general characteristics that set the two apart. And parallel corpora, i.e., corpora of source texts aligned with their respective translations, enable us to uncover prevailing trends that cannot be identified when examining individual source texts and translations one by one.

In this chapter, I use corpora to look at general tendencies that characterize what translators do when they translate from English into Portuguese, and to examine what makes Portuguese translations read differently from texts originally written in Portuguese. With this, my aim is to describe a number of distinguishing features of translated Portuguese in terms of lexis, syntax and discourse.

Because the focus of the study is on what is normal rather than on what is anomalous or stands out as a mistake, many of the issues raised here often go unnoticed. However, it is precisely by noticing them that we can obtain a better understanding of Portuguese translation. This awareness is not only of purely academic interest, but can also have important implications for translator education and the development of translation software.

## 2. Translating from English into Portuguese

This section focuses on what is normal the translation of English into Portuguese. In particular, using a parallel corpus of English source texts aligned with their Portuguese translations, I will report on three short studies that were carried out to uncover trends that are not immediately visible to the naked eye.

### *2.1 Counting words*

Counting words (or characters or lines) is an important concern of many translators because it is the foundation for the way they price their work. The number of words in source texts is rarely identical to the number of words in translated texts. In educated circles, it is often claimed that changes in text length are to a large extent due to cultural and rhetorical differences between languages, with some languages being naturally wordier than others. With regard to English and Portuguese, the general belief is that

Portuguese is the more verbose of the two. According to McKenny and Bennet (2011:248), ‘While English values succinctness, clarity and objectivity, [...] Portuguese [...] is characterised by a general “wordiness” and redundancy’. If one follows this line of thought, then Portuguese translations should contain more words than English source texts. However, I believe the most significant changes affecting word counts in translation do not have to do with contrastive rhetoric, but with the morphological and syntactic characteristics of the language pairs involved.

When compared to English, Portuguese is a morphologically very rich language. This is not the place for an in-depth contrastive analysis of the two linguistic systems, but it is obvious that the number of verb inflections possible in Portuguese is substantially greater than those possible in English, that unlike English, Portuguese common nouns are marked for gender and adjectives inflect accordingly, and so on. The consequence is that many ideas that have to be expressed by a sequence of words in English can be converted into fewer Portuguese words. For example, *it was raining* [three words] becomes *chovia* [one word], *I know* [two words] becomes *sei* [one word], *a female cat* [three words] becomes *uma gata* [two words], *he tried it* [three words] becomes *experimentou-o* [one word]<sup>1</sup>, and so on. Therefore, when translating from English into Portuguese, provided there are no major changes in content, it is unlikely that the overall number of words should expand in the process. If anything, the total number of words should in fact decrease. To verify this, in the present study I looked at word counts in COMPARA<sup>2</sup>, a bidirectional parallel corpus of English and Portuguese fiction (Frankenberg-Garcia and Santos 2003). The 34 bitexts and around 1.5 million words involving English into Portuguese translation in the corpus representing the work of 25 professional Portuguese translators indicated that there was on average a 2.5%

---

<sup>1</sup> One orthographic word in this case, i.e, a string of characters separated by spaces.

<sup>2</sup> Available online at [www.linguateca.pt/COMPARA](http://www.linguateca.pt/COMPARA). Version 13.1.22 dating from 01/05/2011 was used in the present study.

decrease in the number of words from source texts to translations. To verify whether the fact that the overall number of words did not increase could be a general trend rather than a distortion caused by the idiosyncratic behaviour of just a few translators, a Wilcoxon Matched-Pairs Signed-Ranks Test was applied to the word counts pertaining to the 34 bitexts considered in the analysis<sup>3</sup>. As shown in table 1, the differences between the number of words in the source texts and in the translations were not significant. In other words, these results indicate that it is unlikely that word counts will increase when translating from English into Portuguese. Thus even if Portuguese is held to be more verbose, one cannot say that translation will expand the text in terms of word counts, at least as far as translated fiction is concerned. When pricing their work, English-Portuguese translators therefore needn't be overly concerned about whether they should count the number of words in the source text or in the translation in order for their work to be more lucrative<sup>4</sup>.

Table 1 – Comparison of word counts in English source texts (ST) and Portuguese translations (TT) in COMPARA

	MEAN	NO. OF BITEXTS	STD. DEV.	STD. ERROR MEAN
EN ST words	24075,12	34	10795,665	1851,441
PT TTwords	23700,15	34	11144,387	1911,247

RANKS	N	MEAN RANK	SUM OF RANKS
TTwords < STwords	21	17,43	366,00
TTwords > STwords	13	17,62	229,00
TTwords = STwords	0		
Total	34		

<sup>3</sup> This test was chosen because the data did not meet the normality assumption required for the paired t-test. The Wilcoxon matched-pairs test focuses on the differences between paired data – in this case, the differences between the number of words in source texts and translations for each separate bitext – but, unlike the paired t-test, it does not assume normal distribution

<sup>4</sup> However, the same does not appear to be true for translators working in the opposite, Portuguese-English direction, where a reverse analysis drawing on the remaining 41 PT-EN bitexts in COMPARA indicates that there is an average 13% increase in the number of words and that the differences between the mean number of words in STs and TTs is highly significant.

