

# Testimonies of the evolution of corpus-based Translation Studies in the last decade (2003-2013)

## 1. A kaleidoscopic reality

The tenor, somehow kaleidoscopic, of the present volume is a good indicator of the evolution of corpus-based Translation Studies in their second decade of existence (2003-2013). Laviosa (2014) refers to this decade as a period in which the Descriptive approach to this discipline has been marked by the persistent interest in translation universals and by research on the features of translational language – in this case, paying special attention to the study of translators’ (mainly literary translators) style and performance –, as well as by the consolidation of the relation between corpus-based Translation and Interpreting Studies and Contrastive Linguistics. Within the Applied branch of the discipline, Laviosa remarks on the contribution of corpora in this decade to “a culture of research in education” (Laviosa 2014), the compilation of many – and new – corpora, the emergence of the first corpora in Interpreting, together with their related research initiatives and, finally, the use of corpora as a tool for quality assessment and for developing computer-assisted translation tools.

Another element that has characterized the final part of this second decade described by Laviosa (2014) is the emergence of critical voices that stress the need for greater methodological transparency in corpus-based research in general, and in translation universals in particular. An example of this stand is presented in Sutter et al. (2012), in their introduction to number 13, in *Across Languages and Cultures*. Marco (2014), for his part, provides an excellent analysis of these critical voices and outlines possible tendencies for future research on corpus-based Translation Studies, taking into account the useful reflections that these critical voices provide.

The aim of the first part of the present Introduction is to show how the different papers collected in this volume fit into Laviosa’s (2014) representation of corpus-based Translation Studies, which has consolidated itself over the last decade. A brief overview of the contributions, which are described and commented in further detail in sections 2 to 6, is provided below.

(i) As part of the descriptive branch of corpus-based Translation Studies, the present volume compiles a number of works that address indicators for research on translated language (Ruiz and Motoki, Etxeberria), studies on the style of literary translators (Novodvorski, Humblé) and contributions that analyze translators’ practices (Tonin and Castillo, Chretariu-Braescu). As for the Applied branch, it should be noted that three out of the four papers in this volume that provide didactic proposals have emerged from Interpreting Studies (Sánchez



Esta obra se publica con una licencia Creative Commons **Reconocimiento – NoComercial – SinObraDerivada (by-nc-nd)**: No se permite un uso comercial de la obra original ni la generación de obras derivadas.

and Lázaro, Santamaría and Fernández, Araújo and Correia), while Trujillo and Cabrera's contribution is the only one that focuses on the didactics of translation (tourism and commercial translation in this case). Also within the Applied branch, special attention is given to the incorporation of corpora in (meta)lexicography, focusing on the assessment of general bilingual dictionaries (Ortego and Fernández).

(ii) Most of the papers included share a common interest in providing methodological data. In fact, Novodvorski's work includes a step-by-step description of the methodology adopted for his doctoral dissertation. In this vein, we can also find the works of Vilar, Ortego and Fernández, Ruiz and Motoki, Sánchez and Lázaro, Humblé or Tonin and Castillo – the latter with a *corpus-driven* methodology applied to the history of Translation. Among the methodological aspects, Novodvorski's contribution is a clear example of how the linguist-translation scholar develops their own tagging system according to their research interest. Furthermore, within this “kaleidoscopic reality” we want to highlight the coexistence of qualitative and quantitative approaches to research with corpora: Humblé's contribution illustrates how certain quantitative analyses (namely, the *type/token* relation) can give clues about future qualitative research lines. Vilar, for her part, adopts a qualitative methodology and shows how qualitative analysis, adopting CAQDAS tools, becomes an alternative to tagging concordances (or in-corpus annotation according to Smith et al., 2008) in the case of small corpora. Finally, Tonin and Castillo show how corpora (a collection of synoptic editions **in this case**) are but one of the elements subject to qualitative analysis, while paratexts of the editions are also analyzed. The combined analysis of paratexts and the corpus helps to explain the causes and effects behind different features of translations. Tonin and Castillo's contribution proves that the combination of qualitative and quantitative methodologies proposed by Olohan (2002) has yielded results in the last decade.

