The influence of translation technologies on language production

Chapter · January 2016

1 author:

Claudio Fantinuoli
Johannes Gutenberg-Universität Mainz
19 PUBLICATIONS   **85** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Information Technology and Interpreting   View project

CHAPTER NINETEEN

THE INFLUENCE OF TRANSLATION TECHNOLOGIES ON LANGUAGE PRODUCTION

CLAUDIO FANTINUOLI[1]

1. Introduction: Aims and Motivation

The study of translated language on the basis of corpus data has become a major paradigm and methodology that has transformed the way research is done within the discipline of translation studies. Using data-driven techniques and statistics, much research in the field aims at investigating special properties of translated texts as opposed to non-translated texts on the one hand, and at shedding light on the nature of translation itself on the other.

Most of these studies have focused on the search for so-called translation universals (Baker 1996; Neumann 2011; De Sutter et al. 2012), or for the linguistic features of translated language in specific language combinations (Ondelli and Viale 2010). There are only a few studies that take into account the different translation editing environments, that is, the technological setting in which translations are produced, and how this may influence the translation product. In this respect, research in the field of translation technologies is very relevant as these tools have become ubiquitous for almost every professional translator today.

Generally speaking, we can differentiate between three main settings on an imaginary scale of technological development: pure human translation at one end, machine translation at the other, and computer-assisted translation somewhere in between. In the last decades, research has been primarily focused on the comparison of human-translated texts vs machine translation (White 1994; Popovic and Burchardt 2011), with the main purpose of evaluating machine translation. Research on texts translated under other translation editing environments has been virtually

---

[1] Johannes Gutenber Universität. Email: fantinuoli@uni-mainz.de

neglected. In particular, little research seems to have been done regarding the comparison of texts translated with and without CAT tools (Torres-Hostench et al. 2010; Lapshinova-Koltunski 2013, 2015).[2] Furthermore, almost no study has applied a corpus-based approach.

With CAT tools playing an important role in the modern translation profession (with some few exceptions, for example, in the field of literary translation), it is our hypothesis that the use of CAT tools may influence not only the translation process, that is, the way translations are produced nowadays, as pointed out by several scholars (LeBlanc 2013, Teixeira in press), but also the translation product itself in terms of its linguistic features. In particular, these features may diverge from comparable non-translated texts, as generally demonstrated by corpus-based translation studies, and also from translated texts elaborated without the use of CAT tools.

As this aspect has so far received little attention in literature, gaining more insight into the role played by the computer-mediated translation environment could help us to better understand how translation technologies influence language production in translated texts. This will have implications on the design of corpora for the study of translation and will possibly lead to improvements as applicable to the next generation of CAT tools.

The remainder of the paper is structured as follows: Section 2 gives a brief overview of related work; Section 3 introduces the methodology used for this paper and describes the corpus created for the purpose of this study; and Section 4 discusses the results. Finally, in Section 5 we present our conclusion and make some suggestions for further work.

---

[2] A CAT (Computer-Aided Translation or Computer-Assisted Translation) tool is a piece of software that generally includes at least three modules: a translation memory (TM), a terminology recognition module and an editor. CAT tools are generally seen as a means to support professional translators in delivering high quality translations in a short turn-around time. They basically split up texts in segments and compare them with the content of the TM, a database of already translated sentences. Proposing identical or similar segments to the one currently edited by the translator, they allow the reuse of past translations in similar contexts. The terminology module operates at word level and identifies terms saved in a glossary, suggesting them for integration in the target text. As Reinke (2013: 27) puts it, a CAT tool "increases the translator's productivity and helps ensure that the same terminology and expressions are used consistently throughout translations. Thus, TMs facilitate and speed-up the translation of a rapidly growing amount of specialised texts".

2. Related works

In recent years, the widespread use of translation technologies has motivated several researchers to analyze how the translation environment has been affected by the use of CAT tools and, in particular, the impact of TMs on the translation product and on the translation process itself. A first line of research has questioned the assumption that CAT tools help improve the quality of translated texts and the productivity of translators (Bowker 2005; Yamada 2011). Others have concentrated on the mental processes occurring while translating with the support of translation technologies (Christensen 2011). Recently, much interest has been devoted to ethnographic aspects, that is, to the influence of tools on the status and working conditions of translators (LeBlanc 2013; Teixeira in press). Some researchers have analyzed the advantages and disadvantages of translation technologies in a didactic perspective (Hazbavi 2011).