WILCOXON SIGNED RANKS TEST	TTWORDS - STWORDS
Z (Based on positive ranks)	-1,171
Asymp. Sig. (2-tailed)	,242

## 2.2 Inserting words

The fact that the total number of words in texts that are translated from English into Portuguese does not seem to increase does not mean that Portuguese translators do not occasionally add information that was not present in the originals. As amply debated in the literature (for example, see Blum-Kulka 1986, Klaudy 1998, Pym 2005 and Frankenberg-Garcia 2009), there seems to be a general tendency for translators to make translations more explicit than source texts. Simple word counts do not tell us much about this phenomenon because, as I discuss in Frankenberg-Garcia (2009), it is hard to differentiate between the extent to which changes in word counts are merely due to morphosyntactic differences between languages and the extent to which they can be attributed to actual additions or deletions of propositional meanings. Note that a translation can convey more information than a source text even when it has fewer words. In example [1] below, taken from Frankenberg-Garcia (2009), it can be seen that the English source text is seven-words long, while the Portuguese translation, which contains an extra piece of information not present in the source text (in bold), is only five-words long.

[1]	SOURCE	What have I got to complain about?
	TRANSLATION	De que me queixo <b>então</b> ?
	BACK TRANSLATION	What have I got to complain about <b>then</b> ?

One way to find out whether translated Portuguese contains propositional meanings that are not present in English source texts is to compare aligned texts looking out for particular features that might have been added to the translations. When exploring a

small, bidirectional corpus of 120 thousand running words of economic texts in English and Portuguese, Amador (forthcoming) observed that the adverb *também* [also] stood out in the Portuguese translations because in 12% of its occurrences there were no matching words in the English source texts that gave rise to them. In the present study, I attempted to find out whether it would be possible to observe findings similar to Amador (op.cit) in the much larger COMPARA corpus of fiction. As can be seen in example [2], taken from the COMPARA corpus, the addition of the adverb *também* renders the Portuguese translation more explicit than the English source text. In particular, there is nothing in the source text to indicate that the interlocutor's understanding is shared with anyone else, but in the translated text the addition of *também* gives the idea that there is someone else that shares the same understanding.

[2]	SOURCE	Is that your understanding?
	TRANSLATION	<b>Também</b> é assim que vê as coisas?
	BACK TRANSLATION	Is that how you see things <b>too</b> ?

A search for *também* among the Portuguese translations in COMPARA (796,566 words) disclosed 1029 hits. As can be seen in table 2, brief inspection of the parallel concordances then showed that there are several different English words that can generate *também* in Portuguese: *also, too, either, neither, as well, so does someone, oneself*, and so on. However, there were also many concordances where *também* found no equivalent in the English source texts. Each one of the 1029 concordances for *também* in the Portuguese translations was therefore inspected manually for the absence of an English equivalent in the parallel source text<sup>5</sup>.

---

<sup>5</sup> Although search interface to COMPARA allows for queries involving alignment constraints, where one can look up parallel concordances with *também* on the Portuguese side of the concordance but without *also* or *too* or any other possible equivalent on the English side, because of the variety of matches conceivable, it was not feasible to carry out the analysis automatically.

Table 2 – Parallel concordances for *também* in COMPARA

ENGLISH SOURCE TEXTS	PORTUGUESE TRANSLATIONS
The other guys noticed her <b>too</b> .	Os outros tipos <b>também</b> repararam nela.
It's <b>also</b> incredibly, heartstoppingly beautiful.	E <b>também</b> é lindo de morrer
I heard her breathing settle into a deep, slow rhythm before I dropped off <b>myself</b> .	Ouvi a sua respiração tomar um ritmo lento e profundo, até que adormeci <b>também</b> .
Anyway, I'd better stop, or I'll miss the 5.40 <b>as well</b> .	De qualquer modo, é melhor parar ou ainda perco <b>também</b> o comboio das 5.40.
If you had a Fall, <b>so did</b> we.	Se vocês tiveram um pecado original, nós <b>também</b> .
It's not in Edward's nature <b>either</b>	<b>Também</b> não é da natureza de Edward.