(iii) Regarding resources creation/set-up, and in accordance with Laviosa's analysis (2014), this volume includes a number of papers focused on the compilation of corpora, and either describe the resources used or report on the set-up processes or the overall purposes behind their compilation. In some cases, the technical information provided by the authors on how they compiled their corpora is very inspiring and can be replicated in other research contexts (Araújo, Novodvorski). The nature of the material considered along these lines ranges from specialized *ad hoc* corpora (Arnáiz et al.), to broad, open-access corpora (Bilbao and Makazaga, Tonin and Castillo, Araújo and Correia), corpora currently being analyzed (Medina and Ramírez) or corpora still being compiled (Sánchez and Lázaro). Among the examples of compiled corpora, or corpora in the process of being compiled, there are two specific corpora on interpreting (Sánchez and Lázaro, Araújo and Correia).

(iv) *Interpreting Corpora*: Three of the contributions included in this volume deal with interpreting corpora. One of the papers focuses on conference interpreting (Araújo and Correia), while the other two pieces of work concentrate on public service interpreting (Lázaro and Sánchez, Santamaría and Fernández). The application of corpora to Interpreting Studies was first addressed by Shlesinger in her 1998 article and is now gaining momentum, as Straniero and Falbo's 2012 book shows. Lázaro and Sánchez's and Santamaría and Fernández's contributions represent an important innovation, as they shed light on a type of interpreting that has never been researched with the aid of corpora before. It is worth highlighting the interesting contribution made by Santamaría and Fernández, who comment on an IT linguistic resource which, while not a corpus in a strict sense, is based on a corpus made up of examples and can be applied to the didactics of interpreting for public services in the medical field.

(v) In her review of the last decade of corpus-based Translation Studies, Laviosa pointed to surveys of existing corpora and websites with lists of corpus resources, such as those by Federico Zanettin<sup>1</sup> or Xiao<sup>2</sup>. However, in our view this period has also seen the emergence of a line of reflection and enquiry about the corpus resources themselves, exemplified in the work of Ahmad (2008), Borja (2008), or Zanettin (2012), to name but a few. This reflective work is most probably a consequence of the proliferation of many and very different corpora – a proliferation that, in turn, has been facilitated by the technical developments that facilitate the setting up, storing, analyzing and exploiting of corpora. These reflections seem to focus on two different aspects: the classification of corpora on the one hand, and their evaluation on the other. We can find this tendency to classify and evaluate corpora resources in other sister disciplines: Such is the case of Rojo (2010) or Briz and Albelda (2009) within Corpus Linguistics applied to Hispanic Studies. In this volume, the chapters by Faya, Tomaszewicz and Bilbao and Makazaga could be categorized within this line of investigation.

(vi) Lastly, this volume includes some work on the role of corpora at the interface between Translation Studies and other disciplines. Medina and Ramírez have set up a corpus of localized websites, whose analysis might point to conclusions which could be most useful for other disciplines, such as e-commerce and marketing. As for Tomaszewicz, she includes interesting reflections on the possibility – and even necessity – of analyzing the same corpus with conceptual tools used in Translation Studies, but also in other different disciplines at the same time.

Before moving on to introduce the content of this volume in detail, we would like to briefly draw attention to the label “corpus-based Translation and Interpreting studies” (see Sánchez and Lázaro in this volume). Readers who are familiar with research in this discipline will probably agree that the label “corpus-



based *Translation Studies*” is widely accepted. Having said that, and keeping in mind the growing body of research on interpreting drawing on corpora – a line of investigation represented here by three chapters –, we consider that the label proposed by Sánchez and Lázaro gives a more complete picture of the research panorama in the discipline.

## 2. The characterization of translated language

The chapter by Daniel Ruiz and Chie Motoki, entitled “**The uniqueness of Japanese onomatopoeias in translation and non-translation**”, testifies to the productive debate around the Unique Items Hypothesis (UIH). This hypothesis was initially put forward in corpus-based Translation Studies focusing on the English/Finnish language pair (Tirkkonen-Condit, 2004). Further evidence for the UIH was gathered in a translator training context in the same language pair (Eskola, 2004) and, later, in the English/Spanish language pair (Martínez Vilinsky, 2012). Chie and Motoki are now breaking quite new ground: the study of translated Japanese on the basis of Spanish originals, using onomatopoeia as an indicator. The authors explain in detail the significant challenges that arise when working with comparable corpora of the Japanese language in general, and with the use of lexical units, such as Japanese onomatopoeia in particular. The work they present is intended as an initial preliminary study. The first conclusions lead the authors to put forward the tentative explanation of a *lack of semantic stimulus*, a different concept from the already wellknown “lack of *lexical stimulus*”. Ruiz and Motoki point to the lack of semantic stimulus in the Spanish originals as an explanation for the lower frequency of onomatopoeias in those Japanese texts translated from Spanish. Regardless of its exploratory character, this chapter deserves attention due to the original explanations proposed for the phenomenon under examination.

