

Are translations longer than source texts? A corpus-based study of explicitation

Ana Frankenberg-Garcia

Abstract

Explicitation is the process of rendering information which is only implicit in the source text explicit in the target text, and is believed to be one of the universals of translation (Blum-Kulka 1986, Olohan and Baker 2000, Øverås 1998, Séguinot 1988, Vanderauwera 1985). The present study uses corpus technology to attempt to shed some light on the complex relationship between translation, text length and explicitation. An awareness of what makes translations longer (or shorter) and more explicit than source texts can help trainee translators make more informed decisions during the translation process. This is felt to be an important component of translator education.

Introduction

What translators should and what they shouldn't do with texts has been a matter of controversy since Cicero (and later St Jerome) first made reference to the word-for-word versus sense-for-sense dichotomy. In recent years, however, there has been a change of emphasis in translation studies away from the debate of what translators ought to do and towards descriptive studies of what practicing professional translators generally do. The shift of focus is beneficial to translator education. Instead of being swamped with prescriptive dos and don'ts, trainee translators who are made aware of the regular features of translated texts can use this knowledge to make their own conscious and informed decisions during the translation process.

The present study uses corpus technology to revisit one of the more widely discussed characteristics of translated texts: the phenomenon of explicitation. Unlike previous

studies, however, an attempt is made here to analyse explicitation from the perspective of text length. The relationship between translation, explicitation and text length is not simple, and in this study I try to shed some light on the complexity of the matter. In particular, I wish to draw attention to the difficulties of comparing text length across languages, to what happens to word counts in bi-directional analyses of comparable source texts and translations, and to how explicitation appears to be an intrinsic feature of translation even when translations do not have more words than source texts. The analysis carried out in the present study would not have been possible without recourse to corpora, and it is hoped that the results obtained can inform translator education and translation practice.

Explicitation

Explicitation is the process of rendering information which is only implicit in the source text explicit in the target text (Vinay & Darbelnet 1958). Explicitation is obligatory when the grammar of the target language forces the translator to add information which is not present in the source text, but can occur voluntarily when, for no grammatically compelling reason, translators distance themselves from the source text in a way that makes the target text easier to comprehend.

Example 1 below illustrates the obligatory explicitation of gender in the translation of English into Portuguese.¹

(1)	EBJT2 2038
SOURCE	Frances liked her doctor.
TRANSLATION	Frances gostava dessa médica.
BACK TRANSLATION	Frances liked this female doctor.

As Portuguese is marked for gender, the translator in example 1 was forced to discriminate between a female and a male doctor. Obligatory explicitation can also occur in the reverse direction. Example 2 illustrates three different aspects of obligatory explicitation in the translation of Portuguese into English. First, while the Portuguese possessive pronoun *sua* agrees with the object *pele*, the equivalent *her* in English agrees with the subject. This means that while the Portuguese reader has no means of telling that the *skin* in the text belongs to a female, the English translator was forced to make the connection explicit. Second, since Portuguese is a pro-drop language, the reader will read on and still not know whether the person whose nose is ‘the most voluminous one in the world’ is a man or a woman. As English is not a pro-drop language, the translator had to insert the pronoun *she*, making it once again clear to the reader that the person in question is a female. Third, parts of the body do not have to be preceded by the possessive pronoun in Portuguese, but they do in English. The effect is that the person to whom the hair belongs is made more explicit in the English translation.

(2)	PBMRI 575
SOURCE	[...] sua pele lembrava a crosta lunar e tinha o nariz mais volumoso do mundo; o cabelo era cor de fogo [...]
LITERALLY	[...] his/her skin reminded one of the lunar crust and Ø had the most voluminous nose in the world; the hair was the color of fire [...]
TRANSLATION	[...] her skin resembled the lunar crust and she had the most voluminous nose in the world; her hair was the color of fire [...]