Despite this considerable interest in the use of CAT tools in the translation process, their impact on language production has been the focus of only a few studies, and very little is known about the way they influence the language of translated texts. Jiménez-Crespo (2009) analyzed the difference between translated texts using segment-based CAT tools and original texts produced without them. Using the *Spanish Comparable Web Corpus*, Jiménez-Crespo (2008) found that original and translated texts from the same genre show significant differences in their prototypical superstructure, and that CAT-translated texts show lower levels of lexical and typographic consistency compared to original texts produced in the target language. Mesa-Lao (2011) set up a framework for a pilot study on how computer-mediated translation environments may have an impact on the texts they produce. On the basis of this framework, the TRACE group (Torres-Hostench et al. 2010) carried out an experiment to compare translations made with and without the use of CAT tools, with the goal of analyzing the phenomena of explicitation, linguistic interference and textuality. The pilot study does not show any significant differences in explicitation, whereas a tendency to interference and an influence on textuality were partly observed in texts translated with CAT tools.

More recently, Lapshinova-Koltunski (2015) reported on an ongoing project where a comparable translation corpus is being created in order to investigate not only translation variation phenomena in terms of contrasts between languages and text types but also to assess the influence of translation methods, that is, machine, computer-aided and human translation. Her preliminary results show that lexical density and the nominal vs verbal ratio seem to be lower in CAT-translated texts. The

higher "verbality" attested in the translation mode is interpreted by the author as an indicator of "shining through" (Lapshinova-Koltunski 2013: 78), that is, the presence of typical features of the source language in the translation (Teich 2003). Several studies suggest that the use of CAT tools has led to the gradual disappearance of the text as the central textual unit on which traditionally the translation task is based (Jiménez-Crespo 2009: 217), leading to the disruption of text "linearity" (Pym 2011: 1). This forces translators to work with smaller, not always coherent textual units. Such units do not always represent all communicative signals used in a communicative interaction (Nord 1991: 16) and lead to a partial de-contextualization of the components that make up the text (Bowker 2006). Even if CAT tools do not preclude translators from accessing the co-text, that is, the surrounding sentences or paragraphs of the text, several scholars have pointed out that there is an evident predisposition to process the text at sentence level (Dragsted 2004; Shreve 2006). In the words of Pym (2011: 3), "The translating mind is thereby invited to work on one segment after the other, checking for terminological and phraseological consistency but not so easily checking, within this environment, for syntagmatic cohesion". As a consequence, translations might be less coherent and cohesive. Some scholars note (Reinke 2013), one of the main advantages of CAT tools, in particular the terminology recognition module, is to support terminology consistency throughout a text or a set of texts. This is seen as particularly important in complex projects involving multiple texts and several translators, as the use of a consistent terminology ensures good communication, particularly in the technical domains. Nevertheless, scholars have observed a high number of lexical inconsistencies in CAT-translated texts (Jiménez-Crespo 2008), even when terminology databases are used. Many inconsistencies may also depend on the impossibility of the large majority of CAT tools to retrieve subsentence matches, thus leading to the use of different lexical variants. A side effect of the use of terminology databases, which tend to let translators always use the same target, among others, term in the original language, is that typological differences between languages are not taken into proper consideration. As Scarpa (2008: 156) indicates, languages differ by the grade of tolerance in terms of lexical repetitions. In specialised English, for example, the linguistic device of repeating the same term or expression, even in the same sentence, is widely used as a means to avoid referential ambiguity, whereas other languages, for example, Italian, tend to prefer lexical variability and avoid repetitions. The use of CAT tools may therefore also influence the lexical preferences of translated texts.

While our framework is, as far as we know, novel, the specific way in which we tackle the different tasks is standard and based on both a corpus setting and a selection of linguistic variables that are most frequent in standard works on language and translation variation, as Section 3 will show.