Similarly to Ruiz and Motoki, Etxeberria deals with the description of translated language, this time particularly of translated Basque, in the chapter entitled “**Corduroy and velvet in translated and non-translated Basque literature. Linguistic and stylistic aspects resulting from the comparison of two corpora**”. This work draws on Even-Zohar’s (1990) Polysystem Theory, with the actual research being performed on a monolingual comparable corpus of the Basque language named *Ereduzko Prosa Gaur*, which consists of translated and original Basque literary prose (see p. 66 for a detailed description of the corpus). Etxeberria aims at characterizing the translated language of a literary system that she describes as “young, peripheral and weak”. Translated texts play, thus, a crucial role in this system, “accounting nowadays for 30 to 35% of the Basque editorial output” (p. 64. our translation). To that end, Etxeberria develops a fine-tuned, original methodology that, in addition, is carefully explained to the reader. Her methodology comprises (i) isolating four groups of indicators

to analyze four factors: lexical diversity, semantic precision, some unique items of Basque and some loan vs. originally Basque word couples; (ii) investigating the frequencies of the former indicators in the subcorpora of translated and not-translated literary prose, and (iii) analyzing and discussing the results. In our view, this methodology is of a mixed nature: it is corpus-driven as she observes the data and comes to her conclusions inductively. On the other hand, on a higher level, Etxebarria's methodology could be considered *corpus-based*, as she aims at testing the hypothesis formulated following the main concepts of Polysystem Theory (for the *corpus-based* vs. *corpus-driven* opposition, see Tognini Bonelli (2001:17), Marco and Van Lawick (2009) or Saldanha (2009), among others). Etxebarria finds that translated Basque literary prose conforms more closely to the norm than original Basque literary prose. Furthermore, she points to a "clearer tendency towards precision, detail and semantic nuance" in translated Basque (p. 88, our translation), as well as to other interesting findings relating to the original texts.

Several papers discuss issues related to the study of translator style and the behavior of literary translators on the basis of corpora of varying natures and/or corpus linguistics techniques. Philippe Humblé, in his inspiring study entitled "**The translator and the immigrant. Three translations of Emine Sevgi Özdamar's *Life is a Caravasar***", applies these techniques to the research of novels written by immigrants. In these novels, that which has generally been designed as "accent" is part of the author's identity, and, as such, can be employed as a stylistic trait by the latter. Humblé puts the *type/token* ratio into use in order to study how the author's "accent" has been translated. He shows how his quantitative analysis opens up lines of enquiry for further qualitative research. Having some initial objective data established as a result of these first quantitative analyses, Humblé can move on to formulate further hypotheses, thus giving momentum to the research spiral.

Andreea-Anca Braescu-Chretariu works from quite a different, qualitative methodological perspective: using a sample corpus, she examines several translations of the same literary work in order to describe the behavior of the different translators and also to evaluate the quality of the translations. In "**Idioms in *Le rouge et le noir* Romanian translations**", Braescu-Chretariu shows how idioms are translated by the five different translators that have approached the translation of Stendhal's *Le rouge et le noir* during the 20th century. The author comes to the conclusion that the sometimes exaggerated distribution of idioms in some of the translations runs the risk of bringing the source text too close to the target culture. She also detects difficulties when translators handle the ellipsis, a typical feature of Stendhal's literary style.

The main focus of the paper by Ariel Novodvorski, "**Corpus-Based Translation Studies. Methodological aspects for analysis of style**", is not so much on



research results but on research methods. The author is concerned with the “need to define a theoretical and methodological framework for the study of the stylistic profiles of translation” (p. 118, our translation). His proposal consists precisely of a step-by-step account of the methodology that he used for his doctoral research. The parameters adopted by Novodvorski include the use of punctuation and italics, paratexts, keyness and the linguistic devices related to reported speech. Novodvorski also details the techniques and procedures employed (setting up the corpora, tagging them, etc.), which are listed on page 119. He built a parallel corpus consisting of three literary works by the Argentinean writer Ernesto Sábato and their translations into Brazilian Portuguese. What is most striking in this paper is the alignment of the procedures and tools with the research objectives, as well as the development of a special tagging system, which proves to be essential for the analysis of the “reported speech”, “punctuation” and “italics” variables.

