Journal of Translation Studies 2(1) (2018, New Series), 1–4

Introduction: Translation Studies and the Digital

James St. André
The Chinese University of Hong Kong

This special issue of the *Journal of Translation Studies* is devoted to the intersection of digital humanities and translation studies.

The digital and translation have a long history. Some of the earliest experiments using computers to process language digitally were aimed at achieving machine translation, most famously in the partnership between Georgetown and IBM in the 1950s and 1960s (Gordin 2015, 213–217). Although those efforts were ultimately unsuccessful, the dream of digital translation never died, and translators were quick to embrace the advent of personal computers as a way of transforming how they practiced translation, with specialized tools to aid translators evolving in sophisticated but also unexpected directions.

Yet the study of translation was slow to adopt digital methods. In the 1990s, Mona Baker and Maeve Olohan pioneered the use of electronic corpora in translation studies, despite doubts by linguists concerning the value of using texts that might contain interference from other languages, either because the translator was not a native speaker or because, in working from a text in another language, the native speaker might consciously or unconsciously be influenced by the structure and vocabulary of the source text (Olohan 2004, 13–22). Baker and Olohan's work concentrated on universals or norms; in other words, they were concerned with how processing a large amount of data and generating statistics about usage could reveal general patterns in the use of language by translators.

Meanwhile, roughly contemporary to the first attempts to compile and use corpora in translation studies, humanities computing, as it was then known, began to explore the digitization and visualization of data in the humanities. Jerome McGann's *Radiant Textuality* (2001) describes the building of the Rosetti Archive at the University of Virginia in the 1990s and the various reasons it was such an exciting project. Since then, the

Published by The Chinese University Press, ISSN 1027-7978.

field has expanded exponentially, with conferences, edited volumes, and journals dedicated to the field.

Early projects were labor-intensive, slow, and expensive. The exponential growth of computing power and the concomitant decrease in price helped to bring down expenses. The formation of consortia to share the financial burden, the advent of crowdsourcing, and partnerships with private corporations helped with labor costs, although in the last case sometimes resulted in material being placed behind paywalls that could be prohibitive for individuals or smaller institutions.

Two of the articles in this issue deal with the gathering together of data from existing sources to compile corpora, traditionally a bottleneck both in terms of time and funding. Nason Anran Cao explores a variety of databases that contain materials from the late Qing through the early Republican Era in China (roughly 1870–1930), which was a period of intense translation activity and great social and cultural foment. In particular, Cao zeroes in on the pitfalls of OCR (Optical Character Recognition) software, an area where Chinese has lagged behind English and which has meant that many Chinese corpora are still typed manually. James St. André deals with mining existing databases in English for texts relating to translation in order to build both translation corpora and comparable corpora of pseudotranslations and texts originally composed in English. Both articles point to the growing number of resources, but the need for care when using these databases that were originally created for different purposes, in particular, the percentage of error in OCR conversion of texts may call into question statistical results.

The article coauthored by Stacey Triplette, Elisa Beshero-Bondar, and Helena Bermúdez Sabel, starts from traditional corpora compilation but then seeks to use new techniques to investigate Robert Southey's 1806 translation *Amadis of Gaul*. This is made possible by the use of TEI encoding to "stitch together" the source and target texts at the clause level. One surprising result is that they demonstrate how Southey, a Romantic poet and so therefore usually associated with emotional language, actually dampened down emotional passages in his translation. Thus the use of new digital approaches may help us to revisit previously held assumptions about texts and their translations.

The field of digital humanities encompasses much more than just corpus studies as developed from linguistics in the 1990s, however. Perhaps most prominent are attempts at data visualization. In particular, the advent of big data means that new ways of reading texts and

interpreting data need to be found. Moretti (2013) has championed "distant reading" of texts, in contradistinction to the traditional "close reading" so popular since the New Critics. Aggregating large amounts of data, Moretti and his allies have analyzed literary texts for sometimes surprising features, such as the frequency of some common words, and used network analysis to analyze plot. But the idea of distant reading has drawn criticism for abandoning what makes the humanities human, that is, qualitative analysis of texts.

This brings us to David D. Kim's essay in this special issue. Kim marries distant reading techniques with more traditional close reading of important texts. He uses the vast amount of secondary literature stored in JSTOR, as well as bibliographic records from WorldCat, to visualize the dissemination of Goethe's works worldwide and to map out connections between world literature, postcolonial studies, and orientalism. But he also provides close and sympathetic reading of key texts, and key passages, in a sense triangulating between qualitative and quantitative techniques, showing how Goethe's formulation of world literature was a "subversive and aspirational imaginary in opposition to the increasingly dominant Hegelian world vision." This means that the relationship between Goethe's original vision of world literature and the field as it has emerged today is not a simple or linear one, but rather a complex story that demands careful unpacking through the use of various conceptual and mechanical tools.

Michelle Jia Ye also uses a network visualization tool, Gephi, in her research on Republican era (1911–1949) translations appearing in the journal *New Youth*. Her research concentrates on early issues of the journal, tracing the development of intertextual networks in both "recognized" translations (the translation of full or abbreviated texts, explicitly marked as translations) and unmarked translations (the use of translated material in fragmentary quotations, summary, citations, or notes). Building up a visual map of these references in a network allows the translations to be contextualized in a variety of different ways. One result is that a rupture between early and later issues of the journal is revealed, challenging the long-held assumption that translation in *New Youth* had a consistent and strategic bias toward modernist texts.

Finally, Lorenzo Andolfatto uses a different sort of visualization tool: word-cloud generators. These tools give a simple graphic representation of word frequency in texts; Andolfatto uses this property to explore the layering of different English translations of the same source text (a

poem by Li Shangyin 813–858). This allows him to highlight translation idiosyncrasies, reveal convergences and divergences in interpretation, and point at potentially problematic junctures in the source text, thereby implementing a "thick translation" approach that also combines close and distant reading.

We hope that the research presented in this special issue, along with the book reviews of some recent publications in translation studies that focus on corpus-based research, will stimulate other researchers to adopt digital humanities approaches in their future projects.

References

Gordin, Michael D. (2015). *Scientific Babel: How Science Was Done Before and After Global English*. Chicago: University of Chicago Press.

McGann, Jerome (2001). *Radiant Textuality: Literary Studies after the World Wide Web*. New York: Palgrave Macmillan.

Moretti, Franco (2013). *Distant Reading*. London: Verso.

Olohan, Maeve (2004). *Introducing Corpora in Translation Studies*. London: Routledge.