In contrast to obligatory explicitation, voluntary explicitation occurs when, for no grammatically compelling reason, translators distance themselves from the source text in a way that makes the target text easier to comprehend.² In example 3, the translator introduced the adverb *so* at the beginning of the English sentence, although it is

neither present in the Portuguese source text, nor there is anything about the grammar of English that makes it compulsory. The effect is that the connection between the event described by that sentence and a previous one in the text is made more explicit in the translation.

(3)	PBAD1 435
SOURCE	Você também gosta dela?
LITERALLY	You like her too?
TRANSLATION	So you like her too?

As shown in example 4, exactly the same can occur in the translation of English into Portuguese.

(4)	EBDL3T2 799
SOURCE	"It's probably Rummidge.
TRANSLATION	-- Então é provável que seja Rummidge.
BACK TRANSLATION	" So it's probably Rummidge.

There is abundant evidence of voluntary explicitation in literature. Vanderauwera (1985), for instance, described numerous examples in the English translation of Dutch novels. Blum-Kulka (1986) found cohesive devices in Hebrew translations that were not present in English source texts. Séguinot (1988) found non-obligatory connectives in translations from English into French and from French into English. Based on studies such as these, voluntary explicitation has come to be viewed as one of the universals of translation (Vanderauwera 1985) and as something inherent to the nature of the translation process (Séguinot 1988). After a systematic study of the phenomenon from a perspective of discourse, Blum-Kulka (1986) put forward the explicitation hypothesis, which holds that translations tend to be more explicit than

source texts, regardless of the increase in explicitness dictated by language-specific differences.

In the beginning of the nineties, Baker (1993) predicted that qualitative studies such as the above could be greatly enhanced by quantitative, corpus-based analyses of translations. Indeed, Øverås (1998) examined explicitation and implicitation shifts in the English-Norwegian Parallel Corpus, and found that there was more explicitation than implicitation in both Norwegian translated from English and English translated from Norwegian. Using two comparable corpora, Olohan and Baker (2000) analysed the insertion of the optional *that* following the reporting verbs *say* and *tell* in data from the Translational English Corpus (TEC) and the British National Corpus (BNC), and found that the explicitation of *that* is more frequent in the English translations from the TEC than in the English originals from the BNC.

The present study is an attempt to analyse voluntary explicitation from the perspective of text length. Because voluntary explicitation is generally achieved by the addition of extra words in the translation text, this study seeks to test whether translations are likely to be longer than source texts, regardless of the languages concerned. Using the COMPARA corpus (Frankenberg-Garcia and Santos 2003), the length of original English and Portuguese language fiction text extracts was compared with the length of their respective translations into Portuguese and English.³

Text length in COMPARA 5.2

COMPARA is a parallel, bi-directional and extensible corpus of English and Portuguese fiction currently in version 6.7.1, with 2.8 million words. In this study, an earlier version of the corpus was used. Version 5.2, accessed in November 2003, contained 37 source texts (25 in Portuguese and 12 in English) and 40 translations (the corpus admits the alignment of more than one translation per source text). The texts extracts varied from just under 2000 to over 42000 words. The work of twenty-seven different authors and thirty-one different translators was represented, with some authors and translators being represented more than once. The overall distribution of Portuguese and English words in COMPARA at the time is summarized in table 1.

Table 1 Distribution of Portuguese and English words in COMPARA 5.2

Words	Source texts	Translations
Portuguese	388452	384285
English	388430	431691

The above figures indicate that while the English translations in the corpus contained on average 11% more words than their source texts in Portuguese, the Portuguese translations contained 1% fewer words than their source texts in English. All these numbers tell us is that translators working from Portuguese into English will probably earn more if they base their fees on the number of words in the translation text, while those working from English into Portuguese might be better off if they get paid by the number of words in the source text. The above distribution of words does not, however, shed any light on the relationship between translation and explicitation, for it is impossible to tell the extent to which the differences observed are due to differences between Portuguese and English or differences between source texts and translations.