The results indicate that there were 279 cases of explicitation, i.e., 27.1% of the occurrences of *também* in translated Portuguese found no equivalent in their corresponding English source texts. In order to find out whether the addition of *também* to Portuguese translations was not something just a handful of translators represented in the corpus did, but rather a general trend of translated Portuguese fiction, it was hypothesized that the occurrences of *também* in the translations would be significantly greater than the occurrences of English equivalents in the source texts. To test this hypothesis, a Wilcoxon Matched-Pairs Signed-Ranks Test comparing the occurrences of *também* and its English equivalents in the 34 English-Portuguese bitexts in the corpus was carried out<sup>6</sup>. The results are summarized in table 3. As can be seen, the figures indicate that the addition of *também* with no corresponding matches in the source texts was highly significant, with an over 99.9% probability that these differences did not occur by chance. These results suggest that even if Portuguese translations are not significantly longer than English source texts, they can contain additional cohesive devices that contribute to rendering translated Portuguese more explicit than English source texts.

<sup>6</sup> Again, this test was used because the data was not normally distributed.

Table 3 – Comparison of *também* in Portuguese translations (TT) in COMPARA and their equivalents in English source texts (ST).

	MEAN	N	STD. DEVIATION	STD. ERROR MEAN
também_ST	22,06	34	14,719	2,524
também_TT	30,26	34	17,908	3,071

RANKS	N	MEAN RANK	SUM OF RANKS
tambémTT < tambémST	0	,00	,00
tambémTT > tambémST	33	17,00	561,00
tambémTT = tambémST	1		
Total	34		

WILCOXON SIGNED RANKS TEST	TAMBÉMTT - TAMBÉMST
Z (Based on negative ranks)	-5,016
Asymp. Sig. (2-tailed)	,000

### 2.3 Changing the position of words

Having analysed word counts and the addition of the adverb *também* in English translated into Portuguese, I would like to finish this section with an example of a translation change affecting discourse. Whereas in English the unmarked position of adverbs of time is at the end of the clause, in Portuguese there seems to be a preference for inserting time adverbs in sentence-initial position. I was therefore interested to find out whether Portuguese translators tend to move time adverbs to the front of the clause in the process of translation. Using the Portuguese translations in the COMPARA corpus again, a search for *hoje* [today] was carried out in order to retrieve parallel concordances like the ones in table 4, and the position in the clause of *hoje* and adverbial expressions containing this word like *hoje em dia* [nowadays], *hoje à tarde* [this afternoon] and so on was compared with the position of their equivalents in the English source texts.



Table 4 – Parallel concordances for *hoje* in COMPARA

ENGLISH SOURCE TEXTS	PORTUGUESE TRANSLATIONS
I couldn't decide what tie to wear <b>this morning</b> .	<b>Hoje de manhã</b> , não conseguia decidir que gravata usar.
Bobby Moore died <b>today</b> , of cancer..	<b>Hoje</b> morreu Bobby Moore, de câncer.
Even <b>now</b> , I can't think of myself like that.	Mesmo <b>hoje</b> , não consigo imaginar-me como tal.
I think people pay a lot of money for old things <b>these days</b> .	Acho que <b>hoje em dia</b> se paga bom dinheiro por essas coisas velhas.
<b>Nowadays</b> I hardly ever ride on a bus.	<b>Hoje em dia</b> , quase não tomo ônibus.
who were those zombies out there <b>tonight</b> ?	quem, eram aqueles zumbis lá no auditório <b>hoje</b> ?

Both the concordances where the English time adverb was already in clause-initial position, as in [3], and the concordances with no English equivalent to *hoje* adverbs, as in [4], were disregarded. Thus the analysis only took into account the cases in which the *hoje* adverbs were brought forward in relation to where their English equivalents were placed, as in [5], and the cases where they remained in the same non-clause-initial position as their English counterparts, as in [6].<sup>7</sup>

[3]	SOURCE	but <b>today</b> I could rule the world
	TRANSLATION	mas <b>hoje</b> sinto-me uma rainha
	BACK TRANSLATION	but <b>today</b> I feel like a queen
[4]	SOURCE	It's one of those days
	TRANSLATION	<b>Hoje</b> é um daqueles dias <sup>8</sup>
	BACK TRANSLATION	<b>Today</b> is one of those days
[5]	SOURCE	I won't go in <b>today</b> , I'll ring.
	TRANSLATION	<b>Hoje</b> não vou. Vou telefonar
	BACK TRANSLATION	<b>Today</b> I won't go. I'll ring
[6]	SOURCE	You must get a man here <b>today</b> , Sergeant.
	TRANSLATION	Tem que aqui mandar um homem <b>hoje</b> , sargento
	BACK TRANSLATION	You must send here a man <b>today</b> , Sergeant

