Big translation history

Data science applied to translated literature in the Spanish-speaking world, 1898–1945

Diana Roig-Sanz[1,2] & Laura Fólica[1]
[1] Universitat Oberta de Catalunya | [2] ICREA, Institució Catalana de Recerca i Estudis Avançats

This article proposes the term Big Translation History (BTH) to describe a translation history that can be analysed computationally and that we define as involving: (1) large-scale research (geographical and chronological); (2) massive data understood as big data, accompanied by little data, and drawing on a wide range of often heterogeneous and non-structured sources; and (3) the use of computational techniques as part of the research process, and for the production of knowledge, rather than helping only with visualisation of data. We advance the hypothesis that one of the main possibilities of BTH, as a conceptual framework and a methodology, is to help decentralize translation history and literary and cultural history, in a broad sense. The article goes on to present an analysis of the circulation of literary translations and the agents involved in the Spanish-speaking world between 1898 and 1945 as a case study in BTH.

Keywords: big translation history, literary translation, sociology of translation, digital humanities, social network research, circulation, publishing history, Spain, Spanish-speaking Latin America

1.   Introduction

In the last two decades, computational tools have been used in the humanities and the social sciences to study patterns of cultural change, both in the present and in the past, in a growing and interdisciplinary field (Jockers 2013; Borgman 2015). While the main goal has been to measure culture in an innovative way (Bail 2014), it has evolved differently among the wide range of disciplines that study the human condition (Aiden and Michel 2014) and at various university departments all over the world. At the same time, we are only starting to become fully aware of

the potentials and pitfalls of using machine learning and artificial intelligence in the humanities and social sciences.

In the field of translation studies, digital research in the analysis of translation history remains scarce, and we are only beginning to understand how digital tools and quantification can be applied to the study of translated literature and literary translation flows. True, statistical and quantitative approaches have already been advanced in the field, with Heilbron (1999) Heilbron and Sapiro (2002), for example, employing a quantitative approach to analyze a world system of translation that has been propelled or constrained by the dynamics of global publishing and entrenched hierarchies between languages, which range from the hyper-central, as in the case of English, to the semi-peripheral and peripheral languages. Pym (1998, 2001) also proposed using statistics and discussed the notion of cultural sovereignty, as well as alternative ways of understanding the always-problematic idea of the border. Furthermore, Pym renewed translation history by focusing on cultural systems and on the people (translators) who embodied them, as did Chesterman (2009), who proposed "translator studies" as a discipline. Taking a cultural transfer perspective, d'Hulst (2012) and Meylaerts et al. (2018) have likewise put forward the role of cultural mediators, as well as international and intra-cultural networks. All these works have been relevant to translation studies and to the sociology of translation, and have applied both quantitative and qualitative perspectives (Pegenaute 2018), but they do not yet use machine learning and artificial intelligence, and they leave ample room for further discussion of the different ways of quantifying histories of translation – ways that would allow us to move between different scales (local/national/regional/global), between geographical spaces and through time, and from small to large datasets. At this juncture, we need to consider which aspects of translation history can be measured and with which new tools, data sources and methods. What does it mean to represent translation history with data (Manovich 2015; Schäfer and van Es 2017)? What are the opportunities and pitfalls of computationally analyzing translation flows and large cultural datasets? How can we combine quantification, the statistical study of literary translations in an historical period, and data visualization on a large scale with qualitative methods, including textual analysis of selected translations? And how should we combine the analysis of larger patterns with the analysis of agents of translation (Simeoni 1995; Milton and Bandia 2009; Chesterman 2009), or cultural mediators (Meylaerts, Gonne, Lobes, Roig-Sanz 2016; Roig-Sanz and Meylaerts 2018). At the intersection of translation and literary studies, publishing history, media studies and computational technologies, this article encourages reflection on the theoretical and methodological basis of translation

historiography, proposing the term Big Translation History (BTH from now on),[1] which we define as a translation history that can be analyzed computationally. In our view, BTH represents a breakthrough in the sense that it challenges previous research on the circulation of translations, by exploiting the metadata of large datasets, the increasing digitization of texts and the expansion of archives (Algee-Hewitt et al. 2016) that allow us to engage in large-scale analysis, both geographically and over time. The comparison of datasets and transnational translation flows on a scale that was previously unimaginable also reveals both the limits of BTH and the strengths of rewriting translation history through digital archives and metadata. A general hypothesis is that one of the main possibilities offered by BTH is to help decentralize translation history and literary and cultural history, in a broad sense, by breaking with national historiographies in translation. This might be particularly significant for researchers working on periods in which borders have changed, those dealing with translated literature in the diaspora, and those working on translations of regional literatures (such as Flemish) and non-state literatures (as with Catalan), for example. Macro-analyses of national translation programs might also benefit from such a decentralizing approach. BTH can also help us identify centres of production, the transit of translations, and connectivity between actors. BTH is therefore not only a methodological tool to show patterns of translation, the centrality (or not) of specific translators or translated authors (women among them), or how translations are organized in different ways, but it can also contribute to conceptualization in global approaches to literature and translation history.

As we understand it, BTH can be grounded on three fundamentals: (1) large-scale research (geographical and chronological); (2) massive data, understood using a two-pronged approach involving both big data and little data, and drawing on a wide range of often heterogeneous and non-structured sources; and (3) the use of computational techniques as part of the research process and for the production of knowledge, rather than helping only with visualization (Drucker 2014). We understand "little data" as involving both "small numbers" (Terian 2019) and data that is heterogeneous in the sense that it comes from different sources and has not been homogenized and structured.[2] The application of any or all

---

1. This term was first presented at the EST Congress 2019 in Stellenbosch, South Africa, during the panel organized by Diana Roig-Sanz, Laura Fólica and Ondrej Vimr and entitled "Big Translation History and the Use of Data Mining and Big Data Approaches."

2. Terian (2019) refers to "small numbers" for specific datasets in national literary archives (Romanian in this case). Borgman's (2015) reflections on how the definition of big data or "little data" depends on the discipline and field are also of relevance here. By establishing a parallelism with big and little science, Borgman stresses how some disciplines and fields have more homogeneous data and also more solid and international research infrastructures, whereas other

three of the proposed fundamentals ("large-scale research", "massive data" and "computational techniques"), can elucidate, for example, intercultural relations that have been overshadowed or remain unknown (indicators of the emergence, decline, or lack of intercultural relations) and patterns of cultural change in the circulation of translations, and provide visual tools for the processing of relevant contents.

With a relational and agent-driven approach, our research focuses on the possibilities that quantitative analysis opens in the study of translated literature (whether it circulated or not), and the roles played by translators, publishers, critics, literary agents, institutes for the promotion of national literatures, and book fairs. While we anticipate complementing this research with qualitative microtextual analysis, in this article we elaborate on the macro-textual (distant reading) and contextual approach as it can help in the selection of relevant case studies based on both the most widely translated books and the most prolific translators.

In the next sections, we will present our definition of BTH as a conceptual framework and exemplify its usefulness with a case study involving the circulation of literary translations in the Spanish-speaking world, that is, Spain and Spanish-speaking Latin America, between 1898 and 1945.

2.   Big translation history as a conceptual framework

There is still much controversy regarding the definition of big data and how 'big' big data is (De Mauro et al. 2015; Fisher et al. 2012; Manovich 2020),[3] but there is also much consensus regarding the fact that our understanding of big data can vary from field to field and that having "the right data" is more important than the size of the data. Indeed, as Borgman (2015, 6) notes, "data are big or little in terms of what can be done with them, what insights they can reveal, and the scale of analysis required relative to the phenomenon of interest." Thus, BTH as a conceptual framework draws on big data and little data, distant reading, cultural analytics, machine learning, data mining, and artificial intelligence. It is not our purpose to discuss all these perspectives, but to show how they can be applied to translation history and used as exploratory tools that can help us to generate knowledge.