Text length across languages

Claims about the relative length of texts across languages are extremely difficult to put to test. In a recent discussion on the Corpora List, there were over twenty postings on the subject.⁴ The main problem seems to be that, because of the diverging lexicogrammatical characteristics of languages, it is complicated to decide on what scale to use. Different measures will affect different languages differently. If text length is measured in terms of number of words, for example, it is not hard to see that whatever the criteria for counting words are, they might make some languages seem wordier than others. Table two illustrates this by means of a few examples of how word processors count equivalent meanings in Portuguese and English.

Table 2 Word processor word counts in English and Portuguese

English	Portuguese
isn't (1)	não é (2)
teapot (1)	bule de chá (3)
gave him (2)	deu-lhe (1)
Did you like it? (4)	Gostou? (1)

As can be seen, English allows for contractions like *isn't*, which are not possible in Portuguese: *não é*. A word processor counts the former as one word and the latter as two words. Even if contractions were to be counted as separate words, however, there are other problems. For example, there are many compound words in English, like *teapot*, which have to be written separately in Portuguese: *bule de chá*. But then not everything in English is more economical than in Portuguese. Portuguese clitics are often attached to verbs, making separate words in English, like *gave him*, count as a single one in Portuguese: *deu-lhe*. Also, because Portuguese is a pro-drop language, it

is often the case that only one word is required to say things that would take three or four words in English. For example, to ask the four-word question *Did you like it?* in Portuguese, only one word is required: *Gostou?*

This is not the place for an extensive contrastive analysis of the lexico-grammatical characteristics of the two languages. The examples seen, however, show that word counts per se are not enough to compare text length across languages, let alone analyse the relationship between translation and explicitation. In fact, as example 5 below indicates, a translation can be more explicit than a source text even when it has fewer words.

(5)	EBDL1T1 670
SOURCE	What have I got to complain about? (7 words)
TRANSLATION	De que me queixo então ? (5 words)
BACK TRANSLATION	What have I got to complain about then ?

Conversely, example 6 illustrates how there can be an increase in words in translation without any explicitation whatsoever:

(6)	PBRF1 1299
SOURCE	Fui visitá-lo. (2 words)
LITERALLY	I went to visit him.
TRANSLATION	I went to visit him. (5 words)

Some postings on the Corpora List argue that character counts constitute a better measure for comparing text length across languages inasmuch as they disregard the morphological and syntactic problems of word counts. However, as shown in table 3, equivalent meanings in two languages can also vary in terms of character length.

Differences in the number of characters in source texts and translations can therefore not help analyse the question of explicitation any more than word counts can.

Table 3 Character counts (with spaces) in English and Portuguese

English	Portuguese
isn't (5)	não é (5)
teapot (6)	bule de chá (11)
gave him (9)	deu-lhe (7)
Did you like it? (16)	Gostou? (7)

Another method for comparing text length across languages suggested in the discussion list is morpheme counts. Indeed, as can be seen in table 4, counting the number of morphemes of equivalent meanings in two different languages does seem to flatten out many of the differences of word and character counts.

Table 4 Morpheme counts in English and Portuguese

English	Portuguese
isn't (3)	não é (3)
teapot (2)	bule de chá (3)
gave him (4)	deu-lhe (4)
Did you like it ? (4)	Gostou? (3)

However, morphemes are not only extremely difficult to count, but they are also sensitive to increases in explicitness dictated by language-specific differences. Thus in the examples given, *teapot* is made up of two morphemes, but its Portuguese equivalent, *bule de chá*, is made up of three because the preposition *de* has to be inserted to link the nouns *bule* and *chá*. Likewise, the English sentence *Did you like it?* has one morpheme more than its Portuguese equivalent *Gostou?* because the English verb *like* has to be followed by an object, while its Portuguese equivalent, *gostar*, doesn't. As morpheme counts do not discriminate between the addition of morphemes dictated by language specific differences and the extra morphemes that are a product of voluntary explicitation, they too are not appropriate for analysing the

differences between source texts and translations independently of the differences between languages.