3. Methodology

The methodology adopted in our study is related to the assumption taken from corpus-based translation studies (Baker 1996; Laviosa 1998), according to which translations have specific properties that distinguish them from original, non-translated texts. As Hansen and Teich (2001: 45) suggest,

> It is commonly assumed in translation studies that translations are specific kinds of texts that are not only different from their original source language (SL) texts, but also from comparable original texts in the same language as the target language (TL).

We extend this assumption and hypothesize that texts translated with CAT tools share specific features that distinguish them from texts translated without them.

We apply a corpus-driven approach (Tognini-Bonelli 2001) in search of features that are expected to be relevant to the CAT/NO-CAT translation variants. We compare the set of features that will be described in Section 3.1 and analyze their distribution across the translation variants of our corpus, evaluating them with statistical methods. The selected features are generally used to demonstrate the so-called "translation fingerprint" (Gellerstam 1986) or "shining through" (Teich 2003), which is meant to describe the source language interference, that is, the fact that the translation is oriented more towards the source language than to the target language, or the translation universals proposed in literature (Baker 1996), for example, simplification, explicitation and normalization. We use them as a means of detecting differences among our translation variants.

3.1. Hypothesis and Variables Selection

As already mentioned, it is our hypothesis that the use of CAT tools impacts the language of translated texts. We test this hypothesis by comparing the distribution of a set of linguistic features between two subcorpora, namely, a subcorpus containing texts translated with CAT

tools and a subcorpus containing texts translated without CAT tools. We also triangulate these measures with the frequencies obtained from a corpus of non-translated texts, which serve as a reference. In order to test our hypothesis, we selected the following set of shallow features derived from studies on language variation and translation:

1. Sentence length: Sentence length (expressed in characters or in tokens) has been used to characterize translated vs non-translated language. Some scholars have pointed out that translated texts seem to have shorter sentences than comparable non-translated texts (Laviosa 2002). This has been interpreted, for example, as an indication of the simplification of translated language (Baker 1996: 181). Even if scholars have called attention to the many factors that may play a role in the length of translated sentences, such as the typological differences between the two languages and the need to explicitate things to "help" the reader understand the text (Baker 1996; Garzone 2004; Frankenberg-Garcia 2009), sentence length has generally been used to ascertain whether or not the source text interferes with the target text production. The fact that CAT tools work on a text-segmented basis (see Section 2) may leave its traces in the length of translated language.

2. Lexical density: Lexical density (LD) is generally seen as a measure of informational load as it measures the proportion of lexical (content) vs grammatical (function) words. The use of more grammatical and fewer lexical words is one of the ways to make the text more redundant and therefore easier to process (Baker 1996: 183). LD also seems to be related to explicitness (Steiner 2012: 77). Translated texts are hypothesized to have a simplified lexicon, thus a higher ratio of grammatical words and more repetitions, whereas corpora of non-translations have a higher lexical density. The use of CAT tools, and in particular the use of the terminology recognition module, could manifest itself on the LD measure.

3. Pronoun distribution: Pronouns have often been used in literature to characterize translated vs non-translated language. For example, different studies have pointed out that pronouns are overrepresented in translated texts because of a possible effect of interference or of explicitation (Baker 1996; Ondelli and Viale 2010). Demonstrative pronouns, in particular, are a typical means of textual cohesion. As CAT tools change the translator's approach to the text, turning it from a traditionally holistic and linear one to sentence-oriented (see Section 2), it can be hypothesized that the number of demonstrative pronouns will change according to the translation environment used. In particular, CAT-translated texts could use fewer demonstrative pronouns, reducing the cohesiveness between sentences and favouring explicit lexical solutions.

4. Nominal/verbal ratio: As introduced by Biber (1988), a high proportion of lexical words can be interpreted as an indicator of richness and density of experiential meanings. As pointed out by Steiner (2012), if the two compared languages have a typological difference in the distribution of verbal/nominal classes, this value can also be used to find indications of "shining through". As for the LD, this ratio may be influenced by CAT tools, in particular by the terminology recognition module.

The selected variables are among the most frequently used in corpus-based studies characterizing linguistic variation (Laviosa 1998; Biber 1988; Ondelli and Viale 2010 and many others). Even though they are of a rather basic nature, they can help us test for signs of CAT tools influencing the target text production.