---

fields (for example, the humanities) are characterized by a larger amount of heterogeneous, less structured data, which means that researchers in the humanities need to agree on the boundaries between "big" and "little" according to their practices and objects of study.

3. According to Fisher et al. (2012) big data is so large that cannot fit on a single hard drive. In the case study presented below, we account for 258,675 books in our database, which is already massive data for our field and time period.

Distant reading, for its part, is associated with Moretti (1998, 2005, 2013), who pioneered the approach in the field of literary studies. As is well-known, Moretti attempted to analyze patterns in massive datasets of novels, as well as to study the development of the genre. He also gave voice to the thousands of books that were published in the nineteenth century, and not just the canon. Although controversial because of his somewhat simplistic results (Da 2019), the work done by Moretti and colleagues at the Stanford Literary Lab has catalyzed a revolution in literary studies and beyond.

Manovich (2015, 1) defines cultural analytics as "the analysis of massive cultural data sets and flows using computational and visualization techniques." Manovich (2015, 2020) pays specific attention to visual media and distinguishes cultural analytics from digital humanities, claiming with Champion (2017) that the latter is text heavy, visualization light, and simulation poor. In our research, we see cultural analytics as nourishing BTH, since translation is also a cultural practice and we also use machine learning and visualizations to explore our data, even though our main concern is with texts and not images. The processing of texts and images encourages the uncovering of objects that were hitherto unavailable to human readers because of their size, scale, or wide geographical and chronological spectrums. Moretti's distant-reading approach also connects with scholarship related to the analysis of social networks such as Twitter and Instagram, the study of book ratings, as with Goodreads, consumers' cultural interests and likes (Blank 2007), and methods that have been used to quantify literary prestige (Verboord 2003; Verboord, Kuipers and Jansenn 2015; Porter 2018), highlighting the intersection of multiple methodologies and fields.

Tim Hitchcock (2013) proposed the term Big History in the context of studying history computationally (see also Graham et al. 2015; Liu 2018). This notion has inspired our proposal of BTH as a new research avenue that can apply the main principles of Moretti's, Manovich's and Hitchcock's work, all the while remaining aware of the potential and pitfalls of the approach. In this respect, it behoves us to think critically about our methodology and to ask, for example: How should we collect data? How do we gauge how representative it is? How can we reflect the way the data was gathered? How collaborative is BTH? BTH projects already belong to a new generation of researchers who are aware of the dangers of reinforcing previous ideas by using only readily available sources. The Empires were very good at creating data, but we remain ignorant of much data that is not readily available, and of voices that have not yet been heard. In this regard, BTH also connects with the trans-cultural perspective of entangled history, as well as adjacent approaches such as *histoire croisée*, global history, cultural transfer, or shared history from a digital and relational perspective.

Indeed, the digital turn has heightened interest in several disciplines that could benefit from dealing with large-scale data and a transnational approach, but these distance-reading perspectives have not evolved in the same way across all geographical and scholarly contexts. Likewise, the cultural datasets that researchers have at their disposal, and their skills and levels of awareness regarding the Eurocentric or Anglo-American points of view inherent in these datasets can vary. Data infrastructure and policies for open data sharing can also differ. Finally, researchers have tended to prioritize the literature of specific periods (nineteenth-century or medieval texts), specific genres (the novel), and datasets from the Anglo-American, French, and, to a lesser extent, German literary and cultural worlds. In what follows, we give examples of the use of data-driven approaches in a wide range of projects and disciplines in the humanities and the social sciences to better frame similar endeavours in the translation history field.

Anglo-American institutions such as the Stanford Literary Lab and the Cultures of Knowledge project in Oxford paved the way in digital literary history. In recent years, Underwood (2017), Piper (2018), and So and Long (2013) have also contributed to the cultural analytics turn in the study of literature and culture. Quantification has also been very relevant to digital history (Cohen and Rosenzweig 2005) and the study of memory in the digital age (Shandler 2017), as well to the development of a social-network approach to intellectual history (Grandjean 2018) or the recognition of faces in historical photographs (Düring et al. 2021). The possibilities of a digital history of periodicals has also drawn the attention of Schelstraete and van Remoortel (2019), Fólica, Roig-Sanz, and Caristia (2020), Ehrlicher (2020), and Ortuño (2020), while a very interesting gender perspective has been taken on by Jofre et al. (2020) regarding female representation through the analysis of faces extracted from historical journals. The digital history approach has also contributed to other fields, including art history (Rodríguez Ortega 2013; Joyeux-Prunel 2015) and initiatives by Hagener (2016), Treveri Gennari (2018) and Clariana (forthcoming), who apply digital humanities tools and ANT to film studies.

When it comes to translation history, TS has an established tradition of research based on corpus analysis (Bernardini and Kenny 2020). Related work has also been conducted in machine translation (Kenny 2020). Translation history has already experienced the rapid growth of small databases, often related to: the translation of a national literature into other languages, for example German translations into Italian (Sisto 2020) or book translations from Dutch (McMartin 2020; Brems et al. 2020); translations conducted over a specific period, such as in the case of French Romantic translations into Spanish (Lafarga 2014); or translations published by a specific publisher, as with the magazine and Argentinean publishing house *Sur* (Wilson 2004). However, while these are valuable works,

these databases often fall short of enabling users to conduct automatic or relational searches, because translation databases are often designed for manual or individual rather than collaborative research. Work based on translation databases also often lacks reflection on how we might create knowledge beyond providing a list of historical translations. For instance, some translation catalogues offer relevant print-media documents as PDFs, but said format precludes metadata extraction and reuse, both of which would allow for comparisons with other databases.

Despite the fact that some databases do not offer automatic or relational searches, there is no doubt that in the last few years, translation history has been driven by new research projects applying data mining (the automatic retrieval of large sets of data to discover patterns and trends that conventional research cannot find), data collection and cleaning, visualization, computational analysis, and interpretation. The ERC project MapModern Social Networks of the Past and related research undertaken by Fólica, Roig-Sanz and Caristia (2020), Ashrafi et al. (2019) on publishing in Iran, Vimr (forthcoming) on peripheral relations between Swedish, Norwegian and Czech, and Meylaerts and Brems' analysis of intra-Belgian literary translations since 1970 (https://www.kbr.be/en/projects/beltrans/), are some of the first endeavours in this regard.

Within this general context, research projects applying a BTH perspective can promote large-scale analyses of translations, considering not only translated texts but also all related metadata. Thus, a BTH approach can allow us to focus on (1) the analysis of translations from a textual point of view, by applying, for example, topic modelling; or (2) the study of the material conditions of book production and circulation, to analyze literary translation flows, networks of publishers, authors and translated works, the role of given languages and the agents involved, and the uncovering of creative networks of agents of translation. This will certainly help us to document historical patterns of translation and cultural change (for example, the lack or presence of the translator's name in translated books or translated literature in journals, which may show the professionalization of the field), as well as local, national, regional, and global milestones in the history of translation and publishing. We maintain that metadata offers great potential in such analyses and note how bibliographic data can be extracted from library catalogues, archives, databases such as WorldCat, national-library catalogues worldwide, digital collections (Terras 2017) such as the Hathi Trust Digital Library, and newer sources such as Wikipedia, Wikidata, and VIAF (Virtual International Authority File). The Web archive has also become one of the most important digital resources for these analyses (Bruegger 2018). We also maintain that any project on BTH should take the following issues into account:

a. The multiple scales and layers of data collection and analysis.
    In a similar vein to Freitag and von Oppen (2010, 19) regarding the study
    of globalizing processes in the Global South, we consider different scales
    in the analysis of translation developments, since these processes cannot be
    seen as separately and exclusively local, national, regional, or global but can
    instead be observed at the intersection of multiple scales in which transna-
    tional actors with multiple roles and transnational networks may also be
    involved (Fólica forthcoming). The use of BTH both as a methodology and
    a research perspective may allow us to examine neglected aspects of multiple
    histories of translation that are not always framed within national historiogra-
    phies and fixed borders. BTH can also enable us to unearth understudied
    networks, transnational actors (Dietze and Naumann 2018), cultural media-
    tors, translation dynamics in less-established cultural capitals, and the rela-
    tionship between connectivity and intercultural exchange at diverse scales.
    Furthermore, we might retrace translation zones and shed new light on the
    too-often labelled "innovative centres" and "imitative peripheries" (Roig-Sanz
    and Meylaerts 2018). A review of the existing literature on translation history
    and the history of books and publishing shows the clear predominance of
    single monographs focused on an individual translation, translator, publisher
    (Vega and Pulido 2013), or specific period of translation. However, there is not
    much of a tradition when it comes to studying the big picture or multiple lay-
    ers.

