

AN INVESTIGATION OF A LITERARY TRANSLATOR'S STYLE IN A NOVEL WRITTEN BY JORGE AMADO

Diva Cardoso de CAMARGO (Universidade Estadual Paulista)*

ABSTRACT: *Our study is based on Berber Sardinha's and Baker's proposals in order to create a Brazilian Portuguese parallel corpus built in electronic format as well as carry out observation of a literary translator's use of linguistic patterns in terms of type/token ratio in a translation of Jorge Amado's novel.*

KEYWORDS: *Translation; Literary translation; Corpus linguistics; Corpus-based translation research; Translator's style.*

0. Introduction

Although corpus-based research fulfils several potentialities in the discipline of translation studies, this has been only lately recognised, due to the marginal condition translation had had to cope with not only imposed by corpus linguistics (which had formerly viewed translated texts as non authentic texts) but also going backwards to the low status imposed by its first theoretical matrixes. From literary studies, translation inherited sacramentalisation of the 'original' and the author's style; from linguistic studies, a preoccupation with analyses of register or style of social groups of language users. Underlying such associations with the 'original', translators were given the [impossible] task of not having their own styles and, at the most impersonal level, simply reproducing the style of the author. Within that perspective, scholars had developed criteria for evaluating translated text (TT) good or bad quality based on the relative success of original text (OT) 'style' reproduction. Lately, several translation theorists have emphasised the translator's presence but without showing any actual TT features.

For this reason, our small-scale study intends not to gear towards the traditional sense of examining whether the author's style is adequately transposed into the translation. From a different perspective, this paper puts the focus on the question of a literary translator's linguistic patterns, in order to investigate, by means of type/token ratio variation, whether he would individually present some evidence of use of style viewed as distinctive, recurring and preferred marks of linguistic behaviour, that are related to lexical items.

1. Theoretical basis

Our investigation has its starting point based on Berber Sardinha's (1997, 1999, 2000, in press) and Mona Baker's (1993, 1995, 1996, 2000) proposals in order to carry out a study of a translators' use of stylistic patterns by means of observing a parallel corpus.

Based on Berber Sardinha's (1997) investigation on comparison of lexical cohesive patterns in original business report in Portuguese with its respective translated report into English, translation can be considered a different text from the original. On one hand, in his study differences in the lexical cohesive patterns in TTs in relation to OTs were assessed by computer; on the other hand, a high level of proximity was also found. More specifically, the scholar points out that:

(...) as [the] results suggest, if taken at the level of the text as a whole, these apparently isolated changes end up creating appreciable differences in the way a translated text is organised lexically. Nevertheless, all of these differences were hiding a major similarity between the two versions, namely the fact that both versions had the same most central sentence. (1977:152)

With reference to the language of TTs, the use of a computerised parallel corpus allows a larger range of data and more functionality for the study of the nature of translation. Investigations done at the Centre for Translation and Intercultural Studies – CTIS, recently merged into The University of Manchester, have detected certain characteristics that would typically occur in translation (Baker 1996:180-184). Among TT recurrent features, the one more specifically related to our research study is simplification, which may be identified as a tendency to make the TT language simpler and easier to understand.

One expression of simplification is the overall type-token (tt) ratio, a measure that gives the range and diversity of vocabulary used by a writer or a translator in a given corpus. It is the ratio of different words to the overall number of words in a text or collection of texts. In simple terms, by counting the individual words (tokens) and their word forms (types), if TT less varied vocabulary is observed, it can be viewed as an easier way of processing it.

Another procedure for calculation is the standardized type/token (stnd tt) ratio, which, by not being sensible to corpus size, gives more reliable figures than raw type/token ones. The software used to obtain the present figures is Wordsmith Tools, which computes tt ratio (mean) every 1000 words as it goes through each subcorpus. In other words, the ratio is calculated for the first 1000 running words (tokens), then calculated afresh for the next 1000, and so on to the end of each subcorpus. A running average is then computed, which means that the final figure represents an average tt ratio based on consecutive 1000-word chunks of OT and TT, individually.

If data obtained by our corpus investigation seem to show frequency oscillation in TT recurring features related to more simplification, less reduction, more distanciation and less involvement, we may make the hypothesis that a significantly higher or lower variation would allow the identification of preferred choices that together would reveal a translator's use of linguistic patterns in terms of type/token relation in the Brazilian Portuguese→ English direction.

Within the framework of our investigation, the research question may be formulated as “how much of the English TT would present distinctive linguistic patterns related to type/token ratio?”

With reference to the material used in the OT subcorpus, the selection criteria favoured a work chosen from Brazilian Literature fictional texts, written by a contemporary male author enjoying a distinguished international reputation, and having several novels translated into many languages. Born in 1912 and recently died in 2001, he lived in the Northeastern region of Brazil, and his writing is normally associated with regionalism and exoticism. In the TT subcorpus, the criteria are that the respective work is translated into English, in the last two decades (in order to avoid diachronical variation), by an English native-speaking professional male translator (in the American English variant).

