

# Testing the validity of translation universals for Brazilian Portuguese by employing comparable corpora and NLP techniques

Natália Resende

## Abstract

The present study investigates the lexical features of three of the so-called translation universals, namely, *simplification*, *explicitation* and *levelling out*, for the Brazilian Portuguese language. The aim was to test their validity and to see whether the features change over time. To this end, comparable corpora were used containing 9 pairs of specimens of translated and original Brazilian Portuguese narrative prose (fiction) dating from the 19<sup>th</sup>, 20<sup>th</sup> and 21<sup>st</sup> centuries. The methodology consisted in the employment of NLP techniques and statistical measures as a complementary method of analysis. The results obtained confirmed only the validity of the *simplification* hypothesis.

## 1. Introduction

The quest for typical patterns of translated texts is not a new concern among scholars. The birth of the polysystem theory in the late 1970s revolutionized research on translation by shifting the emphasis of the investigations from a prescriptive approach to a more descriptive one. Gellerstam (1986) introduced the term *translationese*, which refers to the specific language of translations. Later on, influenced by the polysystem concepts, Toury (1995) put forward “laws” or “norms” of translations such as the law of standardization and interference. In 1996, Baker elaborated the four so-called translation universals, namely *explicitation*, *simplification*, *normalisation* and *levelling out*. According to Baker (1996), translation universals can be defined as hypotheses on some features that must be present in all translated texts regardless of the source and target languages, the type of translator, and the textual genre.

The present study aims to examine the lexical manifestation of three translation universals, namely *simplification*, *explicitation*, and *levelling out* as described by Baker (1996). *Simplification* postulates that translated texts are simpler, easier to understand; *explicitation* claims that translators tend to make statements more explicit rather than leave them implicit; *levelling out* hypothesizes that translated texts might be more similar to each other in comparison to original texts in the same language. Baker (1996) proposed the use of corpus linguistics techniques and tools to investigate these hypothetical features in order to understand what really happens during the

process of translation as well as to know the distinctive features of translated texts.

In order to test the validity of the translation universals, this study uses comparable corpora of Brazilian Portuguese composed of specimens from narrative prose (fiction) dating from the 19<sup>th</sup>, 20<sup>th</sup> and 21<sup>st</sup> centuries. Some authors have discussed the changes that occur in translations over time and relate them to the literary “norms” of each period (Basnett & Lefevere, 1990). Therefore, the purpose of this study is twofold: (i) to check whether the lexical features of the *simplification*, *explicitation* and *levelling out* hypotheses are in some way manifested in the translated texts, and (ii) to analyse whether these features change from one time period to another. To this end, three experiments were performed by employing NLP techniques as well as statistical measures for the analysis.

Even though the methodology adopted here had already been employed before to study the universals in other languages (e.g. Corpas Pastor, 2008; Corpas Pastor et al., 2008a; Corpas Pastor et al., 2008b; Laviosa, 1997; Laviosa, 1998) the present study is a novelty because: (i) it consists in the investigation of the lexical features of three universals at the same time; (ii) it is a first attempt to investigate the universals for the Brazilian Portuguese language and (iii) because it adopts a diachronic approach.

Knowing the distinctive features of translated texts can help in translators’ training (McEnery & Xiao, 2005) and, from the NLP perspective, this knowledge can be used to enhance the output of NLP applications such as Machine Translation systems (Corpas Pastor et al., 2008b). In the following sections, we will describe work that is related to the present study, the structure of the corpora used for comparison, our experiments, and provide a discussion of our results and conclusions.

## **2. Related Work**

Initial research on translation universals focused on small-sized parallel corpora and the examination of shifts between original and translated texts (see Gellerstam, 1986; ØVERÅS, 1998; Helgegren, 2005; Santos, 1995 and Santos, 1997). Later on, investigations started to use comparable corpora in order to find patterns that differentiate translated from original texts in the same language (Laviosa, 1997; Laviosa, 1998; Laviosa, 2002; Frankenberg-Garcia, 2009).