3.2. Corpus Architecture

During the past years, many comparable and parallel corpora with different language combinations have been developed, for example, the *European Parliament Proceedings Parallel Corpus* – EUROPARL (Koehn 2005), the *Cross-linguistic Corpus* – CroCo (Hansen-Schirra, Neumann and Vela 2006), the *JRC-ACQUIS Multilingual Parallel Corpus* (Steinberger et al. 2006) and the *DARPA-94 MT Evaluation Corpus* (White 1994). As far as we know, only two corpora have been compiled specifically to analyze CAT-translated language: the *VARiation in TRAnslation* – VARTRA (Lapshinova-Koltunski 2013) and the *TRaducción Asistida, Calidad y Evaluación* – TRACE (Torres-Hostench et al. 2010). As these corpora are not available to the public, we decided to compile our own composite bilingual corpus for German and Italian, an under-represented language combination in corpus-based translation studies. In our corpus, German represents the source and Italian the target language. The use of a single source language allows us exclude the influence of the source language and compare the different translation variants. The corpus structure is shown in Figure 1.
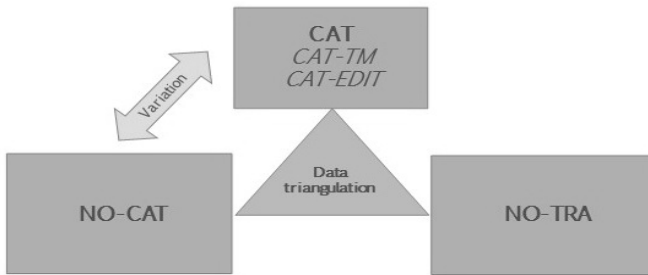
Figure 1. Corpus structure.

The corpus comprises three subcorpora: i) NO-TRA, non-translated texts originally written in Italian; ii) CAT, texts translated into Italian with the use of CAT tools; and iii) NO-CAT, texts translated into Italian without CAT tools. Furthermore, the subcorpus CAT is divided into two components: i) CAT-TM contains textual material extracted directly from the TMs of five international companies, and ii) CAT-EDIT contains texts published by the same five companies, thus translated with CAT tools, but which have undergone a revision phase outside the CAT environment.

Text source	
CAT-TM	Directly from TM
CAT-EDIT	Edited outside CAT
NO-CAT	Translated without CAT
NO-TRA	Original, non-translated texts

Table 1 Corpus components.

The distinction between the two components in the CAT subcorpus takes into account some features of the translation workflow with CAT tools. It is rather common in the translation industry that CAT-translated texts are reviewed before publication outside the CAT environment, generally by company employees in the target markets who are experts in the subject and therefore accustomed to the corporate language of their respective company. This adds a further layer of intervention on the texts. With the two components, it is possible to consider this aspect and see whether there are significant differences between the target language obtained working within the CAT environment and after a revision process, which is normally CAT-independent.

In its actual form, the corpus includes (for practical reasons) only one text type, namely "external communication".[3] In this category we summon up texts such as presentations of organizations, informative documents, press releases and promotion/advertising material in the subject domain of economics.

3.3. Collection of Corpus Data

All corpus data have been obtained by automatically processing the textual data selected for inclusion in the corpus. A description of the content of each subcorpora follows.

CAT: The CAT-TM component of the CAT subcorpus was obtained by extracting the textual elements from the TMs of five international companies operating in the economic and financial sector. Apart from cleaning the segments of any formatting tags, we retained only well-formed sentences in an effort to get rid of all instances of non-running text, including incomplete sentences caused by formatting problems (i.e., erroneous segmentations of the original texts) and segments containing terminology alone. The exact number of translators involved in the translation process is unknown but the examination of the metadata of each TM reveals that the number exceeds 30 units. This reduces the biasing effect of personal choices and translator style. The CAT-EDIT component was created by downloading texts (flyer, prospects, etc.) from the webpages of the same five companies involved in the creation of the CAT-TM component. As for the CAT-EDIT, the number of translators and editors involved is unknown, but in analogy to the CAT-TM component, it should exceed 30 units.