Notwithstanding these limitations, the present study works on the assumption that language-dependent biases can be controlled in bi-directional analyses. In other words, when comparing source texts and translations to find out whether text length increases in translation, it is assumed that an analysis of the translations from language y into language z combined with an analysis of the translations from language z into language y may shed some light on the extent to which differences in text length are due to language-dependent factors alone.

If word, character or morpheme counts happen to make one language seem shorter than the other, it is assumed that this will affect both the translations *and* the source texts in that language, in the same way as it will make both the translations *and* the source texts in the other language seem longer. A carefully balanced, bi-directional sample of source texts and translations will therefore enable one to filter out language-dependent biases, and find out whether translations are longer than source texts regardless of the changes in text length dictated by language-specific constraints.

A balanced corpus

Although COMPARA 5.2 contains a similar amount of Portuguese and English words (c.f. table 1), it is not a balanced corpus. According to Frankenberg-Garcia and Santos (2003:74), the responsibility of achieving balance, if balance is necessary for a

particular study, "is left entirely in the hands of the user" of the corpus. In the present study, as discussed in the previous section, balance was deemed essential. It was important to take care that neither Portuguese nor English, nor any particular author or translator, was over-represented. To ensure this, the starting point for the analysis was the selection of a sub-corpus of sixteen source texts by eight different native-English authors and another eight different native-Portuguese authors translated into Portuguese and English by sixteen different translators. The texts used in the analysis are identified in table 5.

Table 5 Source texts and translations selected for text length analysis

Text ID	Author	Translator
EBDL2	David Lodge	M. Carlota Pracana
EBJB1	Julian Barnes	Ana M. Amador
EBJT1	Joanna Trollope	Ana F. Bastos
ESNG1	Nadine Gordimer	Geraldo G. Ferraz
EUHJ1	Henry James	M.F. Gonçalves
EBLC1	Lewis Carrol	Y. Arriaga, N.Videira & L.Lobo
EBOW1	Oscar Wilde	Januário Leite
EURZ1	Richard Zimler	José Lima
PBPC1	Paulo Coelho	Alan Clarke
PBMR1	Marcos Rey	Cliff Landers
PMMC1	Mia Couto	David Brookshaw
PPMC1	Mário de Carvalho	Gregory Rabassa
PPSC1	Sá Carneiro	Margaret J. Costa
PBAD1	Autran Dourado	John Parker
PBMA3	Machado de Assis	John Gledson
PPCC1	C. Castelo Branco	Alice Clemente

Another crucial aspect of balance was the size of each source text. In order to assign equal weight to the English-Portuguese and Portuguese-English translations, it was important to take as a starting point for the analysis source-text extracts of the same length in the two languages. COMPARA's Complex Search facility was used to retrieve a random selection of sentences from each of the source texts in table 5 aligned with their corresponding translations. Because of copyright restrictions, some of the samples obtained were much shorter than others. To correct this imbalance, all source texts were reduced to around 1500 words each, which was the approximate

size of the smallest source-text sample obtained. This was done simply by cutting down on the number of sentences for each source text until what was left added up to or near 1500 words. It was then possible to find out how many words there were in each corresponding translation. To be extra rigorous in the analysis, translators' notes were excluded, and only the words in the main translation text were taken into consideration.

Results

The number of words in the 16 English and Portuguese source texts analysed and the number of words in their corresponding translations into Portuguese and English are summarized in table 6.

Table 6 Distribution of words in source texts and translations of a balanced, bi-directional sample of comparable Portuguese and English text extracts

Text ID	ST words	TT words
EBDL2	1501	1585
EBJB1	1499	1467
EBJT1	1501	1538
ESNG1	1498	1441
EUHJ1	1499	1364
EBLC1	1499	1321
EBOW1	1498	1299
EURZ1	1500	1550
PBPC1	1499	1682
PBMR1	1499	1714
PMMC1	1502	1867
PPMC1	1501	1726
PPSC1	1502	1714
PBAD1	1501	1675
PBMA3	1500	1753
PPCC1	1502	1583
Total	24001	25279
Mean	1500	1580

According to the above figures, while five translations had fewer words than their corresponding source texts, the remaining eleven translations were all longer. Put together, the translations contained on average 5% more words than the source texts.