The contribution made by Castillo and Tonin, “**Corpora for the History of Translation: *Officina Barezzi* and the synoptic edition of Beccaria's *On Crimes and Punishments* translations into Spanish**” will be of interest to researchers of the History of Translation, and especially to researchers of “the history of Cultural Mediation between Italy and Spain” (p. 140) wishing to analyze examples of how to combine corpus-driven methodology with tools and techniques normally used in the History of Translation, such as the study of paratexts. *Officina Barezzi* is a website which hosts parallel corpora, specifically the “synoptic digital edition of earliest translations of significant works for both Spanish and Italian cultures”. It is an ambitious project, as well as an example of the creation of new and free resources, in an electronic format and freely accessible, in line with the true spirit of digital humanities. This paper includes interesting considerations on important methodological decisions, such as the selection of the original texts for the synoptic editions. But one of the most important points of this paper is certainly how the authors specify and give examples of the two approaches to the comparative study of translation methods from the material existing in *Officina Barezzi*: (1) from paratext to corpus (deductive method), drawing conclusions from the reading of the paratexts and searching for examples that confirm such conclusions, or (2) from corpus to paratext (inductive method), comparing translations in order to infer the method used by each translator and confirming the hypothesis in the paratexts.

### **3. Interpreting Studies and the Applied branch of corpus-based Translation and Interpreting Studies**

In “**Corpora and Interpreter Training: an exploratory study**”, Araújo and Correia introduce Per-Fide/EPIC, a multimedia corpus of bilingual and bilateral interpreting (EN-en /PT-pt) as an extension of the Per-Fide corpus (<<http://per->



[fide.ilch.uminho.pt](http://fide.ilch.uminho.pt)), as well as the results of an exploratory analysis which gives an idea on how to take advantage of the Per-Fide/EPIC corpus for research purposes, but also, and specially for training purposes. The authors maintain that complementing the cognitive paradigm with findings from Contrastive Linguistics would be very appropriate in simultaneous interpreting teaching: using examples from the Per-Fide/EPIC corpus, they establish especially problematic fields in simultaneous interpreting in general, and between Portuguese and English, in particular, that could be focused on in simultaneous interpreting lessons (such as anaphoric relations, tense/aspect/mode, interphrastic cohesion, semantic role of arguments). Using these contrastive analyses of the interpreter's production, the student could "develop metalinguistic awareness" (p. **¡Error! Marcador no definido.**, our translation), as well as "identify peculiarities of languages" (p. **¡Error! Marcador no definido.**, our translation). The value of the interpreting testimonies included in this corpus, all of which are real and multimodal, lies not only in the fact that they are real pieces of data, but in their availability to be researched by the student as part of autonomous work.

This volume contains two papers on public service interpreting, also known as community interpreting. It is remarkable that both papers include a short discussion about (the lack of) a definition of and the best way of labelling this type of interpreting, although each paper is based on different resources, which makes it even more interesting to read both of them alongside one another. More specifically, Santamaría and Fernández state that labelling this type of interpreting as "public service interpreting" excludes this type of interpreting from contexts in which it is carried out in private organizations, such as in NGOs.

After revising the reasons for the dearth of corpus-based community interpreting studies, Santamaría and Fernández infer that the lack of knowledge and regulation of the profession is reflected in the lack of an agreement about which terminology to use to refer to this kind of interpreters. In their paper "**Applicability Of Corpus-Based Interpreting Studies To Academic Training. The Use of Universal Doctor Project In The Academic Training Of Community Interpreters**", the authors start from the premise that documentation is also essential for professionals working in the healthcare field. Then, they analyze a tool based on a corpus. They propose to use this tool in the documentation stage, as well as for the students when receiving training in this kind of interpreting. The authors make interesting proposals to improve the tool, focusing on the need to include some characteristics related to cultural aspects existing within the interaction between doctor and patient, such as the patient's "cultural background" and the patient's "level of cooperation".