NO-CAT: This subcorpus contains texts translated with the use of a simple Word Editor. The subcorpus is modelled on the basis of the CAT subcorpus and mirrors its composition in terms of text type and subject. As it is currently almost impossible to find a sufficient number of texts belonging to the same text type and subject translated without the use of CAT tools, we chose to integrate a small set of texts provided by a translation agency; the texts were translated by fifteen professional translators and final-year MA students. As a consequence, the dimension of this subcorpus is smaller than that of the other subcorpora. Its dimension is comparable with that of the corpora used in similar studies based on supervised experiments (Torres-Hostench et al. 2010) and should

---

[3] For the role of text type on both source and target language, see for example De Sutter et al. (2012) and Neumann (2013).

be sufficient to draw first conclusions on possible differences between the subcorpora.

NO-TRA: The subcorpus of non-translated texts has been constructed by collecting texts from five companies; the texts were written directly in the target language of our study. The companies operate in the same sector as the five international companies that provided the texts for the CAT subcorpus. In order to create a subcorpus that can be considered comparable, the texts were selected according to criteria such as text type, length, etc.

The corpus has been enriched with basic linguistic annotations, namely part-of-speech tags and lemmas. The tagging was performed with the TreeTagger (Schmid 1997), using the standard parameter files for German and Marco Baroni's parameter file[4] for Italian, which performs better than the standard parameter file on pronouns.

	CAT-TM	CAT-EDIT	NO-CAT	NO-TRA
Tokens	1,410,719	78,022	78,022	152,798

Table 2. Corpus dimensions.

4. Results

In this section we present the results of our analysis and make some suggestions on their interpretation.

4.1. Sentence Length

The sentence length was computed on a token basis. Table 3 gives a summary of the means and medians.

	CAT-TM	CAT-EDIT	NO-CAT	NO-TRA
Mean	20.27	23.26	25.47	36.05
Median	18.36	21.48	23.88	32.52

Table 3. Sentence length.

Sentences in translated texts are shorter than sentences in non-translated texts. Among the translated variants, sentences of texts translated with CAT tools that were not externally revised (CAT-TM) are the shortest, whereas the sentences of text translated without the use of

---

[4] Available at http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

CAT tools are the longest (NO-CAT). Sentences translated with CAT tools, but revised externally, are in between (CAT-EDIT).

The results confirm the tendency of translated sentences to be shorter than the originals (Section 3.1). As pointed out in the previous section, the comparison between translated and non-translated texts can be influenced by the source language. In the three translated variants, however, the comparison cannot be biased by this variable as they all share the same source language. Therefore, the differences measured among the three variants are supposed to be related to the computer-mediated translation environment. The sentences are shorter in the texts edited only with the CAT tool, and longer if the translation occurred without tools. Interestingly, a further layer of external intervention on the CAT-translated texts (the final editing process, generally performed outside the CAT environment) reduces the gap in sentence length.

4.2. Lexical Density

According to Stubbs (1986: 17), LD is expressed as a percentage and is calculated by counting the number of lexical words and dividing them by the number of running words. We considered nouns, adjectives and verbs as lexical words (Ondelli and Viale 2010: 16). The computed LD is shown in Table 4:

	CAT-TM	CAT-EDIT	NO-CAT	NO-TRA
LD	58.79%	58.42%	58.19%	57.19%

Table 4. Lexical density.

The analysis of LD does not appear to validate the hypothesis expected according to the literature on the subject (Laviosa 1998; Salsnik 2007): content words have a similar percentage in both translated and non-translated subcorpora, being slightly higher in translated language. The difference among the CAT variants is very small but shows a tendency for a higher LD in CAT-translated texts. This could be seen as an indication of the use of more explicit cohesion instruments, such as lexical words, in the segment-oriented processing of texts in a computer-aided environment.

4.3. Demonstrative Pronouns

Table 5 shows the distribution of demonstrative pronouns in the subcorpora.

	CAT-TM	CAT-EDIT	NO-CAT	NO-TRA
Demonstrative pronouns	0.14%	0.20%	0.25%	0.29%

Table 5. Demonstrative pronouns.