The necessity to establish a more rigorous methodological status to the investigations led to the use of large amounts of textual data, believed to be more representative of a given language, together with robust NLP techniques to search for universal patterns of translated texts (see Corpas Pastor, 2008; Corpas Pastor et al., 2008a; Corpas Pastor, 2008b; Ilisei et al., 2010; Cheong, 2006). Currently, investigations on translation universals are also adopting machine learning methods (e.g. see Ilisei et al., 2010). The findings of some of these studies (Corpas Pastor et al.; Ilisei et al. 2010; Helgegren, 2005) did validate the *simplification* hypothesis by showing that translated texts exhibit lower lexical density when compared to original texts. On the other hand, the convergence (*levelling out*) hypothesis was not confirmed by other studies (Corpas Pastor et al. 2008; Corpas Pastor 2008a). As regards the *explicitation* hypothesis, some studies (e.g. Cheong, 2006) showed that both translation contraction and translation expansion co-occur throughout translated texts even though translation expansion occurs more often. These investigations also added other hypotheses to the list of translation universals such as the *transfer* and *implication* hypotheses which postulate, respectively, that translators tend to transfer syntactic and lexical features from the source language to the target language and that translated texts tend to leave concepts from the source texts implicit rather than explicit in order to adapt the message into a different language system (Corpas Pastor et al., 2008a; Cheong, 2006).

Although translation universals have been investigated for different languages, studies on universals for the Portuguese language are still in their infancy, specifically with regard to the Brazilian Portuguese variety.

### **3. Resource Description**

The corpora we have established for the comparison are composed of two collections of texts, namely, translations from American and British English into Brazilian Portuguese and texts originally written by Brazilian authors. The two groups of texts comprise 18 prose narratives (fictional). All translations were carried out by professional translators and the original Brazilian texts were written by highly reputed authors. The texts are comparable in terms of:

- genre: all texts belong to the same genre and sub-genre. They are all literary narratives (novels), originally written in Brazilian Portuguese or translated into Brazilian Portuguese;

- size: the translations and the original texts used for comparison are roughly the same size;
- time span: the pairs of novels were published in the same period. The majority of the pairs were published in the same year for the first time; only two pairs differ by one year between the dates of their first publication, and one pair, by two years;
- type of translator: all translations were performed by professional translators;
- audience: all the texts were written for male and female intellectual adults as their target audience.

Table 1 displays the size of each text in terms of tokens.

Table 1:Corpus size in tokens

<b>Year</b>	<b>Original</b>	<b>Year</b>	<b>Translation</b>
<b>1875</b>	59,648	<b>1875</b>	61,689
<b>1880</b>	45,704	<b>1880</b>	47,321
<b>1882</b>	57,960	<b>1882</b>	53,186
<b>1943</b>	49,326	<b>1943</b>	52,010
<b>1966</b>	106,845	<b>1966</b>	109,761
<b>1984</b>	69,251	<b>1985</b>	67,370
<b>2003</b>	149,534	<b>2003</b>	151,592
<b>2005</b>	40,903	<b>2006</b>	39,317
<b>2008</b>	50,378	<b>2006</b>	45,924
<b>Total word-count</b>	629,549	<b>Total word-count</b>	628,170

#### 4. Hypotheses to be checked:

##### 4.1 Simplification

The *simplification* universal postulates that translated texts are simpler, easier to understand. They contain simplified language compared to original texts. Low lexical density (type-token ratio) is an indicative feature of this universal (Baker, 1996; Corpas Pastor, 2008; Corpas Pastor et al., 2008b). Low lexical density means a less varied vocabulary (more repetition of words). Thus, it is expected that the *simplification* universal is lexically

manifested in all translated texts in terms of low lexical density in comparison to the comparable original texts.