The analysis disclosed 181 expressions equivalent to *hoje* adverbs that were in non-clause-initial position in the English source texts, of which 81 were fronted in the

<sup>7</sup> In the entire corpus, there were only two instances in which *hoje* adverbs were actually moved to the end of the clause in translation. They were considered to be too marginal and were therefore also disregarded in the analysis

<sup>8</sup> Note that the insertion of the adverb here is another indication of explicitation.

process of translation. To test whether the fronting of *hoje* adverbs in Portuguese translations was not something just a handful of translators represented in the corpus did, but rather a general trend, it was hypothesized that if nothing noteworthy was happening to these adverbs, the number of *hoje* adverbs in non-clause-initial position in English source texts would not be significantly different from the number of adverbs that remained in that same position in the Portuguese translations. A Wilcoxon Matched-Pairs Signed-Ranks Test was thus applied comparing the two, the results of which are presented in table 5. As can be seen, the mean number of *hoje* adverbs in non-clause-initial position in source-texts was 5.32, and an average of only 2.94 remained in the same place in translation. The differences between the two were highly significant, which leads one to reject the hypothesis that nothing special was happening to the position of these adverbs in the process of translation. Because the actual changes observed involved fronting *hoje* adverbs, one can conclude that there seems to be a general trend for this to occur in the translation of English into Portuguese.

Table 5 – Comparison of *hoje* adverbs in non-clause-initial position in English source texts and adverbs that remained in that same position in corresponding Portuguese translations

<i>HOJE</i> ADVERBS	MEAN	N	STD. DEVIATION	STD. ERROR MEAN
Non-clause initial in ST	5,32	34	8,601	1,475
Same position in TT	2,94	34	5,548	,952

RANKS	N	MEAN RANK	SUM OF RANKS
Non-clause initial in ST < Same position in TT	0	,00	,00
Non-clause initial in ST > Same position in TT	25	13,00	325,00
Non-clause initial in ST = Same position in TT	9		
Total	34		

WILCOXON SIGNED RANKS TEST	NON-CLAUSE INITIAL IN ST - SAME POSITION IN TT
Z (based on negative ranks)	-4,423 <sup>a</sup>
Asymp. Sig. (2-tailed)	,000

In this part of the study I used a parallel corpus to identify three features which appear to be the norm in the translation of English into Portuguese fiction: texts do not tend to expand in terms of number of words, the cohesive device *também* is often added to the translated text when there is no corresponding term in the source text, and there is a tendency to front *hoje* time adverbs. The next section focuses on what distinguishes translated and non-translated Portuguese.

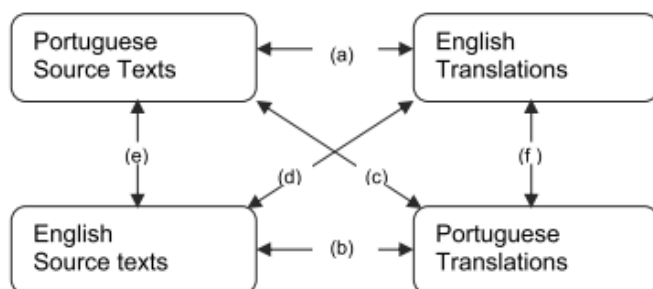
### 3. Comparing translated and non-translated Portuguese

People familiar with reading translations and texts that are not translations may realize that there is something different about them, but it is often not very easy to pin down what exactly makes translated texts sound different from texts that are not translations. Drawing on a couple of recent corpus-based studies that compare translated and non-translated Portuguese, and then carrying out a contrastive analysis of prepositions, I will describe a number of features that set the two apart.

While the analyses described in the previous section were only feasible with recourse to a parallel corpus of English source texts aligned with their corresponding Portuguese translations, parallel alignment is not necessary here. Instead, the studies described in this section draw on comparable corpora of translated and non-translated Portuguese. More specifically, while in the previous section I looked at the English source texts and Portuguese translations in the COMPARA corpus, in this section I will report on findings from that same corpus that draw on original Portuguese fiction and Portuguese fiction translated from English. This is only possible, of course, because, as shown in figure 1, COMPARA is a bidirectional corpus. Thus while the previous section drew on data from (b) in figure 1, this section is based on data from (c). The

translated Portuguese component contains 796,566 words (41 translations), and the non-translated Portuguese part of the corpus is made up of 639,360 words (34 source texts).