In our experiment, translated language makes use of fewer demonstrative pronouns compared to non-translated language. This is contrary to the results of other studies (see Section 3.1). Interesting for the purpose of our study is the distribution of demonstrative pronouns among the different translation variants. The lower frequency is found in the CAT-TM corpus. The higher frequency in the NO-CAT corpus is closer to the non-translated subcorpus than to the CAT-translated variant. This seems to confirm our initial hypothesis of the tendency of texts translated with CAT tools to use fewer pronouns in favour of a more explicit mechanism, such as word repetition, and as an indication of the translator focusing much more on isolated sentences.

4.4. Nominal/Verbal Ratio

Table 6 shows the distribution of the nominal/verbal ratio in the subcorpora: the two CAT subcorpora have a lesser tendency to nominal style than both the non-translated texts and the texts translated without CAT tools.

	CAT-TM	CAT-EDIT	NO-CAT	NO-TRA
Nominal/verbal ratio	2.29	2.25	2.56	2.54

Table 6. Nominal/verbal ratio.

It is very interesting to note that the computed ratios can be found at the two extremes of an imaginary scale, with the two CAT variants at the lower, and the non-translated as well as the NO-CAT-translated texts at the higher, end. This result seems to indicate that the use of CAT tools has an influence on the production of translated language; among the translated variants, the ones translated with CAT tools (CAT-TM and CAT-EDIT) are indeed less similar to the original, non-translated language than the one produced without CAT tools (NO-CAT).

4. Conclusions and Suggestions for Future Work

In this paper we have presented the preliminary results of a study on translation variation between two types of translation modes, namely,

between texts translated with and without CAT tools. The linguistic features extracted from a subcorpus consisting of TMs were compared with the features extracted from a subcorpus of translations made without the use of CAT tools, and from a corpus of comparable non-translated texts that served as a reference.

One significant difference between the two modes was clearly attested by three of the four linguistic features selected for the study. A fourth feature did not present any significant difference but seems to indicate a slight tendency towards a possible impact of CAT tools on the target language. The findings therefore seem to confirm our initial hypothesis regarding the influence of the use of CAT tools on the language production of translated texts.

These results have at least two major implications:

(i) If texts translated with CAT tools have features that differ from translations done without, corpus-based translation studies should pay attention to corpus design and compilation in terms of the included translation variants as this will influence corpus-based analysis.[5]

(ii) Software companies designing and selling CAT tools should consider how the use of their tools influences target text production and design tools that have less impact on text production, allowing translators to deliver texts that are similar to translations performed without CAT tools.

Considering some obvious limitations of the corpus used in this study (the lack of control on the content of real TMs and the revision stages involved in the different subcorpora, etc.), we suggest replicating the same experiment with a more controlled treatment of variables and thus more methodological rigour, as suggested, for example, by Becher (2010) and Mesa-Lao (2011). Better control of the variables involved in the experiment can be achieved by building a comparable corpus containing the same set of texts translated in a supervised experiment with and without CAT tools. Furthermore, working with such a corpus will allow us to measure, using a quantitative as well as a qualitative approach, linguistic features that cannot be measured with the actual corpus, such as more elements of textual cohesion (lexical repetition, pronouns, etc.), considered crucial in differentiating CAT-translated from NO-CAT-translated texts. In this second phase, it would also be desirable to adopt a

---

[5] The number of corpus-based analyses conducted in a controlled experiment is still limited (Mesa-Lao 2011).

more sophisticated method of data analysis, for example, "profile-based correspondence analyses" (Plevoets et al. 2012), as this would allow "findings that are not confined to the individual behaviour of one or two linguistic variables and are therefore more broadly applicable" (De Sutter et al. 2012).

References

Baker, M. 1996. "Corpus-Based Translation Studies: The Challenges that Lie Ahead". In *LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, ed. by H. Somers, 175-186. Amsterdam: John Benjamins.

Becher, V. 2010. "Towards a More Rigorous Treatment of the Explicitation Hypothesis in Translated Studies". *Trans-kom* 3(1): 1-25.

Biber, D. 1998. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Bowker, L. 2005. "Productivity vs. Quality? A Pilot Study on the Impact of Translation Memory Systems". *Localisation Focus* 4(1): 13-20.