b. The dimensions of space and time.
    The spatial turn (Soja 1996; Buell 2005; Harvey 2009; Casey 1993) has placed
    space and movement at the core of many current challenges in a wide range
    of fields, pushing translation historians to review classic dichotomies such as
    centre and periphery, dominant and dominated, global and local, and North
    and South. A consensus on the significant role of Southern Europe and non-
    European regions in modern cultural processes has been established. How-
    ever, we still need source-based, empirical research and appropriate tools and
    methods to historically analyze their connections and roles in specific peri-
    ods and disciplines, between spaces and over time. We thus call for a flexible
    periodization of both the corpus and the spaces of comparison (transnational
    approaches are sometimes difficult because they have to do with specific
    national histories that may differ) and a *longue durée* account, in Braudel's
    terms (1958), of how translations and translators circulate (or not) over time,
    extending far beyond borders. This would mean focusing on the longer term
    and taking a broad view of historical milestones of translations in large-scale
    spaces, as we do in our case study below, which aims to build a transnational
    pattern of translation flows in a large region and across a long time span

through the comparison of metadata from the catalogues of different national libraries.

c. The incompleteness and inherent bias of available data.

It is worth recognizing that the sources and catalogues available to us are imperfect given the genealogy, evolution, and expansion of cataloguing practices and standards. In this sense, researchers in BTH share, to a certain extent, the challenges faced by mainstream historians when it comes to working with partial sources, pushing researchers to find ways to fill in the gaps. In this respect, the valuable sources we have mentioned above do not draw enough attention to the differences between geographical contexts. Neither do they highlight the lack of or weak representativity of sources and digital archives that many researchers face.

d. The knowledge divide.

A greater awareness of how, and from where, we produce knowledge should help us realize the unequal representation of university departments from the Global South (Río Riande 2016). This limitation can certainly introduce many biases and reinforce the apparent centrality of some translations or translators, instead of reducing the knowledge divide (Meyer and Schroeder 2015) and making knowledge more diverse and equal. For example, the *Journal for Cultural Analytics* boasts a focus on cultural equality, but most of the papers and datasets currently being published still espouse a predominantly Anglo-American approach. The same goes for digitization by Google Books, which mainly involves books in the English language.

e. The relationship between qualitative and quantitative analysis.

There has been some resistance to using quantification and scientific methods in the study of culture and translation. However, as we will see in the next section, a BTH approach allows us, at this stage, to answer quantitative questions: How many books are literary and how many are not?[4] What percentage of literary titles were reissued? Or how many literary books were published by place or time period? Quantitative analysis will also enable researchers to identify the most common languages and genres among translated books, and how many literary works have been catalogued with data that identifies the translator – an often-neglected field in most databases.

f. Interdisciplinarity, data sharing, and open science. The ethics of open preservation and publication force us to rethink both the field and the profession. What kind of reputation does digital scholarship have? Certainly, we need standards for digital scholarship, and we push for their development. Digital

---

4. In our case study, we understand as "literary" as those books that are classified as "literature" in the Universal Decimal Classification (UDC).

translation historians must support discussions on methodology, ethics, the cost and funding of large projects and their sustainability. What infrastructure do we have at our disposal? How do we maintain it? BTH is not established yet, and there is an urgent need to foster interdisciplinarity and data sharing, as well as to promote training and the use of different sources that can be increasingly combined. Interdisciplinarity will be crucial if we want BTH to be accepted as part of mainstream translation history. New sources develop new standards. At the same time, we also should respect the standards of our discipline. We need to get to know each other's projects and reflect upon our limitations as translation historians. In short, BTH can be used to visualize translation flows and produce new knowledge. But it can also help build research infrastructure and new data, and even though there are considerable differences between academic cultures and open data policies of institutions from different countries, collaboration stands out as a fundamental development – not just for technical support, but for encouraging interdisciplinarity and different ways of thinking among different disciplines.[5] Data practices also vary enormously across disciplines and fields, and we need knowledge structures that we can keep over the long-term. If we can grasp some of the basic terminology and train translation historians in the methodology, computational skills, collaborative platforms, and crowdsourcing projects, our thinking will be transformed, and new research questions and answers will emerge.[6]

3.   A BTH case study: Translated literature in the Spanish speaking-world (1898–1945)

To illustrate the potential of BTH, in this section we present an empirical case study of translated literature in the Spanish-speaking book market. The general purpose is to present a large-scale analysis of all translations into Spanish completed between 1898 and 1945, analyzing the distribution and circulation of trans-

---

5. As an example of collaboration that can help advance this agenda, in our project we have established a data-sharing agreement with Hanno Ehrlicher, from the University of Tübingen, in order to exchange data about literary translation in a corpus of Latin American periodicals digitized by the Ibero-American Institute of Berlin.

6. In the case of BTH, it is worth noting that a variety of big data analytical methods can be applied, for example, association rule learning, classification, cluster analysis, data fusion, data integration, machine learning, natural language processing, network analysis, pattern recognition, predictive modelling, sentiment analysis, spatial analysis, statistics, time series analysis and visualization.

lated literature in Spain and Latin America. This period is bounded by events of historical significance in the Spanish-speaking world: the loss of Cuba by the Spanish empire in the Spanish-American war in 1898, and the end of the Second World War in 1945.

To this end, we have worked with the catalogue of the National Library of Spain (BNE, to use its Spanish initials), which is available for open-access download at datos.gob.es. Founded in 1712 by King Philip V, it receives a copy of every book printed in Spain, given its role as the Spanish legal deposit. Its catalogue counts around 28 million volumes. It launched its pioneering digital library in 2008. We need, nonetheless, to be aware of certain limitations in the BNE catalogue. For instance, the digitization of the catalogue introduced changes or bias in relation to the previous 'analogue' catalogue. Furthermore, there is missing or non-homogeneous data in their records, because the records are not always filled exhaustively by archivists. Finally, we must note that unofficial and underground translations are not registered in this catalogue, because they do not necessarily have a legal deposit number and their circulation follows other paths. Thus, we cannot assume that all books that were published between 1898 and 1945 in Spain are recorded in its catalogue ("the archive"), in contrast to what is suggested by Algee-Hewitt et al. (2016), who posit the following three notions in the digital era: the archive, the research corpus, and what has been published.

Furthermore, since we aim to develop a relational analysis of the Spanish-speaking translation field, we also need to compare the datasets from Latin American national libraries in countries such as Argentina, Mexico, and Colombia, countries where the translation market grew significantly at the outbreak of the Spanish Civil War. Spatial demarcations are thus also central to the coherence of translation history research, which is, however, usually limited to the national or to a specific language within a nation. As noted above, our research problematizes methodological nationalism and moves toward transnational studies that take linguistic diversity within nations into account, alongside the transnational character of certain languages. The large-scale spatial work and general categories such as "Ibero-America" and "Latin America" do not necessarily imply the existence of a homogeneous or monolingual identity. Quite the contrary, BTH grapples with the tension between the general space and the historical-political circumstances conditioning each of the national case studies.