## 4.2 Explicitation

The *explicitation* universal can be defined as the tendency for translated texts to “spell things out rather than leave them implicit” (Baker, 1996: 180). The lexical manifestation of this universal can be verified by the “use and overuse of explanatory vocabulary” (Baker, 1996: 181). To investigate this feature, Baker suggests to compare the frequency of explanatory words using corpora of original and translated texts in the same language and same domain. Thus, we expect to encounter more explanatory vocabulary (i.e. more words such as *because*, *therefore*, *in the sense that*, *consequently* and so on) within the set of translated texts from all centuries than within the set of original texts.

## 4.3 Levelling out

*Levelling out* is defined as the tendency for translated texts to be more similar to each other in terms of lexical density (type-token ratio) and sentence length in comparison to original texts (Baker, 1996). To verify whether this universal is present in the translated corpus, we propose to examine it by comparing similarities (in these cases differences) of translated texts concerning the lexical density. We expect to encounter smaller differences within a set of translated texts in comparison to greater differences within the set of original texts in terms of lexical density. All translated texts, from all three centuries studied, should consistently present a lower lexical density variance, this being indicative of their higher homogeneity in comparison to a higher heterogeneity within the set of original texts, which corresponds to a greater variance value.

## 5. Our Experiments

In order to verify the *simplification*, *explicitation* and *levelling out* universals, this study tested the validity of the universals by comparing the lexical features of translated and original in terms of lexical density, frequency of explanatory vocabulary and lexical density variance.

### 5.1: Lexical density:

The lexical density was extracted from the corpora using the *Word List* tool from *Wordsmith Tools*<sup>1</sup> software, which provides statistical data of texts selected for analysis. The texts were split into sections and the type-token ratio was calculated for each section of each text. In order to compare the lexical density between the two collections of texts, we calculated the overall average of the individual results obtained for the type-token ratio from the sections of each group of texts, and the partial averages from each time period (Table 3). The unpaired two tailed t-test was also used to calculate the statistical differences between the translated *versus* the original texts in terms of lexical density (Table 4). The calculation was performed for each time period in order to compare the p-values obtained.

## 5.2 Frequency of explanatory vocabulary:

To compare the frequency of explanatory words, we built our own list of explanatory vocabulary of Brazilian Portuguese language (Table 2). A program was written using the *Python* programming language in order to extract the frequency of those words from both collections of texts. Then we calculated the ratio of total occurrences of explanatory words for each text. The frequency of the explanatory vocabulary was compared by calculating the overall average of the ratios for the translated and original collections and the average of the ratios for each time span.

Table 2: List of explanatory vocabulary

<b>Explanatory vocabulary in Portuguese</b>	<b>Translation</b>
<i>a razão pela qual</i>	<i>the reason why</i>
<i>assim</i>	<i>hence, thus</i>
<i>dado que</i>	<i>given that</i>
<i>de modo que,</i>	<i>given that</i>
<i>devido a</i>	<i>due to</i>
<i>já que</i>	<i>since, once</i>
<i>na medida em que</i>	<i>as</i>
<i>no sentido em que</i>	<i>in the sense that</i>
<i>pela simples razão</i>	<i>for the simple reason</i>
<i>de</i>	<i>that</i>
<i>pelo fato</i>	<i>by the fact</i>
<i>pois</i>	<i>because</i>

<sup>1</sup> [http://www.oup.com/elt/catalogue/guidance\\_articles/ws\\_form?cc=global](http://www.oup.com/elt/catalogue/guidance_articles/ws_form?cc=global)

<i>por causa de</i>	<i>because of</i>
<i>por isso</i>	<i>therefore</i>
<i>por motivo de</i>	<i>for the reason</i>
<i>porque</i>	<i>because</i>
<i>portanto</i>	<i>therefore</i>
<i>posto que</i>	<i>given that</i>
<i>sendo assim</i>	<i>hence, thus</i>
<i>sendo que</i>	<i>since, because</i>
<i>uma vez que</i>	<i>since, because</i>
<i>visto que</i>	<i>since, because</i>

### 5.3 Lexical density variance:

The homogeneity of the translated texts versus the heterogeneity of original texts in terms of lexical features was examined by measuring the lexical density variance of the two collections of texts. The results obtained were compared, whereby time span differences were taken into consideration (Table 7).