In this way, the corpus consists of: Jorge Amado's – *Tocaia Grande: a face obscura* (1984), Rio de Janeiro: Editora Record, 421 pages → *Showdown*, translated by Gregory Rabassa, Toronto/New York: Bantam Books, 1988, 432 pages.

At the present stage of our study, the initial corpus composition and size seem to be adequate as a workbench to carry out a pilot research project on a literary translator's style in fictional prose in the Brazilian Portuguese→English direction. The corpus meets Marchuschi's (2001:8) requirements of authentic data of written mode, accordingly selected to the research objectives, being representative and machine-readable. On the one hand, the selected OTs/TTs are not intended to allow generalizations about a stylistic pattern variability in literary translation at this stage of our investigation. On the other hand, the nature of the problem requires a specific initial corpus.

With regard to the procedures, the following steps were taken: 1st) selection of both Brazilian Portuguese OT subcorpus and English TT subcorpus, as well as digitalisation of the whole original and translated works, and saving in txt; 2nd) use of WordSmith Tools – word lists and statistics in both OT and TT subcorpora for observation of linguistic pattern distribution in terms of tt ratio and stnd tt; 3rd) TT/TO comparisons by tokens (frequency of words) and types (word forms); 4th) comparison between TT tt ratio and stnd tt with the British National Corpus – BNC tt ratio and stnd tt; and 5th) additional comparison between OT tt ratio and stnd tt with the Banco do Português – BP tt ratio and stnd tt.

2. Results and discussion

In order to analyse the type/token distribution in the OT and TT subcorpora, we extracted *Table I* below:

<i>Table I</i>	<i>Tocaia Grande</i>	<i>Showdown</i>
Tokens	159,440	141,608
Types	18,880	11,564
tt ratio	11.84	8.17
Stnd tt	53.59	48.19

By processing this comparison in order to detect if the use of vocabulary in the translator's work is more or less varied than in the original novel, some differences may be observed. In absolute terms, *Showdown* registers a lower number of tokens (141,608) and types (11,564) in relation to *Tocaia Grande* (159,440 tokens and 18,880 types). In the pair of works in question, the translator shows a type/token (tt) ratio of 8.17 while the author presents a tt ratio of 11.84, which means a TT lesser vocabulary variation than the one found in OT.

With reference to the translator's use of stylistic patterns in relation to the author's style, we may raise a hypothesis the translator has a lower tt ratio in the order of 3.67.

A further indication that there is a use of less varied vocabulary in the translated work can be found in an examination of type/token standardisation. For the translation, the stnd tt (48.19) is lower than the original work stnd tt (53.59); although the difference is small (5.4), there are actually less unique words in the TT, which means that there are more repetitions done by the translator.

However, it is necessary to have in mind that, on the one hand, the word lists and statistics generally available for this kind of study, provided by WordSmith Tools, version 3, are still rather crude ones, because the program only allows the identification of exact repetitions of words without being sensible of word types resulting from morphological variants. On the other, this software makes possible to look for lexical patterns taking into account the overall words and word forms in the whole source text and target text comprised in the corpus, what, most probably, would not be feasible to be performed to the same extent by a manual analysis only.

As the next step of this investigation was to draw a comparison between the translator's language use and the general originally written English language use by means of a reference corpus, we accessed the British National Corpus – BNC. Data extraction from this reference corpus may be seen in *Table II* below:

<i>Table II</i>	BNC	BNC Fiction
Tokens	90,748,880	19,444,150
Types	377,784	101,577
tt ratio	0.42	0.52
Stnd tt	44.04	41.54

In the BNC, there are 90,748,880 tokens and 377,784 types, and a tt ratio of 0.42.

As the BNC has a fiction (fn) subcorpus composed of 485 files consisting of samples of novels originally written in English, we also compared the data found in the BNC fn with our data. In this subcorpus, there are 19,444,150 tokens and 101,577 types, and a tt ratio of 0.52.

In this way, we may observe that the literary translator in question shows a richer and more varied language use (tt ratio: 8.17) than the BNC (tt ratio: 0.42) and even the BNC fn (tt ratio: 0.52).

Concerning the computation of type/token standardization, the translator's work shows that the stnd tt of 48.19 is higher than the BNC stnd tt of 44.04 and also the BNC fn stnd tt of 41.54. These relevant higher figures for Gregory Rabassa reveal that he uses a wider range of vocabulary in his TT than the originally written English texts contained in the reference corpus (BNC) and in its fiction subcorpus (BNC fn).

In order to have a similar parameter of comparison in the source language, the same was done between the OT and a reference corpus in Portuguese. From the Banco do Português – BP, housed at the Catholic University of São Paulo – PUC/SP, Brazil, we extracted *Table III*, below:

<i>Table III</i>	BP
Tokens	230,460,560
Types	607,392
tt ratio	0.26
Stnd tt	46.08

In the BP, there are 230,460,560 tokens and 607,392 types, and a variation of the tt ratio of 0.26. At the moment of data extraction for this investigation, the BP did not contained a fn subcorpus.

Based on these figures, we may observe the author uses more vocabulary variation (tt ratio: 11,84) than what is found, in general, in the Brazilian Portuguese originally written texts stored in the BP (tt ratio: 0.26).