Our methodology relies on the use of a variety of information tools to extract data from library catalogues. We have searched for different types of data, including bibliographic (exploring libraries' online public access catalogues) and geospatial data (geolocating translations allows us to explore alternative capitals for the international book market, such as Buenos Aires and Mexico City, or smaller ones such as Cusco and Puno). Undoubtedly, extracting knowledge from

library catalogues is difficult because of their unstructured and heterogeneous archives. This is the case when conducting large-scale work with various national libraries from, for example, Latin-America, as its libraries rely on disparate ways of indexing their material, digitization and preservation policies, and budgets. For this reason, and despite the great potential they hold for researchers in translation history, these resources remain underexploited. Indeed, if we attempt to measure literary production over time, we will find ourselves confronted with a lack of publication records that could point to publishing activity, such as the International Standard Book Number (ISBN), which was adopted internationally in 1970 but does not gather information for our period of study. As an alternative, using a large corpus of digitized texts can provide a quantitative description of cultural trends. For instance, Google Ngram allows us to find the frequency of terms over time within the corpus of digitized books on the Google Books platform, and the metacatalogue WorldCat, which aggregates information from various library archives, enables the simultaneous consultation of content from each archive.[7] Even though Google Books and the WorldCat online catalogue present some methodological problems, such as inconsistent data and a lack of genre identifiers, by gathering indicators of misclassifications or omissions, we can begin to describe production in a general way (Ikoff and Martínez 2020).

We started working with the BNE's original catalogue, as it offered more detailed information than we could obtain from WorldCat's online catalogue. After automatically extracting the BNE catalogue, we cleaned the data in order to establish consistent categories for the following fields: author/s, author dates, author role (translator, preface writer), (target-text) publisher, publisher place, book date, book place, language, original language, language of translation, Universal Decimal Classification (UDC). The cleaning process was time consuming but fundamental to ensuring a curated database.

As far as the methodological challenges are concerned, any research working with data from similar sources may find that the more relevant a publication, the less consistent the data. In contrast, information on minor or less-renowned works is more trustworthy, with only one or two entries in the catalogue. For instance, *Don Quixote*, the most famous novel in the Spanish language, boasts 4,473 entries in the BNE, but how can we recognize that all these entries refer to the same book? Do different titles such as *El ingenioso caballero don Quijote* and *El Quijote* refer to the same book? A number of tools may be used to measure the degree of proximity between different entries: we have used a variety of techniques such as regular expressions (a sequence of characters defining a search

---

7. Google Ngram Viewer can be accessed at: https://books.google.com/ngrams; and World-Cat, at: https://www.worldcat.org/.

pattern), Levenshtein distance (the minimum number of single-character insertions, deletions or substitutions required to change one word into another), phonetic similarity (comparing factors such as pronunciation, rather than spelling), and n-grams (from which probabilistic models for predicting the next item in a sequence can be generated). For this research, we tested various methods to identify titles which may vary but still refer to the same original book. We used an algorithm that basically compares each title in the database against every other title and calculates the distance between the pair. For the purposes of the comparison, all titles were normalized, i.e., they were lowercased, and all diacritical characters were mapped to their simple ASCII equivalents (e.g., é -> e, ñ -> n). We also used n-grams in cases where a more succinct title was used or where parts of the title appeared in a different position to the comparable title, e.g., E*l ingenioso caballero don Quijote de la Mancha*, and *El Quijote.*

Other types of metadata can also present problems. Oftentimes the names of the publishers are inconsistent, and place and date can also be missing, perhaps more so in the case of the publications by short-lived and smaller publishing houses. In contrast, the bigger the publishing house, the better the information in the catalogue. As part of our cleaning, we normalized publishing-house names, eliminating different endings such as "comp.," "cía.," "coop.," "ed.," "e hijos," "y sus amigos," "hnos.," and "Imp." But this does not resolve this problem entirely, as some publishers are registered under different names, including, in this case, "Compañía Iberoamericana de Publicaciones", "CIAP," and "Iberoamericana de Publicaciones."

Places of publication are missing in some cases, and in others need to be distinguished from the location of the printer, as opposed to the publisher, of a book. In addition, we needed to standardize names of cities, such as "New York," "Nueva York," and "N. Y.," or "México, D.F." and "Ciudad de México".

Another important challenge we faced in the cleaning process was the lack of information regarding literary genre, which is only occasionally filled out in catalogues. In the BNE, the genre is based on the Universal Decimal Classification (UDC), a bibliographic classification for all branches of knowledge. The UDC classification is not always given in the catalogue however: in our database around 45% of entries have this information (or 111,523 out of 245,178 records). It is also worth emphasizing that genre is an arbitrary and historical category that changes with time. For instance, the "biography" was considered a literary genre at the beginning of the 20th century in Spain, and many authors wrote biographies as a literary genre.

Another especially relevant challenge for BTH involves establishing whether a book is a translation. In general, the label "translation" is not explicit in the BNE catalogue but may be deduced from other more-reliable fields: "language code of

original" and "language code of text" or "language" of the catalogued object,[8] as illustrated by several entries for Gustave Flaubert (Figure 1).

	Author	Title	Lang_O	Lang_T	Lang
1	Flaubert, Gustave	La leyenda de San Julián	fre	spa	spa
2	Flaubert, Gustave	Salarnbó	NULL	NULL	spa
3	Flaubert, Gustave	Pages choisies des grands écrivains	NULL	NULL	fre
4	Flaubert, Gustave	Salammbô	NULL	NULL	fre
5	Flaubert, Gustave	Salammbô	NULL	NULL	fre
6	Flaubert, Gustave	La educación sentimental	fre	spa	spa
7	Flaubert, Gustave	Herodías	NULL	NULL	spa
8	Flaubert, Gustave	La Señora de Bovary	NULL	NULL	spa
9	Flaubert, Gustave	Madame Bovary	fre	spa	spa
10	Flaubert, Gustave	Salammbó	NULL	NULL	spa
11	Flaubert, Gustave	Salambó	NULL	NULL	spa
12	Flaubert, Gustave	Atentaçao de Santo Antao	NULL	NULL	por
13	Flaubert, Gustave	Salammbô	NULL	NULL	por
14	Flaubert, Gustave	Un corazón sencillo	fre	spa	spa
15	Flaubert, Gustave	L'education sentimentale	NULL	NULL	fre
16	Flaubert, Gustave	La légende de Saint-Julien l'Hospitalier	NULL	NULL	fre
17	Flaubert, Gustave	La premiére tentation de Saint Antoine (1849–1856)	NULL	NULL	fre
18	Flaubert, Gustave	Herodías	NULL	NULL	spa
19	Flaubert, Gustave	Madame Bovary	fre	spa	spa
20	Flaubert, Gustave	Tres cuentos	fre	spa	spa

Figure 1. Flaubert's works, including translations (Trad-data DB based on BNE catalogue)

Another issue is how to define the author's gender, which is not mentioned in the BNE catalogue. In this case, the simplest and most effective approach would be to match the list of authors against an existing authority's database, such as the Virtual International Authority File (viaf.org), which provides data on gender as well as author nationality and usual language of expression, or to use information from the Spanish national statistics institute to distinguish male and female names. A

---

8. The information is retrieved from the MARC21 fields 041$ h, 041$ a and 008 of the record in the BNE catalogue. For convenience and simplicity, we have renamed these in our database as 'lang_t', 'lang_o' and 'lang', respectively.

final challenge revolves around the uncertain dates in the catalogue. Figure 2, for example, shows regular peaks at the beginning of each decade (1900, 1910, etc.), reflecting imprecise dates recorded for many entries.
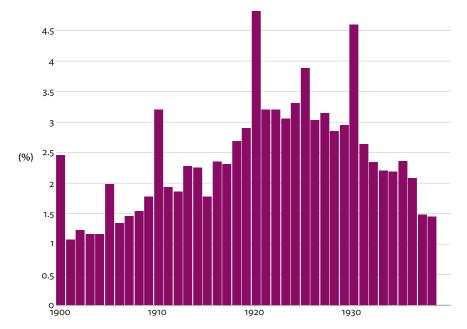

Figure 2. Uncertain dates (Trad-data DB based on BNE catalogue)

Once cleaned, the database we derived from the BNE contained: 258,675 entries (corresponding to 245,178 different books since there are some duplicate entries), of which 38,936 are identified as "literature" (using the UDC); 88,467 contributors (that is, persons who have contributed as authors, translators, compilers, prologue authors, illustrators or in other roles); and 62,958 publishers, publishing houses, and printers. In order to test the reliability of our database, we explored some basic features by means of SQL[9] queries. Note that for this first step of quality control we proposed general and quantitative queries, not to ask ground-breaking new questions but rather to ask simple ones to which answers are already available from previous research, which allowed us to confirm the consistency of our database. These basic questions include:

a.   When were literary books produced?

