
Corpora and Literary Translation

By Titika Dimitroulia & Dionysis Goutsos (University of Thessaloniki & University of Athens, Greece)

Abstract & Keywords

Keywords:

Digital Humanities is increasingly becoming the centre of attention of disciplines and communities in the humanities, the social sciences and the arts, often leading to transnational, collaborative, interdisciplinary projects, which are also progressively more based on corpora (Crompton, Lane and Siemens 2017; Bernard et Bohet 2017; Longhi 2017; Klein and Gold 2016; Jockers 2013; Warwick, Terras and Nyhan 2012). Digital Literary Studies, which range across a number of varying methodologies, explore the latent hermeneutic potential of related fields, by the use of qualitative and quantitative means which privilege corpora (Earhart 2015; Ganascia 2015; Hoover, Culpeper and O'Halloran 2014; Price and Siemens 2013; Schreibman and Siemens 2008; Marchionne, online). This preference is, in a way, reminiscent of the 'prehistory' of corpora, which was rooted in concordances of religious and sacred texts rather than texts with a practical purpose (Gigot 1910; Jones 2016).

The resistance of literary translation to integrating corpora in its practice, analysis and teaching is due both to the relative absence of literary parallel corpora (Olohan 2004) and to the reluctance of researchers in both literary and literary translation studies to investigate the quantitative analysis of literary, either original or translated, texts (Porsdam 2011). Even so, a corpus-based approach constantly gains ground in literary translation studies (Zubillaga, Sanz and Uribarri 2015; Rybicki 2012; Ji 2012), after being explored in its inception by such researchers as Bernardo (1981), Baker (1996, 1999, 2000), Kenny (2000), Zanettin (2001 and 2000) and Bosseaux (2004), among others. This development is also supported by new tools, like CATMA (<http://www.catma.de>), the cultural texts annotation tool, TRADUXIO (<http://traduxio.hypertopic.org>; Goncharova and Lacour 2011), participative platform for cultural texts translators or QU.IT (<http://www.quit.unibo.it>; Zotti, 2016), a parallel database useful both to translation practice and teaching. Thus, literary bi-text interfaces and tools can lead the way to complex approaches to corpora use in literary translation, while the traditional concordances and parallel corpora continue to be of great help not only to translators, but also to researchers and translation teachers, as this issue highlights.

This is a period when textual material, and particularly literary texts, flourishes on the internet and can be used for parallel corpora compilation (Zanettin 2002). Tools also become increasingly simpler and more effective to use, ranging from OCR to alignment, concordancing, annotating and visualizing, while the new perspectives offered by interdisciplinary projects in Corpus-based Translation Studies make the use of corpora in literary translation practice, research and teaching both a choice and a necessity. Corpus-based and corpus-driven approaches (see below) to literary translation, beyond linguistic and stylistic research, are found in macro-level and/or diachronic analyses, among else, of norms, translation history and terminology management in literary or comparative studies and can reveal the relation between original and translated literary production at many levels. Furthermore, they can contribute to research on literary and cultural transfer by doing justice to its complexity and its multiple relationships to culture and society in a global context.

The global context is also shaped by the discussion on World Literature or world literatures (Apter 2006, 2013; Damrosch 2006). In this discussion, firstly, friction is created with regard to the place or even the possibility of translation in it (Apter 2013) and, secondly, the sociological translation perspective applied to macro-level has to be completed by a close analysis of texts (Casanova 2004). Parallel and comparable literary corpora, explored from different points of view, and, most importantly, including all genres of literary texts, can contribute to this discussion, dealing with the relation between language, nation, minorities and all kinds of collectivities, their linguistic and cultural universals and particularities. Research on literary translation based on or driven by big corpora is an unexplored domain, which can enrich Translation Studies in their relation to other disciplines (Ganascia, 2015).

This issue aims at highlighting the challenges of corpus-based literary translation, as well as its possible itineraries in the future, from a historical, theoretical or practical perspective. In the opening paper Gerard Lynch sets the scene by placing the analysis of literary texts through corpora in the broader line of research involving the computational and statistical analysis of written texts. He starts from a historical overview, by identifying an earlier phase that focused on the so-called translationese or translation's interlanguage, following Baker's (1993) pioneering work. The next phase includes, according to him, stylometric studies on literary translation, giving emphasis to micro-features such as those author used in author attribution or stylistic profiling, more generally. Current work belongs to the third phase of research, which involves big data and machine learning methods. Lynch discusses relevant work on literary translation in each phase in order to point out important methodological and theoretical considerations, including the need for more widely accessible machine learning platforms, the restrictions imposed by the limited availability of relevant literary material and the necessity to combine macro-level approaches with more fine-grained analyses. In his view, the interaction of literary translation projects with other research in digital humanities bears the potential of investigating "trends in stylistic variation that may transcend individual translator's choices".