Bowker, L. 2006. "Translation Memory and Text". In *Lexicography, Terminology and Translation*, ed. by L. Bowker, 175-187. Ottawa: University of Ottawa Press.

Christensen, T. P. 2011. "Studies on the Mental Processes in Translation Memory-Assisted Translation - The State of the Art". *Trans-kom* 4(2): 137-160.

De Sutter, G., I. Delaere and K. Plevoets. 2012. "Is Translated Language More Standardized than Non-Translated Language? Using Profile-Based Correspondence Analysis for Measuring Linguistic Distances between Language Varieties". *Target* 24(2): 203-224.

Dragsted, B. 2004. *Segmentation in Translation and Translation Memory Systems: An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM System into the Translation Process*. København: Copenhagen Working Papers in LSP.

Frankenberg-Garcia, A. 2009. "Are Translations Longer than Source Texts? A Corpus-Based Study of Explicitation". In *Corpus Use and Translating*, ed. by A. Beeby, P. Rodriguez and P. Sanchez Gijo, 47-58. Amsterdam: John Benjamins.

Garzone, G. 2004. "Traduzione e Interferenza: Il Punto di Vista della Traduttologia". In *Lingua, Mediazione Linguistica e Interferenza*, ed. by G. Garzone and A. Cardinaletti, 105-127. Milano: Franco Angeli.

Gellerstam, M. 1986. "Translationese in Swedish Novels Translated from English". In *Translation Studies in Scandinavia*, ed. by L. Wollin and H. Lindquist, 88-95. Lund: CWK Gleerup.

Hansen, S. and E. Teich. 2001. "Multi-Layer Analysis of Translation Corpora: Methodological Issues and Practical Implications". In *Proceedings of Eurolan 2001 Workshop on Multi-Layer Corpus-Based Analysis*, 44-55. Iasi.

Hansen-Schirra, S., S. Neumann and M. Vela. 2006. "Multi-Dimensional Annotation and Alignment in an English-German Translation Corpus". In *Proceedings of the Workshop on Multi-Dimensional Markup in Natural Language Processing (nlpxml-2006)*, 35-42. Trento.

Hazbavi, A. A. 2011. "Investigating the Effect of Translation Memory on English into Persian Translation". *World Applied Sciences Journal* 15(5): 683-689.

Jiménez-Crespo, M. 2008. *El Proceso de Localización Web: Estudio Constrastivo de un Corpus Comparable de Género. Sitio Web Corporativo* (Master's Thesis). Departamento de Traducción e Interpretación, Universidad de Granada: Granada.

Jiménez-Crespo, M. A. 2009. "The Effect of Translation Memory Tools in Translated Web Texts: Evidence from a Comparative Product-Based Study". *Linguistica Antverpiensia, New Series - Themes in Translation Studies* 0(8), 213-232.

Koehn, P. 2005. "Europarl: A Parallel Corpus for Statistical Machine Translation". In *Proceedings of the Tenth Machine Translation Summit, Phuket*.

Lapshinova-Koltunski, E. 2013. "Vartra: A Comparable Corpus for Analysis of Translation Variation". In *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora*, 77-86. Sofia, Bulgaria: Association for Computational Linguistics.

Lapshinova-Koltunski, E. 2015. "Variation in Translation: Evidence from Corpora". In *New Directions in Corpus-Based Translation Studies*, ed. By C. Fantinuoli and F. Zanettin, 93-114. Berlin: Language Science Press.

Laviosa, S. 1998. "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose". *Meta: Translators' Journal* 43: 557-570.

Laviosa, S. 2002. *Corpus-Based Translation Studies. Theory, Findings, Applications*. Amsterdam: Rodopi.

LeBlanc, M. 2013. "Translators on Translation Memory (TM). Results of an Ethnographic Study in Three Translation Services and Agencies". *Translation and Interpreting* 5(2): 1-13.

Mesa-Lao, B. 2011. "Explicitation in Translation Memory-Mediated Environments. Methodological Conclusions from a Pilot Study". *The International Journal for Translation & Interpreting Research* 3(1): 44-57.

Neumann, S. 2013. *Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German*. Berlin and New York: Mouton de Gruyter.

Nord, C. 1991. *Text Analysis in Translation*. Amsterdam: Rodopi.