---

9. SQL (Structured Query Language) is a programming language that allows the user to manage data held in a relational database with structured data.

Figure 3, for example, confirms that during the Spanish Civil War (1936–1939) literary production fell considerably in Spain, as corroborated by our secondary sources (Martínez Martín 2015).
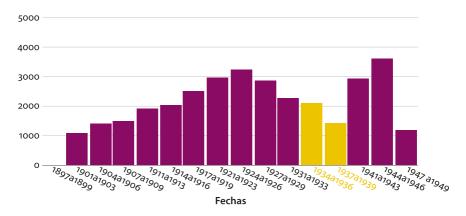

Figure 3. Distribution of literary books by date (Trad-data DB based on BNE catalogue)

b.   Which cities were home to the bulk of literary publishing between 1898 and 1945 in Spanish, Catalan, Galician and Basque according to the BNE catalogue?

Figure 4 places Madrid, Barcelona, and Buenos Aires at the top. Curiously, Mexico City, whose Latin American publishing market was significant throughout the Spanish Civil War, appears in a minor position, after Valencia. Thus, this data should be checked against that of other libraries, such as the Mexican Library, to avoid misjudging the publishing field of the time. We must not forget that the BNE catalogue does not represent the real literary-production market, but only a selection as depicted by a Spanish public cultural institution.

In this respect, it is worth mentioning that the Spanish-speaking market was not limited to Spain and Latin America; France and the United States also published a significant number of translations into Spanish. For instance, Garnier Brothers, P. Lethielleux, and P. Ollendorff published Spanish translations in Paris from the French and English. We note also that publishers of non-Spanish-language origin, such as the American Appletton, the German Brockhaus, the English Ackermann, and the French Garnier and Bouret, controlled the Latin American market at the beginning of the century, posing a stalwart barrier for Spanish publishers (such as CIAP and Espasa Calpe), due to their greater commercial and financial capacities (Fernández Moya 2009, 27).
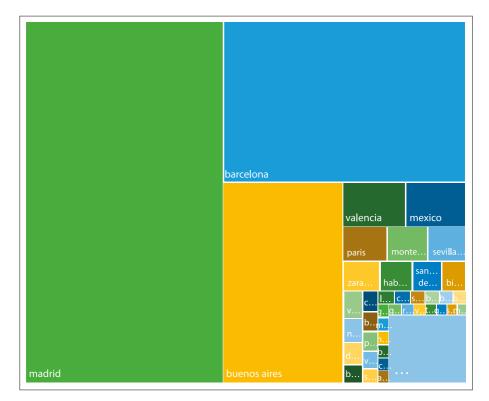
Figure 4. Top ten cities in the Spanish-speaking publishing market (1898–1945) (Trad-data DB based on the BNE catalogue)

c.   What languages were most translated from, according to the BNE catalogue?

We chose the five most translated of 50 source languages because our aim is to study the circulation of translations in the publishing industry. We could study the less-translated languages in the catalogue, but this would tell us less about general patterns of translation flows. For the literary works in our period, the most frequent source language was French, with 1,214 titles and reflecting the well-known *francophilia* of the time, followed by English (1,183 titles), Italian (232), German (208), and Russian (174).

d.   Who were the most translated foreign authors during the period in question?

Our investigation's central goal is to identify the most-translated foreign authors in the BNE catalogue, with the goal of delving into the relationships said authors maintained with their publishing houses in order to establish potential categorizations of the translating publishing houses of the time. We note that graphs consti-

tute useful tools for measuring relationships according to less-traditional notions within the humanities, such as "connectivity" and "centrality", where a node's centrality within a graph depends on the number and diversity of connections it has with the rest of the network. These kinds of visualizations allow us to draw connections between elements in flat databases, as each element in the database may constitute a node and thus be considered in relation to other nodes regarding the presence or absence of connectivity.

In this case, we have detected the BNE's most-translated authors by searching for the 100 authors who appear most frequently in the BNE catalogue. We then checked which of these authors wrote in a language other than Spanish.

The 100 most frequent authors in the BNE include 15 foreign authors, who are, by deduction, the most-translated authors during the period in question. They are listed here in descending order of entries: Shakespeare (257), Verne (168), Dumas (150), Tolstoj (139), Dostoevskiï (137), Balzac (134), Dickens (134), Salgari (134), Wallace (133), Wilde (133), Maurois (132), Goethe (122), Hugo (113), Queirós (109), and Gor'kiï (104).[10] In Figure 5, nodes representing these fifteen authors are shown in red, the nodes of their translators in green, and their publishing houses in purple. The size of the nodes and thickness of the lines vary according to the number of entries. At a glance, we might notice that these authors were published by a wide variety of publishing houses. Some "publishers" are publishing houses, such as Espasa Calpe and Aguilar, while others are magazines that publish books such as *La Revista Literaria* and *Prensa Moderna*, and still others are publishing houses and presses like Imprenta Helénica. Indeed, this variety stands as a heritage to the dispersion of the publishing market of the time. Nonetheless, we can establish a ranking of the top 15 "publishers" as follows: R. Velasco (2684), Espasa Calpe (2561), Labor (978), La Escena (878), Los Contemporáneos (826), Sociedad de Autores Españoles (800), Aguilar (766), Saturnino Calleja (763), *Bambalinas* (760), Ramón Sopena (726), Maucci (720), *La Prensa Moderna* (719), La casa del Patufet (646), Sucesores de Hernando (632), and Renacimiento (610).

If we focus on the most-translated French authors, we may note that the most popular authors – the likes of Dumas, Balzac, and Hugo – had their works published by a wide variety of publishing houses, including some of the most central publishing houses mentioned above – Espasa Calpe, Sopena, and Maucci. Meanwhile, Maurois (on the lower left) seems more tangential to the "French group," having published his works in smaller presses (understood as those with fewer connective relationships with other agents): Juventud, Clarassó, Naúsica, and J. Pueyo. This might lead us to believe that there was a certain divide between the

---

10.   Note that the spelling used by the BNE is respected here.

types of foreign literatures published by the "great" and "small" presses of the time, with the latter being more prone to publish then-contemporary literature, as the Maurois case shows.[11]
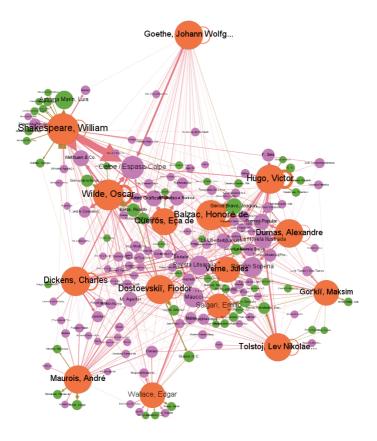

Figure 5. Graph of the BNE's 15 most-translated authors, generated by V. Ikoff using Gephi (Bastian et al. 2009) and based on the BNE catalogue

e.   How do publishing houses feature in the network?