The following papers of the issue offer four case studies in the analysis of literary translations. Mojca Schlamberger Brezar focuses the discussion on how parallel literary translation corpora can be exploited for a variety of purposes, by showing how corpora like FraSloK, a literary translation parallel corpus involving French

and Slovene, are placed within the larger frame of language resources available for the two languages in question. She then moves on to a minute analysis of the frequency of connectives such as the French *mais*, *cependant* etc, which have been extensively discussed in the linguistic literature under various names such as discourse, functional or pragmatic markers etc. (see e.g. Fedriani and Sansó 2017), along with their translation variants in the texts of the corpus. Her analysis suggests how quantitative considerations can shed light to an author's poetics and points out the usefulness of literary corpora, which, because of their nature, include several registers and thus offer a wider spectrum of lexical choices both to native speakers and L2 learners of a language.

In her contribution Adriana Mezeg singles out a specific construction, namely the use of past participles in sentence initial non-finite clauses, in order to identify translation strategies with reference to the same parallel corpus. What is particularly important is the combination of automatic and semi-automatic methods e.g. for the identification of the construction in question or its frequency with a qualitative functional analysis e.g. of its syntactic and semantic properties. The translation process seems to favour, in her findings, explicitation strategies both in syntactic terms and semantic relations, the latter bringing to the fore implicit relations through translation. The implications are obvious for the linguistic analysis of each language and their contrastive relations, as well as pedagogical and professional applications.

Elaine Ng's contribution is a close analysis of four Chinese translations of Hemingway's *The Old Man and the Sea*. She specifically employs Simpson's (1993) model of point of view and focuses on the expression of modality through deontic, epistemic and other modals, as found in the original and the translated texts. Her meticulous analysis bears out the translation shifts occurring in the rendering of a literary text in Chinese and allows the identification of translation strategies, such as omission or modification that may be crucial for the overall patterning of speech and thought presentation in the translated texts. Although, as pointed out, extra-textual information is needed in order to better understand the translators' choices, quantitative corpus analyses such as this can reveal recurring patterns of translation activity and stylistic choice. Despite its small sample, this is a good example of how close textual analysis can draw from quantitative, and more specifically, corpus-based methods in order to reveal what takes place in the translation process, both on the level of individual translators and that of a particular language.

Finally, Rudy Loock shifts the attention to a rather neglected type of parallel corpora, namely a learner corpus which involves translation tasks, in his case from English to French performed by advanced students of English, having French as their native language. He thus brings the discussion back to the question of translation's interlanguage or the first phase of relevant research in Lynch's taxonomy, by carefully comparing student texts with English and French corpora as regards two specific linguistic features, namely derived adverbs in *-ly* vs. *-ment* and existential *there* vs. *il y a* constructions. The findings are also correlated with the assessment of translation quality by independent evaluators. This exploratory and thus highly original type of research is valuable not so much for its definitive results, but rather for the methodological and theoretical issues it raises, including the question of the correlation between intra-language differences as a whole and translation quality.

This special issue on the use of corpora in the study of literary translation concludes with a paper by Federico Zanettin, which aptly wraps up the issues involved. In particular, Zanettin places the corpus analysis of literary translation within the tradition of computer-assisted literary studies, pointing out the difference between qualitatively- and quantitatively-driven or stylometric approaches, in a way reminiscent of Tognini-Bonelli's (2001) famous distinction between corpus-based and corpus-driven analysis. His extensive overview of corpus studies that focus on literary texts indicates the spectrum of methods employed, the range of linguistic features examined and the diverse findings produced. His paper draws the implications of these studies for exploring literary translators' style by means of corpora and the question of translation universals (see Baker 1993), as well as translation criticism.

References

- Baker, Mona (1993) "Corpus Linguistics and Translation Studies — Implications and Applications" in *Text and Technology: In Honour of John Sinclair*, Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds), Amsterdam/Philadelphia, Benjamins, 233-50.
- , (1996) "Corpus-based Translation Studies: The Challenges that Lie Ahead" in *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, Harold Somers (ed.), Amsterdam/Philadelphia: Benjamins, 567-86.
- , (1999) "The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators", *International Journal of Corpus Linguistics*, 4(2): 281-98.
- , (2000) "Towards a Methodology for Investigating the Style of a Literary Translator", *Target* 12 (2), 241-66.
- Bernard, Michel et Bohet, Baptiste (2017) *Littérométrie. Outils numériques pour l'analyse des textes littéraires*, Paris, Presses Sorbonne Nouvelle.
- Bernardo, Aldo (1981) "Maximizing Computer Assistance in Literary Translation: Petrarch's Familiars" in *Translation Spectrum: Essays in Theory and Practice*, Marilyn Gaddis Rose (ed.), New York, State University of New York Press, 74-80.
- Bosseaux, Charlotte (2004) "Point of View in Translation: A Corpus-based Study of French Translations of Virginia Woolf's *To the Lighthouse*", *Across Languages and Cultures*, 5(1): 107-22.
- Crompton, Constance, Lane, Richard J., Siemens, Ray (2017) *Doing Digital Humanities. Practice, Training, Research*, London/New York, Routledge.
- Earhart, Amy E. (2015) *Traces of the Old, Uses of the New The Emergence of Digital Literary Studies*. Michigan, University of Michigan Press.
- Fedriani, Chiara & Sansó, Andrea (eds) (2017) *Pragmatics Markers, Discourse Markers and Modal Particles: New Perspectives*, Amsterdam/Philadelphia, Benjamins.
- Ganascia Jean-Gabriel. (2015) "The Logic of the Big Data Turn in Digital Literary Studies", *Frontiers / Digit. Humanities*, 2 (7), online. DOI: 10.3389/fdigh.2015.00007
- Gigot, Francis E. (1910) "Hugh of St. Cher" *Catholic Encyclopedia*. New York, Robert Appleton.