## 6. Results

### 6.1. Simplification

The investigation of the *simplification* hypothesis compared two groups of results obtained from the sections of the translated and original texts of the three centuries under concern. The assumption was that the lexical density of the translated texts would present smaller values in comparison to the values obtained from the comparable original texts, regardless of the time period.

Table 3 displays the lexical density average calculated for sections of the texts from each century. The results reveal that all translated texts present a smaller average for the type-token ratio in comparison to the average obtained from the comparable original texts, i.e., all translated texts from all centuries present a lower rating result in relation to the corresponding original texts.

Table 3 also reveals that the lexical density average of the texts from the 19<sup>th</sup> century is greater than the average of the texts from the 20<sup>th</sup> and 21<sup>st</sup> century. A greater value obtained for the type-token ratio of a given text implies that

this text contains more varied vocabulary than a text that presents a smaller type-token ratio. More varied vocabulary means more complexity to the reader. Thus, the translations from the 19<sup>th</sup> century contain more varied vocabulary in comparison to the later time spans but less varied vocabulary in comparison to the comparable original texts.

To compare the statistical differences between the translated and original texts we applied the student’s t-test. The confidence level chosen for this study was 0.05. The p-values in Table 4 reveal that there is a statistically significant difference between the translated texts versus the original texts from all centuries. However, this difference is smaller in the 19<sup>th</sup> century texts than in the 20<sup>th</sup> and 21<sup>st</sup> century texts . Thus, the p-value reveals that the translations from the 19<sup>th</sup> century are statistically more similar to the comparable original texts in terms of lexical density than the translations from the 20<sup>th</sup> and 21<sup>st</sup> centuries .

Table 3: Lexical density average of sections

<b>Century</b>	<b>Original Texts</b>	<b>Translated Texts</b>
<b>19th</b>	41.43	39.28
<b>20th</b>	44.96	33.73
<b>21st</b>	41.869	31.55

Table 4: Lexical density: Translated texts versus **original** texts

<b>Century</b>	<b>p-values</b>	<b>Result</b>
<b>19th</b>	0.02	<b>SD</b> <sup>2</sup>
<b>20th</b>	0.000000000000000129	<b>SD</b>
<b>21st</b>	0.000000000232	<b>SD</b>

## 6.2. Explicitation

Table 5 shows the normalized individual results, which confirm that all translated texts from the 19<sup>th</sup> century contain, proportionally to their length, more explanatory vocabulary than the comparable original texts. However, for the 20<sup>th</sup> century this feature was not borne out by all texts. The translated text dating from 1943 presents a smaller amount of explanatory vocabulary in relation to its comparable original text, but the other two translated texts from the same century present a larger amount of explanatory words.

---

<sup>2</sup> SD= Statistically different



Among the translated texts from the 21<sup>st</sup> century, only one of three contains proportionally more explanatory vocabulary when compared to the comparable original text. The overall average of the translated texts from the 21<sup>st</sup> century reveals a lower value than the overall average obtained for the set of original texts from the same time span (Table 5). Therefore, the hypothesis formulated as the *explicitation* universal could only be confirmed for all texts from the 19<sup>th</sup> and 20<sup>th</sup> centuries if we consider the time spans individually (Table 6). However, the overall average of the results of all texts from all centuries is greater for the collection of translated texts, i.e., the overall rate reveals a greater proportion of explanatory vocabulary within the set of translated texts (Table 5).