Figure 1- Possible Directions of Analysis in the COMPARA corpus (Frankenberg-Garcia 2006)



### 3.1 Foreign words in original and translated Portuguese fiction

One of the most immediately obvious features that sets original and translated texts apart is the use of foreign words. Because translators often resort to loans as a strategy when certain words in the source texts find no equivalent in the translation language, or when translators purposefully wish to confer a foreign flavour to the translation (Vinay and Darbelnet 1995), readers often realize that what they are reading are translations rather than source texts. But as I argue in Frankenberg-Garcia (2005), the use of foreign words is not a prerogative of translated texts. People writing directly in their native languages may also resort to foreign words. So what makes translations read differently cannot be merely the presence of foreign words. It must be something about the kind of foreign words that are found and the extent to which they are used. In Frankenberg-Garcia (2005), I examined the use of foreign words in the translated and non-translated Portuguese texts in the COMPARA corpus, focusing on the frequency and language distribution of those words. It was found that in original Portuguese there were on average only 1.5 foreign words for every ten thousand words, compared to 24.3 in translated Portuguese. This means translated Portuguese contained on average around

16 times more foreign words than original Portuguese. It is therefore not just the presence of foreign words in Portuguese translations, but rather the presence of considerably more foreign words that seems to make Portuguese original and translated fiction read differently. Note, however, that around one third of the foreign words and expressions in the Portuguese translations were already present in the English source texts that gave rise to them, as illustrated in example [7] (foreign words in italics). So it was not just a question of Portuguese translators using foreign words as a strategy, as in example [8]. In around one third of the cases, the foreign words were already there, and the translators simply preserved them.

[7] SOURCE Well, *c'est normal*  
TRANSLATION Bem, *c'est normal*

[8] SOURCE I'm writing this on my laptop on the train to London.  
TRANSLATION Estou escrevendo isto no meu *laptop* no trem para Londres.

Another feature of foreign words which may contribute to Portuguese translated fiction reading differently from original fiction in Portuguese is the choice of loan languages used. In Frankenberg-Garcia (2005), it was found that the Portuguese original fiction texts in COMPARA contained foreign words from only four different languages, namely, English, Latin, French and German. The loans were used in very few texts and none of the loan languages seemed to prevail. In contrast, the translated Portuguese texts contained loans from fifteen different languages, with English (the language of the source texts) being by far the most prevalent one. There were also loans from French, German and Latin, but what can make these translated fiction texts sound distinctively different is presence of loans from Italian, Spanish, Hebrew, Afrikaans, Greek, Japanese, Hawaiian, Czech and Yiddish, which were not detected in any of the comparable original Portuguese fiction texts analyzed.

Thus the fact that Portuguese translators resort to loans as a translation strategy only explains in part why translated and non-translated fiction differ with regard to the use of foreign words. What is not so evident is that the existing differences can be exacerbated by the fact that original English fiction seems to tolerate more foreign words and more loan languages than original Portuguese fiction.

### 3.2 Lemmas in original and translated Portuguese

It is not only the comparatively high frequency of foreign words and the presence of loan languages that are not normally used in original Portuguese fiction that may confer a different feel to translated Portuguese fiction. The choice of Portuguese words themselves can differ in original and translated texts. While the words employed in translations are constrained by the words used in the source texts that give rise to them, people writing directly in Portuguese are free to use whatever Portuguese words come to mind. The question is whether there are remarkable differences between the two.

In Frankenberg-Garcia (2008), I compared the distribution of noun, verb, adjective and adverb lemmas in the original and translated Portuguese texts in the COMPARA corpus. With regard to nouns, the main differences found were that nouns used for classifying things, such as *gênero/espécie/tipo* [type], *membro* [member], *grupo* [group], *lista* [list] and *maioria* [majority], and nouns used for specifying manner like *tom* [tone], *modo* [manner], *expressão* [expression], *aspecto* [aspect] and *atitude* [attitude] were at least twice as frequent in translated Portuguese. In contrast, nouns referring to human beings like *sobrinho* [nephew], *moço* [young man], *menino* [boy], *velha* [old lady], *soldado* [soldier], *menina* [girl], *velho* [old man], *padre* [priest], *senhora* [lady], *dono* [owner], *senhor* [gentleman], *primo* [cousin] and *rei* [king], and nouns closely associated with the Portuguese psyche like *lembrança* [memory], *saudade*

[nostalgia], *alma* [soul] and *tristeza* [sadness] were at least twice as frequent in non-translated Portuguese.

The analysis of adjectival lemmas revealed that evaluative adjectives reflecting personal opinions and feelings like *maravilhoso* [wonderful], *evidente* [obvious], *especial* [special], *horrível* [horrible], *suficiente* [enough] and *principal* [main] were at least twice as frequent in translated Portuguese, while adjectives describing facts like *gordo* [fat], *grosso* [thick], *igual* [equal], *nu* [naked], *rico* [rich] and *morto* [dead], and emotions like *triste* [sad] and *alegre* [happy] were at least twice more common in non-translated Portuguese.