- Goncharova, Yuliya and Lacour, Philippe (2011) "TraduXio : Nouvelle expérience en traduction littéraire", *Traduire* online, 225 | 2011. DOI : 10.4000/traduire.94
- Ji, Meng (2012) "Hypothesis Testing in Corpus-based Literary Translation Studies". In *Quantitative Methods in Corpus-based Translation Studies*, Michael Oakes and Meng Ji (eds), Amsterdam/Philadelphia, Benjamins, 53–72.
- Jockers, Matthew L. (2013) *Macroanalysis: Digital Methods and Literary History*, Urbana-Champaign, IL, University of Illinois Press.
- Jones, Steven E. (2016) *Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*, London/New York, Routledge.
- Kenny, Dorothy (2000) *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester, St. Jerome Publishing.
- Klein, Lauren F. and Gold, Matthew K. (2016) "Digital Humanities: The Expanded Field", *Debates in the Digital Humanities*, <http://dhdebates.gc.cuny.edu/debates/2>
- Longhi, Julien (2017) "Humanités, numérique : des corpus au sens, du sens aux corpus" *Questions de communication* 31 (1) : 7-17. <https://www.cairn.info/revue-questions-de-communication-2017-1-page-7.htm>.
- Marchionne, Ilaria (online) "I quattro livelli della rivoluzione digitale in ambito letterario. Capitolo uno: i corpora linguistici", Digital Writing Lab, Gruppo di ricerca del Communication Strategies Lab, <http://www.csl.unifi.it/dwl/riflessi/i-quattro-livelli-della-rivoluzione-digitale-in-ambito-letterario-capitolo-uno-i-corpora-linguistici/>
- Olohan, Maeve (2004) *Introducing Corpora to Translation Studies*, London/New York, Routledge.
- Porsdam, Helle (2011) "Too much 'digital', too little 'humanities'? An attempt to explain why many humanities scholars are reluctant converts to digital humanities". In *Arcadia Project*, Cambridge, MA, Cambridge University Laboratory, University of Cambridge, <http://arcadiaproject.lib.cam.ac.uk/docs/DigitalHumanities.pdf>
- Price, Kenneth M. and Siemens, Ray (2013) *Literary Studies in the Digital Age. An Evolving Anthology*, New York, NY, MLA Commons, Modern Language Association, <https://dlsanthology.mla.hcommons.org/>
- Rybicki, Jan (2012) "The Great Mystery of the (almost) Invisible Translator. Stylometry in Translation" in *Quantitative Methods in Corpus-Based Translation Studies*, Michael Oakes and Meng Ji (eds), Amsterdam/Philadelphia: Benjamins, 231-48.
- Schreibman, Susan & Siemens, Ray (eds) (2008) *A Companion to Digital Literary Studies*. Oxford, Blackwell, <http://www.digitalhumanities.org/companionDLS/>
- , and Unsworth, John (eds) (2004) *A Companion to Digital Humanities*, Oxford, Blackwell, <http://www.digitalhumanities.org/companion>
- Simpson, Paul (1993) *Language, Ideology and Point of View*, London/New York, Routledge.
- Tognini-Bonelli, Elena (2001) *Corpus Linguistics at Work*, Amsterdam/Philadelphia, Benjamins.
- Warwick, Claire, Terras, Melissa and Nyhan, Julianne (eds) (2012) *Digital Humanities in Practice*, London, Facet.
- Zanettin, Federico (2000) "Parallel corpora in translation studies. Issues in corpus design and analysis" in *Intercultural Faultlines*, Maeve Olohan (ed.), Manchester, St Jerome, 105–18.
- Zanettin, Federico (2001) *IperGrimus*, inTRAlinea Monographs, <http://www.intralinea.org/monographs/zanettin>
- Zotti, Valeria (2016) "QU.IT une plateforme électronique d'aide au travail des traducteurs littéraires. La Francesistica italiana à l'ère du numérique", *Publiforum*, 25, http://publiforum.farum.it/ezine_articles.php?id=337.
- Zubillaga, Naroa, Sanz, Zuriñe and Uribarri, Ibon (2015) "Building a trilingual parallel corpus to analyse literary translations from German into Basque" in *New Directions in Corpus-based Translation Studies*, Claudio Fantinuoli and Federico Zanettin (eds.), Berlin, Language Science Press, 71–93.

©inTRAlinea & Titika Dimitroulia & Dionysis Goutsos (2017).

"Corpora and Literary Translation", inTRAlinea Special Issue: Corpora and Literary Translation.

Stable URL: <http://www.intralinea.org/archive/article/2260>