Table 5: Proportion of explanatory vocabulary per text

<b>Century</b>	<b>Original Text</b>	<b>Explanatory vocabulary proportion</b>	<b>Translated Text</b>	<b>Explanatory vocabulary proportion</b>
<b>19<sup>th</sup></b>	<b>1875</b>	<i>0.0045</i>	<b>1875</b>	<i>0.0092</i>
	<b>1881</b>	<i>0.0065</i>	<b>1881</b>	<i>0.0068</i>
	<b>1882</b>	<i>0.0054</i>	<b>1882</b>	<i>0.0097</i>
<b>20<sup>th</sup></b>	<b>1943</b>	<i>0.0066</i>	<b>1943</b>	<i>0.0053</i>
	<b>1966</b>	<i>0.0047</i>	<b>1966</b>	<i>0.0055</i>
	<b>1984</b>	<i>0.0035</i>	<b>1985</b>	<i>0.0051</i>
<b>21<sup>st</sup></b>	<b>2003</b>	<i>0.0074</i>	<b>2003</b>	<i>0.0052</i>
	<b>2005</b>	<i>0.0036</i>	<b>2006</b>	<i>0.0061</i>
	<b>2007</b>	<i>0.0066</i>	<b>2006</b>	<i>0.0053</i>
<b>Overall average</b>	<i>0.0054</i>		<i>0.0065</i>	

Table 6: Average of results per time span

<b>Century</b>	<b>Explanatory vocabulary proportion: Original texts</b>	<b>Explanatory vocabulary proportion: Translated texts</b>
<b>19<sup>th</sup></b>	<i>0.0054</i>	<i>0.0085</i>
<b>20<sup>th</sup></b>	<i>0.0049</i>	<i>0.0053</i>
<b>21<sup>st</sup></b>	<i>0.0058</i>	<i>0.0055</i>

### 6.3. Levelling out

The starting assumption for the validation of this universal was that the collection of translated texts will present consistently lower values of variance when compared to the variance values obtained for the original texts. Thus, the verification of this hypothesis required the application of a variance test.

Table 7 shows the results. The variance is lower within the set of translated texts dating from the 20<sup>th</sup> and 21<sup>st</sup> centuries if we compare it to the results obtained for the original texts from the same periods. However, if we compare the results of variance between translations and original texts from the 19<sup>th</sup> century, the translated texts present a greater variance. These results reveal that the translated texts dating from the 19<sup>th</sup> century are less similar to each other in terms of lexical density (vocabulary variety) than the comparable set of non-translated texts. Thus, the *levelling out* hypothesis cannot be confirmed since we did not find consistently lower variance values for the collection of translated texts regardless of the time span.

Table 7: Results for *levelling out*. Values obtained for lexical density variance

Century	Lexical density variance: Original texts	Lexical density variance: Translated texts
19 <sup>th</sup>	21.87	38.5
20 <sup>th</sup>	87.21	41.22
21 <sup>st</sup>	68.33	46.,11

## 7. Discussion of results

The results obtained with the aim to validate the starting assumptions could not be confirmed for all universals here examined. However, the results reveal that the features of translated texts change from one time period to another. While the 20<sup>th</sup> and 21<sup>st</sup> centuries present similar results, the 19<sup>th</sup> century differs from these periods in terms of lexical density average, proportion of explanatory vocabulary and lexical density variance. The results show that the 19<sup>th</sup> century possessed more varied vocabulary, more explanatory words and they are less similar to each other than the comparable original texts.

The *simplification* universal was the only universal investigated whose results revealed a predictive probability of its validity, given the consistency of the results for all the translated texts. The average of values obtained for all sections from all translated texts as well as the values obtained with the application of the t-test show, respectively, that translated texts present less varied vocabulary than original texts in the same language, and there is a statistically significant difference between the set of translated and original texts in terms of lexical density. However, the 19<sup>th</sup> century presents a greater lexical density average in comparison to the 20<sup>th</sup> and 21<sup>st</sup> centuries.