With regard to verbs, linking verbs like *encontrar-se* [find oneself], *tornar-se* [become] and *sentir-se* [feel], reporting verbs like *revelar* [reveal], *exclamar* [exclaim], *lamentar* [regret], *sugerir* [suggest], *comentar* [comment] and *replicar* [reply], verbs used to indicate movement like *inclinarse* [incline oneself], *regressar* [return], *dirigir-se* [turn to], *baixar* [lower], *virar-se* [turn], *apanhar* [catch], *apoiar* [lean], *voltar-se* [turn], *acenar* [nod] and *abandar* [shake], and verbs that frequently precede other verbs like *tentar* [try], *conseguir* [manage] and *permitir* [allow] were at least twice as frequent in translated Portuguese. On the other hand, highly lexical verbs closely related to the dramatic language of literary texts like *vencer* [win], *fugir* [run away], *beijar* [kiss], *cantar* [sing], *quebrar* [break], *sonhar* [dream], *amar* [love], *roubar* [steal], *chorar* [cry], *matar* [kill], *morrer* [die] and *nascer* [be born] were at least twice as frequent in non-translated Portuguese.

The analysis of adverbial lemmas, in turn, showed that adverbs with the *mente* suffix like *profundamente* [deeply], *absolutamente* [absolutely], *completamente* [completely], *simplesmente* [simply], *perfeitamente* [perfectly] and *imediatamente* [immediately], and other adverbs of manner like *demasiado* [too] and *bastante* [rather]

were at least twice as frequent in translated Portuguese, while in original Portuguese the adverbs that stood out were mostly adverbs of time and frequency, like *enfim* [finally], *logo* [soon], *ontem* [yesterday], *jamaiz* [never], *amanhã* [tomorrow] and *hoje* [today].

Interestingly, the analysis also revealed that some of the lemmas that were noticeably more frequent in the translated Portuguese had near synonyms that were markedly more frequent in non-translated Portuguese. For example, the following pairs show the preferred lemma in translated Portuguese followed by the preferred synonym in non-translated Portuguese: *recordação/lembrança* [souvenir], *escola/colégio* [school], *edifício/prédio* [building], *enorme/imenso* [enormous/huge], *compreender/entender* [understand], *recordar/lembrar* [remember], *reparar/notar* [observe], *observar/examinar* [observe/examine], *decidir/resolver* [decide], *obrigar/mandar* [force/order], *manter/guardar* [keep], *apanhar/recolher* [pick/gather], *completamente/todo* [completely] and *finalmente/enfim* or *afinal* [finally].

While the above synonymous pairs seem to reflect mostly linguistic differences, with translated Portuguese appearing to express preference for the more formal option, many of the noun, verb, adjective and adverb lemma differences described earlier seem mirror actual contrasts in culture.

### 3.3 Prepositions in original and translated Portuguese

The analysis carried out in Frankenberg-Garcia (2008) focused on contrastive lexis. However, there may also be grammatical features that distinguish translated and non-translated Portuguese. In the present study, I conducted an exploratory analysis to examine prepositions in translated and non-translated Portuguese. A search for the overall distribution of a random selection of core Portuguese prepositions in COMPARA revealed that three of them might be over-represented in translated



Portuguese. The results of this summary analysis are presented in table 6. Although there was nothing particularly remarkable about most prepositions, *após* [after], *durante* [during] and *perante* [before] were selected for further analysis because (1) they had a frequency of over 10 hits per 100 thousand words in at least one corpus (the prepositions with lower frequencies were not considered to be representative enough) and (2) their distributions seemed to be markedly different in translated and non-translated Portuguese.

Table 6 - Distribution of a selection of core prepositions in translated and non-translated Portuguese (per 100 thousand words)

PREPOSITION	TRANSLATED PORTUGUESE	NON-TRANSLATED PORTUGUESE
a + ao(s) + à(s)	5069.5	4608.2
ante	1.6	3.6
após	21.3	12.8
até	164.6	148.0
com	1125.1	952.5
conforme	2.9	6.7
de + da(s) + do(s)	6725.8	6328.7
desde	33.9	41.9
durante	101.1	33.3
em + num + numa	1116.8	1012.9
entre	99.9	102.6
exceto + excepto	7.0	1.6
mediante	0	0.5
para	1276.7	937.7
perante	11.7	5.2
por	707.5	577.8
sem	166.7	228.5
sob	27.0	23.8
sobre	147.8	140.3

To find out whether these distinctive distributions were not being distorted by just a handful of translators or authors, a more fine-grained analysis was carried out by investigating their distributions separately in the 41 translated and 34 non-translated Portuguese texts in the corpus, the results of which are summarized in table 7. Next, for

each separate preposition, a Mann-Whitney Test<sup>9</sup> was applied to find out whether the differences observed could be significant. The results are summarized in tables 8, 9 and 10.