#### 4. Other applied research

In “**Building a Virtual Corpus-Based Bilingual Accommodation Glossary**”, Socorro and Cabrera present an activity within a subject using a competency-based learning approach in the European Higher Education Area framework. This activity ensures an adequate response to meet an existing demand in the social environment in which learning takes place (the autonomous community of the Canaries, Spain, whose economy is based on the service sector) and, specifically, in tourist accommodation: the translation of accommodation offers. In this case, students compiled a corpus of online Spanish accommodation offers found in any type of tourist establishment. At a later stage, the corpus was used as a documentary resource in an activity about specialised tourism and commercial translation. Authors devote themselves to describing the methodology used in the activity and the difficulties that arise from drawing up their glossary (presence of technical terms, dialect words, culturemes and loan words).

The research conducted by Arnáiz, Álvarez and Corell also includes the compilation of an *ad hoc* corpus. On this occasion the purpose is not documentation in the framework of translation training, but in professional translation. In “**Audi-visual Corpora vs. Written Corpora. Translating Knowledge Pills**”, authors reflect on the parameters which should prevail when compiling a corpus to translate examples from a very recent audiovisual genre, that is, knowledge pills. Moreover, the study also includes interesting data about the background in which the audiovisual genre was born and a detailed description of it.

In “**Computing equivalents in general bilingual dictionaries: A Reality or a Utopia?**”, M<sup>a</sup> Teresa Ortego Antón and Purificación Fernández Nistal use two general Spanish corpora (*CREA* and *Corpus del Español*) and two specialised corpora (*Open Office* corpus in the *OPUS* system and *Corpus Tècnic*, from the Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, Spain) as reference corpora to confirm whether the use of equivalents gathered in two of the most widely used general bilingual dictionaries is reflected in the aforementioned corpora. In our opinion, from another approach, the methodology used by the authors could also be used to assess the resources. In this sense, the case of the *OpenOffice* corpus (*OPUS*) is particularly interesting, since researchers could not find evidence of some computing terms which could be considered to be “classical ones”. The careful description of the methodology applied in this study allows it to be developed in other contexts, especially if one takes into account the fact that the research was carried out using only online resources.

Although the study led by Karin Vilar has been placed in the section on applied research, the truth is that we understand it to be a testimony of how corpus-based Contrastive Studies can inform or complement Translation Studies. The



paper entitled “**Qualitative Text Analysis with a CAQDAS<sup>3</sup>**” aims to study the micro-functions of a particular text genre, **scholar** reviews, and its linguistic realization in German and in Spanish. The author wants to explain the “betrayed” reading expectations when a Spaniard once read a review written by a German speaker. Vilar qualitatively analyses data (two representative sets of scholar reviews, one of them written in German and the other in Spanish) in order to identify the most frequent micro-functions. Then, she compares how representative each micro-function is, the variety and the frequency of each micro-function and which linguistic devices are associated with them. The study reveals that negative criticism is a representative of German reviews but not of Spanish ones. Vilar also proves that, although micro-functions are the same in reviews written in German or in Spanish, the use of some of them are preferred in Spanish language (i.e. the micro-function “*algo no es como debe ser*”). In addition, in the German texts the distribution of micro-functions seems to be more homogenous. Finally, conclusions about the difficulty of being completely objective in the analysis are extremely interesting.

For the purpose of this introduction, we have included the research of Medina and Ramírez, “**A methodology for evaluating web localization in Spain: initial results from a corpus of websites for Andalusian agri-food SMEs**”, under the label “applied research”, as it focuses on the evaluation of quality in localization. In their research, the authors present some partial results of a wider study which includes (i) the compilation of a corpus of websites from the Andalusian SMEs in the agri-food industry, (ii) the implementation of an audit model, and (iii) the analysis of the localized product, with the aim of making some suggestions for improvement. The research highlights the importance of interdisciplinarity in Translation and Interpreting studies, and more particularly, the way corpus-based Translation Studies and marketing studies can mutually benefit from each other.

## 5. Resource creation / set-up

In practically all the notes to the chapters we have discussed so far, we have referred to the resources that have been specifically created to carry out the respective studies, be they large compilation projects or *ad-hoc* corpora. However, we will now take a further look at some chapters that focus specifically on the creation of online, generally available resources.