A Paired Student's t-test was applied to the above figures in order to test whether this

overall increase in words from source text to translation was significant. The *t*-value obtained for a one-tailed test at the 95% significance level enabled one to reject the null hypothesis. In other words, it can be said with 95% confidence that the translations in this sample contained on average significantly more words than the source texts.

Conclusions

Assuming that the balanced, bi-directional sample of comparable Portuguese and English source texts and translations used in the present study constituted an effective means of cancelling out the language-dependent biases of word counts, it is possible to conclude that the overall increase in the number of words observed in the translations is more likely to be due to differences between source texts and translations than due to lexico-grammatical differences between Portuguese and English. Given that voluntary explicitation often takes the form of the addition of extra words in the translated text, the present results provide quantitative evidence in support of the idea that translations tend to be more explicit than source texts, regardless of the changes in explicitness dictated by language-specific differences.

Since the present analysis was based on only a small sample of Portuguese and English source texts and translations, in the future it would be necessary to carry out additional comparisons of source texts and translations using more texts. As only fiction texts were used, it would also be important to find out if different genres

render similar results. Another essential research question for the future would be to find out if the present results can be replicated using different language pairs.

Implications for translator education

It is not uncommon to overhear in educated circles claims that some languages are “wordier” than others, and that this is the reason why translations are longer or - depending on the language direction – shorter than source texts. Trained translators should know better. An important goal of translator education is achieved when trainee translators become aware of the complexity of translation. This includes becoming aware of the reasons why text length can vary from source texts to translations. As I hoped to have shown in this paper, the relationship between translation and text length is not dictated just by the morphological and syntactic differences between languages, and obligatory explicitation is something quite different from voluntary explicitation. Translators who become aware of issues such as these can make more conscious and more informed decisions during the translation process. Understanding what makes translations longer or shorter and what makes them more explicit than source texts is one of the factors that differentiate trained translators from bilinguals who are not translators.

Notes

1. All examples were taken from the COMPARA corpus. Letter and number codes identify source/translation pair plus alignment unit in question.
2. Voluntary explicitation is being used here as an all-embracing term that covers all explicitation that is not obligatory, from the explicitation of cultural information to the explicitation of syntactically optional elements.

3. Available at <http://www.linguateca.pt/COMPARA/>

4. Available at <http://helmer.aksis.uib.no/corpora/>

References

- Baker, M. 1993. "Corpus linguistics and translation studies. Implications and applications." In *Text and Technology: In Honour of John Sinclair*, M. Baker, G. Francis and E. Tognini-Bonelli (eds), 233-250. Amsterdam/Philadelphia: John Benjamins.
- Blum-Kulka, S. 1986. "Shifts of cohesion and coherence in translation". In *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*, J. House and S. Blum-Kulka (eds), 17-35. Tübingen: Gunter Narr.
- Frankenberg-Garcia, A. and Santos, D. 2003 "Introducing COMPARA, the Portuguese-English Parallel Corpus". In *Corpora in Translator Education*, F. Zanettin, S. Bernardini and D. Stewart (eds), 71-87. Manchester: St. Jerome.
- Olohan, M. and Baker, M. 2000. "Reporting *that* in translated English: Evidence for subconscious processes of explicitation?" *Across Languages and Cultures* 1(2): 141-158.
- Øverås, L. 1998. "In Search of the Third Code: an investigation of norms in literary translation". *Meta*, XLIII, 4.
- Séguinot, C. 1988. "Pragmatics and the Explicitation Hypothesis". *TTR: Traduction, Terminologie, Rédaction* 1 (2):106-114.
- Vanderauwera, R. 1985. *Dutch Novels Translated into English: the transformation of a 'minority' literature*. Amsterdam: Rodopi.
- Vinay, J. P. and Darbelnet, J. 1958. *Stylistique Comparée du Français et de l'Anglais: Méthode de Traduction*. Paris: Didier.