Table 7 - Distribution of *após*, *durante* and *perante* in translated and non-translated Portuguese

	MEAN NO. OF HITS/100 K WORDS IN TRANSLATED PORTUGUESE	N	STD. DEV.	MEAN NO. OF HITS/100 K WORDS IN NON-TRANSLATED PORTUGUESE	N	STD. DEV.
<i>após</i>	20.7	41	18.7	12.1	34	16.5
<i>durante</i>	92.2	41	52.0	32.6	34	31.0
<i>perante</i>	11.3	41	14.6	4.8	34	8.1

Table 8 - Comparison of *após* in translated and non-translated Portuguese

TEXT TYPE	N	MEAN RANK	SUM OF RANKS
translated PT	41	43,70	1791,50
non-translated PT	34	31,13	1058,50
Total	75		

Mann-Whitney U	463,500
Wilcoxon W	1058,500
Z	-2,497
Asymp. Sig. (2-tailed)	,013

Table 9 - Comparison of *durante* in translated and non-translated Portuguese

TEXT TYPE	N	MEAN RANK	SUM OF RANKS
translated PT	41	50,10	2054,00
non-translated PT	34	23,41	796,00
Total	75		

Mann-Whitney U	201,000
Wilcoxon W	796,000
Z	-5,280
Asymp. Sig. (2-tailed)	,000

Table 10 - Comparison of *perante* in translated and non-translated Portuguese

Text type	N	Mean Rank	Sum of Ranks
translated PT	41	43,12	1768,00
non-translated PT	34	31,82	1082,00
Total	75		

<sup>9</sup> This test is used to compare two independent samples of data that does not satisfy the criterion of normal distribution, as was the case here.

Mann-Whitney U	487,000
Wilcoxon W	1082,000
Z	-2,337
Asymp. Sig. (2-tailed)	,019

As can be seen, in all three cases the differences observed were statistically significant. In other words, it was highly unlikely that the differences between the use of *após*, *durante* and *perante* in translated and non-translated Portuguese were merely due to chance or the individual choices of just a few. Interestingly, all three prepositions sound rather formal and have less formal syntactic options with similar meanings that could be used in their place: *após* is equivalent to the less formal prepositional phrase *depois de*; *perante* is equivalent to the less formal prepositional phrase *diante de*; and *durante* can be replaced by a less formal temporal clause with *enquanto*. Thus just like with the synonymous pairs of lemmas discussed in the previous section, this seems to be yet another indication that translated Portuguese fiction tends to be more formal than its non-translated counterpart.

#### 4. Conclusions

Corpora enable us to access large quantities of text and describe language from the viewpoint of a vast number of users. In this chapter I used a bidirectional parallel corpus of English and Portuguese of three million words to describe a selection of distinguishing features of translated Portuguese fiction. The trends unveiled in this study are not idiosyncrasies, but rather the result of an analysis of the combined work of the 25 Portuguese translators represented in the corpus.

In the first part of the study, the analysis was based on English source texts aligned with their Portuguese translations, and it was possible to reach three conclusions. First, despite the general belief that Portuguese tends to be less concise

than English, the translation of English into Portuguese does not seem to result in a significant increase in the number of words. Second, despite the fact that the translated Portuguese texts do not seem to expand in terms of number of words, Portuguese translators may nevertheless add words to the translation that will find no equivalent in the source texts. In particular, this study showed that there is a strong tendency for Portuguese translators of fiction to insert the adverb *também* out of their own accord, rendering the resulting translation more explicit than its source text. Third, there seems to be a propensity for Portuguese translators to front adverbs of time when they are in non-clause initial position in the source text. More specifically, the analysis showed that when adverbs like *nowadays, these days, today, this morning, tonight, now*, and so on are in non-clause initial position in English source texts, there is a tendency for Portuguese translators to bring them forward.

The second part of the present chapter did not look at what happened from source text to translation. Instead, it summarized the findings of two previous studies comparing the translated and non-translated Portuguese texts in the COMPARA corpus and reported on new findings from this same corpus in order to illustrate factors that can contribute to making translations read differently from texts that are not translations. The first study showed that translated fiction tends to contain considerably more foreign words and tends to make use of a wider range of loan languages than non-translated Portuguese fiction. The fact that translated fiction contains more loans is not too surprising, given that the use of foreign words is a common translation strategy. However, what would not have been easily observable without looking at the English source texts as well was that one third of the foreign words present in the Portuguese translations and the unusual loan languages used actually originated in the English source texts. The comparatively extremely high number of foreign words detected in

translated fiction in Portuguese is thus only partly due to translation itself. Much of the foreignness observed stems from the fact that original English fiction seems more permeable to foreign words and different loan languages than original Portuguese fiction.