In the contextualization of their paper, **Araújo and Correia** present the *PerFide* project, which includes the *Per-Fide/EPIC subcorpus*, on which the authors then carry out their specific analyses. The authors provide the reader with clear and useful information about this important compilation project that supports corpus-based studies of the Portuguese language, for which not many resources



of this kind have existed until now. *Per-Fide* is a multilingual and free downloadable corpus (PT / ES / RU / IT / FR / EN / DE). This corpus comprises texts from different “knowledge areas” (literature, religion, legal-administrative, technical, journalism). Simultaneous bilingual searches are allowed, and it is also possible to search in probabilistic translation dictionaries. In the second part of their chapter, Araújo and Correia move on to describe the *Per-Fide/EPIC* subcorpus, a multimodal corpus of simultaneous interpretations between European Portuguese and British English of speeches delivered at the European Parliament. In this corpus, the transcriptions of the interpretations are aligned with the original speeches, both elements are further synchronized with the videos containing the recordings of the interpreted speeches. The authors then refer to the problems that they have encountered when designing *Per-FIDE/EPIC*. The greatest obstacle was the impossibility of including relevant metadata about the interpreters and, secondly, the largely fictive character of the orality in this kind of speech. In addition, they briefly yet comprehensively describe the compilation process. The latter details are, in our opinion, very interesting, as they render the compilation process quite transparent for the reader and can thus serve as a stimulus for those researchers who are/could be involved in the creation of similar resources.

Bilbao and Makazaga, in their paper “*EHUskaratuak: a corpus of academic translations in a minority language*” present the *EHUskaratuak* corpus from a historical and institutional perspective and outline the main objectives of their project. Firstly, the authors of the paper describe the main characteristics of this resource, which has been constructed from the translations completed by the Translation Service of the Basque Department of the University of the Basque Country (UPV/EHU). These translations, carried out by applying a specific collaborative methodology, are made up of translations of academic textbooks, popular science, prose writings and legislation. The corpus also includes a terminological database which will be published in 2014 (*Gaika*) and a service of linguistic queries and advice to promote the correct and appropriate usage of the Basque language (*Ehulku*). Secondly, Bilbao and Makazaga present other Basque-language corpora and describe them according to different parameters: their institutional authorship, their chronological dimension (present-day language corpus vs historical corpus), their linguistic dimension and their level of specialization (general corpus vs specialized corpus). When defining these corpora, the authors also explain whether they have been lemmatized and whether they include original and/or translated texts. Finally, the authors of this paper describe the *EHUskaratuak* corpus in detail and present some interesting search possibilities. As a conclusion, Bilbao and Makazaga point out that the main aim of this corpus is to include all the academic translations coordinated by the Translation Service of the Basque Department of the University of the Basque

Country (UPV/EHU) (2007-2015) and emphasize the importance and usefulness of this resource not only for researchers but also for translators.

## 6. Evaluation and classification of corpora

The research by Bilbao and Makazaga could also be included in this section, because of its methodology for classifying corpora. However, as it has already been presented in section 5 (Resource creation/set-up), we will now focus solely on the contributions made by Goretti Faya and Teresa Tomaszkievicz.

In “**Proposal of Textual Corpora Classification**”, Faya compiles the most varied terminology used to categorize corpora, and branch out the underlying concepts in a new classification based on a hierarchical tree structure. This integrative and forward-looking classification is useful when describing the kind of corpora that is being used in a particular project in depth, with the advantage that the concepts used can be traced in the relevant bibliography and are widely accepted in specialized bibliographies.

In “**The Use of Corpora in the Context of MA Seminars in Translation Studies**”, Tomaszkievicz briefly reviews the concept of **corpus** and comes to the conclusion that it is not a unified concept because it is used differently depending on the discipline (p. 361). Corpora in Translation Studies are samples of data originating in a given context, which can give rise to interesting questions for more than one discipline, even if the analysis is then carried out using linguistic or literary conceptual tools. The reflection of the author on the different ways in which a linguist and a translator scholar can use corpora is particularly interesting (p. 360), as is her consideration of the false belief that the concept of corpus is something new, issued from the modern-day research. For Tomaszkievicz, this false impression can be explained by the fact that the technological resources that are available nowadays to compile and analyse corpora are mixed up with the theoretical idea of corpus itself (p. 360). In a brief section devoted to historiography, Tomaszkievicz states that in Poland, before 1990, the topics for master’s thesis research were very limited, as were the corpora, the reason being that the compilation of corpora with the resources available and the political context in that period were very difficult. Finally, she analyses the different kinds of corpora that have been used for master’s theses at Adam Mickiewicz University in Poznań.