The information can also be organized according to publishing houses. If an author was translated by various publishing houses, they would appear in a position closer to the node for the publishing house with which they have most relations, and in the colour of the community of which they form a part and with

---

11. See also the work by Ronen et al. (2014), which shows the role of mediating nodes in the international creation of literary prestige.

whose members they are most connected.[12] We tested the usefulness of such an analysis by selecting authors with more than ten relationships with the publishing house at hand, gathering 436 authors with ties to the 15 publishing houses with the most entries in the BNE. The ensuing graph is very large and complex and best viewed on a computer screen. What it allowed us to observe, however, is that two publishing houses, Biblioteca Renacimiento (1909–1927) and Saturnino Calleja (1876–1958), and two magazines, *Los Contemporáneos* (1909–1926) and *Prensa Moderna* (*circa* 1920–1940), take up the graph's centre space. All four seem to publish Spanish authors. The "translating" publishing houses, like Espasa Calpe, Maucci, Aguilar, and Sopena lie further away from the graph's centre. Two other interesting groups appear in highly decentred positions within the graph: La casa de Patufet and *La Escena* and *Bambalinas*. The former was a popular Catalan-language publishing house for children's literature, which explains why it boasted fewer ties to more general publishing houses. The latter group, *La Escena* and *Bambalinas*, published playwrights. *Bambalinas*, an Argentine theatre magazine launched in 1918, published monographs with plays. In terms of the *La Escena* theatre magazine, we find two different sets of works in the BNE catalogue: one published in Buenos Aires (with 815 entries between 1918 and 1933), and the other in Barcelona (with 66 entries of Spanish popular drama between 1941–1944). This is problematic since graphically the publisher in all these cases seems to be the same and it is not possible to disambiguate by looking at the name only. While we should look into this specific case in more detail, we believe that, even though both publications have the same name, they are in fact independent of each other – the kinds of plays published in each were quite different. In Argentina, the magazine published popular country- and gaucho-related plays or *sainetes*, while the Barcelonan magazine published popular Spanish literature (and some translations of plays such as an adaptation for theatre of Brontë's *Wuthering Heights*).

The peripheral group of playwrights around *Bambalinas* and *La Escena* may prove quite interesting, as we might home in on figures, such as Wilde, who operated as links between the theatre publishing world and general publishing houses. For instance, according to the BNE's records, *El abanico de Lady Windermere* was first published in *La Escena* (Buenos Aires) in 1919, having been translated by Francisco José Bolla, and was later translated by Ricardo Baeza in 1920, and published in Madrid (publishing house unknown), and in 1930 in Barcelona by El Cisne (translator unknown). Baeza's translation was republished in Madrid in 1931, and by Espasa Calpe in 1942 in Buenos Aires, among other editions. As is

---

12. The layout and position are determined by the ForceAtlas2 algorithm (Jacomy et al 2014), and the community is detected through the method developed by Blondel, Guillaume et al. (2008).

clear from the BNE catalogue, other plays by Wilde were translated by renowned translators for well-known publishing houses: examples include Baeza's *La importancia de llamarse Ernesto* (published by Espasa Calpe in 1920) and Julio Goméz de la Serna's *El retrato de Dorian Gray* (published by Biblioteca Nueva in 1918).

We might also identify certain lesser-known mediators (see Figure 6), such as the playwright Leandro Navarro, who published his works with *La Escena* in 1941, Velasco in 1928, Prensa Moderna in 1930, imp. Rivadeneyra since 1933, and Sociedad de Autores in 1928. This is just one of many examples of non-canonical authors who, nonetheless, served as links between various groups in the editorial map of the time. Fortunately, network analyses using graphs allow us to cast light on such cases.
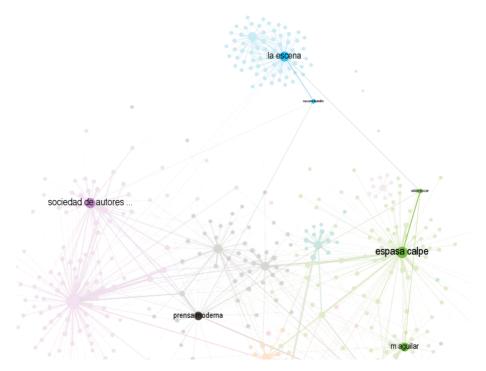

Figure 6. Relationship detail for publishing houses related to Wilde and Navarro, generated by V. Ikoff using Gephi (Bastian et al. 2009) and based on the BNE catalogue

Thus, in this section, we have presented findings based on quantitative queries in our database regarding translation flows and the time and place of publication of different translations, as a good way to test the reliability of our database. Our findings are consistent with the consensus in specialized literature on

publishing history. In a slightly more innovative move, we have also proposed the use of graphs to explore the relationships between foreign authors, translators, publishers, and books. This second part, it is hoped, served to highlight, publishers or cultural mediators who may have been overshadowed in the past.

4.   Conclusions

There is no doubt that, today, many researchers around the world are already working on the study of cultural phenomena on a large-scale, and from a transnational perspective. Translation studies and, more precisely, translation history are currently being changed by data mining and data-driven approaches. We have proposed BTH as a new, interdisciplinary field of study that allows us to analyze large translation datasets and flows using computational and visualization techniques. We have also included a case study to illustrate some of the challenges we still face. In doing so, we have strived to shed light on the following issues: (1) methodologies for the analysis of translation history on a large scale; (2) methodological challenges and pitfalls in the collection, analysis, and visualization of data and metadata on and from a significant number of translations registered in the catalogue of the National Library of Spain; (3) the complementarity between quantitative and qualitative approaches in the analysis of the big picture, as well as in the study of the agents involved; and (4) the challenges of historical depth when applying cross-border research and using large-scale data that touches upon multiple translation histories and translation processes encompassing various parts of the world. As noted, large-scale research requires quantitative analyses – often through visualizations allowing for data exploration – that cannot always reach exhaustivity when using a distant reading approach. This large scale means that researchers must be selective, but they are also constrained by what is available. In that respect, the translation datasets that we have at our disposal are in some cases small or otherwise lacking, and this should also reinforce our awareness of the reliability of our results. To be sure, our findings are often indicative rather than exhaustive, and the data can also introduce some bias depending on the chosen sources and their accessibility and completeness. Thus, it is worth recognizing that the sources and catalogues available are often imperfect and that their construction always involves a wide range of human criteria and factors. Alongside the relative lack of archives, we also must face the problem of specific national criteria in the building of the catalogues. As we have also shown, we do not yet have common standards for the quantification of culture, nor for the quantification of translations, for that matter. However, if we establish a rigorous methodological process, the knowledge we can generate will prove more significant than

ever before. This will certainly allow us to recover singular and representative histories (micro-histories) of translations and mediators, as well as to study the big picture and the multiple layers and relations that are behind it. As stated above, we are convinced that one of the main possibilities offered by the analysis of BTH is to help decentralize translation history and literary and cultural history, in a broad sense, while breaking with national historiographies in translation. In this sense, data science and data-driven approaches seem to be very exciting research avenues.

Funding

Acknowledgements

References

Aiden, Erez, and Jean-Baptiste Michel. 2014. *Uncharted: Big Data as a Lens on Human Culture*. New York: Riverhead Books.

Algee-Hewitt, Mark, Sarah Allison, Marissa Gemma, Ryan Heuser, Franco Moretti, and Hannah Walser. 2016. "Canon/Archive. Large-scale Dynamics in the Literary Field". *Literary Lab. Pamphlet* 11: 1–13.

Ashrafi, Nasrin, Mohammad Reza Hashemi, and Hossein Akbari. 2019. "Towards a new methodological approach to social historiography of translation. A social network analysis perspective." *Translation Spaces* 8 (2): 231–56. Accessed August 30, 2021. https://doi.org/10.1075/ts.19003.ash

Bail, Christopher A. 2014. "The cultural environment: Measuring culture with big data." *Theory and Society* 43 (3–4): 465–82. https://doi.org/10.1007/s11186-014-9216-5

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. "Gephi: an open source software for exploring and manipulating networks." Paper presented at the International Association for the Advancement of Artificial Intelligence on on Weblogs and Social Media, held in Pasadena, California (USA).