The second study compared the distribution of noun, verb, adjective and adverb lemmas in translated and non-translated Portuguese and unveiled a series of lexical contrasts between the two. A few of the differences observed could have been easily anticipated: for example, the conspicuous absence of the very Portuguese word *saudade* [nostalgia] from translations, and, probably because of the influence of the English cognate *finally*, the preference for the adverb *finalmente* instead of its synonyms *enfim* or *afinal*, which are more common in non-translated Portuguese. However, the analysis also disclosed a number of unexpected and remarkable linguistic and cultural contrasts and a tendency for formality, all of which may play an important role in making translated and non-translated Portuguese fiction sound different.

Finally, an analysis was carried out in to explore whether prepositions too might have distinctive distributions in translated and non-translated Portuguese. The results revealed that the prepositions *após*, *durante* and *perante* were significantly more frequent in translations. Because these prepositions have less formal syntactic equivalents which could have been used instead, the observations made add strength to the idea that there is a propensity for a more formal register to be used in translated Portuguese.

This study used corpora to uncover a series of trends portraying what is normal (rather than what is unusual) in Portuguese translations. While it is relatively easy to spot what stands out as anomalous, it is only possible to capture what is standard practice when analysing large quantities of text. I hope the findings reported here will

not only contribute to our general understanding of Portuguese translation, but also provide valuable insights to those working in translator education and the development of translation software. There are countless other analyses that can be carried out using parallel and comparable corpora, and I would like to think that this chapter will inspire others to develop further research in this area.

## References

- Amador, P. (forthcoming), Universais da Tradução: a explicitação através de um estudo de textos económicos. MA Dissertation, ISLA Campus Lisboa, Portugal
- Baker, M. (1993), 'Corpus linguistics and translation studies. Implications and applications', in M. Baker, G. Francis and E. Tognini-Bonelli (eds) Text and Technology: In Honour of John Sinclair. Amsterdam and Philadelphia: John Benjamins, pp. 233-250.
- Blum-Kulka, S. (1986), 'Shifts of cohesion and coherence in translation', in J. House and S. Blum-Kulka (eds) Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies. Tübingen: Gunter Narr, pp. 17-35.
- Frankenberg-Garcia, A. (2005), 'A corpus-based study of loan words in original and translated texts', in P. Danielsson and M. Wagenmakers (eds) Proceedings of the Corpus Linguistics 2005 conference, Birmingham, UK, 14-17 July. Online [<http://www.corpus.bham.ac.uk/pclc/index.shtml>]
- Frankenberg-Garcia, A. (2006), 'Using a parallel corpus in translation practice and research'. Proceedings of Contrapor 2006, 1ª Conferência de Tradução Portuguesa, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 11-12 September 2006, pp. 142-148.
- Frankenberg-Garcia, A. (2008), 'Suggesting rather special facts: a corpus-based study of distinctive lexical distributions in translated texts'. Corpora, 3(2), 195-211.
- Frankenberg-Garcia, A. (2009), 'Are translations longer than source texts? A corpus-based study of explicitation', in A. Beeby, P. Rodríguez, and P. Sánchez-Gijón (eds) Corpus use and translating. Amsterdam and Philadelphia: John Benjamins, pp. 47-58.
- Frankenberg-Garcia, A. and Santos, D. (2003), 'Introducing COMPARA: the Portuguese-English parallel corpus', in F. Zanettin, S. Bernardini and D. Stewart (eds) Corpora in Translator Education. Manchester: St. Jerome, pp. 71-87.

Frawley, W. (1984), 'Prolegomenon to a theory of translation', in W. Frawley (ed) Translation, literary, linguistic and philosophical perspectives. London and Toronto: Associated University Presses, pp. 159-175.

Klaudy, K. (1998), 'Explicitation', in M. Baker (ed) Encyclopedia of Translation Studies. London: Routledge, pp. 80-85.

McKenny, J. and Bennet, K. (2011), 'Polishing papers for publication: palimpsests or procrustean beds?', in A. Frankenberg-Garcia, L. Flowerdew and G. Aston (eds) New Trends in Corpora and Language Learning. London: Continuum, pp. 247-262.

Pym, A. (2005), 'Explaining explicitation', in K. Karoly and A. Foris (eds) New Trends in Translation Studies. In Honour of K. Klaudy. Budapest: Akadémia Kiadó, pp. 29-34.

Vinay, J. and Darbelnet, J. (1995), Comparative Stylistics of French and English: A Methodology for Translation. Amsterdam and Philadelphia: John Benjamins.