## 7. References

Ahmad, Khurshid. 2008. Being in Text and Text in Being: Notes on Representative Texts. @ G. Anderman and M. Rogers, eds. *Incorporating Corpora. The linguist and the translator*. Toronto: Multilingual Matters. ISBN 9781853599859, pp. 60-94.



- Borja Albi, Anabel. 2008. Corpora for Translators in Spain. The CDJ-GITRAD Corpus and the GENTT Project. @ G. Anderman and M. Rogers, eds. *Incorporating Corpora. The linguist and the translator*. Toronto: Multilingual Matters. ISBN 9781853599859, pp. 253-265.
- Eskola, Sari. 2004. Untypical frequencies in translated language. A corpus-based study on a literary corpus of translated and non-translated Finnish. @ A. Mauranen and P. Kujamäki, eds. *Translation universals: Do they exist?* Amsterdam: John Benjamins. ISBN 9789027216540, pp. 83-99.
- Even-Zohar, Itamar. 1990. Polisystem Studies. @ *Poetics Today* 11 / 1.
- Laviosa, Sara. 2014. Corpora and Holistic Cultural Translation. @ M.T. Sánchez Nieto, ed. *Corpus-based Translation and Interpreting Studies: From description to application / Estudios traductológicos basados en corpus: De la descripción a la aplicación*. Berlín: Frank & Timme Verlag für wissenschaftliche Literatur. ISBN 9783732900848, [forthcoming].
- Marco, J. 2014. Taking stock: A critical overview of research on (universal) features of translated language. @ Sánchez Nieto, M.T., ed. *Corpus-based Translation and Interpreting Studies: From description to application / Estudios traductológicos basados en corpus: De la descripción a la aplicación*. Berlín: Frank & Timme, Verlag für wissenschaftliche Literatur. ISBN 9783732900848, [forthcoming].
- Marco, Josep. and van Lawick, Heike. 2009. Using corpora and retrieval software as a source of materials for the translation classroom. @ A. Beeby, P. Rodríguez-Inés and P. Sánchez-Gijón, eds. *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. Amsterdam: John Benjamins. ISBN 9789027224262, pp. 9-28.
- Martínez Vilinsky, Barbara. 2012. On the lower frequency of occurrence of Spanish verbal periphrases in translated texts as evidence for the unique items hypothesis. @ *Across Languages and Cultures* 13 / 2: 197-210.
- Olohan, Maeve. 2002. Corpus Linguistics and Translation Studies: Interaction and Reaction. @ *Linguistica Antverpiensia* 1: 419-429.
- Saldanha, Gabriela. 2009. Principles of corpus linguistics and their application to translation studies research. @ *Tradumàtica* 7. Disponible en <<http://www.raco.cat/index.php/Tradumatica/article/view/154828/206722>>. Consultado el 20/9/2014.
- Shlesinger, Miriam. 1998. Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies. @ *Meta* 43 / 4: 486-493.
- Smith, Nicolas, Sebastian Hoffmann and Paul Rayson. 2008. Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations. @ *Literary and Linguistic Computing* 23 / 2: 163-180.

- Straniero Sergio, Francesco and Caterina Falbo, eds. 2012. *Breaking Ground in Corpus-based Interpreting Studies*, Peter Lang, Berna, etc.
- Sutter, Gert, Patrick Goethals, Thorsten Leuschner & Sonja Vandepitte. 2012. Towards methodologically more rigorous corpus-based translation studies. @ *Across Languages and Cultures* 13 / 2: 137-143.
- Tirkkonen-Condit, Sonja. 2004. Unique items—over-or under-represented in translated language? @ A. Mauranen and P. Kujamäki, eds. *Translation universals: Do they exist?* Amsterdam: John Benjamins. ISBN 9789027216540, pp. 177-184.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Ámsterdam: John Benjamins. ISBN 9789027222763.
- Zanettin, Federico. 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome. ISBN: 9781905763290.

---

<sup>1</sup><https://sites.google.com/site/federicozanettinnet/cl-htm#TOC-Translation-driven-Bilingual-and-Multilingual-Corpora>. [Last accessed 16/09/2014].

<sup>2</sup><http://www.lancaster.ac.uk/fass/projects/corpus/cbls/corpora.asp>. [Last accessed: 16/09/2014].

<sup>3</sup> Computer Assisted Qualitative Data Analysis Software.