The results obtained for the investigation of the *explicitation* universal did not reveal consistency. The overall average of the results obtained from all translated texts indicates that these texts contain more explanatory vocabulary in comparison to original texts; however, the individual results do not present a greater amount of explanatory vocabulary for all translated texts in comparison to the comparable original texts. A consistently larger amount of explanatory vocabulary was found only within the set of texts from the 19<sup>th</sup> century. Hence, we argue that the manifestation of this universal can be neither validated nor discarded. Actually, the investigation of this universal deserves more attention in future work, especially on the basis of a larger amount of data.

The similarity within a set of translated texts *versus* the dissimilarity within the set of comparable original texts could not be confirmed at all. The results obtained for the heterogeneity test revealed inconsistency. This situation implies that not all sets of translated texts present lower variance for the lexical density parameter, i.e. not all translated texts under analysis are similar to each other when compared to original texts. Lower variance values in comparison with greater values obtained for the original texts could be found within the set of translated texts dating from the 20<sup>th</sup> and 21<sup>st</sup> centuries, but a higher value was found within the set of translated texts dating from the 19<sup>th</sup> century. Therefore, we can say that the universal *levelling out* could not be verified for the lexical density parameter in the corpora under analysis even though the results obtained for the translated texts dating from the 20<sup>th</sup> and 21<sup>st</sup> centuries are lower. This universal would be confirmed if the heterogeneity test had presented results consistently lower for all texts from all centuries for the translated collection in comparison to the original collection.

Regarding the variance value obtained for the texts from the 19<sup>th</sup> century, a possible explanation for this higher level of heterogeneity within the texts

from this period might be related to different strategies adopted by the translators.

The t-test revealed differences between the two groups of texts, too. The p-values obtained for the 20<sup>th</sup> and 21<sup>st</sup> centuries shows that the translated texts are statistically very different from original texts in terms of lexical features. However, the p-value obtained for the 19<sup>th</sup> century reveals a smaller statistical difference when compared to the p-values from the other two centuries. Hence, the variance test and the t-test proved that the texts from the 19<sup>th</sup> century exhibit different and peculiar lexical features when compared with the texts from the 20<sup>th</sup> and 21<sup>st</sup> centuries, due to a higher value obtained for variance and a smaller p-value obtained using the t-test. These results mean that they are statistically different from the original texts in terms of lexical features, but in a smaller proportion, and that they are less similar to each other than the translated texts from the 20<sup>th</sup> and 21<sup>st</sup> centuries. The explanation of these phenomena will probably be found in future research in which a corpus alignment would help to understand and differentiate the strategies adopted by translators from one period to another.

## **8. Conclusion**

The validity of the universals examined in the present study was confirmed only for the *simplification* hypothesis. As regards the *explicitation* and *levelling out*, in general, the results obtained do not provide enough support to confirm the validity of these universals. These findings are consistent with previous studies on translation universals. However, the corpora used in the experiments presented here are limited by the restricted number of texts and by the genre. Thus, the validity was tested but the results obtained could not provide conclusive answers to the hypotheses elaborated with respect to the manifestation of the universals in translated Brazilian Portuguese. Therefore, there are many other possibilities of investigation remaining for this language. The present study aimed to be a starting point for investigations into the translation universals for Brazilian Portuguese, but improvements as to the methodology used as well as studies using more textual genres, other source languages and other universal features remain important desiderata.

## Acknowledgements

Research for this paper was funded by a European union grant, under the *Erasmus Mundus Programme*, within the International Master in Natural Language Processing and Human Language Technology. I offer my gratitude to Ph.D. student Iustina Ilisei for invaluable assistance in bringing this study to its final form. Special thanks go to professor Jorge Baptista from the University of Algarve and professor Pieter Seuren from Max Planck Institute for Psycholinguistics in Nijmegen for having contributed to the realization of the present study.

## References

Baker, Mona. (1996), "Corpus-based Translation Studies: The Challenges that Lie Ahead" In: H. Somers (ed.) (1996), *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sage*. Amsterdam & Philadelphia, John Benjamins. 175-186.