Bernardini, Silvia, and Dorothy Kenny. 2020. "Corpora." In *Routledge Encyclopedia of Translation Studies*, edited by Mona Baker and Gabriela Saldanha, 110–15. London/New York: Routledge.

Blank, Grant. 2007. *Critics, Ratings, and Society*. Lanham: Rowman and Littlefield.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment* (October): P10008.
https://doi.org/10.1088/1742-5468/2008/10/P10008

Borgman, Christine. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9963.001.0001

Braudel, Fernand. 1958. "Histoire et Sciences sociales: La longue durée." *Annales. Histoire, Sciences Sociales* 13 (2) (October – December): 725–53.
https://doi.org/10.3406/ahess.1958.2781

Brems, Elke, Theresia Feldmann, Orsolya Réthelyi, and Ton Van Kalmthout. 2020. "The Transnational Trajectories of Dutch Literature as a Minor Literature: A View from World Literature and Translation Studies." *Dutch Crossing – Journal of Low Countries Studies* 44: 2, 125–35. https://doi.org/10.1080/03096564.2020.1747005

Bruegger, Niels. 2018. *The Archived Web: Doing History in the Digital Age*. Cambridge MA: MIT Press. https://doi.org/10.7551/mitpress/10726.001.0001

Buell, Lawrence. 2005. *The Future of Environmental Criticism: Environmental Crisis and Literary Imagination, USA-UK*. Victoria, Australia: Blackwell Publishing.

Casey, Edward S. 1993. *Getting Back into Place: Toward a Renewed Understanding of the Place-World*. Indianapolis: Indiana University Press, 2nd edition, 2009.

Chesterman, Andrew. 2009. "The name and nature of translator studies." *Hermes: Journal of Language and Communication Studies* 42: 13–22.

Clariana Rodagut, Ainamar. forthcoming. "Ibero-American Women Mediators and their Networks in Early Film Clubs: Lola Álvarez Bravo, Victoria Ocampo y María Luz Morales." *Modernism/Modernity, special issue Global South Cinephilias*, edited by Rielle Navistki, 2022.

Cohen, Daniel, and Rosenzweig, Roy. 2005. *Digital History*. Philadelphia: University of Pennsylvania Press.

Champion, Eric Malcolm. 2017. "Digital humanities is text heavy, visualization light, and simulation poor." *Digital Scholarship in the Humanities* 32 (1): 25–32.

Da, Nan Z. 2019. "The Computational Case against Computational Literary Studies." *Critical Inquiry* 45: 3. https://doi.org/10.1086/702594

Dietze, Antje, and Katya Naumann. 2018. "Revisiting transnational actors from a spatial perspective." *European Review of History: Revue européenne d'histoire* 25 (3–4): 415–30.
https://doi.org/10.1080/13507486.2018.1439892

D'Hulst, Lieven. 2012. "(Re)locating translation history: From assumed translation to assumed transfer." *Translation Studies* 5 (2): 139–55. https://doi.org/10.1080/14781700.2012.663597

Drucker, Johanna. 2014. *Graphesis: visual forms of knowledge production*. Cambridge, MA: Harvard University Press.

Düring, Marten, Roman Kalyakin, Estelle Bunout, and Daniele Guido. 2021. "Impresso. Inspect and Compare. Visual Comparison of Semantically Enriched Historical Newspaper Articles." *Information* 12 (348): 1–12. Accessed August 30, 2021. https://doi.org/10.3390/info12090348

Ehrlicher, Hanno. 2020. "The magazine as a medium of cultural translation. First steps towards transnational research." In *Literary Translation in Periodicals. Methodological Challenges for a Transnational Approach*, edited by Laura Fólica, Diana Roig-Sanz, and Stefania Caristia, 273–96. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.155.11ehr

Fernández-Moya, María. 2009. "Editoriales españolas en América Latina. Un proceso de internacionalización secular." *ICE. Revista de Economía* 849: 66–77.

Fisher, Danyel et al. 2012. "Interactions with big data analytics." *Interactions* 19 (3): 50–59. Accessed August 30, 2021. https://doi.org/10.1145/2168931.2168943

Fólica, Laura. forthcoming. "Digital Humanities and Big Translation History in the Global South: the Spanish-Speaking Case." *World Literature Studies*.

Fólica, Laura, Diana Roig-Sanz, and Stefania Caristia (eds). 2020. *Literary Translation in Periodicals: Methodological challenges for a transnational approach*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.155

Freitag, Ulrike, and Achim von Oppen (eds). 2010. *Translocality. The Study of Globalising Processes from a Southern Perspective*. Leiden/ Boston: Brill. https://doi.org/10.1163/ej.9789004181168.i-452

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. London: Imperial College Press. https://doi.org/10.1142/p981

Grandjean, Martin. 2018. Les réseaux de la coopération intellectuelle. La Société des Nations comme actrice des échanges scientifiques et culturels dans l'entre-deux-guerres. PhD thesis, Université de Lausanne.

Hagener, Malte. 2016. "Cinephilia and Film Culture in the Age of Digital Networks." In *The State of Post-Cinema. Tracing the Moving Image in the Age of Digital Dissemination*, edited by Malte Hagener, Vinzenz Hediger, and Alena Strohmaier, 181–94. London/New York: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-52939-8_11

Harvey, David. 2009. *Cosmopolitanism and the Geographies of Freedom*. New York: Columbia University Press.

Heilbron, Johan. 1999. "Towards a Sociology of Translation." *European Journal of Social Theory* 2 (4): 429–44. https://doi.org/10.1177/136843199002004002

Heilbron, Johan, and Gisèle Sapiro. 2002. "La traduction littéraire, un objet sociologique." *Actes de la recherche en sciences sociales* 144: 3–5. https://doi.org/10.3917/arss.144.0003

Hitchcock, Tim. 2013. "Confronting the digital: or how academic history writing lost the plot." *Cultural and Social History* 10: 1, 9–23. https://doi.org/10.2752/147800413X13515292098070

Ikoff, Ventsislav, and Pablo Martínez. 2020. "Quantitative analysis of translations in Spanish-language periodical publications (1900–1945): A methodological proposal." In *Literary Translation in Periodicals. Methodological challenges for a transnational approach*, edited by Laura Fólica, Diana Roig-Sanz, and Stefania Caristia, 313–24. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.155.13iko

Jacomy, M., T. Venturini, S. Heymann, and M. Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS ONE* 9(6): e98679. Accessed August 30, 2021. https://doi.org/10.1371/journal.pone.0098679

Jockers, Matthew. 2013. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press. https://doi.org/10.5406/illinois/9780252037528.001.0001

Jofre, Ana, Vincent Berardi, Carl Bennett, Michael Reale, and Josh Cole. 2020. "Faces extracted from *Time Magazine*, 1923–2014." *Journal of Cultural Analytics* 1 (1): dataset. Accessed August 30, 2021. https://doi.org/10.22148/001c.12265

Joyeux-Prunel, Béatrice. 2015. "Provincializing Paris. The Centre/Periphery Narrative of Modern Art in Light of Quantitative and Transnational Approaches." *Artlas Bulletin* 4 (1): 40–64.

Kenny, Dorothy. 2020. "Machine Translation." In *Routledge Encyclopedia of Translation Studies*, edited by Mona Baker and Gabriela Saldanha, 305–310. London/New York: Routledge.

Lafarga, Francisco. 2014. "Tendencias de la traducción de autores del XVIII francés en España (1975–2013). Consideraciones en torno a un catálogo." *Anales de filología francesa* 22: 161–75.