Ondelli, S. and M. Viale. 2010. "Evidenze Quantitative sull'Italiano Tradotto in un Corpus Giornalistico". In *Statistical Analysis of Textual data. Proceedings of 10th International Conference Journees d'Analyse Statistique des Donnees Textuelles 9-11 June 2010*, ed. by S. Bolasco, I. Chiari, L. Giuliano, 573-584. Milano: Intercultural Studies Group.

Plevoets K, I. Delaere and G. De Sutter. 2012. "Lexical Lectometry in Corpus-Based Translation Studies. Combining Profile-Based Correspondence Analysis and Logistic Regression Modelling." In *Quantitative Methods in Corpus-based Translation Studies. A Practical Guide to Descriptive Translation Research*, ed. by M. P. Oakes and J. Meng, 325-345. Amsterdam: John Benjamins.

Popovic, M. and A. Burchardt. 2011. "From Human to Automatic Error Classification for Machine Translation Output". In *15th International Conference of the European Association for Machine Translation (EAMT 11). Leuven*.

Pym, A. 2011. "What Technology Does to Translating". *The International Journal for Translation & Interpreting Research* 3(1): 1-9.

Reinke, U. 2013. "State of the Art in Translation Memory Technology". *Translation: Computation, Corpora, Cognition* 3(1): 27-48.

Salsnik, E. 2007. "Dagli Universali Traduttivi all'Italiano delle Traduzioni". In *I saperi del Tradurre*, ed. by C. Montella and G. Marchesini, 101-132. Milano: Franco Angeli.

Scarpa, F. 2008. *La Traduzione Specializzata. Un Approccio Didattico Professionale*. Milano: Hoepli.

Schmid, H. 1997. "Probabilistic Part-of-Speech Tagging Using Decision Trees". In *New Methods in Language Processing*, ed. by D. Jones and H. Somers, 154-164. London: UCL Press.

Shreve, G. 2006. "Corpus Enhancement and Computer-Assisted Localization and Translation". In *Perspectives on Localization*, ed. by L. Bowker, M. Cronin, D. Kenny and J. Pearson, 309-331. Amsterdam: John Benjamins.

Steinberger, R., B. Pouliquen, A. Widiger, C. Ignat, T. Erjavec and D. Tufi. 2006. "The Jrc-Acquis: A Multilingual Aligned Parallel Corpus with 20+ Languages". In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, 2142-2147. Bruxelles: European Language Resources Association (ELRA).

Steiner, E. 2012. "A Characterization of the Resource Based on Shallow Statistics". In *Crosslinguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*, ed. by N. Hansen-Schirra Silvia and E. Steiner, 71-90. Berlin, New York: Mouton de Gruyter.

Stubbs, M. 1986. "Lexical Density: A Technique and Some Findings". In *Talking about Text. Discourse Analysis*, ed. by M. Coulthard, 27-42. Birmingham: English Language Research.

Teich, E. 2003. *Cross-Linguistic Variation in System and Text. A Methodology for the Investigation of Translations and Comparable Texts*. New York: Mouton de Gruyter.

Teixeira, C. 2014. "The Handling of Translation Metadata in Translation Tools". In *Post-Editing of Machine Translation: Processes and Applications*, ed. by S. O'Brien, L. Balling, M. Carl, M. Simard and L. Specia, 109-125. Cambridge: Cambridge Scholars Publishing.

Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Torres-Hostench O., J.R. Gil, P.C. Leal, A.M Mor, B. Mesa-Lao, O. Orozco and P. Sanchez-Gijon. 2010. "TRACE: Measuring the Impact of CAT Tools on Translated Texts". In *Linguistic and Translation Studies in Scientific Communication*, ed. by M. L. G. Valor, I. G. Izquierdo and M. J. Esteve, 255-276. Frankfurt: Peter Lang.

White, J. 1994. "The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Further Approaches". In *Proceedings of the 1994 Conference of the Association for Machine Translation in the Americas*, 193-205.

Yamada, M. 2011. "The Effect of Translation Memory Databases on Productivity". In *Translation Research Projects 3*, ed. by A. Pym, 63-73. Tarragona: Intercultural Studies Group.