Bassnett, Susan; Lefevere, André. (1990), *Translation, History & Culture*, Great Britain: Pinter Publishers.

Cheong, Ho-Jeong. (2006), "Target Text Contraction in English-into-Korean Translations: A Contraction of Presumed Translation Universals?" in: *Meta: Translator's Journal* vol. 51, n° 2, 343-367.

Corpas Pastor, Gloria. (2008), *Investigar con Corpus en Traducción: Los Retos de un Nuevo Paradigma*, Frankfurt am Main, Berlin & New York: Peter Lang.

Corpas Pastor, Gloria; Mitkov, Ruslan; Afzal, Naveed.; Garcia, Moya Lisette. (2008a), "Translation universals: do they exist? A corpus-based and NLP approach to convergence", in *Proceedings of the LREC'2008 Workshop on Building and Using "Comparable Corpora"*, LREC-08, Marrakesh, Marroco.

Corpas Pastor, Gloria; Mitkov, Ruslan.; Afzal, Naveed.; Pekar, Viktor. (2008b), "Translation Universals: Do they exist? A corpus-based NLP study of convergence and *simplification*" in *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas (ATMA)*, Waikiki, Hawaii, 21-25.

Frankenberg-Garcia, Ana. (2009), "Are translations longer than source texts? A corpus-based study of *explicitation*" in: BEEBY A., RODRÍGUEZ

Patrícia., SÁNCHEZ-GIJÓN, Pilar. (eds.) *Corpus use and translating*. Amsterdam & Philadelphia: John Benjamins, 47-58.

Gellerstam, Martin. (1986), "Translationese in Swedish novels translated from English" in WOLLIN L., LINDQUIST H., *Translation Studies in Scandinavia*, CWK Gleerup, Lund, 88-95.

Helgegren, Sofia. (2005), *Tracing Translation Universals and Translator Development by Word Aligning a Harry Potter Corpus*, Master's dissertation, University of Linköpings.

Ilisei, Iustina; Inkpen, Diana; Corpas Pastor, Gloria; Mitkov, Ruslan. (2010) "Identification of Translationese: A Machine Learning Approach" in: GELBUKH A. (ed.) *CICLing 2010, LNCS 6008*, Springer, Heidelberg, 503-511.

Laviosa, Sara. (1997) "Investigating *Simplification* in an English Comparable Corpus of Newspaper Articles" in: KLAUDY K., KOHN J. (eds.) (1997) *Transferre Necesse Est. Proceedings of the 2nd International Conference on current trends in Studies of Translation and Interpreting 5-7*, Budapest, Hungary, 531-540.

Laviosa, Sara. (1998) "Core patterns of lexical use in a comparable corpus of English narrative Prose" in *The Corpus-Based Approach*, Volume Special Issue of Meta, Montreal: Les Presses de L'Universite de Montreal, 557-570.

Laviosa, Sara. (2002) *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam & New York: Rodopi.

Mcenery, Anthony; Xiao, Zhonghua. (2005) "Parallel and comparable corpora: What are they up to?" in JAMES, G. (ed) *Corpus Linguistics and Translation Studies*. Clevedon: Multilingual Matters.

Øverås, Linn. (1998) "In Search of the Third Code: An Investigation of Norms in Literary Translation", Cand. Phil. Degree thesis, University of Oslo.

Santos, Diana. (1995) "On grammatical translationese", *Short papers presented at the Tenth Scandinavian Conference on Computational Linguistics*, (Helsinki, 29-30<sup>th</sup> May 1995), compiled by Kimmo Koskenniemi, 59-66.

Santos, Diana. (1997) “O tradutês na literatura infantil publicada em Portugal”, in *Actas do XIII Encontro da Associação Portuguesa de Linguística*, Vol.II (Lisboa,1-3 October 1997), 259-74.

Toury, Gideon. (1995) *Descriptive Translation Studies and Beyond*, Amsterdam/ Philadelphia: John Benjamins.