Liu, Alan. 2018. *Friending the Past: The Sense of History in the Digital Age*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226452005.001.0001

Manovich, Lev. 2015. "The Science of Culture? Social Computing." *Digital Humanities and Cultural Analytics*. Accessed August 30, 2021. https://doi.org/10.22148/16.004

Manovich, Lev. 2020. *Cultural Analytics*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/11214.001.0001

Mauro, André de, Marco Greco, and Michele Grimaldi. 2015. "What is big data? A consensual definition and a review of key research topics." AIP Conference Proceedings 1644, 97. Accessed August 30, 2021. https://doi.org/10.1063/1.4907823

McMartin, Jack. 2020. "Dutch Literature in Translation: A Global View." *Dutch Crossing-Journal of Low Countries Studies* 44 (22): 145–64. https://doi.org/10.1080/03096564.2020.1747006

Martínez Martín, Jesús A. (ed.). 2015. *Historia de la edición en España 1939–1975*. Madrid: Marcial Pons. https://doi.org/10.2307/j.ctt20fw6qd

Meyer, Eric, and Schroeder, Ralph. 2015. *Knowledge Machines. Digital Transformations of the Sciences and Humanities*. London/ Cambridge (US): MIT Press. https://doi.org/10.7551/mitpress/8816.001.0001

Meylaerts, Reine, Maud Gonne, Tessa Lobes, and Diana Roig Sanz. 2016. "Cultural mediators in cultural history: what do we learn from studying mediators' complex transfer activities in interwar Belgium?." In *Doing Double Dutch. The International Circulation of Literature from the Low Countries*, edited by Elke Brems, Ton Van Kalmthout, and Orsolya Rethelyi, 51–75. Leuven: Leuven University Press.

Meylaerts, Reine, Lieven d'Hulst, and Tom Verschaffel (eds). 2018. *Cultural mediation in Europe, 1800–1950*. Leuven: Leuven University Press. https://doi.org/10.11116/9789461662408

Milton, John, and Paul Bandia (eds). 2009. *Agents of Translation*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.81

Moretti, Franco. 1998. *Atlas of the European Novel, 1800–1900*. London: Verso

Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.

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

Ortuño, Rocío. 2020. "Challenges and strategies for beginners to solve research questions with DH methodologies on a corpus of multilingual Philippine periodicals." In *Literary Translation in Periodicals. Methodological Challenges for a Transnational Approach*, edited by Laura Fólica, Diana Roig-Sanz, and Stefania Caristia, 247–72. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.155.10ort

Pegenaute, Luis. 2018. "Translation and cultural development: Historical approaches." In *The Routledge Handbook of Translation and Culture*, edited by Sue Ann Harding, and Ovidi Carbonell i Cortés, 177–206. London/New York: Routledge.

Piper, Andrew. 2018. *Enumerations: Data and Literary Study*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226568898.001.0001

Porter, J. D. 2018. "Popularity/Prestige." *Pamphlet* 17: 1–24. Stanford: Stanford Literary Lab.

Pym, Anthony. 1998. *Methods in Translation*. London: Routledge.

Pym, Anthony. 2001. "Alternatives to borders in translation theory." *Althanor* 12: 172–82.

Río Riande, Gimena del. 2016. "De todo lo visible y lo invisible o volver a pensar la investigación en humanidades digitales." *SIGNA* 25: 95–108. https://doi.org/10.5944/signa.vol25.2016.16943

Rodríguez Ortega, Nuria. 2013. "Digital Humanities, Digital Art History and Artistic Culture: Relationships and Disconnections." *Artnodes* 13. Accessed August 30, 2021. https://doi.org/10.7238/a.v0i13.2017

Roig-Sanz, Diana, and Reine Meylaerts. 2018. *Literary Translation and Cultural Mediators in 'Peripheral' Cultures. Customs Officers or Smugglers?*. London/New York: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-78114-3

Ronen, Shahar, et al. 2014. "Links that Speak: The Global Language Network and Its Association with Global Fame." In *Proceedings of the National Academy of Sciences* 111 (52): E5616–E5622. Accessed August 30, 2021. https://doi.org/10.1073/pnas.1410931111

Simeoni, Daniel. 1995. "Translating and studying translation: The view from the agent." *Meta* 40: 3, 445–60. https://doi.org/10.7202/004146ar

Sisto, Michele. 2020. "Literary journals and book series as agents of consecration. Thomas Mann and Franz Kafka in the Italian literary field (1908–1938)." In *Literary Translation in Periodicals. Methodological Challenges for a Transnational Approach*, edited by Laura Fólica, Diana Roig-Sanz, and Stefania Caristia, 69–92. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.155.03sis

Schäfer, Mirko Tobias, and Karen van Es. 2017. *The Datafied Society. Studying culture through data*. Amsterdam: Amsterdam University Press.

Schelstraete, Jasper, and Marianne van Remoortel. 2019. "Towards a Sustainable and Collaborative Data Model for Periodical Studies." *Media History* 25 (3): 336–54. Accessed August 30, 2021. https://doi.org/10.1080/13688804.2018.1481374

Shandler, Jeffrey. 2017. *Holocaust Memory in the Digital Age*. Stanford: Stanford University Press. https://doi.org/10.1515/9781503602960

So, Richard Jean, and Hoyt Long. 2013. "Network Analysis and the Sociology of Modernism." *Boundary* 2 40 (2): 147–82. https://doi.org/10.1215/01903659-2151839

Soja, Edward. 2008. "Taking Space Personally." In *The Spatial Turn: Interdisciplinary Perspectives*, edited by Santa Arias, and Barbara Warf. London: Taylor and Francis, 11–35.

Terian, Andrei. 2019. "Big Numbers. A Quantitative Analysis of the Development of the Novel in Romania." *Transylvanian Review* 28 (1): 55–71.

Terras, Melissa, et al. 2017. "Enabling complex analysis of large-scale digital collections: Humanities research, high-performance computing, and transforming access to British Library digital collections." *Digital Scholarship in the Humanities* 33 (2): 356–466.

Treveri Gennari, Daniela. 2018. "Understanding the Cinemagoing Experience in Cultural Life: The Role of Oral History and the Formation of 'memories of Pleasure'." *TMG Journal for Media History* 21 (1): 39–53. https://doi.org/10.18146/2213-7653.2018.337

Underwood, Ted. 2017. "A genealogy of distant reading." *Digital Humanities Quarterly* 11 (2).

Vega, Miguel, and Martha Pulido. 2013. "The History of Translation and of the Theory of Translation in the Context of Translation Studies." *MonTI* 5: 39–70.

Verboord, Marc. 2003. "Classification of Authors by Literary Prestige." *Poetics* 31: 259–81. https://doi.org/10.1016/S0304-422X(03)00037-8

Verboord, Marc, Giselinde Kuipers, and Susanne Janssen. 2015. "Institutional Recognition in the Transnational Literary Field." *Cultural Sociology* 9 (3): 447–65. https://doi.org/10.1177/1749975515576939

Vimr, Ondrej. forthcoming. "Geographies and Networks: Spatial Aspects of Global Literary Circulation." In *Global Literary Studies. Key Concepts*, edited by Diana Roig-Sanz and Neus Rotger. Berlin: De Gruyter, 2022.

Wilson, Patricia. 2004. *La constelación del* Sur. *Traductores y traducciones en la literatura argentina del siglo XX.* Buenos Aires: Siglo XXI.

Address for correspondence

Diana Roig-Sanz
Internet Interdisciplinary Institute (IN3)
ICREA, Institució Catalana de Recerca i Estudis Avançats
Universitat Oberta de Catalunya
Avinguda Carl Friedrich Gauss, 5
Castelldefels
08860 Barcelona
Spain

dsanzr@uoc.edu
https://orcid.org/0000-0002-6353-7708

Co-author information

Laura Fólica
Universitat Oberta de Catalunya

lfolica@uoc.edu
https://orcid.org/0000-0002-0262-3864

Publication history