With regards to type/token standardization, the author's range of variation corresponding to a stnd tt of 53.59 is higher than the BP stnd tt of 46.08. Somehow, these higher figures for Jorge Amado were already expected, as his well-known novels dealing with regionalism usually require a high demand for diversified culturally marked terms.

As I have already acknowledged the unreliability of the lists and statistics owing to the crude nature of the software I have available, I stress here we cannot rely on automatic comparisons of word forms generated by the program in the present investigation; however, the tt ratio and stnd tt do suggest a significant dissimilarity among the OT/TT subcorpora as well as respectively among the BNC and the BP reference corpus that appears to validate close examination.

In this way, we may say the access to a parallel corpus made possible to detect significantly lower frequency of overall TT stylistic patterns in relation to OT ones, which mean that, even though Gregory Rabassa shows a high tt ratio (8,17) and stnd tt (48.19), he is still far from Jorge Amado's abilities to use different word forms (tt ratio: 11,84 and stnd tt: 53.59).

By contrast, if we allow for the process of simplification, then we would not expect a higher frequency of types and tokens in a given translator in relation to the BNC. Even being restricted to one source text and one target text in the present small-scale study, the use of a computerised parallel corpus allowed to observe that Gregory Rabassa presents much higher diversified uses of linguistic

patterns and much less vocabulary repetition (tt ratio: 8,17 and stnd tt: 48.19) than what is found in the variety of text-types represented in the BNC (tt ratio: 0.42 and stnd tt: 44.04) and in the language of fictional texts originally written in English stored in the BNC fn (tt ratio: 0.52 and stnd tt: 41.54).

Nevertheless, there are other variables which are difficult to be accounted for concerning the poetics of Brazilian narrative, author's different stylistic resources, translation norms, publishing coercion, translator's choice of the type of material to translate. In the present case, a variable to be taken into consideration is related to word types in the linguistic pair involved, which is a result of the larger number of inflections in Portuguese as compared to English. Because of these interferences that may be occurring in the translator's own linguistic choices, the next step in our investigation will be the use of a moving tt ratio calculation and Bick's VISL lemmatizer in both subcorpora in an attempt to develop further observation of the translator's individual preferred linguistic marks.

Despite all possible variables, if we consider the tt ratio and stnd tt as one large-scale indication of the translator's language use, we may say, at this point of our investigation, that Rabassa's lower 3.7 tt ratio and 5.4 stnd tt differences in relation to Amado's (tt ratio: 11,84 and stnd tt: 53.59) may be pointed out as one of the manifestations of this literary translator's distinctive features, although perhaps not a feature that is unique to him, but still revealing the impact of his extent of difference in *Showdown* in relation to *Tocaia Grande*.

NOTA

* This paper comments some findings of our post-doctoral research project developed at LAEL, PUC/SP, during the period of October 2002 to March 2003, under the supervision of Dr. Antonio Paulo Berber Sardinha.

REFERENCES

- BAKER, M. Corpus Linguistics and Translation Studies: implications and applications. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (eds.). *Text and technology: in honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 1993, p. 233-250.
- _____. Corpora in Translation Studies: an overview and some suggestions for future research. *Target*, v. 7.2, 1995, p. 223-243.
- _____. Corpus-based translation studies: the challenges that lie ahead In: SOMERS, H. (ed.). *Terminology, LSP and Translation Studies in language engineering, in honour of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamins, 1996, p. 175-186.
- _____. Towards a methodology for investigating the style of a literary translator. *Target*. v. 12.2, 2000 p. 241-266.

- BERBER SARDINHA, A. P. Patterns of lexis in original and translated business reports: textual differences and similarities. In: SIMMS, K. *Translating Sensitive Texts: linguistic aspects*. Amsterdam/Atlanta: Rodopi, 1997, p. 147-153.
- _____. Ferramentas de busca e de exploração de corpora. Paper presented at the I Seminário de Estudos de Corpus, São Paulo: USP, 14-15 out. 1999.
- _____. Lingüística de Corpus: histórico e problemática. *D.E.L.T.A.* São Paulo: PUC/SP, 2000, p. 323-367.
- _____. Corpora eletrônicos na pesquisa em tradução. *Cadernos de Tradução*. Florianópolis: UFSC, in press.
- BICK, E. *VISL – Visual Interactive System Learning*. Morphological analyser. Project at the Institute of Language and Communication – ISK, University of Southern Denmark, Syddansk Universitet – SDU, Odense Campus, Denmark, since September 1996. <<http://visl.hum.sdu.dk/visl/pt/parsing/automatic/upload.php>>.
- BRITISH National Corpus: written corpus design specification. OUP Promotional Document dated 2 September 1991.
- MARCUSCHI, L. A. Um corpus lingüístico para a análise de processos na relação fala e escrita. Paper presented at Simpósio Corpus Lingüístico, 11th. InPLA, São Paulo, PUCSP, 4-5 May 2001.
- SCOTT, M. *WordSmith Tools*. Software for Windows 3.1, 95 e 98. Oxford University Press. Demo version, with limited functions, available in: <<http://www.liv.ac.uk/~ms2938>>.